Comments on SP study for ANASE: Brett Day

Table of contents

1. Overall Comments

1.1 Is the work technically and analytically sound?

1.1.1 This report details the design, implementation and results of two valuation exercises carried out on the same sample. The first of those exercises was a contingent valuation (CV) exercise in which respondents were asked directly how much they would be willing to pay to reduce aircraft noise at their home. The second of those exercises was a choice modeling exercise (referred to in the report and, consequently, in this document as an SP exercise) in which respondents willingness to accept compensation for increases in aircraft noise at their home are inferred from their responses in a series of choice tasks.

1.1.2 My overall impression is that both of these valuation exercises have been designed and implemented to a very high standard.

1.1.3 In general, the analysis of the SP data survey is also of a very high standard with model development following a logical progression and resulting in a specification that captures the key elements of the noise valuation function in a succinct and theoretically defensible manner.

1.1.4 In contrast, the analysis of the CV data does not pass muster. I have some serious reservations concerning the specification of the econometric model and the choice of estimator used to derive those parameters, reservations that have been expressed at previous rounds of review.

1.2 Do the draft conclusions follow from the analysis?

1.2.1 The key conclusions of this research are four fold;

- The identification of an absolute value for a 1dB Leq change in noise exposure at home from the CV exercise.
- The identification of a series of relative weights for different types of aircraft over-flights from the SP exercise.
- The identification of a series of relative weights for noise exposure experienced at different times of the day from the SP exercise.
- Identification of the implied relationship between disutility and the number and sound exposure level of aircraft over-flights.

1.2.2 As documented in sections 2.3 and 2.4 of my comments, I have reservations concerning the modeling of the CV data. Accordingly, I am of the opinion that the values reported for a 1dB Leq change coming from that study are not reliable.

1.2.3 The analysis of the SP exercise has been carried out to a high standard. As such the relative weights placed on over-flights of different aircraft types are empirically defensible and seem to be of plausible magnitude.

1.2.4 As discussed in comment 3.4.11, I believe that the parameters used by the consultants to establish the relative annoyance of aircraft noise at different times of day are confounded by the fact that they do not account for presence or absence from the home. In a previous note, I have outlined how I believe such confounding could be handled in the analysis. Accordingly, I do not believe that the relativities reported here are accurate.

1.2.5 Whilst I am no expert on the measurement of noise, I have reviewed the analysis regarding the relationship between the Leq measure of noise and the disutility of noise implied by the SP study, and believe it to be sensible.

1.2.6 Worryingly, the consultants find a very large disparity in the values for noise avoidance implied by the SP study and those recorded in the CV study. In my comment 3.4.9 I make clear that I do not believe the disparity in values is a result of poor experimental design or of study implementation on behalf of the consultants. However, I would like to see the consultants provide a more cogent explanation for the difference in values even if that explanation is to question the efficacy of one or both of the valuation methods. The current appeal to a part-whole effect does not hold water (see my comment 3.4.8).

1.2.7 The consultants appear to indicate that the CV and SP results might be combined such that the absolute values for reductions in aircraft noise exposure are derived from the CV exercise whilst the relativities with regards to different types of aircraft and different times of day are derived from the SP exercise. Unfortunately, the consultants provide no indication as to how the relativities might be used in practice (see my comment 3.4.10). I reserve judgment on the appropriateness of this proposal until it is made clear how the two sets of results are to be applied in practice.

1.3 Optionally, suggestions for where further research might be useful

1.3.1 I make numerous suggestions in my comments as to how the current analysis might be minorly improved.

1.3.2 One very important issue that is ignored by the current study is that of self-sorting with regards to noise exposure. In particular, households with relatively high tolerance for noise will choose to locate in areas with relatively high levels of aircraft noise exposure since they are compensated through cheaper housing costs. Conversely, noise intolerant households will be prepared to take on higher housing costs so that they can locate in relatively quiet areas.

