DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

Some methodological issues in the
implementation of subjective poverty
definitions

by Arie Kapteyn
    Peter Kooreman
    Rob Willemse   FEW 245

SOME METHODOLOGICAL ISSUES IN THE

IMPLEMENTATION OF SUBJECTIVE POVERTY DEFINITIONS

by

Arie Kapteyn (Tilburg University)
Peter Kooreman (Tilburg University)
Rob Willemse (Federal Reserve Bank of Richmond, Virginia)

For all correspondence write to:

Department of Econometrics
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands

CONTENTS

Abstract

The paper contains an investigation of the effects of systematic under-reporting of income and of sample selectivity on the estimated levels of two subjective definitions of poverty: the so-called subjective poverty line and the Leyden poverty line. Both problems turn out to have substantially biasing effects. Methods are presented that remedy the biases. The resulting adjusted poverty lines prove to be quite accurate. Furthermore, suggestions are made for the design of questionnaires that are used in the surveys on which these poverty definitions are based.

1.   Introduction

Whatever poverty definition one adheres to, a proper implementation is prerequisite for its possible use in social policy. In this paper we pay attention to a number of methodological issues that arise in the empirical implementation of certain subjective definitions of poverty. In particular, we are concerned with the so-called subjective poverty line (SPL) and Leyden poverty line (LPL), which have been introduced and discussed in a string of papers by, mainly, Dutch and American authors (see references). Both approaches are subjective in that they are based on responses to survey questions which try to elicit either a respondent's evaluation of income levels or his judgment about minimum needs. Furthermore, both approaches are model based, in the sense that the responses themselves do not generate poverty lines immediately. One needs to estimate a model that explains interhousehold variation in the responses to the survey questions. These two aspects identify two crucial methodological issues in the implementation of the SPL and LPL: the responses should measure what they are supposed to measure and the model should be correctly specified and estimated.

In this paper we present some empirical evidence on the sensitivity of the poverty line definitions to systematic biases in the responses and to an incorrect estimation method for the model. The systematic response bias is a result of the rather general tendency of respondents to underestimate their own after tax household income. The incorrect estimation method stems from a disregard of sample selectivity due to partial non-response.

In Section 2 we briefly explain the SPL and LPL and we discuss and estimate the model that is assumed to generate the responses. We point out some implausible outcomes.

In Section 3, we present evidence that respondents severely underestimate their after tax household income, which results in a downward bias of the poverty lines. We also present a method to correct this bias. In Section 4 we take up the problem of selectivity bias. It turns out that selectivity is a statistically significant problem. In Section 5 we compare the poverty lines that result from the various models. Both the income correction and the correction for sample selectivity bias lead to sizable changes in the estimated level of the poverty line.

In Section 6 we draw conclusions from our findings for the design of questionnaires used in the empirical work that underlies the subjective poverty line definitions considered here. All in all, the problems considered can be remedied rather easily.

2. Two subjective definitions of poverty and their implementation

The so-called subjective poverty line (SPL) is based on the following survey question, posed to the head of the household:

"Which after tax monthly income do you, in your circumstances, consider to be absolutely minimal? That is to say that with less you could not make ends meet".

Absolutely minimal per month $...

don't know

A respondent's answer to this minimum income question (MIQ) will be referred to as his minimum income y_{min}. It turns out that y_{min} depends on the respondent's actual after tax income and a number of other factors, including family composition. See, for instance, Kapteyn, Van de Geer, Van de Stadt [1985] for details. In formula:

(2.1) $\quad y_{min} = f(y;x)$,

where y is the respondent's actual income[1] and x is a vector of other factors. The function f is monotonically increasing in y and there exists an income level y_{min}^{*} defined by

(2.2) $\quad y_{min}^{*} = f(y_{min}^{*};x)$,

1) In this paper "income" is always after tax household income.

such that for all incomes y less than y_{min}^{*} we have that $y<y_{min}$ and for all incomes y greater than y_{min}^{*}, $y>y_{min}$. See Fig. 2.1. The income level y_{min}^{*} is
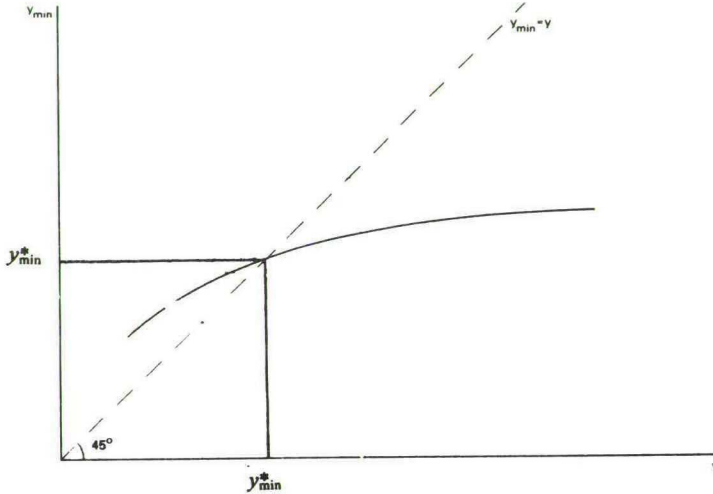

Fig. 2.1 The subjective poverty line

the SPL, as it is the point where families can just make ends meet; with less income they cannot make ends meet and with more income than y_{min}^{*} they can. Since the position of the function $f(y;x)$ depends on x, it is clear from (2.2), and from Fig. 2.1 that the SPL depends on x. That is, if families have different characteristics they will require different amounts of money to make ends meet.

