European Economic Review 8 (1976) 181–191. © North-Holland Publishing Company

TESTING AND APPLYING A THEORY OF UTILITY

An Attempt to Decompose Income in Compensatory and Scarcity Rents*

Nienke BOUMA and Bernard M.S. VAN PRAAG

Economic Institute of Leyden University, Leyden, The Netherlands

Jan TINBERGEN

Emeritus Professor, Erasmus University, Rotterdam, The Netherlands

Received April 1976

Data on income after tax, schooling completed, job held, age, and 'level of satisfaction' of 2663 members of the Dutch Consumer Union have been used to estimate regression equations of two types. Type I may be called a specification of a utility function, Type II an 'earnings function' (where income after tax was used as earnings). For both types a number of alternatives were estimated both with regard to mathematical shape and with regard to variables included. Defining equitable or justified income differences as differences which do not change the level of satisfaction, a formula for equitable incomes for given combinations of job, schooling and age can be derived from Type I equations. All regression coefficients are found to be lower than the corresponding earnings function coefficients. The latter can then be decomposed into a 'compensatory' component and a 'scarcity rent' component.

1. Introduction

Elsewhere one of us [Tinbergen (1975)] submitted a theory of utility or welfare characterized by the following assumptions:

(i) Utility is a quantitative concept.

(ii) It depends on entities to be distinguished as *parameters* and *variables*. Ideally parameters constitute constants for a given individual (or, alternatively, household), characterizing its capabilities or its needs. Variables affecting utility may vary either independently of the individual's will or may be under his or her control. Independent variables may be the tax system; variables under control are, for instance, the individual's occupation and the income that goes with it.

(iii) The mathematical form of the utility function and its coefficients are the same for all human beings considered, since apparently different coefficients for different individuals may be taken to constitute parameters.

*We are grateful to an anonymous referee for his comments on an earlier draft. The authors alone are responsible for this text.

Of course the 'theory' so presented has been adhered to by others before. In the form chosen it may indicate a programme of measurements with decreasing oversimplification. Measurement has not so far been undertaken by many authors; one of us made some serious attempts [Van Praag (1971, 1973)], and there are a few more.

In the present paper some or all of the following parameters were used:

v = years of schooling, considered to represent mainly intellectual capabilities;

w = capability to take independent decisions;

t = age, believed to represent also experience.

The variables considered are:

X = income after tax;

s = occupation. From other data available to us for a smaller sample we derived the conclusion that an important job characteristic consists of years of schooling required, as distinct from actual years of schooling. For lack of direct data on schooling required we used three alternative measures of s, indicated as s_1, s_2 and s_3, and standing for the lower quartile of the distribution of v for each group of occupations considered, the median and the upper quartile.

The scope of the present article is twofold, as indicated by its title. First, we used material collected by Van Praag and Kapteyn (1973) on utility or welfare as declared by a sample of 2663 members of the Dutch Consumer Union and made an attempt to measure the coefficients of a utility function. Among the determinants of that function (covering both parameters and variables as defined above) we included some or all of the entities just summed up, plus a 'tension term' supposed, in the original theory [Tinbergen (1975)], to codetermine utility. This 'tension theory' had been mentioned earlier in a theoretical analysis [Tinbergen (1956, 1959)] and upholds that a difference between schooling required and actual schooling is felt to be a burden both when positive and when negative.

A second testing operation whose results are shown in this article consists of an attempt to explain incomes by an 'incomes function' comparable to the 'earnings function' used by several other authors [cf. Mincer (1974)], assuming that in that function the same determinants appear as in the utility function.

We will refer to the two tests just mentioned as Type I and Type II tests. Type I tests have been performed with the aid of a logarithmic and a linear formula, respectively,

$$\log \omega = a_0 + a_1 \log X + a_2 \log s + a_3 \log v + a_4 (\log s/v)^2 + a_5 \log w + a_6 \log t, \tag{1}$$

and

$$\omega = a'_0 + a'_1\ X/1000 + a'_2\ s + a'_3\ v + a'_4\ (s-v)^2 + a'_5\ w + a'_6\ t, \qquad (2)$$

where ω stands for utility or welfare experienced by the individual.

