European Economic Review 15 (1981) 261-286. North-Holland Publishing Company

EMPIRICAL COMPARISON OF THE SHAPE OF WELFARE FUNCTIONS*

Floor G. van HERWAARDEN

Centre for Research in Public Economics, Leiden, The Netherlands

Arie KAPTEYN

University of Southern California, Los Angeles, CA 90007, USA

Final version received June 1980

This study brings together data from 8 samples to compare empirically 12 alternative functional forms of welfare functions to the lognormal welfare function proposed by Van Praag (1968). The comparison comprises 11 different wordings in surveys of, in total, about 14,000 respondents. The lognormal function outperforms 11 of the alternative functions in terms of the residual variance criterion, while the logarithm performs slightly better than the lognormal. On the basis of theoretical and practical considerations it is suggested that the lognormal function may be maintained, although further research into the measurement procedure is needed.

1. Introduction

In 1968 Van Praag formulated a theory which assumes that an individual is able to evaluate income levels on a [0, 1]-scale [Van Praag (1968)]. Making some further assumptions he derives that the resulting evaluation $U(z)$ of an income z follows approximately a lognormal distribution function: $U(z)=\Lambda(z;\mu,\sigma)$.

This lognormal welfare function (WF) has been called the individual *welfare function of income* (WFI). Fig. 1 gives some examples. To avoid confusion: the lognormal distribution function has no statistical connotation

*The research reported in the paper is supported by a grant from the Netherlands Organization for the Advancement of Pure Research (ZWO). The data stems partly from surveys of members of the Belgian Consumer Union in 1969, 1970 and 1973 and a survey of members of the Dutch Consumer Union in 1971. These surveys were designed by Bernard M.S. van Praag. Another part of the data comes from a survey designed by Mr. J. Box of Delft University of Technology. Also a survey is used of Dutch municipalities, designed by Joop Linthorst and Bernard van Praag. Finally, part of the data comes from surveys conducted by the Dutch Central Bureau of Statistics (C.B.S.), mainly designed by Theo Goedhart, Bernard van Praag and the second author. Their willingness to make available the data is gratefully acknowledged. We thank Jan Tinbergen, Bernard van Praag, Tom Wansbeek, Roberto Wessels and an anonymous referee for their helpful suggestions, and Rik Storm for programming the complicated computations. The study is part of the Leyden Income Evaluation Project.

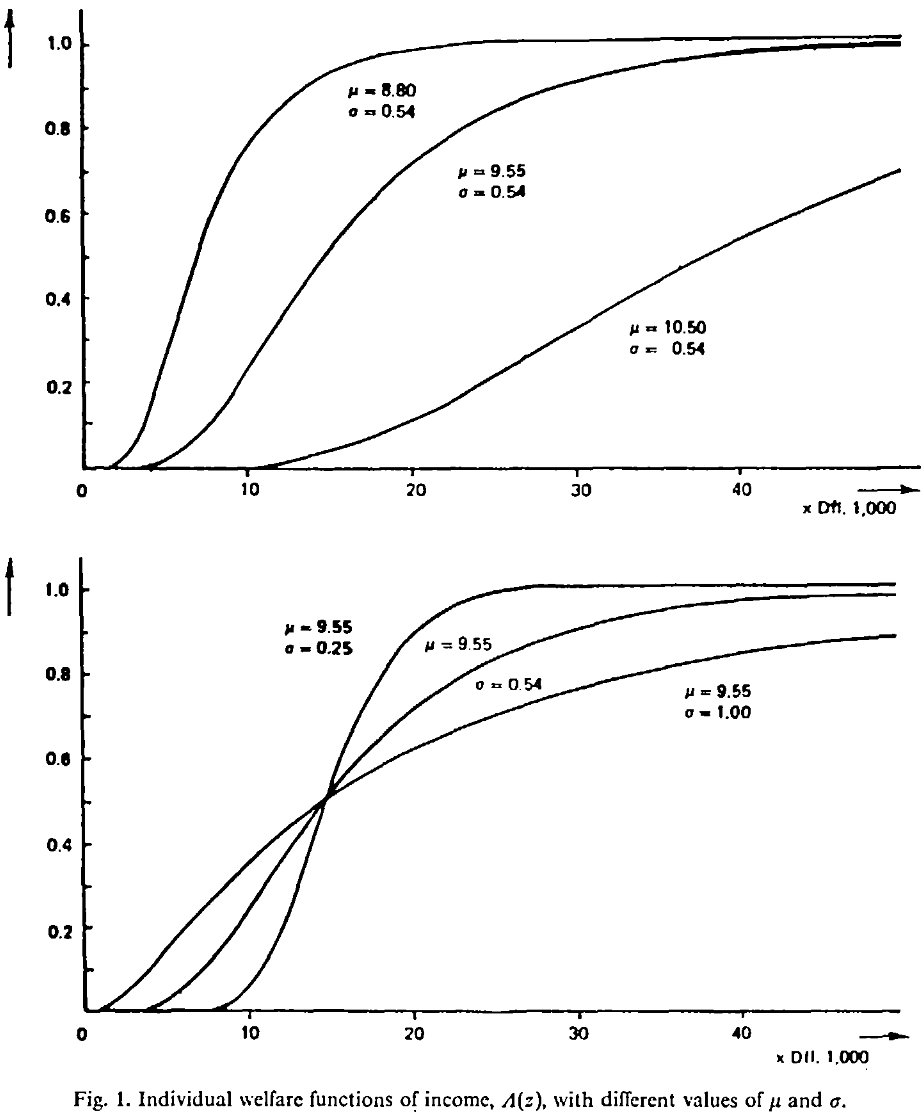

Fig. 1. Individual welfare functions of income, $\Lambda(z)$, with different values of μ and σ.

in the present context. It is merely a mathematical description of how people, supposedly, evaluate income levels on a cardinal, bounded scale.

In the years following the publication of Van Praag's monograph empirical research on the WFI-concept has been carried out. This research corroborates the theory. Hitherto about 14,000 WFIs have been measured and various relationships have been studied between the individual parameters μ and σ on the one hand and socioeconomic characteristics (like income and family size) on the other hand.[1] Gradually, the models used to explain differences in μ and σ between individuals have grown more complex, which has broadened the scope for application of the measured WFs. For instance, tests of the economic theory of consumer behaviour [Kapteyn, Wansbeek and Buyze (1979)], exercises in optimal income distribution [Van Praag (1977, 1978), Kapteyn and Van Herwaarden (1980)], a theory of preference formation [Kapteyn (1977), Kapteyn, Wansbeek and Buyze (1980)] and the analysis of the financial needs of Dutch municipalities [Van Praag and Linthorst (1976)] have been based on the lognormal WF.

The more extensively one uses a certain measuring instrument, the more the instrument itself should be subject to scrutiny. Hence, the necessity was felt to compare other possible functional forms of the WF with the lognormal specification on the basis of the data now available. The present study gives such a comparison. Several of the functional forms we shall investigate have been proposed in the literature, some as early as 1738 [Bernoulli (1738)].

As this paper is exclusively devoted to a statistical comparison of different functional forms we do not give an economic interpretation of the welfare functions, nor do we summarize any of the results obtained in the research so far. For these aspects we refer to the aforementioned papers. Even for a purely statistical analysis, however, one needs a theoretical framework serving as a maintained hypothesis. In particular, we shall assume throughout that welfare functions are bounded, which allows for the normalization of their range to the [0, 1]-interval.

This 'finite bliss, finite agony' assumption is not testable itself by the measurement methods we use to estimate welfare functions (the measurement method is described below). Intuitively, however, it is hard to imagine what, for example, infinite bliss could be, or for that matter how a human being could express feelings of infinite bliss. Words like 'superb' or 'excellent' rather seem to express that the individual cannot imagine to be more delighted about a certain aspect of life and this, according to basic mathematics, entails the boundedness of the experience (even although the individual may express his feelings by claiming to be infinitely happy): If somebody is infinitely happy, it is logically also possible to be twice as happy and this is not what 'superb' or 'fantastic' seem to express.

To be specific: If an individual terms an income of $1,000,000 'excellent' that seems to mean that his evaluation of this income is close to a maximum. Of course, the individual can easily think of an income that is twice as high (*viz.* $2,000,000) but not of an income that would be two times as excellent. Incidentally, if the individual enjoys the $1,000,000 income for a long time, habit formation may lower his evaluation so that *after a while* it does

[1]See e.g. Van Praag (1971), Van Praag and Kapteyn (1973), Van Herwaarden, Kapteyn and Van Praag (1977).

become possible to think of an income that, temporarily, would make him twice as happy. This is the preference drift phenomenon, coined by Van Praag (1971). This phenomenon has to do with changing preferences and does not affect our basic argument.