The Leyden poverty line (LPL) is based on the so-called income evaluation question (IEQ):

"Which after tax monthly income would you, in your circumstances, consider to be very bad? And bad? Insufficient? Sufficient? Good? Very good?

very bad $ ...
bad $ ...
insufficient $ ...
sufficient $ ...
good $ ...
very good $ ...

(we mean after tax household income)

A respondent's answers to the IEQ are used to estimate his <u>welfare function</u> <u>of income</u> (WFI). The measurement method is illustrated in Fig. 2.2.
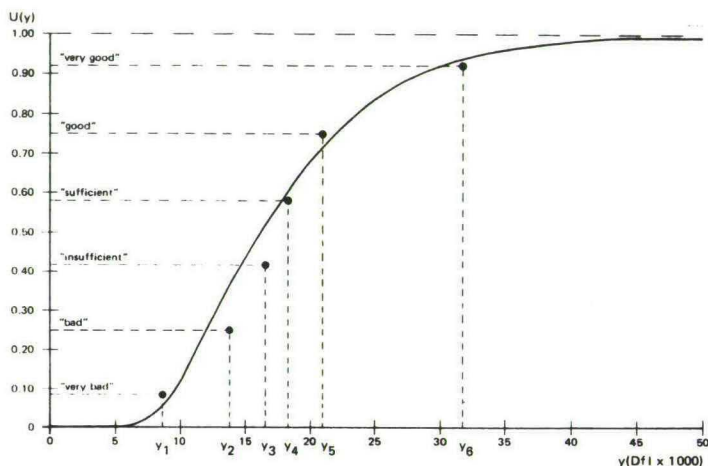

Fig. 2.2   The Leyden poverty line

The verbal labels "very good", "good", etc. have been identified with mid-points of the six equal intervals on a zero-one scale. In this way the verbal evaluations have been transformed into numerical evaluations. (For details and justification see Van Praag [1971], Van Praag and Kapteyn [1973], Kapteyn [1977]). The response to the IEQ can now be represented by a scatter of six points. According to Van Praag [1968] the relation between an income level z and its numerical evaluation on a zero-one scale, $U(z)$, can be approximated quite well by a lognormal distribution function:

(2.3) $U(z) = \Lambda(z,\mu,\sigma)$,

where $\Lambda(.,\mu,\sigma)$ is the lognormal distribution function with median e^{μ} and log-variance σ^2. The parameters μ and σ of a respondent are estimated by

fitting a lognormal function through the scatter of points in $(z, U(z))$-space, as illustrated in Fig. 2.2. Van Praag and Kapteyn [1973] and Van Herwaarden and Kapteyn [1981] provide further details.

The estimated lognormal distribution function represents the respondent's WFI. The quantity e^μ is a location parameter, it is the income level which is evaluated at 0.5 by the individual; σ is a slope parameter. The evaluation of an income z, by an individual with <u>welfare parameters</u> μ and σ is given by

$$(2.4) \quad U(z) = \Lambda(z; \mu, \sigma) = N(\ln z; \mu, \sigma) = N(\frac{\ln z - \mu}{\sigma}; 0, 1),$$

where $N(.; \mu, \sigma)$ is the normal distribution function with mean μ and variance σ^2.

The LPL is based on the notion that poverty is a state of low utility. If the WFI is taken as a cardinal utility function of income, someone is defined as poor, if his income y is such that

$$(2.5) \quad U(y) \leq \alpha,$$

where α is a "welfare level" (a number between zero and one), which has to be set by politicians. Let us define u_α by

$$(2.6) \quad \Lambda(u_\alpha; 0, 1) = N(\ln u_\alpha; 0, 1) = \alpha,$$

then according to the LPL someone is poor if his income y satisfies

$$(2.7) \quad \frac{\ln y - \mu}{\sigma} \leq u_\alpha,$$

or

$$(2.8) \quad y \leq \exp(\mu + \sigma \cdot u_\alpha)$$

It turns out that the parameters μ and σ both depend on a vector of family characteristics x. Thus, just like the SPL, the LPL also depends on x.

2.1. The Model

To make the SPL and LPL operational we have to specify the relation be-
tween y_{min}, μ and σ on the one hand and the vector of household
characteristics x on the other hand. The explanation of μ and σ is derived
from a "theory of preference formation" developed in Kapteyn [1977]; see,
for example, Van de Stadt, Kapteyn, Van de Geer [1985] (SKG hereafter) and
Kapteyn, Van de Geer, Van de Stadt [1985] (KGS hereafter) for details.
Estimation of the complete model requires panel data. Only cross section
data are available, however, for our analysis. Therefore all lagged ex-
planatory variables in the model are ignored. The resulting specification
for μ is

$$(2.9) \quad \mu_n = \beta_0 + \beta_1(1-\beta_2)\ln fs_n + \beta_2 \ln y_n + \beta_3(m_n - \beta_1 hs_n) + \epsilon_n,$$

where μ_n is the value of μ for family n, fs_n is the size of family n, y_n is
its after tax income, m_n is mean log-income in the reference group of
household n, hs_n is mean log-family size in the reference group of household
n, ϵ_n is an error term capturing all omitted factors[1]. Since μ_n is an in-
dicator of the level of financial wants of a family, equation (2.9) says
that a family's financial wants are determined by its income, family size
and by the geometric mean of incomes in the reference group, adjusted for
the geometric mean of family sizes in the reference group.

The theory of preference formation mentioned above implies that σ_n is
determined by the _dispersion_ of incomes and family sizes in family n's
reference group, both present and past, and by the variability of past in-
comes of family n. Although this dependency raises various interesting
policy issues (see KGS) we will ignore it here for simplicity's sake.

---

1) Equation (2.9) can be obtained from (22) in KGS by omitting all lagged
   variables on the right side.

