

THE INDIVIDUAL WELFARE FUNCTION

A Rejoinder

Arie KAPTEYN *

Tilburg University, The Netherlands

Tom WANSBEEK

University of Groningen, The Netherlands

Received July 8, 1985; accepted July 8, 1985

Professor Ratchford's insightful comments provide a nice opportunity to shed extra light on some points. Most of his remarks deal with the estimation of the WFI. He points out correctly that his doubts about the validity of the WFI-concept do not necessarily affect the validity of the preference formation theory. We agree with most of his remarks and disagree with some. In a few cases our review appears to have been too brief to get basic points across. In any case, we have tried to give an overview of an ongoing research project, so that any suggestion which may help in directing future research is welcome.

Let us discuss Ratchford's comments in the same order.

The relation between WFI and utility

'In standard economics, utility refers to a preference ordering...', Ratchford says. This is, of course, true. And if the utility indicator were only an ordinal representation of a preference ordering, direct asking for welfare levels, as is done in the measurement of the WFI, would not make sense. As Ratchford himself points out, however, there are many

* Mailing address: A. Kapteyn, Dept. of Econometrics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands.

choice situations where a preference ordering has to be represented by a cardinal utility indicator. It seems likely that preference orderings are stable across decision problems, in the sense that they don't change whimsically from one problem to the other. So, if a utility function can serve as a representation of preferences in one situation, it can also serve as a representation of preferences in another situation. In other words, if utility is cardinal in one case, then it is cardinal in all cases.

If we then assume that utility can be measured on a cardinal scale, the question arises how it can be measured. The obvious, and most often practiced, approach is to start from the axiomatization that is used to characterize behavior in the decision problem at hand and derive measurement procedures accordingly. For example, if we choose axioms to describe rational behavior under uncertainty, the obvious approach to the measurement of utility is to devise gambling experiments. But we are not forced to approach measurement in this way. If we accept that utility is cardinal and basically invariant across decision situations, one would expect the scale values themselves to have some meaning to individuals. The IEQ, and its variants, is an attempt to tap this meaning.

A check on the validity of the WFI is then whether the measurement results show a pattern that is consistent with the theoretical notion of utility.

Although we return to this last point at various junctures, one observation can be made already: Ratchford contrasts the WFI with the usual economic definition of utility and concludes that 'utility would be defined as the response to a survey question which evaluates various income levels rather than an expression of the ordering or intensity of preferences'. At least in this respect the WFI fits perfectly within mainstream economic theory, because a so-called indirect utility function is also a descriptor of the utility derived from income (with prices given, and generally only up to a positive monotonic transformation).

Construct validity

Ratchford lists a number of criteria for construct validity that should be satisfied by the WFI. Let us briefly see how well the WFI fares on these criteria.

Observational meaningfulness: So far we have simply not encountered an instance where empirical results obtained with the WFI were widely divergent from our theoretical expectations.

Internal consistency: Ratchford suggests that this ‘...could be assessed by determining how well the fitted WFI relations predict income values associated with scale levels which were held out in estimation’. We have not exactly done this, but the very good fit of the regression that is used to estimate a respondent’s WFI (see Kapteyn and Wansbeek 1985: section 4 and fn. 4) implies that such predictions would be quite good.

Convergent validity: It is of interest that family equivalence scales constructed on the basis of WFIs are similar to scales obtained from the (very different) revealed preference approach to the measurement of utility. Thus, two very different approaches to the measurement of utility agree quite well. This suggests that both approaches measure the same thing, viz., utility. Convergent validity for the WFI itself has not been established so firmly. An ongoing project by Dubnoff (Center for Survey Research, Boston), funded by the National Science Foundation, will hopefully shed some more light on this issue before too long.

Discriminant validity: This has, indeed, not been investigated yet.

In sum, previous research has not focussed on all criteria for construct validity. To the extent that it has been done, results are rather reassuring.

Equal intervals

Regarding the equal interval assumption we reproduce here an empirical result obtained by Antonides, Kapteyn and Wansbeek (1980). They consider the case where the labels used in the IEQ are ‘very bad’, ‘bad’, ‘insufficient’, ‘sufficient’, ‘good’ and ‘very good’. Apart from using these labels in the IEQ (in a simpler version than the one quoted in our review), they also ask respondents to assign a number between zero and ten to these verbal labels.

Table 1 gives some results for a sample of 314 respondents. We have divided all numbers by ten, thus transforming the [0,10]-scale into a [0,1]-scale. For comparison, the last column of the table gives the numerical values corresponding to the equal interval assumption.

If we look upon the sample means as unbiased estimates of corre-

Table 1
Numerical values attached to verbal labels. ($N = 314$.)

Labels	Average numerical values	Sample standard deviations ^a	Equal interval values
Very bad	0.1685	0.1268 (0.0072)	0.0833
Bad	0.3221	0.1020 (0.0058)	0.2500
Insufficient	0.4713	0.0670 (0.0038)	0.4166
Sufficient	0.6203	0.0638 (0.0036)	0.5833
Good	0.7823	0.0622 (0.0035)	0.7500
Very good	0.9326	0.0691 (0.0039)	0.9167

^a The numbers in parentheses are sample standard deviations divided by $\sqrt{314}$.

sponding population values, then the numbers in parentheses are the standard errors of the estimates. Obviously, the estimates differ significantly (at any conventional level of significance) from the values implied by the equal interval assumption. Yet, although the differences are significant, they do not seem to be large, except for the labels 'very bad' and 'bad'.