1.3.3 Indeed, one could convincingly argue that the current study is fundamentally flawed because it does not account for self-sorting. In particular, the CV results that provide the key valuation finding of the research are obtained by measuring the relationship between a householdsâ current exposure to aircraft noise and their stated willingness to pay to remove all aircraft noise. Underpinning that analysis is the assumption that there are no systematic differences in the noise tolerance of households at different levels of current exposure. As I have just argued that is not the case. Those households in noisy locations will likely be noise tolerant and express relatively lower values for noise reductions than the general population. Conversely, those households in quiet locations will likely be noise intolerant and express relatively higher values for noise reductions. Overall, the impact of self-sorting is to flatten the observed relationship between WTP and current noise exposure. As such, we would expect that the values for a 1dB reduction in noise exposure derived from the CV exercise in this study will be biased downwards.

1.3.4 In addition, it is important to note that the values reported in this study do not include the price premiums already paid for quiet environments through differences in housing costs in quieter areas.

1.3.5 A much more ambitious study could address these issues by combining a study along the current lines with a study investigating the role of the property market in compensating for noise exposure. Like the current research, that study would be carried out across a number of locations. Hedonic pricing techniques would be applied in each location to isolate the price paid for noise avoidance through differential housing costs in each property market. As well as yielding the possibility of deriving a revealed preference measure of the value of noise, information from the hedonic pricing study could be used to correct for the bias introduced be self-sorting in an accompanying SP study.

2. Specific Comments: CV study

2.1 Scenario

2.1.1 The CV study presents respondents with a scenario (11.1.5) in which they are asked to consider how much they would be willing to pay to eliminate their exposure to aircraft noise at home. The scenario is rather abstract requiring respondents to consider moving to a property identical to their present home, but suffering no exposure to aircraft noise. They are also informed that their living expenses at this new property would be Â£X more a week.

2.1.2 Designing CV scenarios confronts researchers with a difficult trade-off. On the one hand, the scenario can offer the non-market good through a plausible provision mechanism associated with a realistic and credible payment method. Such scenarios are desirable in so much as they maintain realism and mitigate possible hypothetical bias. Realism however, is not always easy to achieve and scenarios that are very realistic may prompt high levels of protest responses. On the other hand, the scenario can be kept deliberately vague in the hope that respondents can be encouraged just to consider the key trade-off between money and provision of the non-market good. Clearly, such âvagueâ scenarios are much easier to construct (e.g. there is no need to explain how aircraft noise is eliminated or why a payment is necessary, to whom it is made and what it is spent on), but may be criticized for a lack of realism, a failing that might induce hypothetical bias or induce large variance in respondents answers as they plug the holes in the scenario with their individual interpretations. In this case, the scenario developed by the consultants is of the âvagueâ type.

2.1.3 Overall, I would acknowledge that there are real difficulties in developing realistic scenarios and payment vehicles in the context of aircraft noise reduction projects and suggest that the scenario is adequate for purpose. Having said that, I am surprised that we have not been presented with a brief account of how this scenario was selected in favour of alternative scenarios. Also, I would consider a discussion of how the scenario was received by respondents in focus group testing an integral part of the write-up of a CV survey. The respondents should include such a discussion in the final report.

2.2 Elicitation Mechanism:

2.2.1 In the CV study, values are elicited through the use of an initial dichotomous-choice question with a further refinement being achieved through a follow-up payment card question (11.1.6 to 11.1.8). The researchers intimate that evidence exists to suggest that the bid amount offered in the dichotomous-choice question may influence responses to the subsequent valuation question. I have two comments here.

2.2.2 The first is a matter of writing style. In particular, academic and professional writing should support claims of âevidenceâ with references. In this case, the appropriate references might be Herriges and Shogren (1996) or the like.

2.2.3 More importantly, the internal validity of the CV analysis might be called into question if problems of starting-point bias are found to dominate responses. Having highlighted the possibility that such biases might exist, the consultants fail to provide an analysis of whether such patterns are observable in the data. I would recommend that such an analysis is undertaken.

2.3 Data Analysis:

2.3.1 The CV data show some nice regularities with stated WTP correlating strongly with reported annoyance levels (11.1.10 to 11.1.15) and with current levels of exposure (11.1.16 to 11.1.18). Both these results are in line with expectations and reassuring regarding the quality of the data.