Empirically, this amounts to taking the dispersion of incomes and family sizes as given and dealing with the observed variation of σ across families as being determined exogenously. Furthermore, as σ appears to be uncorrelated with the explanatory variables on the right hand side of (2.9), we take σ as exogenous and for the purpose of constructing a poverty line according to (2.8) we set σ equal to its sample mean $\bar{\sigma}$.

The explanation of the logarithm of y_{min} is based on a similar specification as the explanation of μ:

$$(2.10) \quad \ln y_{min,n} = \alpha_0 + \alpha_1(1-\alpha_2)\ln fs_n + \alpha_2 \ln y_n + \alpha_3(m_n - \alpha_1 hs_n) + u_n,$$

where u_n is an error term, possibly correlated with ε_n in (2.9).

The specification of the influence of family size so far is primitive for two reasons. First, it is restrictive to simply count the number of family members without regard for their ages. Therefore, $\ln fs_n$ is redefined as follows:

$$(2.11) \quad \ln fs_n := \sum_{j=1}^{I_n} \omega_j f(a_j),$$

where I_n is the number of persons in family n; $\omega_1 = 1$ and

$$(2.12) \quad \omega_j := \ln(j/(j-1)), \qquad\qquad j = 2, \ldots, I_n;$$

$$(2.13) \quad f(a_j) := 1, \qquad\qquad a_j > 18$$

$$f(a_j) := 1 + \gamma_2(18-a_j)^2 + \gamma_3(18-a_j)^2(36+a_j) \qquad 0 \leq a_j \leq 18,$$

so that the age function is a third degree polynomial for $a_j \leq 18$, with $f(18)=1$, $f'(18)=0$; γ_2 and γ_3 are parameters, which have to be estimated. Thus, the logarithmic weighting of family members has been retained but in addition children under 18 are also weighted on the basis of their age. Second, both (2.9) and (2.10) specify the cost of an increase in family

size as a fixed percentage of income irrespective of the level of this income (see below). For example, if a baby costs 15% of household income at a very low income level, then (2.9) and (2.10) imply that it will also cost 15% of household income at a much higher income level. A simple way to relax this restriction is to replace β_1 by $\beta_1 + \delta \ln y_n$ and α_1 by $\alpha_1 + \psi \ln y_n$, as will become clear shortly. If the same adjustment is carried through for the family size in the reference group, this yields

$$(2.13) \quad \mu_n = \beta_0 + \beta_1(1-\beta_2)\ln fs_n + \delta(1-\beta_2)\ln fs_n \ln y_n + \beta_2 \ln y_n$$

$$+ \beta_3 m_n - \beta_3 \beta_1 hs_n - \beta_3 \delta m_n hs_n + \varepsilon_n.$$

with $\ln fs_n$ defined by (2.11). The expression for $\ln y_{min,n}$ is analogous:

$$(2.14) \quad \ln y_{min,n} = \alpha_0 + \alpha_1(1-\alpha_2)\ln fs_n + \psi(1-\alpha_2)\ln fs_n \ln y_n + \alpha_2 \ln y_n$$

$$+ \alpha_3 m_n - \alpha_3 \alpha_1 hs_n - \alpha_3 \psi m_n hs_n + u_n$$

Once the parameters in (2.13) and (2.14) are known, we have our operational version of the LPL and the SPL. With respect to the LPL we note that the poverty line corresponding to a welfare level α is given by (substitute (2.13) in (2.7)):

$$(2.15) \quad LPL(\alpha) = \exp \frac{\beta_0 + \beta_1(1-\beta_2)\ln fs_n + \beta_3 m_n - \beta_3 \delta \, m_n hs_n - \beta_3 \beta_1 hs_n + \bar{\sigma}u_\alpha}{(1-\beta_2)(1-\delta \ln fs_n)}.$$

The SPL is given by (substitute (2.14) in (2.2)):

$$(2.16) \quad SPL = \exp \frac{\alpha_0 + \alpha_1(1-\alpha_2)\ln fs_n + \alpha_3 m_n - \alpha_3 \psi \, m_n hs_n - \alpha_3 \alpha_1 hs_n}{(1-\alpha_2)(1-\psi \ln fs_n)}$$

The error terms ε_n and u_n have been ignored in the derivation of (2.15) and (2.16).

Both poverty lines depend on the family composition and on the distribution of incomes and family compositions in a family's reference group. If δ and ψ are equal to zero, it is clear from (2.15) and (2.16) that both poverty lines increase proportionally with (redefined) family size. For the LPL this is restrictive for the following reason.

The dependence of the LPL on family size defines equivalence scales (cf. e.g., Kapteyn and Van Praag [1976, 1980]), which tell us how family income has to vary with family composition in order to keep the family at welfare level α. Proportionality implies that in order to compensate for the birth of a baby, for example, it takes the same percentage of family income for both a very high welfare level α (and hence a high income prior to the birth of the baby) and a very low welfare level α (with a corresponding low income prior to the birth of the baby). It is more likely, however, that the cost of a baby for a richer family is a higher amount but a smaller proportion of income, than for a poorer family. For $\delta < 0$, this is generally what the model implies (see Section 5).

2.2. Preliminary empirical results

Models (2.13) and (2.14) have been estimated for a sample of 773 households taken from the Dutch population in 1982. This so-called labor mobility survey only samples families whose head is under 65. In other respects the sample is random. Table 2.1. presents some of the estimation results that have been obtained with the LISREL-V maximum likelihood computer program.