Type II tests have been carried out with the aid of three alternative formulae, namely,

$$\log X = b_0 + b_2 \log s + b_3 \log v + b_4(\log s/v)^2 + b_5 \log w + b_6 \log t, \qquad (3)$$

$$\log X = b'_0 + b'_2 s + b'_3\ v + b'_4(s-v)^2 + b'_5\ w + b'_6\ t, \qquad (4)$$

$$X/1000 = b''_0 + b''_2\ s + b''_3\ v + b''_4\ (s-v)^2 + b''_5\ w + b''_6\ t, \qquad (5)$$

The reason why both a utility function and an incomes function have been tested is that, applying another suggestion made by one of us, from the utility function a criterion for an *equitable income distribution* is derived. This is the 'application' announced in the title of this article. The criterion used is that equity requires equal satisfaction or utility for all.

Setting ω constant in (1) or (2) we may derive iso-utility-curves, described by eqs. (6) and (7),

$$\log X = c_0 - \frac{a_2}{a_1} \log s - \frac{a_3}{a_1} \log v - \frac{a_4}{a_1} (\log s/v)^2 - \frac{a_5}{a_1} \log w - \frac{a_6}{a_1} \log t, \qquad (6)$$

$$X/1000 = c'_0 - \frac{a'_2}{a'_1} s - \frac{a'_3}{a'_1} v - \frac{a'_4}{a'_1} (s-v)^2 - \frac{a'_5}{a'_1} w - \frac{a'_6}{a'_1} t. \qquad (7)$$

They describe how earned income has to depend on the variables 'required schooling', 'actual schooling', the tension between them, the capability to take independent decisions and age in order that every individual is satisfied to the same degree. So (6) or (7) represent the 'earnings functions' in an ideal society with an equitable income distribution. (The value of the constant c_0 determines the common utility level.) These equations will be compared with eqs. (3) and (5), respectively, and enable us to find out whether the actual income distribution deviates, and by how much, from our concept of an equitable income distribution. No use has been made of eq. (4), since it did not perform better than (3).

2. Material and units used

As already indicated the material used consisted of the results of an enquiry made with 2663 members of the Dutch Consumer Union by Van Praag and

Kapteyn (1973). In fact, that enquiry showed data for 2815 respondents, but we removed from it the groups of retired, of those without profession, and of those temporarily unemployed. Thus we were left with family heads in ten different occupational groups. For them the material summarized in table 1 is available, as well as an estimate of each individual's satisfaction ω. Put briefly, this figure was derived from the procedure described by Van Praag (1968) and the observations reported on by Van Praag (1971) and Van Praag and Kapteyn (1973).

The sample was fairly large: its 2815 individuals were selected at random from a larger sample of people who filled in a rather elaborate and laborious questionnaire, the answering of which took about 1½ hour, and which had to be sent under anonymous cover to the Union.

Since the Union is not completely representative for the Dutch population as a whole our results are not wholly representative either. However, from cross-comparisons of various frequencies in the survey with material representative for the Dutch population, we do believe that our data form a fair approximation of the Dutch population for the present purpose.

The results of the enquiry may be summarized as follows:

(1) There was ample evidence for the thesis that an individual welfare function can be estimated and is lognormal.

(2) The parameter σ varies about 0.5 and depends hardly on any objectively measurable individual characteristics so far available.

(3) The parameter μ can be explained quite well by objective characteristics such as net income, family size etc.

It is well-known from lognormal distribution theory that

$$\Lambda(X;\mu,\sigma) = N(\log X;\mu,\sigma) = N\left(\frac{\log X-\mu}{\sigma};0,1\right),$$

where $N(\cdot)$ denotes the normal distribution function, and $N(\cdot\,;0,1)$ the standard normal distribution function. Instead of Λ it seems more appropriate to focus on $(\log X-\mu)/\sigma$, being the more handsome expression. It has a one-one relationship to the ophelimity index $\omega=\Lambda(X;\mu,\sigma)$.