A different argument for the boundedness of welfare functions can be based upon Menger's super-St. Petersburg paradox [cf. Samuelson (1977) for a discussion].

Along with the boundedness of welfare functions their cardinality is taken for granted. Again, we refer to the publications mentioned for a discussion of the issue. Notice, though, that the mere fact that we are able to discriminate statistically between different functional forms (as will be seen in the sequel) is a strong argument in favour of the cardinal nature of the welfare function concept analysed in this paper.

In the sequel we take the lognormal specification, Λ, as a reference point to which the other functional forms will be compared. So, in fact, we test the null-hypothesis that Λ provides the correct specification of the WF. The criteria to decide whether or not Λ is better than other functional shapes will be set out in section 3. Before that, we give a brief exposé of the measurement procedure for the lognormal WF.

Section 4 describes the data. Section 5 gives the empirical results. The results are discussed in section 6, and section 7 concludes.

2. The measurement procedure

In this section we describe the methods of measurement for three types of WFs, the aforementioned welfare function of income (WFI), the *partial welfare function* (PWF) and the *municipal welfare function* (MWF).

2.1. *Measurement of welfare functions of income*

An individual's WFI is measured by asking him the following question (we filled in income levels as answered by one arbitrarily chosen individual, from a survey of members of the Dutch Consumer Union in 1971):

*Taking into account your own situation with respect to family and job you would call your net-income (including fringe benefits and after subtraction of social security premiums):**

Per week A
month B
year C

Excellent	*if it were above*	Dfl. 45,000,		
Good	*if it were between*	Dfl. 35,000	*nd*	Dfl. 45,000,
Amply sufficient	*if it were between*	Dfl. 30,000	*and*	Dfl. 35,000,

Sufficient	*if it were between*	Dfl. 25,000	*and*	Dfl. 30,000,
Barely sufficient	*if it were between*	Dfl. 22,000	*and*	Dfl. 25,000,
Insufficient	*if it were between*	Dfl. 20,000	*and*	Dfl. 22,000,
Very insufficient	*if it were between*	Dfl. 17,000	*and*	Dfl. 20,000,
Bad	*if it were between*	Dfl. 12,000	*and*	Dfl. 17,000,
Very bad	*if it were below*	Dfl. 12,000.		

*Encircle your reference period.

We call this question the *income evaluation question.* To measure an individual's WFI from his answer to the income evaluation question the verbal qualifications 'excellent', 'good', 'amply sufficient', etc. have to be transformed into numbers in the [0, 1]-interval. This is accomplished by the following reasoning, due to Van Praag (1971):

The amounts inserted in the income evaluation question furnish a division of the income range $[0, \infty)$ into income brackets $[z_0, z_1]$, $(z_1, z_2], \ldots,$ (z_n, z_{n+1}), where $z_0 = 0$ and $z_{n+1} = \infty$. To fix ideas: For the income evaluation question quoted, $n = 8$ whilst the answer can be summarized by: $z_0 = 0$, $z_1 = 12{,}000$, $z_2 = 17{,}000$, $z_3 = 20{,}000$, $z_4 = 22{,}000$, $z_5 = 25{,}000$, $z_6 = 30{,}000$, $z_7 = 35{,}000$, $z_8 = 45{,}000$, $z_9 = \infty$. The division of the income range differs between individuals, but certainly the division is not being made in a random way. There seems to be a general principle behind the fact that extreme brackets tend to be wider than the brackets in the middle.

It is not unreasonable to assume that the individual tries to inform us as *exactly* as possible about his welfare function, i.e., he attempts to maximize the *information value* of his answer. How can we define the information value?

Consider a particular income bracket $(z_j, z_{j+1}]$. The welfare evaluation of an income in this bracket is on the average

$$U(\bar{z}_j) \equiv \tfrac{1}{2}[U(z_j) + U(z_{j+1})], \tag{1}$$

by which equation $\bar{z}_j$ is defined. For example, $U(\bar{z}_5)$ corresponds with 'sufficient', $U(\bar{z}_6)$ with 'amply sufficient'.

However we cannot say that all income levels in $(z_j, z_{j+1}]$ are evaluated by $U(\bar{z}_j)$. The *average inaccuracy* of evaluating the income levels in $(z_j, z_{j+1}]$ by $U(\bar{z}_j)$ may be measured by a quadratic loss function,

$$\int [U(z) - U(\bar{z}_j)]^2 \, \mathrm{d}U(z), \qquad z_j < z \leqq z_{j+1}. \tag{2}$$

When we have a partition $[0, z_1]$, $(z_1, z_2], \ldots,$ (z_n, ∞) the *total average inaccuracy* of this partition is defined by

$$\sum_{j=0}^{n} \int [U(z) - U(\bar{z}_j)]^2 \, \mathrm{d}U(z), \qquad z_j < z \leqq z_{j+1}. \tag{3}$$

The separate integrals increase with the variation of the U-function on $(z_j, z_{j+1}]$ and with the interval length $(z_{j+1}-z_j)$. Hence, the individual selects narrow brackets where the U-function is steep, and wide brackets where U increases slowly. Mathematically, the individual attempts to choose the z_j-values in such a way that (3) is minimized. Applying the transformation $x = U(z)$ we replace minimization of (3) by the problem

$$\min_{x_1 \ldots x_n} \sum_{j=0}^{n} \int_{x_j}^{x_{j+1}} [x-\bar{x}_j]^2 \, \mathrm{d}x, \tag{4}$$

where $x_j = U(z_j)$, $\bar{x}_j = \frac{1}{2}(x_j + x_{j+1})$, $x_0 = 0$, and $x_{n+1} = 1$.

Integration of (4) yields

$$\min_{x_1 \ldots x_n} \tfrac{1}{12} \sum_{j=0}^{n} (x_{j+1}-x_j)^3, \tag{5}$$

setting $p_j = x_{j+1} - x_j$ we have $\sum_{j=0}^{n} p_j = 1$. So the problem reduces to

$$\min_{p_0 \ldots p_n} \sum_{j=0}^{n} p_j^3 \quad \text{subject to} \quad \sum_{j=0}^{n} p_j = 1. \tag{6}$$

The solution is $p_j = 1/(n+1)$, which implies $x_j = j/(n+1)$ and

$$U(z_j) = j/(n+1). \tag{7}$$

In words, the result can be stated as: *The individual partitions the income range according to equal quantiles of the welfare function.* In the wording quoted the income evaluation question leaves room for nine brackets, so z_j is the jth 11.1%-quantile of the distribution defined by the distribution function U, $j = 1, \ldots, 8$.

The definition of the average inaccuracy by (2) contains an element of arbitrariness. It can be shown that if one replaces $[U(z) - U(\bar{z}_j)]^2$ by any other differentiable function monotonically increasing in the absolute value of $[U(z) - U(\bar{z}_j)]$, one gets the same solution (7). Moreover the notion of an average used in (1) can be generalized while retaining the result [cf. Kapteyn (1977, app. 3A)].

By the method described we have found for the individual a sequence of points $\{(z_j, U(z_j))\}_{j=1}^{n}$ which have to be on the graph of his WFI.

If the points $\{(z_j, U(z_j))\}_{j=1}^{n}$ were points of the graph of a distribution function $\Lambda(z; \mu, \sigma)$, there would hold

$$U(z_j) = N(\ln(z_j); \mu, \sigma) = N((\ln(z_j) - \mu)/\sigma; 0, 1), \tag{8}$$

where N is the normal distribution function.

We know that the logarithms of the z_j's quoted are 11.1%-quantiles, say $w_1, \ldots, w_n$ of the normal distribution, hence there has to hold

$$(\ln(z_j) - \mu)/\sigma = w_j, \tag{9}$$

or

$$\ln(z_j) = \mu + \sigma w_j. \tag{10}$$

It stands to reason that an individual's answers will not strictly satisfy (10), but we may assume that (10) holds approximately; we estimate μ and σ from the linear model

$$\ln(z_j) = \mu + \sigma w_j + \varepsilon_j, \qquad j = 1, \ldots, n, \tag{11}$$

where ε_j is an *i.i.d.* random disturbance term, with expectation zero and variance σ_ε^2.

Applying ordinary least squares to the n observations $(\ln(z_j), w_j)$ we obtain estimates for μ and σ. (For the answer quoted above the estimates are: $\mu = 10.08, \sigma = 0.52$.) If the individual has not inserted all answers but has omitted say, the first and the third, we have still $(n-2)$ observations (z_2, w_2), $(z_4, w_4), \ldots, (z_n, w_n)$ to which we may apply the regression. Only the one- and two-point answers are excluded.

