Journal of Public Economics 24 (1984) 135–151. North-Holland

# REVEALING DEMAND FOR AN ACTUAL PUBLIC GOOD

Peter BOHM*

*University of Stockholm, 10691 Stockholm, Sweden*

Received February 1983, revised version received December 1983

In spite of important theoretical advances and a number of laboratory tests in recent years, no practicable mechanism for revealing demand for public goods has yet emerged. A method is presented which seems capable of meeting the political demands of such mechanisms. It was successfully tested on an actual public-good project provided by the Swedish government. The method is evaluated against the background of the results of this nonlaboratory and nonhypothetical test.

## 1. Introduction

Research on demand-revealing mechanisms for public goods during the last decade has produced important findings and new insight into this field, but few instruments that can actually be used for decision-making in the public sector. The theoretical contributions have been characterized by technical sophistication rather than practicability, and empirical research has relied heavily on hypothetical approaches and laboratory testing. Hence, little evidence has been accumulated that would make politicians willing to apply any of the new ideas for demand revelation to important public-good issues.

This paper reports a nonhypothetical, nonlaboratory application of a demand-revealing mechanism that seems possible to use in practice. The application concerns an actual public good. The good is produced if and only if demand is revealed to be sufficiently high, i.e. aggregate willingness to pay (WTP) exceeds costs. Costs are covered in a way that is determined in conjunction with demand revelation. For reasons spelled out below, the method is intentionally aimed at an interval estimate of the aggregate WTP. The results for our particular test case suggest, however, that the method's degree of approximation is sufficient for decision-making.

In section 2 recent research in this area is evaluated from the point of view of applicability, given a set of requirements necessary for demand-revealing

*I am indebted to Jan-Eric Nilsson and Hans Jelf who helped me carry out the investigation reported here. I have benefited from comments by Sören Blomquist, Roger Pyddoke and other colleagues at the Department of Economics, Stockholm University, on earlier drafts of the paper.

0047–2727/84/$3.00 © 1984, Elsevier Science Publishers B.V. (North-Holland)

mechanisms to be able to replace public-good decisions by political representatives. The method to be tested here is presented in section 3. Certain requirements for testing a demand-revealing method in order to disclose its usefulness for future applications in public decision-making are observed in section 4. In section 5 we present the test case and the test results. The general applicability of the method is evaluated in section 6.

## 2. The applicability of existing methods

The theoretical and practical problems of estimating WTP for public goods have attracted a great deal of attention for quite some time now. The reasons for this development seem to include the following four points. First, recall the conventional wisdom until the beginning of the 1970s:

(a) any individual who has a chance to be a free rider would be one, and

(b) there is no feasible, institutional way to avoid misrepresentation of preferences.

These views eventually became recognized as assumptions rather than facts and prompted an interest in developing alternative hypotheses. Second, the growing public sector in many countries made it important to study public decision-making mechanisms and, in particular, to intensify the search for alternatives to tax financing. Third, from an entirely different angle, economists have shown a growing interest in experimental methods as a complement to a pure econometric approach to empirical knowledge; issues of demand revelation for public goods offered one field of study to which this interest could be channelled. Finally, and perhaps most important, decisions about public goods are made without reliable information about consumer preferences and this constitutes a considerable problem from the point of view of a well-functioning democracy. In other words, there is a need for instruments or institutions that are more appropriate or efficient than those of a representative democracy where the 'representatives' often do not know what to represent on an individual public-good issue.

While it is easy to understand why this field has received so much attention lately, the direction this interest has taken in large parts of the economic literature is more questionable. The main issues dealt with seem to be the following.

(a) Is there a mechanism that would make 'rational' people reveal their preferences for public goods? The general idea here seems to be that, if the answer is yes and even if the mechanism is technically sophisticated, we should simply go ahead and use it for public decision-making. [See, for example, Clarke (1971), Groves and Ledyard (1977), Tideman and Tullock (1976) and Green and Laffont (1979); see also review articles in *Public Choice* (1977), and by McMillan (1979) and V. Smith (1980).]

(b) To what extent do people actually try to be free riders when there are incentives for doing so? The idea here is that if people do not seem to

respond strongly to such incentives, we should not worry too much and simply use the WTP responses people make — to hypothetical questions, for example — as an approximation of their true WTP. [See, for example, Marwell and Ames (1980), Bishop and Heberlein (1979) and Brookshire et al. (1982).]

With respect to finding methods for more immediate application the emphasis on these two issues seems somewhat misplaced. Two reasons can be given in support of this view [see also Johansen (1977, 1981)].

(1) Where are the politicians who would let decisions be made by, say, a Groves and Ledyard/Clarke mechanism that even students of economics need some time to grasp and that moreover is subject to criticism in certain important respects [see, for example, Ng (1979)]?