2.3.2 However, the analysis of the CV data is superficial and, for want of a better word, unsophisticated. I indicated my misgivings concerning the CV analysis in previous comments and am disappointed to see that little has been done to improve the quality of the analysis in the draft final report.

2.3.3 The consultants appear to have engaged in a simple pattern-fitting form of statistical analysis employing ordinary least squares (OLS) as a default estimator. As far as I can tell, there has been little consideration of economic theory in developing their model or econometric theory in selecting the appropriate estimator.

2.3.4 A âbestâ (see comment 2.3.12) model is presented as one in which WTP is modeled as a function of current noise exposure and a dummy variable identifying those with income in excess of Â£40,000 per annum. The justification for this functional form is not made clear.

2.3.5 On what grounds do the consultants justify the use of a functional form that is linear in current noise exposure? This choice results in an estimate of marginal WTP that is independent of noise exposure, a result that runs counter to the predictions of economic theory that envisage marginal WTP as decreasing in quantity.

2.3.6 Also, there are no theoretical grounds for presupposing a structural shift in WTP at an annual income of Â£40,000. Surely, some continuous transformation of the income variable would be more appropriate? I am concerned that an unjustified modeling decision may translate into policy recommendations that are meaningfully different for two artificially defined income groups (as seems to be the recommendation in 11.1.77).

2.3.7 The consultants appear to have made no attempt to account for features of their data which might impact upon the appropriateness and/or efficiency of using an OLS estimator. In this regard, I have two particular concerns.

2.3.8 First, the data in Figure X.2 would seem to indicate that there is a large degree of heteroskedasticity in the data with a greater variance in individualsâ stated WTP at higher levels of current noise exposure. At the very least the standard-errors of the model parameters should be calculated using a heteroskedastic-consistent estimator. Better still, the consultants should employ an estimator of the standard errors that also accounts for the possible correlation in errors between respondents from the same common noise area.

2.3.9 Second, 63% of respondents declare a WTP of zero. In such circumstances it seems very unwise to use an OLS estimator. In particular, if one envisages a structural model of value formation as being one in which an individualâs WTP only becomes positive after some threshold level of noise is exceeded, then the OLS estimator will be inappropriate. Indeed, it is easy to show that the OLS estimate of the slope parameter (defining marginal WTP for noise avoidance) will be biased downwards. I highly recommend that the consultants are asked to re-estimate the model using an appropriate estimator. I direct them to the âCVM data analysis chapterâ in the book by Bateman et al. (2002).

2.3.10 In addition, 13% of respondents are identified as âprotest votersâ and excluded from the analysis. It is standard practice to compare the characteristics of this group to those providing âvalidâ responses in order to examine whether the sample is biased by their exclusion. A particular concern would be if protest voting correlates with current noise exposure?

2.3.11 The consultants do not report whether their analysis makes use of the sample weights. Clearly, it should.

2.3.12 Finally, I draw attention to the lack of rigor in the use of statistical language in the discussion of results (11.1.20 to 11.1.25). I would ask that the consultants consider rephrasing, or expanding upon, the following statements:

- 11.1.20: What exactly is meant by ânoticeableâ in the statement âdid not yield a noticeable improvementâ? Are there statistical grounds on which to reject the arguably more theoretically consistent quadratic model?
- 11.1.24: How exactly is âimportantâ defined in the statement âmost important variable affecting amount willing to payâ?
- 11.1.24: What defines âbestâ in the statement âthe best individual-level modelâ?
- 11.1.25: What makes something âstatistically validâ as expressed in the statement âstatistically valid relationship between willingness to pay to eradicate aircraft noiseâ?

2.4 Results:

2.4.1 There is no theoretical reason to support the use of a model that averages WTP responses by area (11.1.22). I support the consultantsâ decision to take forward results from the individual level model.

2.4.2 With the caveats that;

- the functional form of the model requires justification especially with regard to the linearity in current noise exposure and the artificial introduction of a structural shift in WTP for households with income in excess of Â£40,000 per year (see comments 2.3.4 to 2.3.6),
- the parameters of this model may be downwardly biased through the use of an inappropriate estimator (see comment 2.3.9),
- the standard errors of the model may be incorrect since they do not account for heteroskedasticity or spatial autocorrelation (see comment 2.3.8), the consultants derive estimates of marginal WTP in an appropriate manner from their estimated model.