2.2. *Measurement of partial welfare functions*

An individual's *partial welfare function* (PWF) with respect to a certain commodity group describes how he evaluates expenditures on that commodity group. If the commodity group is 'broad' enough, i.e., a large number of characteristics can be distinguished in it, then Van Praag's theory predicts that the evaluation will approximately follow a lognormal distribution function [Van Praag (1968)]. A commodity group may just comprise a single good.

An individual's PWFs are measured by asking him questions like the following one (we filled in the answer of one arbitrarily chosen individual from the aforementioned survey of members of the Dutch Consumer Union in 1971):

Many people think there is always a connection between price and quality. For example one person expects an armchair to suit him very badly if he pays only Dfl. 100 for it, badly if he pays Dfl. 150 for it, moderately if he pays Dfl. 200 for it, reasonably if he pays Dfl. 400 for it, well if he pays Dfl. 650 for it and perfectly if he is to pay Dfl. 800 or more for it.

Another person may have quite a different opinion and have other prices in mind.

To learn about your opinion, we should like you to mention the amounts of money you have in mind when you think of the articles you plan to buy in the near future. Please mention an amount of money in each row.

What durables you may buy in the near future? DRILL

I strongly suppose the purchase
will not suit me at all *if I would pay about* Dfl. 50,
will not suit me *if I would pay about* Dfl. 80,
will suit me moderately *if I would pay about* Dfl. 140,
will suit me reasonably *if I would pay about* Dfl. 160,
will suit me well *if I would pay about* Dfl. 180,
will suit me perfectly *if I would pay about* Dfl. 240.

We call this question the *partial evaluation question.*

PWFs are measured analogously to WFIs. Rather than asking the individual to divide the range $[0, \infty)$ in a number of intervals $[z_0, z_1]$, $(z_1, z_2]$ etc., we now ask for the midths of these intervals $\bar{z}_0$, $\bar{z}_1$ etc. This affects the conclusions of the information maximization argument only in that $z_0, z_1, \ldots, z_j, \ldots, z_m$ $(m=6)$ do not correspond to the $j/(m+1)$ quantiles but to the $(j-\frac{1}{2})/m$ quantiles.

Given this modification we arrive again at a regression model which reads for the ith commodity group

$$\ln(\bar{z}_j) = \mu_i + \sigma_i \bar{w}_j + \varepsilon_j, \qquad j = 1, \ldots, m, \tag{12}$$

where the $\bar{w}_j$ are appropriately defined quantiles of the normal distribution.

2.3. Measurement of municipal welfare functions

A *municipal welfare function* (MWF) describes the evaluation on a $[0, 1]$-scale by local authorities, like alderman or mayor, of municipal outlays on certain expenditure categories. Van Praag's theory suggests that also these evaluations may be expected to follow a lognormal distribution function. A MWF is measured by asking the authority concerned a *municipal evaluation question.*

As an illustration we present an example given by Van Praag and Linthorst (1976, p. 56) of a municipal evaluation question with respect to the portfolio 'Public Works', answered by an alderman of a Dutch municipality with approximately 28,000 inhabitants.

Taking into account the specific circumstances and needs of your municipality (number of inhabitants, location, etc) you would call the level of welfare as regards public works:

		Expenditure level (×Dfl. 1000)		
Excellent	*if the expenditure level were above*	Dfl. 5,800,		
Good	*if the expenditure level were between*	Dfl. 5,500	*and*	Dfl. 5,800,
Amply sufficient	*if the expenditure level were between*	Dfl. 5,200	*and*	Dfl. 5,500,
Sufficient	*if the expenditure level were between*	Dfl. 5,000	*and*	Dfl. 5,200,
Barely sufficient	*if the expenditure level were between*	Dfl. 4,000	*and*	Dfl. 5,000,
Insufficient	*if the expenditure level were between*	Dfl. 3,800	*and*	Dfl. 4,000,
Very insufficient	*if the expenditure level were between*	Dfl. 3,600	*and*	Dfl. 3,800,
Bad	*if the expenditure level were between*	Dfl. 3,500	*and*	Dfl. 3,600,
Very bad	*if the expenditure level were below*	Dfl. 3,500.		

The measurement of a MWF on the basis of an answer to the municipal evaluation question is analogous to the measurement of WFIs.

2.4. *Differences in wordings of the income evaluation questions*

MWFs and PWFs have been measured in only one survey. On the other hand WFIs have been measured in a number of surveys. Between the surveys the wording of the income evaluation question has varied, mainly because attempts have been made to simplify the respondents' task of answering the income evaluation questions. The main differences are:

(1) The number of income levels to be provided by the respondents was either equal to 8, 6 or 5.
(2) The income evaluation question was worded from 'excellent' to 'very bad' or vice versa.
(3) Instead of asking for intervals (i.e., two income levels per qualification, cf. the income evaluation question cited in subsection 2.1), in some surveys one income level is asked, analogous to the procedure with the partial evaluation question.
(4) In a few surveys some qualifications were underlined, for example 'good' and 'bad'. The respondent was asked to start with providing income levels corresponding to the underlined qualifications.

When describing the data we shall indicate the wording of the evaluation question used in each survey. Since we primarily want to compare Λ to a number of alternative functional forms we shall ignore the distinction between WFI, PWF and MWF. As the particular wording of the evaluation question presumably affects the answers of the respondents, we shall compare Λ to its competitors for each different wording.

3. Selection of alternative functions and the criterion to compare them with Λ

In this section we set out criteria for selecting alternative functions to be compared to Λ. This results in a set of 12 alternative functions. Next we develop a criterion for comparison.

3.1. *Selection of alternative functions*

We restrict the set of alternative functions by requiring that these functions

(a) have 2 parameters,
(b) are monotonically non-decreasing,
(c) are either (i) probability distribution functions on $[0, \infty)$ or (ii) have been used or advocated as WFs in economic research.

Moreover,

(d) estimation of the parameters on the basis of the procedure sketched below should not be excessively costly in terms of computer time.

Table 1 gives a list of the selected functional forms that will be compared to Λ (given in the first row). The meaning of the columns (6) and (7) will be explained in the next subsection. The straight line has been added because it would be conceivable that respondents enter income levels linearly, indicating that the evaluation questions are too difficult to answer. The function proposed by Keller and Hartog (1977) [row (9) in table 1] is derived from the requirement that the elasticity of the relative marginal utility of income is constant: $\partial \ln(m(z))/\partial \ln(z) = \text{constant}$, where $m(z) = \{\partial \ln(U(z))/\partial \ln(z)\}$. One function is conspicuously lacking, i.e., the incomplete Γ function. It has been discarded because this function does not meet requirement (d), i.e., the estimation of its parameters (see the last paragraph of the next subsection) appeared to be prohibitively costly.

3.2. *Criterion for comparison*

Our comparison of the various functions with Λ will be based on Theil's residual variance criterion [Theil (1961, 1971)].

When comparing Λ to other functional forms for the WF we basically compare models like (11) and (12) with alternative models explaining the response sequence $\{z_j\}_{j=1}^{n}$. Hence we specify the alternative models as

$$\ln(z_j) = f(U(z_j); a, b) + \varepsilon_j. \tag{13}$$

Table 1

Selected functional forms.