(2) If a given mechanism can be shown to work (which in itself is difficult to establish) in one, two or three laboratory tests, how can we be sure it will work in the fourth instance when we want an important decision to be determined by it? And if free-rider behavior is insignificant in one, two or three experimental studies, how can we rely on this being consistently so in actual public decision-making? More important, would those who decide if and when the suggested approaches should be used in practice, i.e. politicians, be convinced that these approaches would continue to work and work better than their own wise deliberations?

Thus, it can be argued that neither issue (a) or (b) deals with the implementation aspects and that, in fact, there is an undersupply of ideas addressed to the needs of actual policy-making and present-day government institutions. In particular, research in this area should take into account what is feasible in a political context. As has been pointed out in Bohm (1979), this seems to require meeting at least an 'intelligibility' condition and a 'verifiability' condition, i.e.

(i) the proposed decision mechanism must be *simple and easily understood* by ordinary people; otherwise, it will not work and it is unlikely to be adopted by hesitant, or even hostile, politicians;

(ii) if a mechanism cannot be *guaranteed* to elicit truthful WTP statements, it must be possible to *check* the actual extent of misrepresentation.

The background to the 'verifiability' condition (ii) should perhaps be explained a little further. First, it should be noted that the demand-revealing property of most methods is vulnerable to the formation of coalitions — and guarantees that coalitions cannot be formed can rarely be given. Second, each new practical application of a given method may run into problems of political controversy and may therefore be difficult to evaluate without an instrument to ascertain the extent of truthful reporting. More specifically, those who do not like the outcome may try to disqualify it by making unsubstantiated but also irrefutable references to an allegedly extensive misrepresentation of preferences. Given such risks, politicians are not likely

to accept the use of a demand-revealing method for decision-making on important issues unless its possible failure to reveal preferences can be observed. In other words, for a method to be politically acceptable, it is not enough to have shown that people who in a strict, but limited, sense are *rational* will truthfully report their WTP.

Few, if any, of the methods proposed in the literature meet the 'intelligibility' and 'verifiability' conditions [see Bohm (1979) and Johansen (1981)]. The Clarke/Groves and Ledyard method, however ingenious its basic principle, is not constructed in such a way that it could meet the latter condition. And it is unlikely, although this is a matter of belief, that it even comes close to meeting the former condition. An obvious reason for this insufficiency is, of course, that the two requirements are far from easy to fulfill or that they may be impossible to fulfill without making important sacrifices in other respects. We now turn to an approach which seems capable of meeting the intelligibility as well as the verifiability conditions without sacrificing to much in other important respects (accuracy, costs of administration, equal treatment of equals, etc.).

## 3. The interval method

The 'interval method', discussed in Bohm (1979) is described in more detail below. Its general characteristics are as follows. Instead of using one possibly demand-revealing procedure for eliciting WTP responses from potential consumers of a public good under consideration, two simple, straightforward procedures are employed *simultaneously*. A random sample of, say, 50 percent of the respondents is placed in a position that would give (preferably weak) incentives to *under*-report individual WTP and the remaining respondents are subjected to a situation that would give (preferably weak) incentives to *over*-report individual WTP. Multiplying the average individual WTP from each group by the number of potential consumers and allowing for a sampling-error correction, we get a lower and an upper bound on the true aggregate WTP (a lower bound from the first group which is subjected to incentives to under-report, but not to over-report, etc.). Given the criterion that the good is to be produced if true aggregate WTP exceeds production costs, the good will be produced if the interval which contains the true aggregate WTP as a whole exceeds the cost figure. Similarly, the good will not be produced if the interval falls short of the costs. Moreover, if the interval is small *and* the cost figure ends up inside the interval, the natural (or agreed upon) interpretation is that benefits approximately equal costs and that production is a matter of indifference.

If any of these three situations occurs, the information base will be no worse than if an exact, true WTP figure could be arrived at. One reason why this is so is that we are dealing with a discrete public good. (The case of divisible public goods, requiring the revelation of WTP functions, or at least

WTP for several alternative quantities, is referred to the science fiction department for the time being.) Another reason is that we took the WTP interval to be sufficiently small.

When the interval turns out to be very wide, i.e. when an extensive misrepresentation of WTP is *revealed* to have taken place, we have a problem if the cost figure ends up well inside the interval; in this situation we cannot know if true WTP exceeds or falls short of costs. However, when this is the outcome, the rule could be to return the issue to the politicians, to be determined in the same way as would otherwise have been the case. Thus, no disadvantage (aside from 'referendum' costs) would result from using the interval method, but it is possible that a clear-cut answer would be obtained, as suggested in the preceding paragraph. In addition, the method does not contain complicated 'tricks' to make demand revelation a dominant strategy. Thus, it may possibly meet the intelligibility condition (for details, see section 5 below). And it does not have to rely on respondents 'behaving as they should', as significant misrepresentation would be revealed by this approach, hence meeting the verifiability condition.

