PEER REVIEW OF THE REPORT:

"ESTIMATING MARINE RECREATIONAL FISHING'S ECONOMIC CONTRIBUTIONS IN NEW ZEALAND"

As prepared by Southwick Associates

Review compiled and presented by

Andy Heinemann: NRB Ltd

Alistair Gray: Statistics Research Associates Ltd

Adolf Stroombergen: Infometrics

June/July 2016

National Research Bureau Ltd, PO Box 10118, Mt Eden, Auckland, New Zealand,
Phone (09) 6300 655, www.nrb.co.nz

CONTENTS

Page No.

A. EXECUTIVE SUMMARY AND CONCLUSION

This review assesses the likely accuracy and usability of estimates of the contribution of marine fishing to the NZ economy. Economic estimates were prepared by Southwick Associates in collaboration with Blue Water Marine Research and Moana Consultants over 2015 and reported in 2016.

1. SURVEY PSYCHOMETRICS

The researchers tasked themselves with obtaining estimates of the dollar spending driven by marine fishing for two segments, viz domestic recreational marine fishers and international marine fishers visiting New Zealand. The first of these was addressed with a marine fisher questionnaire hosted in an online community member base. The second made use of a questionnaire addressed to charter boat operators who reported the perceived local vs overseas compositions of their clients, and accounted for the number of charters, clients and associated fees for the 2014 year. Both leveraged larger formally structured surveys, prepared respectively for MPI (the NPS 2011/12) and MBIE (the IVS) to expand from their surveys to the population.

The reviewers recognised that with modest resources available to them, the researchers had attempted one of the most daunting of behavioural survey undertakings. Fisher recall of spending on irregular activities, like marine fishing, distinguishing the sum spent on the 'most recent' occasion versus other occasions, correctly recalling amounts for several types of spend, apportioning correctly between spends, attributing correctly to marine fishing versus other purposes, and scaling for spent on/by self versus for others, is a far more complex task than the researchers have acknowledged in their chosen questionnaire approach. The questioning style to fishers presents simple lists of categories that may be spent on, with space for a numerical dollar figure to be entered. Greatly under-appreciated is that fishers are not pre-primed to recall either the items, or their dollar value or whether part of the value is attributable to non-fishing. Most recent occasion is likely to fuse into "typical" occasion, and prompted by the item list, a proportion of fishers are likely to enter a spend for that line item dating to the time when they did spend on it, rather than the most recent occasion.

Heroic assumptions around the recall of spends and the way respondents' process their recall also arise in the charter operator survey. Charter/operators are asked to say what proportion of their clients were from overseas and amongst those what proportion were in NZ for the primary purpose of fishing. These again are ambush style recall for which operators have no records and did not anticipate having to recall at all, never mind in quantitative terms.

Reliance on common conversational expression and questioning will likely have waylaid the intention of the fisher survey to identify and quantify marine fishing spending. This arises from the implied but weakly articulated distinction between …

> "… a trip on which you (also) went marine fishing" and "a trip whose primary purpose was to marine fish".

The questionnaire appears to use the former when it asks for "thinking about your most recent trip where you saltwater fished …". Only later in the charter questionnaire do the researchers evoke the critical construct of "primary purpose" but unfortunately only in a call for information the operators would likely need to guess at. Without primary purpose in the recreational fishers' questionnaire there is an under-attribution of spend to all non-fishing drivers of spending on each item. Even with marine fishing merely an associated driver, allocation of spend between purposes of the trip would have taken some deliberative questioning.

We could not rule out mischief in the behaviour of some online respondents. While the researchers thought to protect the survey from members who wished to complete the questionnaire purely for the rewards, avid on-liners know the strategy of screening to exclude eligibles and will try to bypass it.

Respondents had to select marine fishing from a list of options as an activity they took part in, before being allowed forward. Clicking several or all options would be a way to raise chances of getting to participate and getting the reward. This suspicion is raised by the researchers' reporting that …

> "Based on the number of fishers, our analysis (sic, data) does not find that fishers prefer fishing during the summer rather than winter, and thus no strong seasonal trend."

This is strongly counter-intuitive for New Zealand and at variance with the NPS 2011/12 finding that over 70% of marine trips are taken in the summer months. Were respondents filling in trips for winter to present a, to them, credible report and maintain their online panel membership? Data interrogation of the winter trips is indicated.

This reviewer, on the basis of the above appreciation of the marine fisher and charter boat operator questioning does not feel that the data can be relied on to reflect its quantitative purpose.