Table 2.1. <u>First estimation results</u> (standard errors in parentheses)[a]

equation (2.13)	equation (2.14)
$\hat{\beta}_1$ = 31.07 (22.05)	$\hat{\alpha}_1$ = 18.64 (11.33)
$\hat{\beta}_2$ = 0.89 (0.06)	$\hat{\alpha}_2$ = 0.85 (0.07)
$\hat{\delta}$ = -2.98 (2.11)	$\hat{\psi}$ = -1.79 (1.03)

$$\gamma_2 = 10.11 * 10^{-2}(2.17 * 10^{-2})$$

$$\gamma_3 = -2.77 * 10^{-3}(0.55 * 10^{-3})$$

Number of observations: 773

a) Only those results are reported which are relevant for the analysis in subsequent sections.

Equations (2.13) and (2.14) have been estimated jointly, allowing for correlation between ε_n and u_n. A likelihood ratio test of the restriction that the age function parameters γ_2 and γ_3 are identical in both the μ- and y_{min}-equation did not lead to rejection. This equality was therefore imposed. Equality of the remaining parameters across the two equations was rejected at the 5%-level by a likelihood ratio test.

The large standard errors of $\hat{\beta}_1$, δ, $\hat{\alpha}_1$, ψ are not necessarily disturbing because one should consider the effects of $\hat{\beta}_1$ and $\hat{\delta}$ (and of $\hat{\alpha}_1$ and ψ) jointly. A closer inspection of the estimated values of β_1 and δ reveals an anomaly, however. For large values of the welfare level α ($\geq$ 0.7) and with all variables in (2.15) set at their sample mean we find that an increase in family size leads to a <u>reduction</u> in the poverty line. More generally, the effects of family size on the cost of living of a family turn out to be quite small. This result has been obtained more often and it has been criticized as being implausible (e.g., by Watts [1985]). This problem will be addressed in the next section.

3.   Measurement of Income

A central notion emerging from the previous Sections is that an in-
dividual's evaluation of income levels strongly depends on the level of his
own income. A complicating factor in this context is that people in general
only know approximately the level of their actual income. When answering the
income evaluation question and the minimum income question, the respondent
will take his estimate of his actual income as a frame of reference. Due to
the fact that in our survey household income is measured twice, however, we
are able to determine the biasing effects of the respondents' systematic er-
rors in "estimating" their own income.

The income evaluation question in the questionnaire is preceded by a
question which asks the respondent to indicate in which one of seven income
brackets net household income falls. It is likely that the income level the
respondent has in mind when answering this question is also the income level
he refers to when answering the subsequent income evaluation questions. The
last section of the questionnaire asks the respondent to provide detailed
information on a large number of different components of the household's net
income, such as earned income, fringe benefits, family allowance, spouse's
income, etc.. The aggregate of these components is likely to provide a much
more reliable measure of the household's income. A comparison of the two in-
come measures suggests that respondents tend to underestimate household
income when answering the income class question; see Table 3.1.

Table 3.1. Comparison of two income measures [a]

income bracket	[b]
< 17,500	8,769
17,500-20,000	23,918
20,000-24,000	28,552
24,000-28,000	30,743
28,000-34,000	36,086
34,000-43,000	41,656
> 43,000	59,189

a) Dfl. per year
b) Average income of all households in the income bracket according to the
second income measure

In order to analyze the systematic difference between the two income measures, we postulate the following relation between the income y_n^* underlying the answer to the income question in brackets and the income components y_{ni} $(i=1,..,m)$ recorded at the end of the questionnaire:

$$(3.1) \quad y_n^* = (\sum_{i=1}^{m} \alpha_i y_{ni}) \cdot e^{\eta_n}$$

where the α_i's are unknown parameters and η_n is a normally distributed error term with zero mean and variance σ_n^2. The n-th respondent's answer falls into the i-th bracket if y^* is between the upper and lower bound of this bracket. The values of α_i are expected to lie in the unit interval. The smaller a parameter α_i, the more the respondent "forgets" the i-th income component when answering the income question in brackets.

The parameters α_i and σ_n^2 can be estimated by means of maximum likelihood. The likelihood function is given by

$$(3.2) \quad L(\alpha_1,..,\alpha_n,\sigma_n^2) = \prod_{n \in \theta_1} \Phi\left(\frac{\log a_1 - \log(\Sigma \alpha_i y_{ni})}{\sigma_n}\right)$$

$$\prod_{n \in \theta_2} \Phi\left(\frac{\log a_k - \log(\Sigma \alpha_i y_{ni})}{\sigma_n}\right) - \Phi\left(\frac{\log a_{k-1} - \log(\Sigma \alpha_i y_{ni})}{\sigma_n}\right)$$

$$\prod_{n \in \theta_3} 1 - \Phi\left(\frac{\log a_6 - \log(\Sigma \alpha_i y_{ni})}{\sigma_n}\right)$$

where

$n\epsilon\theta_1$:   if the n-th household's income falls into the first income
bracket

$n\epsilon\theta_2$:   if the n-th household's income falls into the k-th income bracket
$(k=2,\ldots,6)$

$n\epsilon\theta_3$:   if the n-th household's income falls into the seventh income
bracket

a_k :   upper bound of the k-th income bracket

$\Phi(.)$ standard normal distribution function

and y_{ni} denotes the i-th income component of the n-th household.

The results of the maximum likelihood estimation of the parameters in model
(3.1) are presented in Table 3.2.

Table 3.2. <u>Estimation results equation (3.1)</u>

income component	α_i	standard error
earned income of respondent from main occupation	0.99	0.01
holiday allowance plus fringes of respondent	0.60	0.11
rent subsidy plus income from subletting rooms	0.06	0.14
property income plus income from secondary jobs	0.48	0.08
family allowance	0.17	0.09
income of the spouse	0.77	0.04
income of oldest child of respondent	0.21	0.05
income of other household members	0.29	0.08

Number of observations: 811

Apparently, only the respondent's own earned income is fully taken into account when answering the income question in brackets. All other components are to some extent "forgotten", especially rent subsidies and family allowances.