## 4. Requirements for trial applications to provide relevant information

The interesting question, of course, is how wide the interval is likely to be in real applications. If the design of the interval method and existing incentives are such that a narrow interval is likely to emerge, the method would be more useful than if the interval is likely to be wide. The main purpose of the case study to be discussed below was to estimate the size of the interval in situations where a real public good is to be produced, depending on the actual responses by potential consumers and where consumers, if the good were to be produced, would have to pay according to prespecified rules.

Now, in order for a trial application to give reliable information about the method used, i.e. information relevant for future applications of the method for actual decision-making concerning public goods, it seems necessary that the following requirements be met.

(a) Information given to the respondents should be *correct*. If false information were given — to reduce experimentation costs or to make the good appear more attractive to the respondents than it actually is, etc. — this could become obvious to the respondents and others after the experiment was completed, and hence jeopardize future use of the method.

(b) All *relevant* information about the goods as well as the decision-making and payment procedures should be provided. Not only might the responses otherwise be biased, but in addition, negligence in supplying relevant information would endanger future use of the method. One information item of particular relevance is the existing incentives to

misrepresent preferences, whenever such incentives are present. If these incentives are not brought to the attention of the respondents, future applications may differ significantly from preceding ones due to the extent that respondents are aware of the incentives.

(c) Respondents should be given enough time to consider the question posed and be allowed to speak to whoever they would like to consult before responding. This also means that they should be given the opportunity to form coalitions if they would like to do so. What this amounts to is an attempt to create a situation which resembles, as closely as possible, an open referendum. Otherwise, the method may not be able to survive for long in a democracy.

(d) To improve the performance of any method in which there are possible incentives for misrepresentation, measures should be taken a priori to reduce the inclination to give in to such incentives. This is not as important for the interval method suggested here as it is for approaches aiming at a direct estimate of the true WTP. However, it might reduce the size of the interval, and thus increase the accuracy of the method. But such measures should be undertaken in connection with the interval method only as long as they do not outweigh the existing unidirectional incentives presented to each of the two groups. (For example, the group which is confronted with incentives to under-report their WTP might be subjected to checks which can prevent strong responses to the existing incentives, but should not be confronted with additional incentives which by themselves would work in the direction of an overstatement of WTP.) The important aspect here, of course, is that such measures should be chosen so that they can be used in future applications. The following two counter-strategic measures (used in the test discussed below) can be taken as examples:

(1) Respondents are told that their WTP statements are not anonymous but will be made public — the point here is that other people (friends, neighbors, etc.) may have an idea of what the respondent actually thinks the good is worth, which (or nonanonymity in itself) could make him refrain from extensive misrepresentation; this measure is less appropriate, however, if the public good is such that WTP statements are likely to be judged by others from a solidarity or charity point of view.

(2) Respondents are informed about (i) the importance of being able to base collective decisions on consumer valuations instead of politicians' valuations or their interpretations of consumer valuations, and (ii) the implications of giving in to the mirepresentation incentives, namely that decision-making would be based on unreliable data. This is not to say that such information about the 'social value' of certain conduct (truthful reporting) actually would influence behavior. But the idea here is the same as in general elections when people are told that certain conduct — voting *participation* — is essential for a well-functioning democracy and that they

should overcome the 'free-rider' incentive to abstain from voting (i.e. accept to bear voting costs for, practically speaking, zero influence on the election outcome). Some 50–90 percent of the electorate in democracies appear to be influenced by this kind of consideration.

Before presenting the nonlaboratory application of the interval method, it should be mentioned that certain aspects of the method were tested in a laboratory — but nonhypothetical — experiment in 1969 [see Bohm (1972)]. The relevant characteristics and results of that test were briefly as follows.

A representative sample of the Stockholm population (some 200 persons) was invited to evaluate, ex ante and in money, the opportunity to watch a TV show not yet shown to the public. The participants were to be shown the program on closed-circuit TV if their aggregate WTP exceeded the given costs of showing the program, and if so, were to pay in a prespecified fashion, varying among (at least) four groups of participants as follows:

- people were to pay their stated WTP;
- people were to pay a percentage of their stated WTP such that costs would be covered exactly;
- people were to pay a (low) flat fee;
- people did not have to pay, the state TV company would pay, but their WTP responses would still determine whether the commodity would be produced or not.

There are clear incentives for under-reporting in the first two groups and clear incentives for over-reporting in the last group. However, no significant differences in average stated WTP emerged from the four randomly selected groups, indicating that the belief in misrepresentation of preferences may have been exaggerated, at least when counterstrategic arguments of the type mentioned under (d) were present.