The psychometric values have not enjoyed due process in the creation of the data gathering tools. If the answers turn out to be close to those from a more systematic approach, this would be fortuitous. The researchers however, have performed a valuable service from a seminal perspective. They have identified the steps and the constructs they relied on and the wording they operationalised with. These mark the difficulties that need development when this topic is next addressed.

2. SURVEY STATISTICS

The aspects of sampling, expansion to population from the gathered sample data, and statistical estimation of error margins associated with the expanded figures, were appraised by a survey statistician.

Found worrisome when used in the context of producing figures for use in the public sector was the non-probability sample from two(?) online internet panels. There is direct experience of such samples producing different estimates for behavioural variables even after re-balancing on a range of demographics. In this application there was no exposition of the generation of the panels, the propensity for an individual to belong, or the size and character of non-response to the marine fisher survey. Propensity scores may have been helpful in moving from self-selection to selection probabilities as in a probability based sample. Helpful also would be a discussion of non-response and why it could be expected that the calibration model could control for both the biases in the sampling scheme and the non-response. Avidity weighting made use of the NPS initial self-classified avidity level, whereas the whole-year monitoring would have provided an Actual Avidity level. This may have had a substantial effect on the estimates.

In summary, both the statistical approach to sampling and the strategy for expansion provide at most an equivocal set of estimates for quantification of the value of recreational marine fishing. Future estimates would benefit by noting the compromises that needed to be made on this project, and would seek to remedy them.

3. SURVEY ECONOMETRICS

The econometric expansions of the findings on spending were appraised by an econometrician experienced in this methodology. The application to this marine fishing data to calculate the direct, indirect, and induced economic effects was considered to conform to generally accepted practice. Viewed favourably in this regard was the step in which spending figures were adjusted for direct imports and GST before the multiplier effects were calculated. Similarly approved was the distinguishing between industries that supply margins and industries that actually produce goods in the allocation of spending, prior to applying the multipliers.

A reservation was expressed in relation to the exposition of the meaning of the economic contribution figure arrived at. Specifically, the reviewer felt the distinction around NET economic contributions was desirable for the appreciation of non-technical readers. Instanced was the precaution that the expansion was not to be read to say that 8,000 people would not be employed, were it not for recreational fishing, or that GDP would be $638m lower.

In overview, the econometric expansion was seen to conform to best practice. The figures provided for the analysis however, bear a high level of uncertainty which econometric analysis does not presume to modify.

B. PSYCHOMETRIC CONSIDERATIONS

1. VALIDATION

A high level, umbrella validation is offered the reader by reference to four overseas studies with the same economic valuation purpose. Here it is argued that "minimising the differences between studies (chosen) strengthens the overall comparability of the results presented" (p91). The location of three of the studies in Australia where it is reasonably argued that opportunities and conditions are similar is called in support of this position. A United States marine angler study is the fourth and is included on the basis of similar methodologies having been employed.

The counter to such value as this comparability may have lies in the risk of common methods error. This is where a narrow pool of researchers, following relatively common methodology tend to produce answer sets that display what looks like a family of studies, reinforcing each other. Use of recall at a demanding level from the fishers, and of proportion/percentage style estimates of behaviour from participants in the studies warns of the risk of measurement error in the common approaches. Part of the similarity in findings may be attributable not so much to common spending levels as to common limitations on fishers' ability to recall (numeric) dollar numbers and common limitations on their ability to apportion spend (numerically).

Validation that calls on other research directed toward the same purpose would better serve if quite different techniques at key method steps, could be shown to produce comparable results.

Other validations of the basic data captured are not presented or discussed. The questionnaire forms and their operationalisation rely on a clear intention and pragmatic implementation – the whole being presented implicitly in the spirit of face validity, only.

Inspection of raw data from surveys can give some insight into their likely validity. The researchers report that data cleaning (editing) to deal with missing data, aberrant answers and spurious responding was intensively undertaken. No analysis of this by type and quantity is provided. Grooming rules could be defined and put forward for use in future attempts of this work.

2. DEVELOPMENT OF THE QUESTIONNAIRES

Two questionnaires were constructed for use in the survey. One directed to fishers to report their activity and spending, and one directed to charter boat operators to report the number and areas of their trips and to characterise the overseas traveller proportion of their clients.

Neither cognitive pretesting, nor piloting to allow inspection and/or justification of these two tools is reported. Similarly, no validation of the tools appears to be available.

Notable in the questionnaires is that the two key constructs, ie, "spending" and "primary purpose" are not adequately operationalised, even supposing they could be recalled, attributed, or allocated as required.