We assume that if the respondent would have been aware of the actual value of his household's income as measured by the sum of the income components, the resulting values for e^μ and y_{min} would have been higher by the

same percentage as by which $\sum_{i=1}^{m} y_i$ exceeds $\sum_{i=1}^{m} \alpha_i y_i$. Thus, we adjust the

measured values of e^{μ} and y_{min} for each respondent as follows:

$$(3.3) \quad e^{\hat{\mu}_n} = e^{\mu_n} (\sum_{i=1}^{m} y_{ni} / \sum_{i=1}^{m} \alpha_i y_{ni})$$

$$(3.4) \quad \hat{y}_{min,n} = y_{min,n} (\sum_{i=1}^{m} y_{ni} / \sum_{i=1}^{m} \alpha_i y_{ni})$$

where $e^{\hat{\mu}_n}$ and $\hat{y}_{min,n}$ are the adjusted values for household n.

Subsequently, models (2.13)-(2.14) has been reestimated with μ_n and

$\ln y_{min,n}$ replaced by μ^* and y_{min}^*. Table 3.3 gives some of the results.

Table 3.3. Estimation results after adjustment of μ and $\ln y_{min}$
    (standard errors in parentheses)

equation (2.13)	equation (2.14)
$\hat{\beta}_1 = 1.94$ (1.10)	$\hat{\alpha}_1 = 0.40$ (0.75)
$\hat{\beta}_2 = 0.57$ (0.06)	$\hat{\alpha}_2 = 0.44$ (0.07)
$\hat{\delta} = -0.15$ (0.10)	$\Phi = 0.002$ (0.07)

$$\tau_2 = 0.27*10^{-2}(1.17*10^{-2})$$

$$\tau_3 = -0.12*10^{-3}(0.30*10^{-3})$$

Number of observations: 773

The earlier observed anomaly has disappeared. For a very wide range of incomes, the estimates of δ and β_1 imply an increase of the cost of living when family size increases. These effects are presented in greater detail in Section 5. Furthermore, it appears that Ψ is not significantly different from zero. Imposition of the constraint $\Psi=0$, yields the estimation results reported in Table 3.4. (Since the results of this model will be used for the construction of poverty lines, in Section 5, they are presented in more detail than the previous ones). The estimates in Tables 3.3 and 3.4 are almost identical. A discussion of the economic interpretation of the estimates is deferred to Section 5.

Table 3.4. Estimation results with $\Psi=0$
(standard errors in parentheses)

equation (2.13)   equation (2.14)

$\hat{\beta}_0 = 2.72$ (0.41) $\hat{\alpha}_0 = 3.16$ (0.39)

$\hat{\beta}_1 = 1.92$ (0.78) $\hat{\alpha}_1 = 0.43$ (0.04)

$\hat{\beta}_2 = 0.58$ (0.05) $\hat{\alpha}_2 = 0.45$ (0.03)

$\hat{\delta} = -0.15$ (0.07)

$\hat{\beta}_3 = 0.13$ (0.04) $\hat{\alpha}_3 = 0.19$ (0.05)

$\Upsilon_2 = 0.24 \ast 10^{-2} (1.18 \ast 10^{-2})$

$\Upsilon_3 = 0.12 \ast 10^{-3} (0.3 \ast 10^{-3})$

$\hat{\sigma}_\varepsilon = 0.22$ (0.0056) $\hat{\sigma}_u = 0.25$ (0.0065)

$\hat{\sigma}_{\varepsilon u} = 0.04$ (0.002)

$R^2 = 0.59$ $R^2 = 0.56$

Number of observations: 773

4. Sample Selectivity

Almost any empirical work based on micro-data is confronted with the problem that a number of observations in the available sample cannot be used in the analysis, because of missing information on one or more variables. Usually, these observations are simply left out in the hope that the omitted observations are more or less "random" with respect to the analysis. If the number of deleted observations is relatively large, however, and the dropout is connected with the endogenous variables of the model, the results might be subject to selection bias.

To analyze this problem in more detail, we assume that the process by which observations are removed from the sample can be described by a selection equation of the form:

$$(4.1) \quad y_{3n}^* = X'_{3n}\eta + v_n$$

The household is removed from the sample if $y_{3n}^* \geq 0$ and retained otherwise. The variables in X_{3n} are thought to affect the selection probability of the i-th household[1]. We assume that ε_n, u_n and v_n follow a multivariate normal distribution with zero mean and covariance matrix

$$(4.2) \quad \Omega = \begin{pmatrix} \sigma_\varepsilon^2 & * & * \\ \sigma_{\varepsilon u} & \sigma_v^2 & * \\ \sigma_{\varepsilon v} & \sigma_{uv} & 1 \end{pmatrix}$$

1) We assume that X'_{3n} is observed for all households.

If the researcher ignores the possibility of selection bias, he simply estimates (4.1) and (4.2) under the assumption that $\hat{E}\varepsilon_n = Eu_n = 0$. The expectation of the error terms in (2.13) and (2.14), however, has to be taken conditionally on the household being retained (selected) in the sample, that is

$$(4.3) \quad E(\varepsilon_n | y^*_{3n} < 0) = E(\varepsilon_n | v_n < -X'_{3n}\eta)$$

$$(4.4) \quad E(u_n | y^*_{3n} < 0) = E(u_n | v_n < -X'_{3n}\eta)$$

These expectations are generally unequal to zero. Consequently, β_1 and β_2 will be estimated inconsistently.

We will estimate the selection equation (4.1) jointly with (2.13) and (2.14) and test whether selection bias is present, which is equivalent to testing the null hypothesis $\sigma_{\varepsilon v} = \sigma_{uv} = 0$. We will first give a verbal description of the selection process.