Given these results, it seemed appropriate to investigate whether or not statements from groups with opposite incentives to misrepresent WTP would deviate when (i) larger groups of people were involved, (ii) people were not paid for participating in the test (as they were in this case), and (iii) a real-world public good was involved. Such a nonhypothetical and nonlaboratory test — formalized as an application of the interval method — required cooperation with a government or some voluntary organization, in the hope that they would let the production of a nontrivial public good be determined by such a test.

In 1982 a chance for a real-world test appeared. It concerned a public good in the form of *access* to a future market for a service, where access was nonrival but excludable. The beneficiaries in this case were local governments, not ordinary consumers. But the revelation problem refers to all decision-makers for whom misrepresentation incentives exist and can be expected to influence WTP reponses. Local governments can also be assumed to belong to this category.

## 5. The test

### 5.1. Background

In the general atmosphere of trying to reduce government expenditures, the Swedish government appointed a committee to find ways of cutting down on the 'extensive' production of statistics. One of the committee's tasks was to find instruments for evaluating costs and benefits of statistics produced by the government and to investigate new ways of financing the production of statistics otherwise made available at no cost to users.

The committee, which included members of the Parliament from the political left to the political right (Social Democrats to Conservatives), agreed to let the fate of a project proposal from the Central Bureau of Statistics be determined by using the interval method. The project involved detailed statistics on housing in Sweden, which were assumed to be of interest to local governments (for their housing policy, evaluation of master plans, energy-saving measures, etc.). The basis for the statistics was a nationwide census which contained, among other things, a specification of certain characteristics of all dwelling units. If the project was to be carried out, the Bureau would arrange a system for data processing and data presentation, the overhead costs of which were estimated at $40 000 (SEK 200 000). Once this system had been established, a local government could order data for its own jurisdiction at an additional cost depending on what data it wanted. Thus, a kind of two-part tariff was involved, with fixed costs to be met by payments linked to the WTP responses, provided the project was accepted (aggregate WTP > fixed costs), and variable costs to be covered by user charges.

### 5.2. Payment rules and misrepresentation incentives

The local governments were informed that the project would be carried out if their WTP statements for access to the project services exceeded the fixed costs. The 279 local governments in Sweden were stratified with respect to population size. Half were distributed to group 1 and the other half to group 2 by a random selection process. Group 1 was to state its WTP in a contract and if the good was to be produced, it would pay a *percentage* of its stated WTP so that the fixed costs would be covered exactly. (This percentage figure could be determined only after the responses had been given, of course.) Group 2 was to state its WTP in a slightly different contract which said that if the good was to be produced, it would pay a small fee, $100 (SEK 500), identical for all who reported a WTP of at least that amount. Those in group 2 who stated a WTP below this figure would not be offered the services at all. The same was true for those in group 1 who reported a zero WTP. This implies that, if the services were to be provided, an exclusion mechanism would be used.

The incentives presented to these two groups were as follows. Respondents in group 1 had a reason to under-state, but no reason to over-state, their WTP (the free-rider incentive). Respondents in group 2 with a true WTP above \$100 would have a reason to over-state, but no reason to under-state their WTP. Those in group 2 with a stated WTP below \$100 would be excluded from consumption as well as from payments; hence, they were not exposed to any particular incentive to misrepresent their WTP. This means that group 2 as a whole was left with an over-statement incentive.

It was considered important that the payment for respondents in group 2 be small, so that a sufficiently large number of respondents would have a WTP in excess of the payment and hence an incentive to over-report their WTP. In this way we would also limit the efficiency loss from excluding consumers with a nonzero WTP in this group. On the other hand, an extremely low payment level, e.g. zero, might jeopardize the quality of the responses because of the reduced checks on carelessness (in group 2) and because of feelings of injustice (in group 1).

The sum of \$100 was considered low enough to fall short of the expected WTP of most local governments. Although it is a negligible part of its budget, the amount cannot be spent by a government without a reason. Since the public official who administers these matters in a local government would be forced to do some paperwork as soon as a payment is involved, he might also be expected to have to make some rudimentary calculation of what the offer would be worth to the government.

It should be pointed out here that local governments in Sweden have been under heavy financial pressure for some time now. This is likely to have created a cost consciousness that did not exist a couple of years ago when additional funds were easy to raise. Thus, if it is argued that local governments behave in a way quite different from that of consumers (to which the misrepresentation issue normally refers), it should be observed that this is probably much less true now than in the 1970s.

### 5.3. *A brief outline of the information process*

The formal respondent was the local government, the board of which would either have to approve the WTP figure to be reported or delegate the matter to one of its employees.

The project was introduced to the local governments in a covering letter signed by the chairman of the government committee on statistics. The letter stated that a new approach for decision-making had been proposed, to be used in deciding whether the given project should be carried out or not. The covering letter was accompanied by guidelines for responding to the question posed and a kind of sales brochure from the potential producer which gave the details of the proposed statistical product and a list of the user charges involved.