(1)	(2) Functional form $U(z)=\ldots$	(3) Restrictions[a]	(4) c.d.f.[b]	(5) Concavity[b,c]	(6) f[d] [see (13)]	(7) f linear?	(8) References
(1) Lognormal	$\Lambda(z;\mu,\sigma)$	—	c.d.f.	n.c.	$\mu+\sigma N^{-1}(U;0,1)$	yes	J–K (1970),[e] Aitchison and Brown (1957)
(2) Normal	$N(z;a,b)$	$b>0$	c.d.f.	n.c.	$\ln(a+bN^{-1}(U;0,1))$	no	J–K (1970)
(3) Logarithm	$a+b\ln(z)$	—	—	c.	$(U-a)/b$	yes	Bernoulli (1738), Bartels (1977)
(4) Straight line	$a+bz$	—	—	n.c.	$\ln\{(U-a)/b\}$	no	—
(5) Log-logistic	$\{1+\exp(a+b\ln(z))\}^{-1}$	$b<0$	c.d.f.	n.c.	$\{\ln((1-U)/U)-a\}/b$	yes	J–K (1970)
(6) Logistic	$\{1+\exp(a+bz)\}^{-1}$	$b<0$	c.d.f.	n.c.	$\ln\{\{\ln((1-U)/U)-a\}/b\}$	no	J–K (1970)
(7) Log-hyperbola	$1+b/(a+\ln(z))$	—	—	c.	$-a-b/(1-U)$	yes	Tinbergen[f]
(8) Hyperbola	$1+b/(a+z)$	—	—	c.	$\ln\{-a-b/(1-U)\}$	no	Tinbergen[f]
(9) Keller–Hartog	$\exp\{(a/b)z^b\}$	$b\leqq 0$	c.d.f.	n.c.	$\{\ln((b/a)\ln(U))\}/b$	yes	Keller and Hartog (1977)
(10) Power law[g]	az^b	$a,b>0$	—	c.	$\{\ln(U/a)\}/b$	yes	Stevens (1972)
(11) Pareto	$1-(a/z)^b$	$a,b>0$ $z\geqq a$	c.d.f.	c.	$\ln(a)-\{\ln(1-U)\}/b$	yes	J–K (1970)
(12) Weibull	$1-\exp\{-(z/a)^b\}$	$a,b>0$	c.d.f.	n.c.	$[\ln\{-a^b\ln(1-U)\}]/b$	yes	J–K (1970), Bartels (1970)
(13) Stone–Geary	$b\ln(z-a)$	$z>a$	—	c.	$\ln\{\exp(U/b)+a\}$	no	Philips (1974)

[a]For all functions $z\geqq 0$.
[b]c.d.f. = cumulative distribution function; (n.)c. = (non)concave.
[c]That is: $\partial^2U/\partial z^2<0$.
[d]U is an abbreviation of $U(z_j)$.
[e]J–K (1970) means Johnson and Kotz (1970).
[f]Personal communication with J. Tinbergen.
[g]Also called psychophysical law.

A few observations are in order. First, all evaluation questions provide respondents with verbal qualifications which represent, by assumption, welfare levels. Hence, whatever the functional form, the response z_j has to be seen as an endogenous variable whilst the welfare level $U(z_j)$ is the independent variable. The parameters a and b are unknown. Second, the additive stochastic specification is tantamount to multiplicative response errors. The assumption can be motivated by reference to the Weber–Fechner Law.

Third, the transformation of verbal qualifications into numerical values $U(z_j)$ has been motivated in section 2 by an information-maximization argument. This argument rests upon the boundedness of the range of $U(z_j)$. This creates a problem with the functions 3, 4, 7, 8, 10 and 13 in table 1 as these are not bounded from above and below. We therefore interpret these functions as approximations to some unknown function with range $[0,1]$.[2] This interpretation also entails that $U(z_j)$ in (13) is the same for any functional form f. The values of $U(z_j)$ follow from the argument in section 2.

Our criterion for comparison is, as said, Theil's residual variance criterion. Theil has shown that if a set of rival linear models contains the *true* model, i.e., the disturbances are *i.i.d.* distributed, then the true model will exhibit the lowest disturbance variance. Since in linear models the residual variance is the unbiased OLS-estimator of the disturbance variance, the true model will, on average, exhibit the lowest residual variance. As we estimate models like (11), (12) and (13) are estimated many times (about 25,000), we can also easily determine confidence intervals for the disturbance variances of the competing models and thus, with considerable certainty, choose the true model.

Unfortunately not all functions give rise to a linear specification for model (13) [cf. columns (6) and (7) in table 1]. It can be shown, however, [cf. Kapteyn (1977)] that if the true model is linear then this model will also have lower disturbance variance than other non-linear models.

There is one practical problem left. For the non-linear model the residual variance is a consistent but biased estimator of the disturbance variance. Hence, we make the additional assumption that the bias of the residual variance is sufficiently small, i.e., that the relative ranking of models by residual variance coincides with the relative ranking by disturbance variance. In fact, because Λ will turn out to have lower residual variance than any of its non-linear competitors it suffices to assume that the residual variance is not biased upward.

To apply the residual variance criterion, and because of the motivation given in the first paragraph of this subsection, all models have to be in the

[2]The idea that a functional specification is a local approximation to an unknown utility function is, of course, quite common, especially since the work by Christensen, Jorgenson and Lau [cf., e.g., their (1975) publication].

form (13). This requirement has led us to discarding the incomplete Γ-function. It appears impossible to write the Γ-function in the explicit form (13). Still one can conceive of the z_j as solutions to the implicit equations to which an error term is added. That is, we may assume the z_j to be generated by $\ln(z_j)=\ln(z_j^*)+\varepsilon_j$, where z_j^* is defined by $U_j \equiv \Gamma(z_j^*; a, b)$. In principle it is then still possible to estimate the parameters by non-linear least squares. However, the amount of computer time required for estimation turned out to be excessive. It took about one minute of CPU-time on an IBM 370/158 to estimate a and b for one welfare function. There are about 25,000 welfare functions to be measured.

4. Data

In the empirical analysis evaluation questions are employed from 8 different samples yielding in total some 25,000 answers to either the income evaluation question, or the partial evaluation question or the municipal evaluation question. The samples will be denoted by the country and the year in which the survey was conducted. Information on the samples is given in table 2. A description and further references with respect to the Belgian (1969, 1970, 1973) and the Dutch (1971, 1974a, 1975) samples can be found in Van Herwaarden, Kapteyn and Van Praag (1977). These samples all contain answers to income evaluation questions. The Dutch (1971) sample also contains answers to partial evaluation questions. The ensuing PWFs were measured by Kapteyn, Van Herwaarden and Van Praag (1977). The Dutch (1974b) sample contains answers to municipal evaluation questions. The MWFs were measured by Van Praag and Linthorst (1976). The Dutch (1977) sample contains answers to income evaluation questions and is based on a pilot survey of Dutch citizens aimed at comparison of different wordings of the income evaluation questions.

A classification of the data according to differences in wording of the evaluation questions is given in table 3. In the sequel we shall distinguish the various wordings by referring to the corresponding column number in table 3.

5. Results

In table 4a the average residual variances ($\overline{s^2}$) and their sample standard deviations (in parentheses) are presented for all functional forms that make (13) linear in parameters. The different wordings in tables 3 and 4a have been numbered in such order that the $\overline{s^2}$-values for Λ are descending.

In tables 4b up to 4f inclusive, we successively compare $\overline{s^2}$-values corresponding to Λ with $\overline{s^2}$-values corresponding to one of the non-linearizable functions. One observes that the number of observations vary

Table 2

Dates, sizes, and origins of the samples.

Name of sample	Date of drawing	Size[a]	Drawn from	Way of interviewing
Belgian (1969)	Dec. 1969	2545	Membership of Belgian Consumer Union	Written
Belgian (1970)	Dec. 1970	2293	Membership of Belgian Consumer Union	Written
Belgian (1973)	Dec. 1973	2201	Membership of Belgian Consumer Union	Written
Dutch (1971)	Oct. 1971	2952	Membership of Dutch Consumer Union	Written
Dutch (1974a)	March 1974	878	Both members and non-members of Dutch Consumer Union[b]	Oral[c]
Dutch (1974b)	April 1974	551	Population of 842 Dutch Municipalities	Written
Dutch (1975)	Jan. 1975	1748	Dutch population	Oral[d]
Dutch (1977)	May 1977	574	Dutch population	Oral[d]

[a]Number of respondents who have inserted at least three levels in the evaluation question. The sizes of the Belgian samples are somewhat larger than reported in Van Herwaarden, Kapteyn and Van Praag (1977) (HKP) because there only observations of individuals have been used of whom the family income and family composition is known. The size of the Dutch (1974a) sample is somewhat smaller (41 observations less), because we used a non-screened version of the data-set also used by HKP. We had the choice of either correcting punching errors (as had been done by HKP) or removing the corresponding observations. For technical reasons we decided to remove them.

[b]About 585 respondents are members of the Dutch Consumer Union and about 293 are not. The latter have been chosen in such a way that they exhibit socioeconomic traits similar to the members of the Dutch Consumer Union.

[c]The income evaluation question in this survey has been asked by letting respondents fill in their answers to this question on a card. Afterwards the respondent could insert the card into an envelope, seal the envelope and hand it to the interviewer.

[d]After an oral introduction, the questionnaire with the income evaluation question was left behind with the respondent. The respondent was requested to fill out the questionnaire and to send it back.

between these tables. This is caused by the non-linear nature of (13) for these functions. For each respondent the parameters in (13) are estimated by Marquardt's algorithm [Marquardt (1963)]. This algorithm does not always converge. In view of our sample size it is practically impossible to try new starting-values[3] until convergence is reached. Hence we left out all respondents for whom convergence did not obtain. In order to maintain comparability, the resulting $\overline{s^2}$-values of the non-linear functions are given with the $\overline{s^2}$-values corresponding to Λ of the same respondents. Since convergence problems are most likely to obtain for respondents with high

[3]Starting values were in principle obtained by first fitting Λ, and using the estimated μ and σ to compute the starting values of the parameters of the alternative functions.