In the analysis of Section 2 and 3 a fairly large proportion of the observations could not be used due to deficient information with respect to one of more variables. In 122 cases the respondent did not answer the MIQ or provided insufficient information to estimate μ, whereas in 145 cases net household income could not be calculated due to insufficient information on one or more income components. We assume that the selection of these 267 households can be described by a single selection equation. In addition, 18 observations had to be left out because of various other deficiencies. Since this latter selection pertains to a small number of observations only, we ignore the possible selection bias that may result.

In order to estimate equations (4.1), (2.13) and (2.14), we have to choose a set of variables which are thought to affect the selection probability:

- constant term
- log age of respondent
- squared log age of respondent
- dummy variable = 1 if the respondent is an employed wage earner
  = 0 otherwise
- dummy variable = 1 if the respondent is a higher executive of a firm
  = 0 otherwise

The likelihood function of the observations for model (2.13), (2.14), (4.1) is given by:

$$(4.6) \quad L = \prod_{n \in \theta_0} \int_{-\infty}^{0} f_n(y_1, y_2, y_3) dy_3 \cdot \prod_{n \in \theta_1} \int_{0}^{\infty} f_{3n}(y_3) dy_3$$

where $n \in \theta_0$ if the household is retained in the sample,

$n \in \theta_1$ if the household is removed from the sample,

$f_n(.)$ is the joint normal distribution function of (y_1, y_2, y_3) and $f_{3n}(.)$ is the marginal distribution function of y_3, for the n-th household. The estimation results are presented in Table 4.1. For comparison we also show the estimates from Section 3.

Table 4.1. <u>Results of joint estimation of equations (2.13), (2.14), (4.1)</u>

parameter	estimate	standard error	estimate from Table 3.4	standard error
β_0	2.86	0.41	2.72	0.41
β_1	1.68	0.69	1.92	0.78
β_2	0.55	0.05	0.58	0.05
δ	-0.13	0.06	-0.15	0.07
β_3	0.14	0.04	0.13	0.04
γ_2	$0.185*10^{-2}$	0.012	$0.235*10^{-2}$	0.012
γ_3	$-0.80*10^{-4}$	0.0003	$-0.117*10^{-3}$	0.0003
α_0	3.23	0.39	3.16	0.39
α_1	0.39	0.04	0.43	0.04
α_2	0.43	0.03	0.45	0.03
α_3	0.20	0.05	0.19	0.05
η_0	-8.91	5.07		
η_1	4.18	2.81		
η_2	-0.52	0.38		
η_3	-0.018	0.09		
η_4	0.88	0.27		
σ_ϵ	0.27	0.009	0.22	0.006
σ_u	0.28	0.010	0.25	0.007
$\sigma_{\epsilon u}$	0.06	0.005	0.04	0.002
$\sigma_{\epsilon v}$	0.24	0.012		
σ_{uv}	0.18	0.025		

Although the differences between the first two and the last two columns in Table 4.1 are generally small, a likelihood ratio test strongly rejects the null hypothesis $\sigma_{\epsilon v}=\sigma_{uv}=0$.

Furthermore, it appears that the probability of being removed from the sample increases with age. For example, the probability for a 25 year old respondent is 0.11 on average, whereas it rises to 0.48 for a 55 year old respondent. The probability of being removed from the sample also increases if the respondent is a higher executive.

5. Poverty Lines

To obtain a better feeling for the importance of the issues dealt with
in the previous Sections we compare the poverty lines implied by the three
models of Sections 2, 3, and 4. First the age functions $f(a_j)$ for each of
the three models are considered. These have been drawn in Fig. 5.1.

The age function for the model of Section 2, illustrates the implausible
results obtained on the basis of the incorrect income measures. The age
functions for both other models do not differ widely, and both look
plausible. The preferred model, with correction for sample selectivity,
shows somewhat less steep age effects than the model of Section 3.

weight rela-
tive to adult
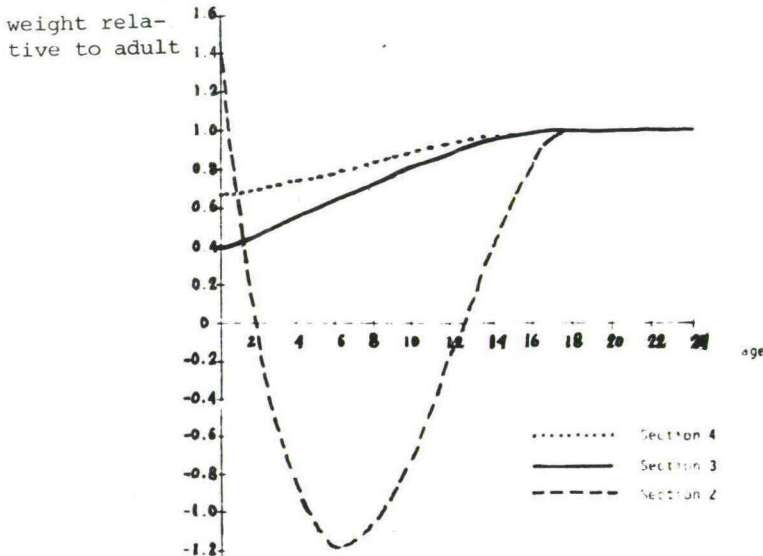
Figure 5.1 **Age functions for three models**

Next we illustrate how the cost of an extra child varies with income and
family size. Fig. 5.2 presents a few examples, based on the estimates of the
third model.