The guidelines included information about:

- the amount and origin of the fixed costs;
- the costs involved for those who place an order for services from the statistics package (if made available) later on;
- the rules for the services to be made available for purchases ('if the aggregate WTP for access to the services turns out to cover fixed costs');
- the random distribution of the respondents into two groups with different payment conditions, both of which were explained;
- the reason for the different payment conditions ('to check possible attempts to respond by tactical WTP instead of the true WTP', thus indicating explicitly the different incentives that existed for the two groups);
- the fact that giving in to the existing incentives to misrepresent WTP would jeopardize the value of the investigation and could give another outcome than that corresponding to the true aggregate valuation of the project;
- the nonanonymity condition, i.e. that all responses would be publicized.

(The letter and the guidelines in extenso are available in English from the author on request.)

### 5.4. The results

As shown in table 1, 274 out of 279 local governments responded (98 percent), the same absolute number in each group. The average stated WTP was SEK 827 in group 1 and SEK 889 in group 2. The standard deviation was SEK 1317 in group 1 and SEK 1361 in group 2.

The distribution of the responses is presented in table 2. We note that, of all local governments, 30 percent stated a WTP equal to zero and 9 percent a nonzero amount below SEK 500 ($100). Twenty-seven percent responded with a WTP equal to the only figure mentioned, SEK 500, the payment relevant for group 2. Thirty-four percent revealed a more noticeable interest in the project by stating a WTP > SEK 500, the maximum figure given being SEK 10 000.

Table 1

Aggregate and average WTP (in kronor (SEK), $1 = SEK 5 approx.).

| | Number of governments | | Total WTP (sum of responses) | Average WTP | $\overline{WTP} \times 274$ | WTP interval | |
|---|---|---|---|---|---|---|---|
| | Total | Responded | | | | In SEK | In percent of lower bound |
| Group 1 | 140 | 137 | 113 350 | 827 | 226 700 | 16 962 | 7.5 |
| Group 2 | 139 | 137 | 121 831 | 889 | 243 662 | | |
| Total | 279 | 274(98%) | (235181) | | | | |

Table 2
Distribution of WTP responses; absolute numbers (percentages).

| SEK | | Both groups | Group 1 | Group 2 |
|---|---|---|---|---|
| 0 | | 81(30) | 49(36) | 32(23) |
| 1–499 | | 26 (9) | 14(10) | 12 (9) |
| 500 | | 74(27) | 25(18) | 49(36) |
| 501–999 | 93(34) | 12 (4) | 5 (4) | 7 (5) |
| 1000 | | 34(12) | 19(14) | 15(11) |
| 1001–5000 | | 44(16) | 24(18) | 20(15) |
| >5000 | | 3 (1) | 1 (1) | 2 (1) |
| | | 274(100) | 137(100) | 137(100) |

What may appear to some as a possible version of fairness, at least to noneconomists, is the total fixed costs (200 000) divided by the total number of governments, i.e. some SEK 700. A response of that order of magnitude would be compatible with the view that all governments (although different in population size) should pay the same 'price'. As we can see, very few (4 percent) gave such a response. Another possibility is that respondents calculated 'their share' of the costs as each local government's share of total population in the country. If so, stated WTP per inhabitant would be around SEK 0.025. As we can see from table 3, this does not seem to have happened to any significant extent. Only some 15 percent of the responses were in the neighborhood of that figure (0.02–0.03).

A large number of governments stated SEK 500 (27 percent), possibly as another version of fairness. The argument may have been that, since SEK 500 was the only figure mentioned explicitly and the one group 2 would have to pay if the project materialized, this was *the* 'price' and of course 'you should say you only want to pay the price, not more'. It should be noted, however, that respondents are unlikely to come up with an uneven figure when asked a question such as this. Thus, 189 (about 85 percent) of the 227 who gave a response of at most SEK 1000 stated an 'even' SEK 0, 500 or 1000 figure. In this perspective, the large number of SEK 500 responses may be less remarkable.

Table 3
WTP per inhabitant.

| | Average | Standard deviation | Percentage of respondents in 0.02–0.03 bracket | |
|---|---|---|---|---|
| | | | Of all respondents | Of nonzero respondents |
| Group 1 | 0.039 | 0.087 | 11 | 17 |
| Group 2 | 0.039 | 0.052 | 11 | 14 |