Table 3

Differences in the wordings of the evaluation questions.

	Different wordings										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Maximum number[a] of income levels	6	6	6	8	8	6	5	6	8	5	8
Hilo/Lohi[b]	Lohi	Hilo	Hilo	Hilo	Hilo	Lohi	Hilo	Lohi	Hilo	Lohi	Hilo
Interval[c]	no	yes	no	yes	yes	no	no	yes	yes	no	yes
Underlining[d]	no	yes (2,4,6)	yes (2,5)	yes (1,4,8)	no	yes (2,5)	yes (2,5)	yes (2,5)	yes (2,5,8)	yes (2,5)	no
Corresponding data sets	Dutch (1971)	Dutch (1977)	Dutch (1977)	Dutch (1974a)	Belgian (1969, 1970, 1973) Dutch (1971)	Dutch (1977)	Dutch (1977)	Dutch (1977)	Dutch (1975)	Dutch (1977)	Dutch (1974b)
Name of WF[e]	PWF	WFI	WFI	WFI	WFI	WFI	WFI	WFI	WFI	WFI	MWF
Number of respondents	9029	95	99	878	9991	94	90	100	1748	96	2675

[a]See subsection 2.4, point (1).

[b]See subsection 2.4, point (2). Hilo (high low) means that the evaluation question was worded from 'excellent' to 'very bad'. Lohi (low high) means that it was worded from 'very bad' to 'excellent'.

[c]See subsection 2.4, point (3).

[d]Notice that with evaluation questions where intervals are asked (cf. the income evaluation question in section 2) the number of verbal qualifications exceeds the maximum number of income levels to be inserted by the respondent by one. Keeping this in mind, the reader can infer from the numbers in parentheses which qualifications have been underlined. These numbers are assigned to the qualifications from the top to the bottom of the list of verbal qualifications. For organizational reasons [cf. Van Praag (1971, p. 346)] the Walloon questionnaires (approximately 1300) in the Belgian (1969) survey left room for nine income levels. These are included in this column.

[e]WF = welfare function, PWF = partial WF, WFI = WF of income, MWF = municipal WF.

Table 4

Average residual variance, s^2.[a]

	Different wordings										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(a)											
(1) Lognormal	0.0217 (0.0519)	0.0170 (0.0700)	0.0149 (0.0550)	0.0097 (0.0163)	0.0096 (0.0149)	0.0079 (0.0100)	0.0079 (0.0111)	0.0069 (0.0119)	0.0061 (0.0116)	0.0049 (0.0067)	0.0038) (0.0126)
(3) Logarithm	0.0196 (0.0506)	0.0162 (0.0653)	0.0141 (0.0647)	0.0086 (0.0160)	0.0089 (0.0151)	0.0066 (0.0080)	0.0067 (0.0109)	0.0077 (0.0138)	0.0063 (0.0124)	0.0042 (0.0064)	0.0035 (0.0120)
(5) Log-logistic	0.0236 (0.0537)	0.0174 (0.0716)	0.0156 (0.0529)	0.0106 (0.0170)	0.0102 (0.0152)	0.0088 (0.0111)	0.0085 (0.0114)	0.0068 (0.0115)	0.0064 (0.0116)	0.0054 (0.0070)	0.0041 (0.0132)
(7) Log-hyper-bola	0.1662 (0.2060)	0.0445 (0.0848)	0.0810 (0.1644)	0.0698 (0.0786)	0.0575 (0.0680)	0.0648 (0.0630)	0.0470 (0.0455)	0.0194 (0.0187)	0.0437 (0.0542)	0.0399 (0.0328)	0.0291 (0.0755)
(9) Keller–Hartog	0.0362 (0.0759)	0.0143 (0.0532)	0.0228 (0.0823)	0.0146 (0.0228)	0.0129 (0.0214)	0.0111 (0.0135)	0.0095 (0.0106)	0.0052 (0.0071)	0.0085 (0.0158)	0.0066 (0.0078)	0.0065 (0.0215)
(10) Power Law	0.0519 (0.0747)	0.0407 (0.1366)	0.0242 (0.0276)	0.0219 (0.0319)	0.0222 (0.0288)	0.0291 (0.0296)	0.0248 (0.0317)	0.0198 (0.0310)	0.0164 (0.0241)	0.0191 (0.0271)	0.0065 (0.0160)
(11) Pareto	0.0683 (0.1138)	0.0178 (0.0458)	0.0380 (0.1154)	0.0260 (0.0364)	0.0226 (0.0345)	0.0228 (0.0256)	0.0175 (0.0199)	0.0070 (0.0083)	0.0161 (0.0262)	0.0141 (0.0149)	0.0119 (0.0359)
(12) Weibull	0.0275 (0.0517)	0.0255 (0.0976)	0.0156 (0.0331)	0.0121 (0.0198)	0.0128 (0.0181)	0.0144 (0.0170)	0.0132 (0.0197)	0.0114 (0.0197)	0.0086 (0.0152)	0.0092 (0.0147)	0.0038 (0.0108)
Number	9029	95	99	9991	878	94	90	100	1748	96	2675
(b)											
(1) Lognormal	0.0160 (0.0322)	0.0082 (0.0127)	0.0150 (0.0550)	0.0093 (0.0142)	0.0091 (0.0146)	0.0078 (0.0100)	0.0079 (0.0111)	0.0061 (0.0088)	0.0058 (0.0093)	0.0046 (0.0058)	0.0035 (0.0115)
(2) Normal	0.0254 (0.0318)	0.0195 (0.0413)	0.0140 (0.0201)	0.0158 (0.0224)	0.0178 (0.0562)	0.0163 (0.0229)	0.0148 (0.0247)	0.0121 (0.0221)	0.0105 (0.0160)	0.0091 (0,0161)	0.0048 (0.0177)
Number	7776	93	99	873	9838	92	90	99	1738	95	2663

(c)											
(1) Lognormal	0.0158 (0.0330)	0.0063 (0.0076)	0.0150 (0.0550)	0.0084 (0.0125)	0.0082 (0.0129)	0.0078 (0.0100)	0.0073 (0.0096)	0.0056 (0.0072)	0.0054 (0.0086)	0.0042 (0.0045)	0.0028 (0.0073)
(4) Straight line	0.0179 (0.0217)	0.0111 (0.0149)	0.0110 (0.0160)	0.0115 (0.0162)	0.0107 (0.0147)	0.0127 (0.0183)	0.0110 (0.0186)	0.0103 (0.0147)	0.0079 (0.0109)	0.0064 (0.0095)	0.0026 (0.0062)
Number	7672	89	99	849	9379	92	89	98	1710	94	2623
(d)											
(1) Lognormal	0.0168 (0.0338)	0.0082 (0.0127)	0.0150 (0.0550)	0.0094 (0.0144)	0.0093 (0.0148)	0.0078 (0.0100)	0.0079 (0.0111)	0.0061 (0.0088)	0.0058 (0.0094)	0.0049 (0.0067)	0.0037 (0.0123)
(6) Logistic	0.0301 (0.0356)	0.0202 (0.0428)	0.0153 (0.0217)	0.0171 (0.0245)	0.0191 (0.0290)	0.0177 (0.0244)	0.0157 (0.0254)	0.0123 (0.0224)	0.0116 (0.0184)	0.0113 (0.0224)	0.0062 (0.0272)
Number	8033	93	99	873	9840	92	90	99	1743	96	2546
(e)											
(1) Lognormal	0.0210 (0.0486)	0.0151 (0.0678)	0.0150 (0.0550)	0.0095 (0.0148)	0.0096 (0.0156)	0.0080 (0.0100)	0.0079 (0.0111)	0.0069 (0.0119)	0.0060 (0.0114)	0.0049 (0.0067)	0.0036 (0.0121)
(8) Hyperbola	0.0858 (0.0767)	0.0251 (0.0713)	0.0568 (0.0979)	0.0343 (0.0379)	0.0355 (0.0330)	0.0433 (0.0349)	0.0339 (0.0290)	0.0122 (0.0122)	0.0256 (0.0251)	0.0294 (0.0231)	0.0152 (0.0255)
Number	8985	94	99	874	9956	94	90	100	1742	96	2664
(f)											
(1) Lognormal	0.0217 (0.0515)	0.0170 (0.0700)	0.0150 (0.0550)	0.0096 (0.0149)	0.0098 (0.0162)	0.0080 (0.0100)	0.0079 (0.0111)	0.0069 (0.0119)	0.0061 (0.0116)	0.0049 (0.0067)	0.0038 (0.0126)
(13) Stone-Geary	0.1262 (0.1432)	0.0624 (0.0850)	0.0908 (0.1716)	0.0744 (0.0815)	0.0376 (0.0438)	0.0742 (0.0684)	0.0602 (0.0550)	0.0334 (0.0280)	0.0565 (0.0647)	0.0521 (0.0413)	0.0232 (0.0558)
Number	9005	95	99	877	9624	94	90	100	1746	96	2674

[a]Standard deviations in parentheses.