Basically, Buyze (1982) comes to a similar conclusion. She also rejects the equal interval hypothesis, but at the same time concludes that the differences are not large. Ratchford correctly observes that Buyze assumes lognormality of the WFI in constructing her test. Thus it is the *joint hypothesis* of lognormality and equal intervals that she rejects and it is also the joint hypothesis of lognormality and equal intervals that may still provide a reasonable approximation to reality.

Ratchford is undoubtedly right in stressing that the labels used in the IEQ do affect the possible validity of the equal interval assumption if they are chosen sufficiently extreme. The information theoretic argument that underlies the equal interval assumption may still hold as long as the labels are not too obvious in contrast with the notion of equal intervals. In any case, the results by Antonides, Kapteyn and Wansbeek suggest that the information theoretic argument gives results that are in the right ball park, for the wording usually employed.

Upper bound

'In the standard theory, the utility function has no upper bound'. Ratchford claims. For ordinal utility functions, one can always find a

transformation that guarantees the existence of an upper bound. So, when mentioning 'standard theory' he presumably does not refer to neoclassical ordinal utility functions. For the rest, whether or not utility is bounded from above is hard to decide on a priori grounds. In any case, the statement that '... a WFI = 1 might imply more total happiness for one person than another', is about interpersonal comparability and not about cardinality. These are two very different concepts (cf. Sen 1974).

Ratchford's argument against the boundedness of utility may be based on a confusion of, in this case, income and the utility of income. If he would trade his current meager professor's salary for an executive's salary at General Motors, his income might become ten times as large, but there is no reason why his utility would increase as dramatically. The statement that 'people always seem to be quite pleased to make more money if they don't have to work for it' is completely consistent with this observation. That we may never be satisfied with what we have, has more to do with habit formation than with the boundedness of the utility function.

To some extent we do define concepts by the way we measure them. Ratchford's statement that an income of \$100,000,000 might give ten times as much satisfaction as an 'excellent' income of \$100,000 is not defined as long as no scale of satisfaction has been introduced.

A rather strong theoretical argument for the boundedness (from above) of utility functions is provided by Menger's so-called super-St. Petersburg paradox (cf. Samuelson (1977) for discussion).

As a final observation, it should be noted that if utility is not bounded from above, then there must exist something like 'infinite bliss'. It is hard to imagine what infinite bliss could be, or 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 entails the boundedness of the experience.

Relative well-being

'The implication that a doubling of all incomes would leave no one better off would seem very implausible... Wouldn't most of us prefer that state of affairs to our current one'. Three points are worth making

here. First, the second statement is irrelevant to the first one. Even if evaluations are completely relative, individuals will still think that a doubling of their income makes them happier. Only quite a while after they (and everyone else) have received the higher income, will their evaluation of their income have dropped to the level previous to the doubling of incomes. And even although the new situation does not produce greater utility than the old situation, individuals will think that going back to the old situation will make them less happy.

This also illustrates the second point to be made. The implication that a doubling of all incomes would leave no one better off is only implausible *ex ante*, i.e., given our present preferences. But *ex post*, after our preferences have shifted, we don't feel better off.

Thirdly, our experiments with WFIs are not the only pieces of evidence in favor of the preference formation theory. In our 1982 article we discuss a number of different investigations, from fields like relative deprivation theory, adaptation-level theory and reference group theory, that all lead to similar conclusions.

Ratchford suggests a couple of times that the method of measurement of WFIs may be responsible for the success of the preference formation theory. Apart from the fact that there is a fair amount of evidence that does not rest on WFIs, it is not clear why the formulation of the IEQ would lead to a measure of relative well-being. In the question there is no reference to the position of others, and there is no apparent reason why someone with a \$35,000 annual income with poor friends would be led by the IEQ to evaluate his income as excellent, whereas someone with very wealthy friends evaluates the same \$35,000 as barely sufficient. Yet, this is what the empirical results bear out.

Overall evaluation and needed research

We agree with Ratchford that especially in the area of measurement of WFIs a lot of additional work remains to be done. Fortunately, some of this work is being done right now: we already mentioned Dubnoff's NSF-project.

Still, we feel that certain empirical regularities are firmly established. In particular, the preference formation theory holds up quite well. The doubts about certain aspects of WFI measurement would suggest, moreover, that the empirical tests of the preference formation theory are biased against it, rather than in favor of it.

References

- Antonides, G., A. Kapteyn and T.J. Wansbeek, 1980. 'Reliability and validity assessment of ten methods for the measurement of individual welfare functions of income. Working paper. Modelling Research Group, University of Southern California, Los Angeles.
- Buyze, J., 1982. The estimation of welfare levels of a cardinal utility function. *European Economic Review* 17, 325–332.
- Kapteyn, A. and T.J. Wansbeek, 1982. Empirical evidence on preference formation. *Journal of Economic Psychology* 2, 137–154.
- Kapteyn, A. and T. Wansbeek, 1985. The individual welfare function: a review. *Journal of Economic Psychology* 6, 333–363.
- Samuelson, P.A., 1977. St. Petersburg paradoxes: defanged, dissected, and historically described. *Journal of Economic Literature* 20, 24–55.
- Sen, A.K., 1974. Informational bases of alternative welfare approaches. *Journal of Public Economics* 3, 387–403.