Looking at the responses from each group, group 2 had a considerably larger number of SEK 500 responses than group 1 (36 percent as compared to 18 percent for group 1). The 'price' interpretation referred to above may be particularly relevant for group 2. (But it should be recalled that we explicitly asked for *maximum* WTP and told the respondents why this figure was needed for decision-making.) The large number of SEK 500 responses from group 2 should also be viewed in the perspective of the over-statement incentive provided to this group. Thus, it may be asked whether this could be interpreted as an expression for over-statements of a true WTP below SEK 500. The answer is no, of course. Since governments in group 2 who gave a response of SEK 500 would have to pay exactly this amount, they would certainly have to be assumed to have a *willingness* to pay SEK 500. Thus, if over-statements occurred in group 2, it would have to show up in the frequency of statements exceeding SEK 500. As we can see from table 2, the number of responses with WTP above SEK 500 was not larger in group 2 and did not increase in any significant fashion for higher WTP levels. Instead, we see that the number of responses at each level was very similar for the two groups, except for the SEK 500 and 0 levels (see fig. 1). Group 1 had a much higher percentage of zero responses than group 2 (36 percent as compared to 23 percent). As respondents in group 1 were confronted with an incentive to under-report their WTP, a natural question is whether the relatively large fraction of zero responses in this group could be a reflection of this fact. Here, however, we must note that all respondents stating a

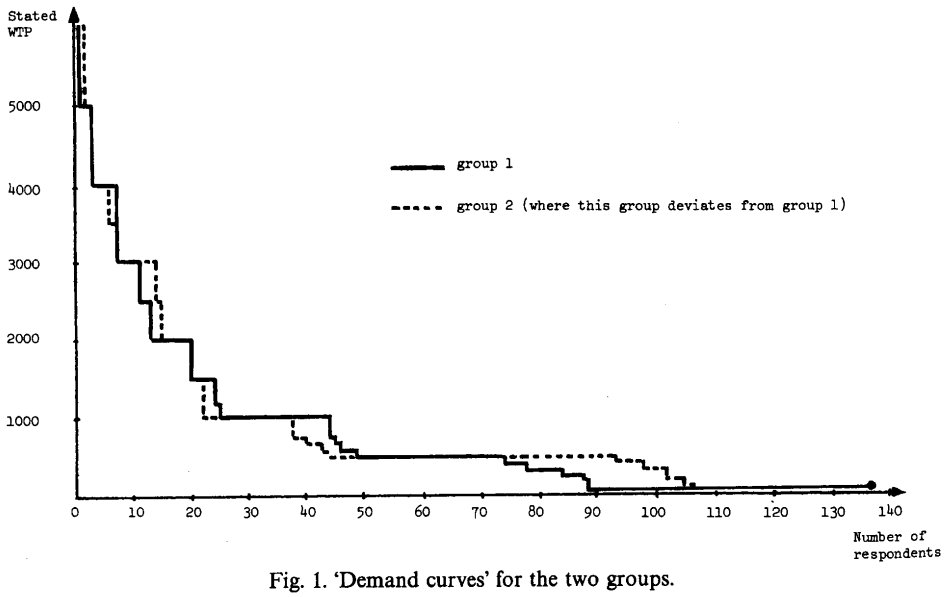


Fig. 1. 'Demand curves' for the two groups.

WTP = 0 were excluded from consuming the good in the future (which they knew in advance). Thus, a zero WTP could not be interpreted as a free-rider response.

### 5.5. *The output decision*

As we have seen, there is little evidence of significant misrepresentation behavior in the two groups. Consequently, the two estimates of average stated WTP were quite close. Multiplying the average figure for each group by the total number of responding governments (274), we get an estimate of the lower bound for the aggregate WTP as SEK 226 700 and an estimate of the upper bound as SEK 243 662 (see table 1). Thus, the directly observable WTP interval is SEK 16 962, or 7.5 percent of the estimated lower bound. But in order to take into account that the composition of the two groups may differ, we have to make an adjustment for the sampling error. With a 95 percent one-sided confidence interval,[1] we get the lower bound as SEK 226 700 − 35 761 and the upper bound as SEK 243 662 + 36 956. Thus, in spite of the fact that the directly observable WTP interval was narrow and clearly exceeded the project costs of SEK 200 000, there is a possibility that the group 1 sample was biased to the extent that the lower bound to aggregate WTP would fall short of the cost figure. As we have seen, there is 'only' a 95 percent confidence that the lower bound exceeds SEK 190 939 (i.e. 226 700 − 35 761). Similarly, there is a 95 percent confidence that the upper bound is at most SEK 280 618 (i.e. 243 662 + 36 956).

This raises the following question with respect to basing an output decision on the results of this test. Should the wider WTP interval including the sampling errors, i.e. SEK 190 939 to 280 618, be *interpreted* to mean that the true aggregate WTP could be anywhere inside this interval? Considering the possibility of the true aggregate WTP coinciding with the lower end of this interval, the implication would be that no respondent in group 1 reacted to the incentive to under-report his WTP and that all misrepresentation took place in group 2 (and vice versa for a true WTP equal to the upper end of the interval). This possibility cannot be ruled out, of course. But given the existence of incentives for misrepresentation in both groups and the fact that these incentives were brought to the attention of all respondents, such an outcome is not very likely. Even less likely is the combination of (1) the samples being extremely skewed so that those with a high true WTP were

[1]The formula used for the lower bound was:

$$N\left(\overline{WTP}_1 - 1.64 s_1 \sqrt{(N-n)/N} \cdot \frac{1}{\sqrt{n}}\right),$$

where $N$ is the total population, $\overline{\text{WTP}}_1$ the average WTP for group 1, $s_1$ standard deviation in group 1 and $n$ the sample size. 1.64 is the $\lambda$-value corresponding to a normal approximation for a one-sided confidence interval of 95 percent.

primarily allocated to group 1 and (b) all misrepresentation taking place in group 2. It is this combination of events that is required for us to treat values at the lower end of the wider interval as a possible true aggregate WTP.