2.4.3 The marginal WTP estimates recorded in 11.1.24 are reported without standard errors. It is a fundamental requirement of any statistical analysis to report the standard errors of estimates. These must be calculated and reported such that the reader can ascertain the precision of the estimates.

2.4.4 As I have pointed out previously, it is categorically wrong to state that recent hedonic price studies assume that WTP is zero below 55dB Leq. In actuality, those studies find that there is no distinguishable impact of noise exposure on *house prices* below 55dB Leq. Households may well value noise reductions below this level, but conditions of supply in the market (i.e. a large housing stock with low noise exposure) might mitigate against substantial price premiums for marginally quieter properties below 55db Leq.

3. Specific Comments: SP study

3.1 Scenario

3.1.1 I agree with the consultants, that the SP (choice modeling) approach is well-suited to the task of identifying the separate effects of aircraft type and time of day on values for reductions in noise exposure (11.1.29). I fully support the adoption of this methodology for this task.

3.1.2 The scenario presented to respondents in the SP study (11.1.34) is far more specific and realistic than that used in the CV study (see comments 2.1.1 to 2.1.3). In particular, respondents are informed that households living near to an airport will receive a compensatory grant. They are then asked to make a series of choices. Each choice presents respondents with three options with each option offering a package coupling different numbers of overhead flights (by different types of plane at a particular time of day) with a specified grant. I find the scenario convincing and believe it is suitable for purpose.

3.1.3 Notice that the SP study differs fundamentally from the CV study in eliciting a willingness to accept (WTA) value rather than a willingness to pay (WTP) value. As noted in 11.1.32, pretesting of the WTA questions revealed that they were less likely to invoke protest responses than WTP questions. Given this finding, I am surprised that the consultants did not attempt to provide a consistent scenario in both CV and SP studies based on a WTA question. This difference in scenarios calls into question the compatibility of the two studies.

3.1.4 The consultants have done an excellent job in designing and conveying the choice tasks. The choice tasks are of cognitively manageable proportions, they include graphic information to aid respondent understanding (11.1.37) and have been tailored to realistically reflect aircraft traffic at each respondents residential location (11.1.35).

3.1.5 Another highly commendable feature of this study is that respondents have been played recordings of locally recorded aircraft noise (11.1.36). This provides respondents with a common understanding of the noise generated by each aircraft type portrayed in the choice tasks.

3.1.6 We are given some details of the thorough pretesting that has informed the development of the SP scenario and choice tasks (e.g. 11.1.32, 11.1.36, 11.1.39, 11.1.40). Again the consultants have done an excellent job in this regard. Indeed, I am aware that the consultants have carried at least one other detailed pretest of the possibility of ânumbers gamingâ. This too was an excellent piece of work and its findings should be recorded (however, briefly) in the final report.

3.2 Experimental Design

3.2.1 The consultants faced a number of issues in deciding upon an experimental design. In particular, they have had to deal with recovery of the noise to money trade-off disassociating that value by both aircraft type and by time of day. The consultants have addressed this problem using a âtwo-stageâ design (reported in 11.1.40 to 11.1.45).

- The aircraft-money trade-off is handled in the standard way through choice sets involving options offering different numbers of aircraft flights paired with different levels of grant. I assume the options are defined according to some standard main-effects design.

- The time of day effect is captured in the design by setting each choice in one of six different four hour divisions of the day. Accordingly, all options in one choice set pertain to the same time of day, but this time of day may differ from one choice set to the next.

3.2.2 The efficacy of the design in recovering the parameters of the utility function has been tested using software that simulates individual choices based on reasonable assumptions concerning the true values of those parameters. In the main, I would conclude that the consultants have done a good job in developing their experimental design. They have applied appropriate methods and constructed a design that is fit for purpose. I have two comments to make on the design.