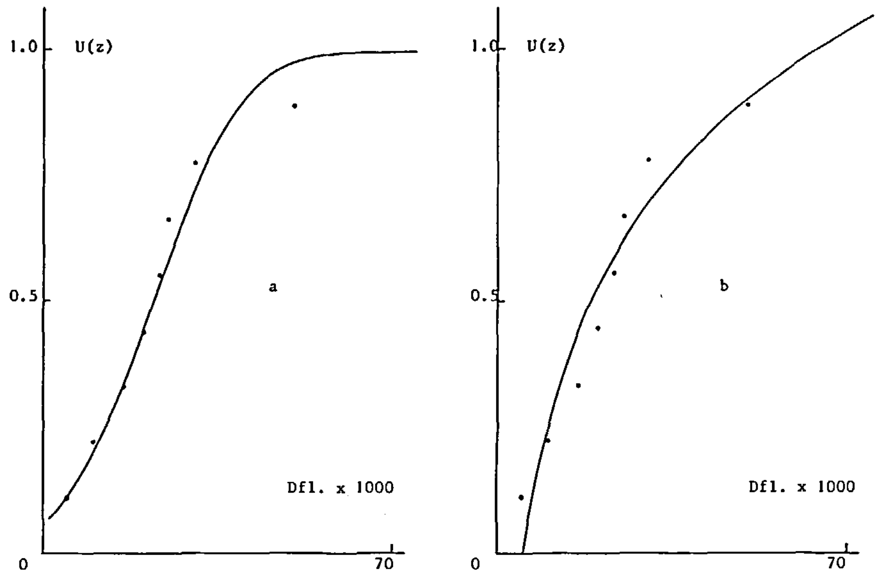

Fig. 2. Thirteen functions on the basis of parameters estimated from the answer quoted in subsection 2.1 and the corresponding scatter. (a) Normal, (b) Logarithm.

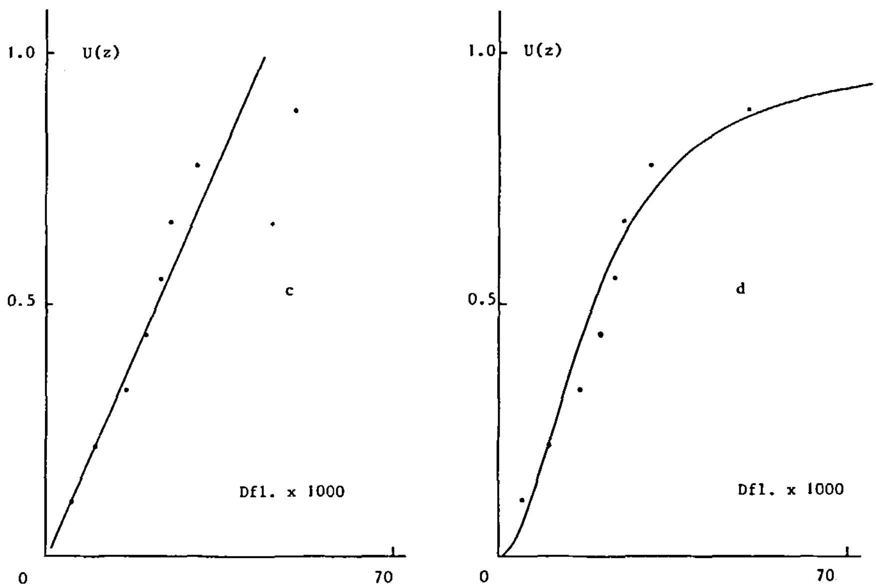

Fig. 2 (continued). (c) Straight line, (d) Log-logistic.

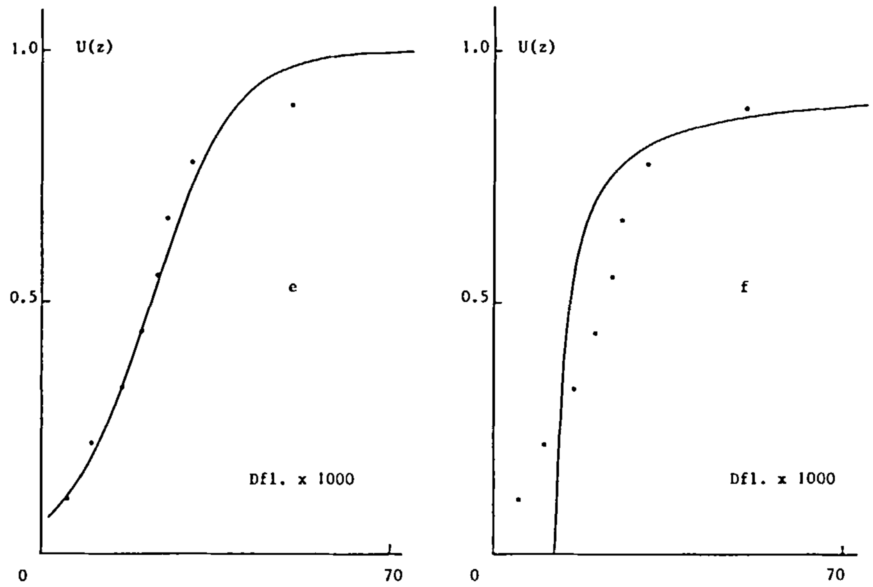

Fig. 2 (continued). (e) Logistic, (f) Log-hyperbola.

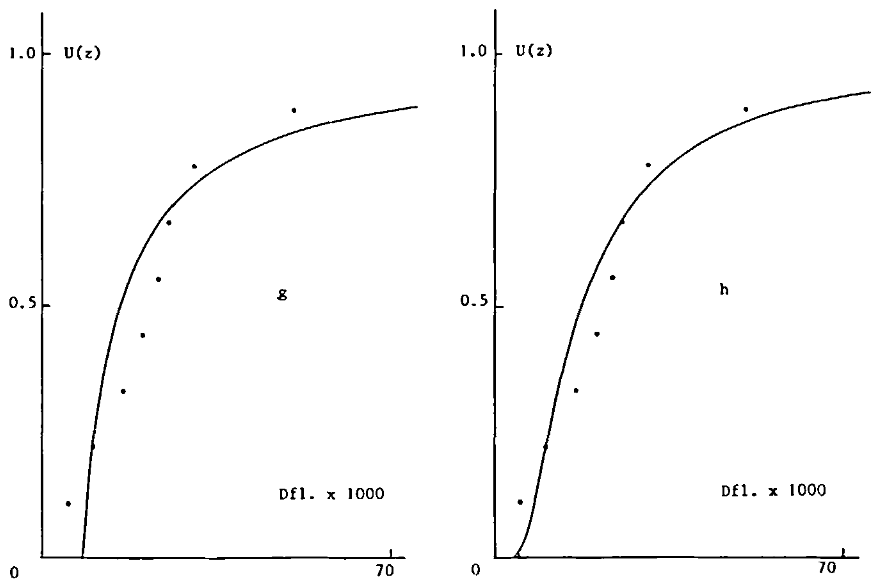

Fig. 2 (continued). (g) Hyperbola, (h) Keller-Hartog.