The units used for the other parameters and variables are net income X: guilders per annum; schooling v: years completed; schooling required s_i ($i=1,2,3$): years; capability to take independent decisions w: a dummy variable to which the values 1, 2 and 3 have been assigned as indicated in table 1; age t: periods of 20 years. This choice was made in order to investigate whether the 'tension' $w-t$ is likely to play a role comparable to the tension $s-v$. The difference supposed to be a quadratic function, symmetric about zero, it seemed obvious to us that w and t had to be measured in such a way that $(w-t)^2$ or $\ln(w/t)^2$ satisfied this property more or less. This is the reason that we took a period of 20 years as the unit of measurement. No influence of $w-t$ was found, however.

3. Calculations carried out and testing results

With the aid of the material described a large number of computations were performed, the main results of which are shown in tables 2 and 3, which are self-explanatory to a large extent.[1] With respect to the quality of the explanation as measured by R^2, we find that the quality is not bad, when we realize that the size of the sample is 2663. For the 'income function' (table 3) these values are comparable to those found by other authors. In both cases the logarithmic formulae fit the material better than the linear formulae. Since the 'mixed' formula (4)—inspired by Mincers' work although in a somewhat different context—did not fit the material better than eq. (3), only a few results of it are shown. A general feature of the results is that R^2 is hardly affected by the choice between the three alternative s values. This is evident in Cases 06 through 08, 13 through 15 and some more groups of three cases. This being so, Cases 02, 12 and 22 have only been shown for s_1. Since inclusion of all regressors led to non-significant regression coefficients (showing standard deviations well over 50 percent) no such cases have been shown. Whereas table 2 generally shows higher standard deviations than table 3, it can be stated nevertheless that X (in table 2), v, w, and t or their logarithms obtain significant and stable regression coefficients, with the intuitively expected algebraic sign.

Turning now to table 2 we find that inclusion of s instead of v (Case 01 and 02) somewhat reduces R^2 and yields a less reliable coefficient to s than to v. The negative coefficient of log (v) in 01 suggests that there is a negative influence on satisfaction emanating from a higher intellectual capability, a conclusion also suggested by Cases 04′ and 09 for a linear utility function. Put otherwise, a higher intellectual capability creates cultural needs. That the coefficient of s in 02 is less reliable may be due to the fact that s is an indirect measure only of schooling required. As for Case 01 it is economically not attractive to have no s-term, s representing the effort needed.

Cases 06 through 08 constitute a test of the 'tension theory', assuming at the same time that there is no direct effect of intellectual capability on satisfaction. Here we observe that the coefficients of s are rather stable and that the selection of different s-values has a considerable influence on the coefficient of the 'tension term' $(\log s/v)^2$. The tension theory is only acceptable if we choose s_1 (Case 06). Inclusion of the tension term, however, only marginally improves R^2. The choice of s_1 can be defended on the argument that it implies that 25 percent of

[1]The numbering of the cases shown is such that the first digit is 0 in table 2 and 1, 2 or 3 in the three compartments of table 3. The second digit characterizes a combination of regressors chosen. Cases with primes are different from those without primes.

Table 1

Average net family income (in ƒ 1000, — per annum) and number of observations (right-hand lower line in each cell), years of schooling v, degree of independent decision making w, and three alternative estimates of years of schooling required s_1 (lower quartile), s_2 (median) and s_3 (upper) for 2663 members of Dutch Consumer Union.

	Primary education $v=6$	Ext. primary education $v=10$	Primary vocational education $v=10$	Secondary education $v=12$	Secondary vocational education $v=13$	Higher vocational education $v=17$	University education $v=18$	w	s_1	s_2	s_3
Unskilled labour	10.84 7	10.27 3	19.80 1	0.0 0	0.0 0	0.0 0	0.0 0	1	2.4	4.7	7.7
Skilled labour	11.72 17	14.90 18	12.00 75	14.40 2	13.34 18	25.00 1	0.0 0	1	9.5	10.0	10.0
Administrative personnel	17.95 8	13.99 140	13.23 7	14.17 89	13.93 34	16.55 20	13.88 6	1	7.9	10.0	11.6
Lower and middle executives	14.25 11	17.72 96	14.71 24	18.81 86	14.77 76	19.19 86	21.77 10	2	9.6	11.5	13.0
Army and police	18.58 6	13.68 28	14.80 12	17.89 21	15.10 9	25.47 26	15.99 8	1	9.1	10.9	14.0
Instructors teachers	22.20 1	19.13 5	10.50 2	15.10 12	14.78 6	19.56 281	24.71 132	2	14.2	15.8	17.2
Professional experts	12.20 2	18.15 43	13.68 5	19.28 101	15.62 144	18.65 464	21.00 152	2	12.5	14.4	16.3
The professions	13.00 1	23.73 10	21.00 2	30.48 30	18.91 7	26.16 80	31.33 219	3	14.9	17.2	17.6
Commercial occup.	13.20 2	16.96 35	14.32 5	20.34 38	18.26 18	19.94 24	24.28 6	3	9.4	11.2	12.9
Agrarians	13.80 1	8.90 2	9.66 3	0.0	18.63 7	29.00 3	0.0 0	2	10.0	12.3	12.9