If the wider WTP interval had exceeded (or fallen below) the cost figure, the output decision would have been unambiguous. As this interval in the present case covers the cost figure (SEK 200 000), some *interpretation* of the results is required in order to reach an output decision. On the basis of the arguments just mentioned, the government committee made the interpretation that the true aggregate WTP was unlikely to be in the lower end of the interval and hence decided that the project should be carried out. (We may note that the consequences of being wrong on this point would be without significance as the interpretation would then be one of indifference between having the project carried out or not.)

The project was carried out in late 1982 with all respondents meeting their payment obligations. It should be added that the closeness of the WTP estimates and the cost figure had one final important implication. Given that respondents in group 2 were to pay a low 'symbolic' amount and that the stated WTP of group 1 was low, the total revenue from the two groups could not be made to cover the costs; total revenue amounted to only SEK 160 000 with the stated WTP of group 1 taken out in full. The method is such that a deficit will arise when the outcome is this close and hence the expected deficit has to be treated as a test cost. However, compared to the alternative of tax financing, the financial outcome in this case meant that only SEK 40 000 instead of SEK 200 000 had to be covered by taxes.

## 6. Future use of the method

In this test the directly observable WTP interval was very narrow. Adding the sampling error, the interval was still quite narrow from the point of view of decision-making. Moreover, it should be kept in mind that in many cases where WTP investigations are called for, that is, when the government is very uncertain about demand in relation to costs for a public good project, this wider WTP interval may well deviate significantly from the cost figure, as long as the directly observable interval is narrow.[2]

In the preceding section it was shown how the results were used for decision-making in the particular case at hand. Here, let us sum up how the

[2]Another factor influencing the decision whether or not to use this kind of demand-revealing instrument is, of course, the size of the administrative costs. Repeating a test of the type reported here would require mainly the input of an experienced administrator and an economist of a total of some six man-months. Investigations involving other kinds of respondents would probably require a larger administrative input. But it should be noted that projects for which this kind of information would be demanded are likely to be much larger than the project discussed here. Thus, administrative costs are likely to be small relative to production costs.

method is envisaged to be used in general. If the WTP interval including the sampling error exceeds (falls short of) the project costs, the project is (not) to be carried out. If costs appear inside this interval, the government will have to make an interpretation of the results, primarily by speculating on the relative extent of misrepresentation in the two samples. In general, this interpretation procedure may be expected to result in the project being refuted unless the costs appear in the lower section of the WTP interval.

In future uses of the method an adjustment of the relative size of the two samples may be called for. Making group 1 larger than group 2 could be motivated on three grounds. First, it would reduce the sampling error where such a reduction is most important, given that it is the lower bound to aggregate WTP that determines the outcome. (If the lower bound exceeds costs the project is automatically accepted; if the lower bound falls short of costs it is not accepted, unless the government wants to interpret an outcome where costs barely exceed the lower bound as one where the project should still be carried out.) Second, since respondents in group 1 may feel that they are treated in an unjust fashion (paying up to the level of their stated WTP, whereas group 2 only pays a low symbolic amount), making group 2 smaller might mitigate such feelings. Third, making group 1 larger would reduce the risk of a financial deficit. Given such considerations, group 2 would only have to be large enough to fulfill its function of checking the extent of misrepresentation. Hence, it would then even more clearly play the role of a 'control group'.

For obvious reasons, public-good projects provided by a government to a new demand-revealing method are likely to be of the excludable kind. But the important case for real-world use is of course that of *nonexcludable public goods.* Applying the method to such cases requires that respondents in group 2, who state a WTP below the group 2 fee, be taxed in the amount of this fee. The reason is, of course, that in contrast to the excludable case discussed above, everybody may now benefit from the public good once it is produced. Since a compulsory fee in group 2 will provide an incentive for those with a true WTP below the fee level to under-report their WTP, thus doing what they can to discourage acceptance of the project, the WTP responses below this level may have to be raised somewhat before the WTP interval is calculated. In this way, we would be sure to get an upper bound estimate of aggregate WTP from this group.