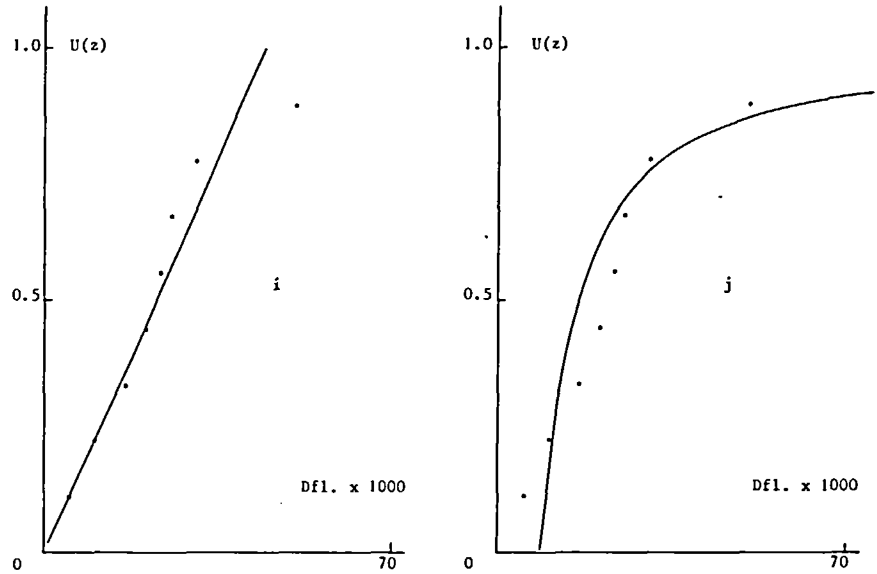

Fig. 2 (continued). (i) Power law, (j) Pareto.

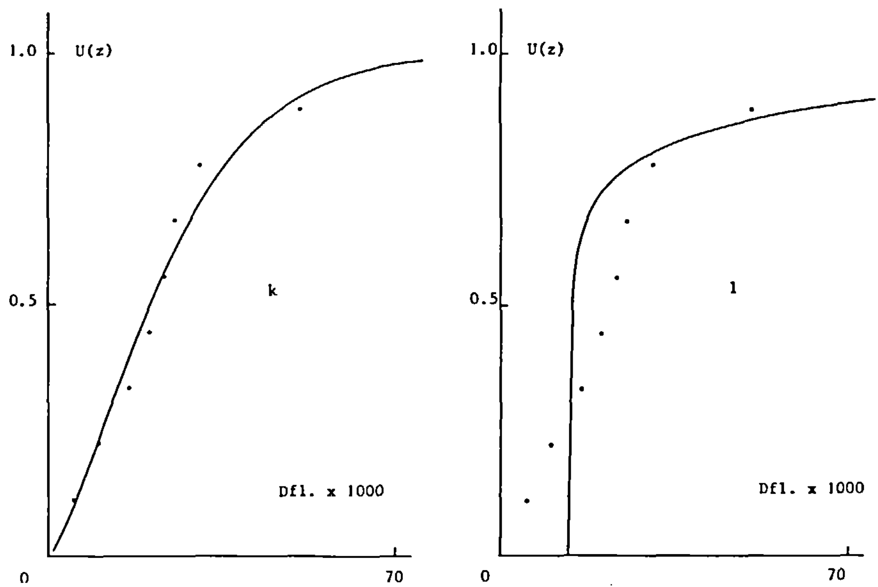

Fig. 2 (continued). (k) Weibull, (l) Stone–Geary.

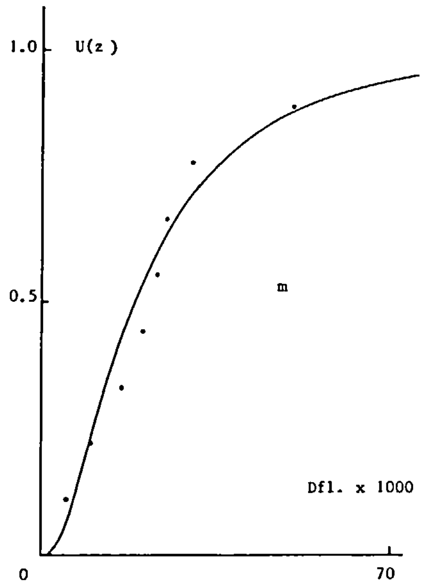

Fig. 2 (continued). (m) Lognormal.

disturbance variance we believe that, if anything, our procedure will bias the results in favour of the non-linear functions.

In figs. 2a up to 2m inclusive the answer to the income evaluation question quoted in section 2 is plotted together with the successively fitted functions.

6. Discussion

In this section we discuss the empirical results, look at some assumptions underlying the measurement method, and view the consequences of the empirical results for the lognormality hypothesis.

Let us first try to develop some intuition by looking at fig. 2. In the region where the data points are (roughly, between welfare levels 0.1 and 0.9) some functions appear to have a shape quite similar to Λ, in particular the log-logistic and to a lesser extent the logarithm and the Weibull. (This can be seen more clearly if one makes drawings of Λ and the other functions in the same figure. For reasons of space, these drawings are omitted.) Even though the data points exhibit an S-shaped pattern, this does not necessarily mean that an S-shaped curve fits the data best. Fig. 2 suggests, for instance, that the logarithm fits as well as Λ. In the example depicted, the reason is that the inflection point corresponds to a rather high evaluation level (about 0.6), whereas most of the S-shaped functions only allow for inflection points below 0.5. It is not the case, however, that the data in general suggest inflection points at high evaluation levels. Inspection of the answers by different respondents show the inflection point to vary substantially. Quite a few of the answers do not exhibit an inflection point at all.

The average unsquared correlation coefficient corresponding to regressions (11) and (12) is 0.98. Obviously the monotonous nature of an individual's response to the evaluation questions more or less guarantees a high correlation coefficient. Still, a value close to one may be considered encouraging. For comparison: for the scatter depicted in fig. 2, Λ gives a correlation coefficient equal to 0.996.

Returning to table 4a, we see that the $\overline{s^2}$-values for the log-logistic and the logarithm are quite close to those for Λ. The other functions usually show $\overline{s^2}$-values that are substantially higher than the corresponding $\overline{s^2}$-values for Λ. These observations are summarized in table 5.

It is clear from table 5 that only the logarithm and the log-logistic are viable alternatives to Λ. The logarithm usually has a somewhat smaller $\overline{s^2}$ than Λ, whereas the $\overline{s^2}$ corresponding to the log-logistic is usually somewhat higher. This pattern is only reversed in column (8), where the logarithm has a bigger $\overline{s^2}$ than Λ and the log-logistic $\overline{s^2}$ is slightly below $\overline{s^2}(\Lambda)$ and in column (9) where the logarithm has a bigger $\overline{s^2}$. Taking as null-hypotheses that the s^2-values for the logarithm and the log-logistic have a probability of 0.5 of being bigger than $s^2(\Lambda)$, a simple sign-test would reject both hypotheses. This indicates that the logarithm provides the model with the lowest disturbance variance.

This conclusion can be made sharper. The logarithm, the log-logistic and Λ all lead to linear specifications of (13). The s^2-values are therefore unbiased estimators of the corresponding disturbance variances. In view of the large number of observations for each wording we furthermore assume that each $\overline{s^2}$ is approximately normally distributed with mean equal to the disturbance variance and variance equal to the square of the standard deviation shown in table 4, divided by the number of observations. We find that the $\overline{s^2}$ for the logarithm is significantly smaller than $\overline{s^2}(\Lambda)$ at at least the 5%-level (one-sided) for wordings 1 and 4. The $\overline{s^2}$ for the log-logistic is significantly (5%-level) larger than $\overline{s^2}(\Lambda)$ for these wordings. For the other wordings the differences are not significant. As one would expect the significant differences are found only for those wordings for which a large number of observations is available. Presumably the other wordings would also show significant differences if the number of observations would increase. To sum up: the logarithm gives a better fit than Λ, whereas all other functions give a worse fit, independent of the particular wording of the evaluation question.

This finding can be interpreted in various ways. If we assume that the set of functions considered contains the true function, then the logarithm

Table 5
Proximity of $\bar{s}^2$-values to $\bar{s}^2(\Lambda)$-values.[a]

	Different wordings										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(2) Normal	++	+++	−	++	++	+++	++	++	++	++	++
(3) Logarithm	−	−	−	−	−	−	−	++	+	−	−
(4) Straight line	++	++	−	++	++	++	++	++	++	++	−
(5) Log-logistic	+	+	+	++	+	+	+	−	+	+	+
(6) Logistic	++	+++	+	+++	++	+++	+++	+++	+++	+++	++
(7) Log-hyperbola	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
(8) Hyperbola	+++	+++	+++	+++	+++	+++	+++	++	+++	+++	+++
(9) Keller–Hartog	++	−	++	++	++	++	++	−	++	++	++
(10) Power Law	+++	+++	++	+++	+++	+++	+++	+++	+++	+++	++
(11) Pareto	+++	+	+++	+++	+++	+++	+++	+	+++	+++	+++
(12) Weibull	++	++	+	++	++	++	++	++	++	++	−
(13) Stone–Geary	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++

[a]Based on tables 4a through 4f. Explanation:

− $\bar{s}^2 \leq \bar{s}^2(\Lambda)$,
+ $\bar{s}^2(\Lambda) < \bar{s}^2 \leq 1.1\bar{s}^2(\Lambda)$,
++ $1.1\bar{s}^2(\Lambda) < \bar{s}^2 \leq 2.0\bar{s}^2(\Lambda)$,
+++ $2.0\bar{s}^2(\Lambda) < \bar{s}^2$,

where $\bar{s}^2(\Lambda)$ is the $\bar{s}^2$-value for Λ.

apparently is the true form of a WF. If we allow for the possibility that the true function is not included in the set of functions considered, we at least have to conclude that Λ is apparently not the true WF. Given the fairly small difference in $\overline{s^2}$ between the logarithm and Λ, the lognormal may however still be close to the true WF.