Table 2

Results of Type I tests: R^2, intercept and regression coefficients (with their standard deviations as a percentage) on a number of combinations of explanatory variables for utility ω, using three alternative s values.[a]

Case	R^2	a_0	log X	log s	log v	$(\log s/v)^2$	log w	log t	s used
01	0.258	−6.53 (4)	0.797 (3)	n.i.	−0.213 (9)	n.i.	−0.079 (36)	−0.40 (9)	n.i.
02	0.251	−6.53 (4)	0.764 (3)	−0.11 (49)	n.i.	n.i.	−0.094 (36)	−0.36 (10)	s_1
06	0.254	−6.43 (4)	.776 (4)	−0.19 (32)	n.i.	−0.22 (35)	−0.081 (43)	−0.37 (10)	s_1
07	0.252	−6.40 (4)	0.769 (3)	−0.18 (38)	n.i.	0.05 (250)	−0.070 (52)	−0.37 (10)	s_2
08	0.256	−6.37 (4)	0.783 (3)	−0.23 (31)	n.i.	0.32 (32)	−0.063 (55)	−0.39 (10)	s_3
Case	R^2	a_0'	$X/1000$	s	v	$(s-v)^2$	w	t	s used
04′	0.199	0.275 (22)	0.029 (4)	n.s.	−0.0067 (55)	n.s.	−0.037 (53)	−0.127 (16)	s_2
09	0.198	0.273 (20)	0.029 (4)	n.s.	−0.0069 (49)	n.i.	−0.037 (45)	−0.127 (16)	n.i.

[a] n.s. means 'non-significantly different from zero', n.i. means 'not included'.

Table 3

Results of Type II tests: R^2, intercept and regression coefficients (with their standard deviations as a percentage) on a number of combinations of explanatory variables for log. income (log X) or income (X/1000), using three alternative s values.[a]