3.2.3 First, (I assume) a main-effects design has been used in constructing choice sets that identify the aircraft-money trade-off. As such the design does not allow for the possibility of interaction effects between numbers of aircraft of different types. Whilst I believe this to be perfectly acceptable given the sample size limitations of the study I would like to see this fact acknowledged in the report.

3.2.4 Second, the design does not allow individuals to report their preferences in choices that involve directly trading-off noise in one time period against noise in another. Accordingly, conclusions drawn regarding sensitivity to aircraft noise by time of day (Section 11.2) are inferred indirectly from comparison of answers in different choice sets. A useful extension of this study would be to derive time-of-day relativities from direct trade-off questions and compare them to those estimated indirectly in this study.

3.3 Data Analysis

3.3.1 The model developed by the consultants (11.1.46 to 11.1.57) follows a logical progression and results in a specification (11.1.55) that captures the key elements of the choice function in a succinct and theoretically defensible manner.

3.3.2 As discussed previously with the consultants, a possible further step would be to allow for the scale of the error term to differ across sites. I was under the impression that the consultants were planning to estimate such a model. Presuming this was done, what were the reasons for not reporting this more general version of the model?

3.3.3 The consultants employ a logit model to estimate the parameters of their model. Whilst this is certainly not state of the art, the logit model is an appropriate estimator for this analysis.

3.3.4 I believe the consultants have corrected the standard errors of their parameter estimates to account for the fact that multiple responses are recorded for each respondent. This is not documented in the report. It should be.

3.4 Results

3.4.1 The results of the pooled multiplicative model are reported in Table 11.2. The coefficients follow plausible patterns in terms of their relative magnitude and are, in the main, statistically significant. Estimates of marginal WTP are calculated from the model parameters in an appropriate manner (11.1.63).

3.4.2 The consultants also report partial results for a model in which the parameter on the money (grant) variable is segmented by household income. In contrast to the specification used in the CV analysis (see comment 2.3.6), separate parameters are estimated for each of five income partitions; <Â£10k, Â£10k to Â£20k, Â£20k to Â£40k, Â£40k to Â£50k, >Â£50k. The coefficients follow a logical sequence in their relative magnitude.

3.4.3 One concern is the somewhat anomalous appearance of a middle income category spanning Â£20k when each of the other partitions identify only Â£10k ranges. As per comment 2.3.6, the consultants need to justify their choice of income partitioning.

3.4.4 It is not made clear how the âprotest votersâ group is defined in the context of the SP study. Are these the protest voters identified in the CV exercise? Why is that it makes sense that these individuals behave like low income households? The discussion of this variable requires expansion and clarification.

3.4.5 The model reported in 11.1.68 makes the simplifying assumption that the welfare impact of an over-flight from each aircraft type is identical at all sites. The consultants acknowledge that this is a gross simplification given that the noise impact of an aircraft over-flight differs greatly across sites. All the same, this simple model allows the calculation of WTP values for the avoidance of one over-flight of each aircraft type. Whilst these values do not reflect differences in the noise from each such flight across sites nor do they reflect differences in values across income groups, they do provide a general indication of the size of the WTP estimates coming out of the SP study.

3.4.6 What are the units of the WTP values in 11.1.70? I presume that they are values per month but this not made clear.

3.4.7 The WTP estimates are reported without standard errors (11.1.70). As per comment 2.4.3 all such estimates should be accompanied by standard errors. It is imperative that the consultants provide these measures of estimate precision.

3.4.7 A crucial finding of the SP analysis is contained in 11.1.72 to 11.1.73. Here the consultants report that the implied values from the SP analysis âseem very highâ. The consultants compare their values with those of an alternative SP study and find the ANASE results to be between 3 and 10 times greater. They also indicate that if the ANASE SP values are used to calculate a value for the complete elimination of aircraft noise at each site (as per the values derived from the CV study), then the resulting estimates are unrealistically large. Unfortunately, we are not provided with details of these calculations. Whilst the comparison may not be pretty I think the consultants are beholden to report these values.

