Empirical comparison of the shape of welfare functions

van Herwaarden, F.G.; Kapteyn, A.J.

Published in:
Economics Letters

Publication date:
1979

Economics Letters 3 (1979) 71–76

EMPIRICAL COMPARISON OF THE SHAPE OF WELFARE FUNCTIONS *

Floor G. VAN HERWAARDEN
CERPEC, Leyden University, 2311 XK Leyden, The Netherlands

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

Received 26 December 1978

This study compares the fit of the lognormal welfare function proposed by Van Praag (1968) with the fit of 12 other functions. The comparison uses a sample of about 14,000 respondents. The lognormal function outperforms 11 alternatives in terms of the residual variance criterion, while the logarithm performs slightly better.

1. Introduction

In this letter we provide a test of Van Praag's hypothesis that individuals are able to evaluate any arbitrary income level z on a [0, 1]-scale and that the resulting evaluation $U(z)$ follows approximately a lognormal distribution function: $U(z) = \Lambda(z; \mu, \sigma)$ [Van Praag (1968)]. The function $U(z)$ has been called the individual *welfare function of income*. The hypothesis also applies to expenditures on commodities, in which case U is called a *partial welfare function*. In a somewhat different context it has been suggested that the evaluation of municipal expenditures by local authorities also will follow a lognormal distribution function, which is called a *municipal welfare function*. For the test we present, the distinction between the three types of welfare functions (WF's) is immaterial.

Our test consists of a comparison of the goodness of fit of the lognormal function, Λ, to 12 other two-parameter functions. These functions are either distribution functions on $[0, \infty)$ or have been proposed as utility functions in the economic literature. Since we have about 25,000 observations, even small differences in fit are highly sig-

* The research reported in the paper is supported by a grant from the Netherlands Organization for the Advancement of Pure Research (ZWO). A first version was written while both authors were at CERPEC. We thank the editor, Bernard van Praag, Tom Wansbeek and Roberto Wessels for their helpful suggestions, and Rik Storm for programming the complicated computations. The complete version of this paper is available under the same title as report 78.02 of CERPEC. The study is part of the Leyden Income Evaluation Project.

nificant. It turns out that 11 functions fit significantly worse than Λ and often substantially so. The logarithm fits slightly, but in view of the large number of observations significantly, better than Λ.

In the sequel we briefly describe the way WF's are measured, the ensuing test and the results. Finally we discuss the implications of the results for research using individually measured WF's.

2. Measurement of WF's and the test

WF's are measured by providing respondents to survey questionnaires with a number of verbal labels like 'excellent', 'good', 'bad'. They are asked to provide for each label i a money amount z_i which, in their opinion, is best described by that label. The crucial step in the measurement procedure is the translation of the words 'excellent', 'good' etc. in numbers between zero and one. On the basis of an information maximization argument [Van Praag (1971)] the labels are identified with equal quantiles. For example, if there are five labels 'excellent', 'amply sufficient', 'so-so', 'very insufficient', 'very bad', then 'excellent' is identified with 0.9, 'amply sufficient' with 0.7, etc. Denoting the numerical value attached to the label corresponding to the answer z_i by $U(z_i)$, one thus obtains for each respondent a sequence of observations $\{z_i, U(z_i)\}_{i=1}^{n}$ from which the parameters μ and σ of the lognormal WF can be estimated. In fact it is easy to see that estimation amounts to running the simple regression

$$\ln z_i = \mu + \sigma N^{-1}(U(z_i); 0, 1) + \epsilon_i, \qquad i = 1, ..., n, \tag{1}$$

per respondent, where ϵ_i is an i.i.d. error term.

In different surveys the number of labels has varied between 5 and 8 and some other variations in wording of the questions have been tried out. In total there are 11 different wordings and we test the goodness of fit of Λ for each of them.

So far there is nothing in the description of the measurement method which requires the lognormality of $U(z)$, except eq. (1). If a different form of welfare function were adopted, eq. (1) would be replaced by

$$\ln z_i = f(U(z_i); a, b) + \omega_i, \qquad i = 1, ..., n, \tag{2}$$

with ω_i an i.i.d. error term. We propose to compare Λ to 12 other functional forms by comparing the goodness of fit of (1) to the goodness of fit of (2) for these other functional forms. The criterion chosen is Theil's residual variance criterion, s^2, which is known to be smallest *on average* for the correct model. As we estimate models like (1) and (2), 25,000 times, the average s^2 yields a powerful criterion to judge the correctness of Λ.

It should be noticed that the stochastic specifications of (1) and (2) are not arbitrary. The z_i are the endogenous variables, being the answers given by a respondent who is confronted with the exogenous variables $U(z_i)$.

Table 1
Selected functions.