These interpretations take model (13) for granted, in particular that the $U(z_j)$ are equal quantiles and that the disturbances of the true model are uncorrelated. With respect to the latter assumption we have carried out some simulations: It turns out that if Λ is the true functional form underlying (13) one should, also with correlated disturbances, find $\overline{s^2}(\Lambda)$ to be smaller than the $\overline{s^2}$ for other functional forms, like the logarithm. The assumption of correlated disturbances, therefore, cannot save Λ. The assumption of the $U(z_j)$ being equal quantiles cannot be tested on the data at hand. Here conclusive testing seems only possible by devising different measurement methods for WFs.[4] Recently a new research project has started at the Centre for Research in Public Economics which aims at the development and comparison of different measurement methods. As long as the results of this project are not available, the equal quantile assumption cannot be tested.

Thus, we are left with a choice between either rejecting Λ and accepting the logarithm or questioning the assumptions underlying the measurement procedure.[5] In our opinion there are a number of reasons to maintain Λ until further research into the measurement procedure has been carried out. First, the lognormal form stems from a well-developed theory [Van Praag (1968)] whereas the theoretical basis for the logarithm is unclear. Remember for instance that the measurement procedure rests upon the finite bliss, finite agony assumption. The logarithm can then only reasonably be interpreted as a local approximation to some unknown true function. To give up a tractable function for an approximation to an unknown alternative is not very attractive.

Second, the research into the determinants of μ and σ has been successful. Significant portions of variance in μ and σ can be explained by factors like income, family size, variation in income, etc. [e.g., Van Herwaarden, Kapteyn and Van Praag (1977)]. The functional specifications of the models that explain μ and σ follow in a natural way from Van Praag's theory and a few simple additional postulates [cf. Kapteyn (1977)]. These functional specifications have passed various tests. Moreover the same models have

[4]Schokkaert (1978) has tested the equal quantiles assumption on the Belgian (1969) sample by making a number of additional assumptions. The conclusions of his test appear to be very sensitive to the additional assumptions made.

[5]Of course a researcher is never forced by outcomes of a statistical test to fully accept a particular conclusion. Cf. Theil (1971, p. 545) who observes: 'The analyst may be convinced on a priori grounds that one specification is more realistic than another, in which case he should feel justified in applying the former even if the latter has a slightly smaller residual variance estimate.'

been used in social policy applications with elegant results [e.g., Goedhart et al. (1977), Kapteyn and Van Praag (1976, 1980), Van Praag, Goedhart and Kapteyn (1980)]. It would seem unwise to give up these advantages without having an attractive alternative.

Third, preliminary experiments with PWFs to use them in predicting buying behaviour have turned out to be promising [Kapteyn, Wansbeek and Buyze (1979)]. Also here the lognormal form of the PWF appears to generate hypotheses in a natural way. It is hard to see how one would arrive at these hypotheses without specific knowledge of the shape of the welfare function.

7. Conclusions

In this paper we compared the lognormal WF Λ to 12 alternative functions. It appears that Λ outperforms 11 of these by the residual variance criterion. Only the logarithm performs slightly, though significantly, better. Naturally, this outcome rests upon the procedure used to measure the WFs. The measurement procedure cannot be justified for unbounded functions like the logarithm. Hence the logarithm is to be interpreted as a local approximation to an unknown WF. Given the theoretical basis for the lognormal form, as compared to the logarithm, and its success in numerous applications, we believe that for the moment it is justified to maintain Λ as the true shape of the WF.

However, our results also indicate the need for further research into the measurement procedure.

References

Aitchison, J. and J.A. Brown, 1957, Lognormal distribution (Cambridge University Press, Cambridge, U.K.).

Bartels, C.P.A., 1977, Economic aspects of regional welfare (Martinus Nijhoff, Leyden).

Bernoulli, D., 1738, Exposition of a new theory on the measurement of risk, in: A.N. Page, ed., 1968, Utility theory, A book of readings (Wiley, New York) 199–214.

Christensen, L.R., D.W. Jorgenson and L.J. Lau, 1975, Transcendental logarithmic utility functions, American Economic Review 65, 367–383.

Goedhart, Th., V. Halberstadt, A. Kapteyn and B.M.S. van Praag, 1977, The poverty line: Concept and measurement, The Journal of Human Resources 12, 503–520.

Johnson, N.L. and S. Kotz, 1970, Continuous univariate distributions, 2 vols. (Houghton Mifflin, Boston, MA).

Kapteyn, A., 1977, A theory of preference formation, unpublished Ph.D. Thesis (Leyden University, Leyden).

Kapteyn, A. and F.G. van Herwaarden, 1980, Interdependent welfare functions and optimal income distribution, Journal of Public Economics, forthcoming.

Kapteyn, A. and B.M.S. van Praag, 1976, A new approach to the construction of family equivalence scales, European Economic Review 7, 313–335.

Kapteyn, A. and B.M.S. van Praag, 1980, Family composition and family welfare, in: J.L. Simon and J. da Vanzo, eds., Research in population economics (JAI Press, Greenwich, CT) forthcoming.

Kapteyn, A., F.G. van Herwaarden and B.M.S. van Praag, 1977, Direct measurement of welfare functions: Methods and results, Report 7702, Centre for Research in Public Economics (Leyden University, Leyden).
Kapteyn, A., B.M.S. van Praag and F.G. van Herwaarden, 1978, Individual welfare functions and social reference spaces, Economics Letters 1, no. 2, 173–178.
Kapteyn, A., T.J. Wansbeek and J. Buyze, 1979, Maximizing or satisficing?, The Review of Economics and Statistics 61, 549–563.
Kapteyn, A., T.J. Wansbeek and J. Buyze, 1980, The dynamics of preference formation, Journal of Economic Behaviour and Organization 1, no. 2.
Keller, W.J. and J. Hartog, 1977, Income tax rates and proportional sacrifice, Public Finance 32, 321–332.
Marquardt, D.W., 1963, An algorithm for least-squares estimation of non-linear parameters, Journal of the Society of Industrial and Applied Mathematics (S.I.A.M.) 11, 431–441.
Phlips, L., 1974, Applied consumption analysis (North-Holland, Amsterdam).
Samuelson, P.A., 1977, St. Petersburg paradoxes: Defanged, dissected, and historically described, Journal of Economic Literature 15, 24–55.
Schokkaert, E., 1978, Gebruik van psychologische schaaltechnieken bij de bepaling van de welvaartsfunktie van inkomen (Use of psychological scaling methods in the determination of the welfare function of income), Masters Thesis (Catholic University, Louvain).
Stevens, S.S., 1972, Psychophysics and social scaling (General Learning Corporation, Morristown, NY).
Theil, H., 1961, Economic forecasts and policy (North-Holland, Amsterdam).
Theil, H., 1971, Principles of econometrics (Wiley, New York).
Van Herwaarden, F.G., A. Kapteyn and B.M.S. van Praag, 1977, Twelve thousand individual welfare functions, European Economic Review 9, 283–300.
Van Praag, B.M.S., 1968, Individual welfare functions and consumer behaviour (North-Holland, Amsterdam).
Van Praag, B.M.S., 1971, The welfare function of income in Belgium: An empirical investigation, European Economic Review 2, 337–369.
Van Praag, B.M.S., 1977, The perception of welfare inequality, European Economic Review 10, 189–207.
Van Praag, B.M.S., 1978, The perception of income inequality, in: A. Shorrocks and W. Krelle, eds., Personal income distribution (North-Holland, Amsterdam).
Van Praag, B.M.S. and A. Kapteyn, 1973, Further evidence on the individual welfare function of income: A empirical investigation in The Netherlands, European Economic Review 4, 33–62.
Van Praag, B.M.S. and J. Linthorst, 1976, Municipal welfare functions, Regional Science and Urban Economics 6, 51–79.
Van Praag, B.M.S., T. Goedhart and A. Kapteyn, 1980, The poverty line, A pilot survey in Europe, The Review of Economics and Statistics, forthcoming.