cost

I: 2 adults, children 14 and 11
II: 2 adults, children 13,16 14 and 11
III:2 adults, children 4 and 2
IV:2 adults, children 14,11,4 and 2

net household income (Dfl.x1000)
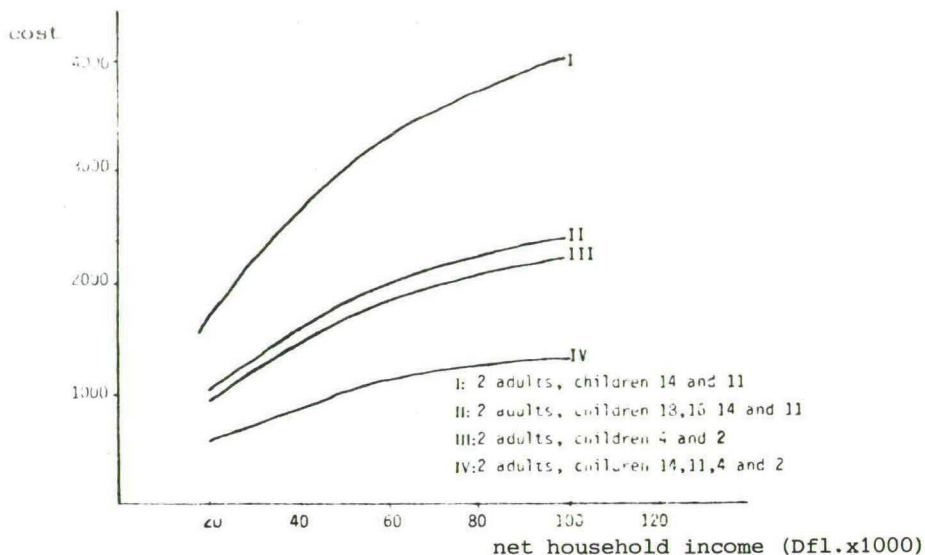
Figure 5.2   Cost of an additional child as a function of income, for various family compositions

Fig. 5.2 shows that the cost of the extra child, that is the youngest child in the family, increases with income but less than proportionally. Furthermore, there are substantial economies of scale. For instance, the two-year old child costs considerably less in the six person family than in the four person family (see cases III and IV). Older children cost more than young children. Despite the economies of scale due to a large family, the eleven-year old child in case II costs more than the two-year old child in case III.

Finally, we have computed poverty lines for some selected family compositions based on the three models for the SPL and two versions of the LPL ($\alpha=0.4$ and $\alpha=0.5$) (see Table 5). In the calculation of the poverty lines both m_n and hs_n have been set equal to the sample means of log-income and log-family size, respectively. The standard errors in parentheses are based on an asymptotic approximation, which follows straightforwardly from the fact that the poverty lines are differentiable functions of the parameters. For comparison, the last column of Table 5.1. gives the money amounts that correspond to the official poverty line in The Netherlands at the time of the survey.

Table 5.1. <u>Poverty lines for various family compositions, according to three</u>
<u>models</u> (Dfl. x 100)

Family composition	model of Section 2			model of Section 3 (correction for income measurement errors)		
	SPL	LPL(0.4)	LPL(0.5)	SPL	LPL(0.4)	LPL(0.5)
1 adult	116(1)	137(9)	167(9)	130(6)	153(7)	180(8)
2 adults	160(6)	185(5)	211(5)	175(7)	202(6)	235(6)
2 adults+6	130(9)	153(9)	182(9)	195(7)	221(6)	258(7)
2 adults+12	157(7)	182(6)	208(6)	204(7)	229(7)	265(7)
2 adults+12,6	136(11)	159(11)	188(10)	220(9)	243(8)	280(8)
2 adults+12,6,1	144(10)	168(1)	196(10)	229(9)	251(8)	289(9)
2 adults+18,12,6,1	169(8)	195(7)	220(7)	254(11)	270(9)	310(10)
2 adults+12,6,2,1,	141(13)	165(14)	193(13)	238(12)	258(10)	296(11)

Table 5.1. <u>continued</u>

Family composition	model of Section 4 (correction for sample selection bias)			official poverty line
	SPL	LPL(0.4)	LPL(0.5)	
1 adult	117(6)	132(7)	154(7)	132
2 adults	154(6)	171(6)	198(7)	189
2 adults+6	174(7)	190(7)	219(7)	199
2 adults+12	178(7)	194(7)	223(7)	199
2 adults+12,6	195(8)	208(8)	238(8)	217
2 adults+12,6,1	207(9)	218(9)	249(9)	235
2 adults+18,12,6,1	225(10)	232(10)	265(10)	257
2 adults+12,6,2,1	217(11)	226(10)	258(10)	257

The official poverty line exceeds the amounts based on the model of
Section 2. This model also suggests substantial economies of scale for large
families, that is the estimated poverty lines increase very slowly with an
increase in family size and sometimes even decrease (compare, for instance,
two adults and two adults with a six-year old child). These findings are
typical for the results that have been obtained in earlier studies. See, for
example, Goedhart et al. [1977], Colasanto, Kapteyn, Van der Gaag [1984].
(In these studies a negative weight for certain ages was not observed,
simply because no weighing on the basis of age took place).

The model of Section 3 yields quite a different picture. The level of
the SPL corresponds roughly with the official poverty line and both versions

of the LPL exceed the official poverty line. The economies of scale for a large family also look more plausible than in the previous model.

The most striking outcome of the model of Section 4 is that the correction for sample selectivity bias leads to a downward revision of approximately 10 percent of the level of the poverty line as compared with the results of the model of Section 3. As a result we see that the official poverty line is generally high enough to allow households to make ends meet. Furthermore, economies of scale are plausible. They are a little larger than implied by the official poverty line.