3.4.8 In 11.1.74 and 11.1.75 the consultants explain the anomalously high values derived from the SP study as resulting from a âpart-wholeâ effect. That is to say, that the values estimated in the study only indicate the value placed on the removal of the marginal over-flight, extrapolating to the removal of many or all over-flights by multiplying up the marginal figure may be inappropriate. Whilst this observation is true, I cannot accept that such an explanation is plausible in this case. In particular, in 11.1.70 we are informed that the marginal value for the removal of one Jumbo over-flight in the middle of the day is Â£6.36 per month. That translates into a value of Â£76.32 a year (the value for the removal of one night time Jumbo flight is Â£172.68 a year). The consultants acknowledge that the reduction in one Jumbo over-flight per day translates into a reduction in noise exposure substantially less than 1dB Leq. However, evidence from the CV study indicates that the same respondents value a 1dB Leq reduction in noise exposure at between Â£11 and Â£18 a year. The data would seem to indicate that respondents value a

larger good (a 1dB Leq reduction in noise) less than a smaller good (a reduction of 1 aircraft over-flight a day). Clearly, the disparity in these results cannot be explained as a part-whole effect.

3.4.9 The disparity in findings is disheartening, though I would contend that this result is not a consequence of a poor experimental design or substandard implementation of the study on behalf of the consultants. Quite the contrary. I believe this to be a very well-executed study maintaining excellent standards in experimental design and implementation. Accordingly, I would like to have heard the consultantsâ thoughts as to why the SP study returned such high values. Are those values a consequence of respondentsâ lack of understanding of the good being offered (as intimated in 11.1.74)? Is the problem a generic one with SP (choice modeling) elicitation of values? Could design changes be implemented that would have mitigated the observed over-valuation?

3.4.10 In 11.1.77 the consultants suggest taking the CV estimates (Â£11 and Â£18 a year) as defining the absolute value of a 1 dB reduction in aircraft noise exposure. They then appear to propose that those absolute values may be disaggregated using the relativities implied by the SP study (11.1.76 and 11.1.77). It is far from clear to me how this would work. What exactly should we take as being the value of a reduction in noise resulting from a Jumbo in the middle of the day? Or a tail-jet at night? Indeed, the consultants are very unclear as to how the outputs from the SP study should be applied in practice. For this research to be of use to the DfT these issues need to be considered and clear advice presented in this report.

3.4.11 As I have indicated in a previous note, I am not in agreement with the statement that the time period coefficients can be interpreted as indicating different sensitivity to the same aircraft noise in different periods (11.2.4 and 11.2.5). The key thing to note here is that those parameters represent the average (relative) sensitivity to noise at each time period across the entire sample. As shown in Figure 2.1, up to 50% of the sample may be absent from the home in any one period. Since it seems reasonable to assume that individuals realize close to zero benefit from aircraft noise reductions at their homes when they are in another location, the trend in the time period coefficients may merely be an artifact of the proportion of individuals present in the home in each period. To be clear, the reported results may well be consistent with a model in which sensitivity to noise is identical across all periods of the day but less people are affected by noise at the home in the middle of the day. Since the measurement of the relative sensitivity to aircraft noise across different day is one of the key objectives of this study, I feel it is important that the consultants address this issue more thoroughly. I direct them to the model specification that I outlined in a previous note in which separate time of day parameters were estimated specific to periods âin the homeâ and âout of the homeâ.

3.4.12 Section 11.3 presents an analysis concerning the relationship between the Leq measure of noise exposure and the implied disutility of that noise exposure derived from the SP study. Whilst I am no expert on the measurement of sound, I have reviewed this analysis in more detail previously, and it appears sensible. My one comment is that the presentation of this material in the draft final report is so terse as to make it very difficult to follow the progression of the analysis to the key recommendation presented in 11.3.14.

References

Herriges, J. and J. Shogren (1996) Starting Point Bias in Dichotomous Choice Valuation with Follow-up Questioning, *Journal of Environmental Economics and Management*, 30, 112-131.

Bateman, I.J., R.T. Carson, B. Day, M. Hanemann, N. Hanley, T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Ãzdemiroglu, D.W. Pearce, R. Sugden and J. Swanson (2002). *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar, Ltd. Cheltenham, 2002, ISBN: 1840649194, 480 pp.