Case	log X	R^2	b_0	log s	log v	$(\log s/v)^2$	log w	log t	s used
11		0.398	8.07 (1)	n.i.	0.44 (6)	n.i.	0.24 (8)	0.71 (3)	n.i.
12		0.363	8.35 (1)	0.38 (10)	n.i.	n.i.	0.23 (11)	0.67 (4)	s_1
13		0.400	7.91 (1)	0.13 (31)	0.40 (8)	n.i.	0.20 (12)	0.70 (3)	s_1
14		0.401	7.73 (1)	0.22 (24)	0.38 (8)	n.i.	0.18 (14)	0.70 (3)	s_2
15		0.400	7.77 (1)	0.17 (33)	0.39 (8)	n.i.	0.20 (12)	0.70 (3)	s_3
16		0.379	7.92 (1)	0.54 (7)	n.i.	0.46 (18)	0.20 (12)	0.67 (3)	s_1
17		0.370	7.88 (1)	0.55 (9)	n.i.	0.07 (115)	0.18 (15)	0.67 (3)	s_2
18		0.375	7.88 (1)	0.52 (9)	n.i.	−0.45 (16)	0.21 (12)	0.69 (3)	s_3
Case	**log X**	**R^2**	**b'_0**	**s**	**v**	**$(s-v)^2$**	**w**	**t**	**s used**
21		0.391	8.34 (0)	n.i.	0.035 (6)	n.i.	0.16 (7)	0.35 (4)	n.i.
22		0.356	8.45 (0)	0.040 (8)	n.i.	n.i.	0.14 (10)	0.33 (4)	s_1
23′		0.393	8.30 (0)	n.i.	0.031 (8)	0.012 (32)	0.13 (10)	0.35 (4)	s_1
24′		0.394	8.27 (0)	n.i.	0.030 (8)	0.015 (27)	0.13 (11)	0.35 (3)	s_2
25′		0.393	8.26 (0)	n.i.	0.031 (8)	0.011 (36)	0.14 (9)	0.35 (4)	s_3
Case	**X/1000**	**R^2**	**b''_0**	**s**	**v**	**$(s-v)^2$**	**w**	**t**	**s used**
33′		0.336	−11.71 (8)	n.i.	0.57 (10)	0.30 (31)	3.09 (10)	7.35 (4)	s_1
34′		0.338	−12.44 (9)	n.i.	0.54 (11)	0.39 (25)	2.85 (12)	7.31 (4)	s_2
35′		0.335	−11.83 (9)	n.i.	0.61 (10)	0.16 (58)	3.37 (9)	7.37 (4)	s_3
36		0.327	−11.05 (8)	0.91 (9)	n.i.	0.070 (13)	3.03 (10)	7.16 (4)	s_1
37		0.319	−11.13 (9)	0.92 (9)	n.i.	0.36 (37)	2.65 (13)	7.00 (4)	s_2
38		0.315	−10.59 (10)	0.70 (11)	n.i.	−0.63 (21)	3.48 (9)	7.18 (4)	s_3
39		0.334	−11.77 (8)	n.i.	0.67 (7)	n.i.	3.63 (7)	7.39 (4)	n.i.

[a] n.i. means 'not included'.

the labour force is not in the possession of schooling required, whereas taking s_3 implies that 75 percent has less schooling than required, which is less probable.

The two linear cases shown (04′ and 09) give no credit to the joint inclusion of s and $(s-v)^2$.

Table 3 gives clear support to the inclusion, in the income function, of v rather than s. Better results for the explanation of incomes are obtained by the inclusion of both s and v, independently of the choice of s, but only marginally better (Cases 13–15 compared with Case 11). Case 16 supports the tension theory, especially when compared with Case 12. Again the algebraic sign is only correct and significant if we select s_1. Here R^2 increases considerably in comparison to Case 12.

Case 21 reconfirms that, for our measure s for schooling required, log X is explained significantly better with the aid of v than with the aid of s (Case 22). Inclusion into Case 21 of the tension term (Cases 23′ through 25′) yields significant and theoretically correct coefficients, but only marginal improvements of R^2. Compared with Case 39, Cases 33′ through 35′, yield the same conclusions. Cases 36 through 38 also confirm what was found in the upper compartment of the table (Cases 16 through 18).

4. Application of the concept of an 'equitable income distribution' to the test results

The highly significant coefficients in the regression estimates of the incomes function and the iso-utility curves enable us to apply the concept of an equitable income distribution as mentioned at the end of section 1. The concept enables us to derive from eqs. (1) and (2) an expression of log X or $X/1000$ in terms of the variables s and the parameters v, w and t which equalizes utility between all individuals (or family heads) considered. In table 4 the first lines repeat some regression results with respect to eqs. (1) and (2) already presented in table 1. The resulting 'equitable' incomes functions or equations are shown in table 4 by the letter E. They are then confronted with the actual incomes functions A using the same regressors. Table 4 contains five such pairs. In all cases we find that E shows lower coefficients than the corresponding A equations. The confrontation enables us to split up the A coefficients into two components:

(i) the numerical value in the E equation which represents a *compensation* for assuming the 'burden' of a higher v, s, w, t or tension, and

(ii) the difference between A and E coefficients which can reasonably be called the coefficients of a 'scarcity rent' R, also shown in table 4.