6. Conclusions

Since income is the central concept in most social security and welfare policies, the implementation of any poverty line definition should be based on an accurate measurement of household income. As the subjective poverty line definitions try to elicit directly which income level is necessary to make ends meet or to guarantee a certain welfare level, it is important that respondents have an accurate knowledge of their income. For the purpose of questionnaire design, this yields two alternatives. The first alternative is the procedure adopted in this paper, where the subjective questions are preceded by a question which measures the respondent's perception of his own after tax household income. On the basis of an accurate measurement of income later in the questionnaire, one can then adjust the response to the subjective questions. A second alternative is to begin with a large number of factual questions about household income, total the components and only then pose the subjective questions.

The finding that sample selectivity creates a significant problem is not surprising. The poor tend to have characteristics that give them a lower probability of participation in surveys. Correction for sample selectivity is important to avoid policies that are aimed at the poor, but which are mainly based on observations from a middle class population.

The poverty lines discussed here, are based on an explicit model of determinants of subjective evaluations. It appears that any poverty line will be based on some implicit or explicit model of behavior or valuation.

To construct an internally consistent or reliable poverty line, the model on which it is based should be correctly specified. The model (2.13)-(2.14) used in this paper is misspecified, because, due to the lack of adequate data, the influence of lagged variables had to be ignored. Hence, the empirical results should not be taken too seriously. The poverty lines that are discussed here require both panel data <u>and</u> a questionnaire design that allows for an accurate treatment of one of the crucial variables in social policy: income.

References

Colasanto,D., A. Kapteyn, and J. van der Gaag. 1984. Two subjective defini-
   tions of poverty: Results from the Wisconsin Basic Needs Study. Journal
   of Human Resources 19: 127-38.

Danziger,S., J. van der Gaag, E. Smolensky, and M. Taussig. 1984. The direct
   measurement of welfare levels: How much does it take to make ends meet?
   Review of Economics and Statistics 66: 500-505.

Goedhart,T., V. Halberstadt, A. Kapteyn, and B.M.S. van Praag. 1977. The
   poverty line: Concept and measurement. Journal of Human Resources 12:
   503-20.

Kapteyn,A., and B.M.S. van Praag. 1976. A new approach to the construction
   of family equivalence scales. European Economic Review 7: 313-35.

Kapteyn,A. 1977. A theory of preference formation. Ph.D.diss., Leyden Uni-
   versity.

Kapteyn,A., and B.M.S. van Praag. 1980. Family composition and family
   welfare. In: J.L. Simon and J. daVanzo (eds.), Research in population
   economics, Vol. II. JAI Press, Greenwich, 77-97.

Kapteyn,A., S. van de Geer, and H. van de Stadt. 1985. The Impact of
   Changes in Income and Family Composition on Subjective Measures of Well-
   being. In: M. David and T. Smeeding (eds.), Horizontal Equity,
   Uncertainty, and Economic Well-Being. University of Chicago Press,
   Chicago, 35-64.

Van de Stadt,H., A. Kapteyn, and S. van de Geer. 1985. The relativity of
   utility: Evidence from panel data. Review of Economics and Statistics
   67: 179-87.

Van Herwaarden,F.G., and A. Kapteyn. 1981. Empirical comparison of the shape of welfare functions. European Economic Review 15: 261-86.

Van Praag,B.M.S., T. Goedhart, and A. Kapteyn. 1980. The poverty line: A pilot survey in Europe. Review of Economics and Statistics 62: 461-65.

Van Praag,B.M.S., A.J.M. Hagenaars, and H. van Weeren. 1982. Poverty in Europe. The Review of Income and Wealth 28: 345-59.

Watts,H.W. 1985. Comment on The Impact of Changes in Income and Family Composition on Subjective Measures of Well-Being. In: M. David and T. Smeeding (eds.), Horizontal Equity, Uncertainty, and Economic Well-Being. Chicago, IL: University of Chicago Press, Chicago, 64-67.

IN 1986 REEDS VERSCHENEN

217 P.M. Kort
The firm's investment policy under a concave adjustment cost function

218 J.P.C. Kleijnen
Decision Support Systems (DSS), en de kleren van de keizer ...

219 T.M. Doup and A.J.J. Talman
A continuous deformation algorithm on the product space of unit simplices

220 T.M. Doup and A.J.J. Talman
The 2-ray algorithm for solving equilibrium problems on the unit simplex

221 Th. van de Klundert, P. Peters
Price Inertia in a Macroeconomic Model of Monopolistic Competition

222 Christian Mulder
Testing Korteweg's rational expectations model for a small open economy

223 A.C. Meijdam, J.E.J. Plasmans
Maximum Likelihood Estimation of Econometric Models with Rational Expectations of Current Endogenous Variables

224 Arie Kapteyn, Peter Kooreman, Arthur van Soest
Non-convex budget sets, institutional constraints and imposition of concavity in a flexible household labor supply model

225 R.J. de Groof
Internationale coördinatie van economische politiek in een twee-regio-twee-sectoren model

226 Arthur van Soest, Peter Kooreman
Comment on 'Microeconometric Demand Systems with Binding Non-Negativity Constraints: The Dual Approach'

227 A.J.J. Talman and Y. Yamamoto
A globally convergent simplicial algorithm for stationary point problems on polytopes

228 Jack P.C. Kleijnen, Peter C.A. Karremans, Wim K. Oortwijn, Willem J.H. van Groenendaal
Jackknifing estimated weighted least squares

229 A.H. van den Elzen and G. van der Laan
A price adjustment for an economy with a block-diagonal pattern

230 M.H.C. Paardekooper
Jacobi-type algorithms for eigenvalues on vector- and parallel computer

231 J.P.C. Kleijnen
Analyzing simulation experiments with common random numbers

IN 1987 REEDS VERSCHENEN