Moving one step further towards the more interesting public-good projects, we should consider such cases where the number of potential consumers is very large, including perhaps all inhabitants in a county or in the nation. One possible institutional arrangement here would be to have all the inhabitants divided into two groups and state their WTP for a certain public good in connection with their income tax returns. Another possible approach would be for the government to take a representative sample of the

population, divide this sample into two groups and have them state their WTP. If the aggregate WTP estimated from this sample is high enough, group 2 will pay its fee as usual while group 1 will pay 'its' share of the costs. That is, if average stated WTP in group 1 is twice the average costs, people in this group pay 50 percent of their stated WTP. If group 1 is, say, 1 percent of the population and group 2 is, at most, the same size, this leaves 98–99 percent of the costs to be covered by payments from the rest of the population, i.e. those not included in the sample. Here, the payment could be either a flat fee or an amount that varies with some easily observable variable (income; car, house or other property ownership, etc.) if the sample has revealed this variable to be correlated with stated WTP.

Investigating the WTP of a sample of the population would increase the sampling error term, hence widening the interval. But, in conclusion, looking at the alternative of a mechanism of the Clarke/Groves and Ledyard type, we may first of all note that a sample would be used here as well, for the same reasons as in connection with the interval approach, and hence create a similar uncertainty problem. Thus, no additional difference between the approaches would arise. Returning to the basic aspects discussed in section 1, we have that none of the approaches provides any guarantee against the formation of coalitions; however, the WTP elicited by a mechanism of the Clarke/Groves and Ledyard type would, in the presence of coalition behavior, have an unknown bias as to both sign and magnitude, whereas the method discussed here would have the misrepresentation from coalition behavior or other sources revealed at least approximately by the size of the interval. Moreover, the fact that the aggregate extent of serious misrepresentation is revealed may act as a deterrent to the formation of coalitions. (In the test case discussed above, coalitions could be formed, but the size of the resulting interval did not reveal any significant behavior of this kind.) Even if coalitions could be avoided — which seems doubtful given the political prerequisites for referenda in a democratic society — mechanisms aiming at one nondistorted WTP estimate will have a hard time convincing decision-makers that a nondistorted estimate will in fact be obtained, especially if the device which is supposed to do the trick adds complexity to the whole procedure. In contrast, the interval method reveals the approximate extent of misrepresentation using two simple questions, one of which is put to what might be called a 'control group'. The reason for this division of the sample or the population into two groups is similar to, and presumably no less hard to understand than, the reason for using control groups in, e.g. testing new drugs or other medical therapies.

## References

Bishop, R. and T. Heberlein, 1979, Measuring values of extramarket goods, American Journal of Agricultural Economics 61, 926–930.

Bohm, P., 1972,
Review 3, 55–(
Bohm, P., 1979,
Economics 81
London, 1981)
Brookshire, D.S.
approaches, A
Clarke, E., 1971, M
Green, J. and
Amsterdam).
Groves, T. and Le
Johansen, L., 19
Economics 7, 1
Johansen, L., 198
decision-makir
Marwell, G. and
Economics 15,
McMillan, J., 197
Ng, Y.K., 1979, W
Public Choice, 19
Smith, V.L., 198(
American Ecor
Tideman, N. and
Journal of Poli

Bohm, P., 1972, Estimating demand for public goods: An experiment, European Economic Review 3, 55–66.
Bohm, P., 1979, Estimating willingness to pay: Why and how?, Scandinavian Journal of Economics 81, 142–153, and in: S. Strøm, ed., Measurement in public choice (Macmillan, London, 1981) 1–12.
Brookshire, D.S. et al, 1982, Valuing public goods: A comparison of survey and hedonic approaches, American Economic Review 72, 165–177.
Clarke, E., 1971, Multipart pricing of public goods, Public Choice 11, 17–33.
Green, J. and Laffont, J., 1979, Incentives in public decision-making (North-Holland, Amsterdam).
Groves, T. and Ledyard, J., 1977, Optimal allocation of public goods, Econometrica 45, 783–810.
Johansen, L., 1977, The theory of public goods: Misplaced emphasis?, Journal of Public Economics 7, 147–152.
Johansen, L., 1981, Review and comments: J. Green and J.J. Laffont, Incentives in public decision-making, Journal of Public Economics 16, 123–128.
Marwell, G. and Ames, R.E., 1980, Economists free ride, does anyone else?, Journal of Public Economics 15, 295–310.
McMillan, J., 1979, The free-rider problem: A survey, The Economic Record 55, 95–106.
Ng, Y.K., 1979, Welfare economics (Macmillan, London).
Public Choice, 1977, Special issue on demand-revelation mechanisms 29.
Smith, V.L., 1980, Experiments with a decentralized mechanism for public good decisions, American Economic Review 70, 584–599.
Tideman, N. and Tullock, G., 1976, A new and superior process for making social choices, Journal of Political Economy 84, 1145–1159.