Function	Functional form $U(z) = \ldots$	Restrictions [a]	P.d.f. [b]	Concavity [c,d]	f^{e} [see (2)]	f linear
(1) Lognormal	$\Lambda(z;\mu,\sigma)$	–	p.d.f.	n.c.	$\mu + \sigma N^{-1}(U;0,1)$	yes
(2) Normal	$N(z;a,b)$	$b > 0$	p.d.f.	n.c.	$\ln(a + bN^{-1}(U;0,1))$	no
(3) Logarithm	$a + b\ln(z)$	–	–	c.	$(U-a)/b$	yes
(4) Straight line	$a + bz$	–	–	n.c.	$\ln\{(U-a)/b\}$	no
(5) Log-logistic	$\{1+\exp(a+b\ln(z))\}^{-1}$	$b < 0$	p.d.f.	n.c.	$\{\ln((1-U)/U) - a\}/b$	yes
(6) Logistic	$\{1+\exp(a+bz)\}^{-1}$	$b < 0$	p.d.f.	n.c.	$\ln\{\{\ln((1-U)/U) - a\}/b\}$	no
(7) Log-hyperbola	$1 + b/(a+\ln(z))$	–	–	c.	$-a - b/(1-U)$	yes
(8) Hyperbola	$1 + b/(a+z)$	–	–	c.	$\ln\{-a - b/(1-U)\}$	no
(9) Keller–Hartog	$\exp\{(a/b)z^{b}\}$	$b \leqslant 0$	p.d.f.	n.c.	$\{\ln((b/a)\ln(U))\}/b$	yes
(10) Power law [f]	az^{b}	$a, b > 0$	–	c.	$\{\ln(U/a)\}/b$	yes
(11) Pareto	$1 - (a/z)^{b}$	$a, b > 0$ $z \geqslant a$	p.d.f.	c.	$\ln(a) - \{\ln(1-U)\}/b$	yes
(12) Weibull	$1 - \exp\{-(z/a)^{b}\}$	$a, b > 0$	p.d.f.	n.c.	$[\ln\{-a^{b}\ln(1-U)\}]/b$	yes
(13) Stone–Geary	$b\ln(z-a)$	$z > a$	–	c.	$\ln\{\exp(U/b) + a\}$	no

[a] For all functions $z > 0$.
[b] P.d.f. = probability distribution function.
[c] (n.)c. = (non)concave.
[d] That is: $\partial^2 U/\partial z^2 < 0$.
[e] U is an abbreviation of $U(z_j)$.
[f] Also called psychophysical law.

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

Function	Different wordings [b]										
	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	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++	+++
Number of observations [c]	9029	95	99	878	9991	94	90	100	1748	96	2675

[a] Based on Tables 4a through 4f of the complete version of the paper.
Explanation:
–: $\overline{s^2} \leqslant \overline{s^2}(\Lambda)$,
+: $\overline{s^2}(\Lambda) < \overline{s^2} \leqslant 1.1\overline{s^2}(\Lambda)$,
++: $1.1\overline{s^2}(\Lambda) < \overline{s^2} \leqslant 2.0\overline{s^2}(\Lambda)$,
+++: $2.0\overline{s^2}(\Lambda) < \overline{s^2}$,
where $\overline{s^2}$ is the average residual variance of the observations in the cell and $\overline{s^2}(\Lambda)$ is the $\overline{s^2}$-value for Λ.
[b] See section 2.
[c] These are the maximum numbers of observations: Some functions do not give rise to a linear specification of (2). These are estimated by Marquardt's iterative non-linear least squares algorithm. The algorithm does not always converge. In view of our sample size it is practically impossible to try new starting-values until convergence is reached. Hence we left out all respondents from whom converge did not obtain.

Table 1 gives a list of the two-parameter functions to be compared to Λ, along with Λ itself. The parameters are denoted by a and b, except for Λ.

3. Data and results

Eight different samples, drawn in Belgium and The Netherlands between 1970 and 1975 and comprising about 14,000 individuals, are used, yielding in total some 25,000 measured WF's (in one sample welfare functions of income and partial welfare functions of a few commodities were measured for the same individuals). A comparison of the goodness of fit of the different functions with Λ for each of the 11 different wordings is provided by table 2.

One observes that the logarithm and the log-logistic are the only viable alternatives to Λ. The logarithm usually has a somewhat smaller residual variance, $\overline{s^2}$, than Λ, whereas the $\overline{s^2}$ corresponding to the log-logistic is usually somewhat higher. Given the large number of observations, the differences are mostly significant. We have to conclude therefore that the logarithm fits slightly but significantly better than Λ.

4. Implications for research

Measured lognormal WF's have been used in a number of studies, like tests of the economic theory of consumer behavior [Kapteyn, Wansbeek, Buyze (1980)], exercises in optimal income distribution [Van Praag (1977, 1978), Kapteyn and Van Herwaarden (1976)], a theory of preference formation [Kapteyn (1977)] and the analysis of the financial needs of Dutch municipalities [Van Praag and Linthorst (1976)]. The more one uses a measuring instrument, the more the instrument itself should be subject to scrutiny. The present study has been motivated by that consideration.

Van Praag's contention that the WF's considered here are approximately lognormal is certainly supported by the test applied, but it seems that the logarithm provides an even better approximation. One may question the assumptions underlying the test, however. The crucial assumption obviously is that verbal labels can be identified with equally spaced points in a [0, 1]-interval. Since the logarithm and Λ are so close, slight departures of the assumption may already affect the results significantly. In view of the successful applications of the lognormal WF hitherto and because of its theoretical underpinning which is largely missing for the logarithm, a hasty discarding of Λ would be unwise. Obviously, though, more research is needed into the properties of the measurement procedure. Currently, such research is being undertaken on the basis of an experimental survey in which different measurement methods are tried out.

References

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.

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

Kapteyn, A. and F.G. van Herwaarden, 1976, Interdependent welfare functions and optimal income distribution, Report no. 76.12 (Center for Research in Public Economics, Leyden University, Leyden).

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, 1979, Family composition and family welfare, in: J.L. Simon, and J. da Vanzo, eds., Research in population economics, vol. II (JAI Press, Greenwich, CT) forthcoming.

Kapteyn, A., T.J. Wansbeek and J. Buyze, 1980, Maximizing or satisficing?, The Review of Economics and Statistics, forthcoming.

Theil, H., 1961, Economic forecasts and policy (North-Holland, Amsterdam).

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 behavior (North-Holland, Amsterdam).

Van Praag, B.M.S., 1971, The welfare function of income in Belgium: An empirical investigation, European Economic Review 2, 337–360.

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 J.M. Linthorst, 1976, Municipal welfare functions, Regional Science and Urban Economics 6, 51–79.