From the figures we see that, roughly speaking, *about one-half of the actual income differences are in this terminology scarcity rents and the other half compensations*. We observe that we did not find negative differences. This indicates

Table 4

'Equitable' (E) and actual (A) income equations ($\log X$ or $X/1000$) and scarcity rents R: their dependence on parameters and variables considered and coefficient ratios CR.[a]

Case	$\log X$	$\log s$	$\log v$	$(\log s/v)^2$	$\log w$	$\log t$
1	0.797	n.i.	0.213	n.i.	0.079	0.40
	E $\log X_E$	n.i.	0.27		0.10	0.50
11	A $\log X_A$	n.i.	0.44	n.i.	0.24	0.71
	R $\log(X_A/X_E)$	n.i.	0.17	n.i.	0.14	0.21
	CR		0.39		0.58	0.30
2	0.764	0.105	n.i.	n.i.	0.094	0.36
	E $\log X_E$	0.14	n.i.	n.i.	0.12	0.47
12	A $\log X_A$	0.38	n.i.	n.i.	0.23	0.67
	R $\log(X_A/X_E)$	0.24	n.i.	n.i.	0.11	0.20
	CR	0.63	n.i.	n.i.	0.48	0.30
6	0.776	0.19	n.i.	0.22	0.881	0.37
	E $\log X_E$	0.245	n.i.	0.28	0.10	0.48
16	A $\log X_A$	0.54	n.i.	0.46	0.20	0.67
	R $\log(X_A/X_E)$	0.295	n.i.	0.18	0.10	0.19
	CR	0.55	n.i.	0.39	0.50	0.28
Case	**$X/1000$**	**s**	**v**	**$(s-v)^2$**	**w**	**t**
4′	0.029	n.s.	0.0067	n.s.	0.037	0.127
	E $X_E/1000$	n.s.	0.23	n.s.	1.28	4.38
34′	A $X_A/1000$	0.000	0.54	0.39	2.85	7.31
	R X_A-X_E	0.0	0.31	0.38	1.57	2.93
	CR		0.57	0.97	0.55	0.40
9	0.029	n.i.	0.0069	n.i.	0.037	0.127
	E $X_E/1000$	n.i.	0.24	n.i.	1.28	4.38
39	A $X_A/1000$	n.i.	0.67	n.i.	3.63	7.39
	R X_A-X_E	n.i.	0.43	n.i.	2.35	3.01
	CR	n.i.	0.64	n.i.	0.65	0.41
Average CR		0.59	0.53	0.68	0.55	0.34

[a]n.s. means 'non-significantly different from zero', n.i. means 'not included'.

that all the variables considered are scarce. We wonder whether this result will change in the near future in view of the fact that there is an enormous increase in supply of higher-educated labour to be expected. In order to approach an equitable income distribution we should, if possible, try to reduce the scarcity rent, either by 'creating' more of the scarce capabilities or by finding a taxation device for the purpose.

5. Conclusions

Keeping in mind that the data used have not been specially gathered for the purpose for which we used them and that the sample is not completely representative for the Dutch population we are inclined to summarize the preliminary results of this paper as follows:

(a) It is desirable to collect direct data on 'schooling required' as distinct from 'actual schooling'; and some material has already been obtained.

(b) With the material of the enquiry considered in this article the tension theory can not completely be dismissed, but it remains a marginal improvement only of any utility function or income function here considered.

(c) In contradistinction to the doubts about the best measure for schooling required and the tension theory, the tests show a clear and stable impact of income, actual or in some cases required schooling, capacity to take independent decisions and age or experience on utility and of the latter four regressors on income. In later studies there will be possibilities to compare the coefficients found for the Netherlands with those for some other countries.

(d) About one-half of the actual net-income differences can be explained as compensations for schooling differences, age differences, and the assumption of more or less responsibilities. The other half has to be considered as scarcity rents generated in a 'seller's market'.

(e) The scarcity rents observed are always positive, indicating that there is no oversupply of education, etc. at this moment.

References

Mincer, J., 1974, Schooling, experience and earnings (N.B.E.R., New York).

Tinbergen, J., 1956, On the theory of income distribution, Weltwirtschaftliches Archiv 77, 155; reprinted with some corrections in: J. Tinbergen, 1959, Selected papers (North-Holland, Amsterdam) 243.

Tinbergen, J., 1975, Income distribution: Analysis and policies (North-Holland, Amsterdam).

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.

Van Praag, B.M.S. and A. Kapteyn, 1973, Further evidence on the individual welfare function of income: An empirical investigation in the Netherlands, European Economic Review 4, 33.