Spending in Q.10 of the fisher questionnaire is asked for that spending on both self and others in the fishing party, with spending by others on the respondent to be excluded. These spends are added to produce a total for the respondent. Q.11 then refers the respondent to this total and asks them to say what percent was for "their share", thereby removing his/her spending for others which was asked to be included in Q.10. Spending for others is therefore stranded, notwithstanding that it is included as trip driven. Gift spending is inadvertently excluded.

The construct "primary purpose" is used to allocate spending by overseas visitors. Curiously "primary purpose" is not similarly applied to the fishers own trips in NZ, nor to persons accompanying the fisher on their recorded trip, only "who went with you?".

Primary purpose is a more nuanced construct than the questionnaire and subsequent extrapolations recognise. Specifically, it may well be the case that without the opportunity to seawater fish, the overseas traveller may not have come to New Zealand. To this extent his/her primary purpose was to fish here. However, having leveraged into NZ for this reason it is not unlikely that they took the opportunity to do much else in the way of diverse tourism to round out the trip. Perhaps, without this broader range of opportunities the primary purpose would not have sufficed on its own to drive the trip. This may mean their spending is associated minorly with seawater fishing and majorly with other tourism.

Similarly, the fact that a New Zealand resident took a saltwater fishing trip does not uniquely fix it as a fishing driven trip. A bundle of motives associated with bach ownership, boat ownership, family preferences for get-away-breaks, may all be drivers for a given trip, on which saltwater fishing is done. At the least this clouds the attribution of spending to the purpose of fishing, and raises fuzzy boundaries in the allocation of parts of a spend between fishing and non-fishing purposes.

'Most recent trip' and 'typical trip' are constructs which appear together in the fisher questionnaire, but which are reasonably interpreted differently. The introduction to Q.5 is headed "typical fishing trip" and goes on to say, "in this section tell us about a typical fishing trip". The computer then selects a region and season to be answered. Q.5 which follows immediately uses the "most recent trip" as the construct to be reported, as do successive questions. Most recent trip likely dominates, with the effect from "typical trip" standing in where the respondent has not stored detail of the most recent trip into memory or is unable or unwilling to retrieve the detail from memory.

3. MEASUREMENT ERROR

Though never eliminated, measurement error (as distinguished from sampling error) is capable of being reduced to the advantage of creating sounder estimates from surveys such as these. Measurement error used here is not a generic reference to all error possibly associated with the production of survey estimates. Rather it is a technical term referring to possible or probable bias in the answers given to the survey questions.

The greatest risk of measurement error in the two survey questionnaires used in this study relate to treatment of the recall of numbers and percentages. If we distinguish reflective questions (what did you get for your birthday this year?) from constructive questions (how many times have you filled your petrol tank this year?), it is clear that two quite different tasks entailing different processes and different levels of effort are required. Reflective questions have a context and a cause to help the person remember them, they refer to answers about things the person is pre-programmed to store in memory. For constructive questions there is rarely (or little) purpose to remembering and mentally collating the information as such, and in most cases this is not done.

Questions 10-13 of the fisher questionnaire ask for a wide range of spending on items related to fishing, which it could not reasonably be expected the fisher was pre-programmed to remember. Being unable to rely for all but some special or eventful items on reflective recall, he/she is obliged to construct a numerical answer for an occasion occurring sometime in the last 12 months. This context is prone to almost all the recall errors found in survey measurement. Forgetting of more distant purchases, telescoping of larger ones, substituting usual spend for actual spend, trying for consistency (as they see it) between items to report what seems right, puzzling over items gifted to and by them, and so on. In these questions the responding fisher is constructing replies, albeit in good faith, which he/she numerically codes to dollars. The replies are about spends he/she had no expectation of recalling or reporting before the questionnaire arrived in the online request. We do not argue that the answers are "all wrong" but that in the light of the reporting difficulties are reluctant to expect that they are "all right".

The biases such recall is prone to are exacerbated in the case of fishing activity by its irregular and unsystematic occurrence for most fishers. Fishing competition schedules for the minority aside, the general case is that an appealing combination of wind-sun-sea conditions prompts a fishing trip. Weekends give freedom to fish but do not provide the basis for any recurring pattern in the way of the organised and structured sports.

While charter boat operators are able to access their Activity Catch Return forms to respond to questions on the number of trips and number of passengers they carried, the charter version of the spending questionnaire moves well outside of the information in these forms. Charter operators, like recreational fishers, are asked to report estimates of information that they had no pre-knowledge of, and would therefore have observed and remembered on only an incidental basis. Questions on the percentage of passengers that were international visitors (Q7) and further the percentage of those in turn (Q8) that were thought to have come to New Zealand with the primary purpose of marine fishing, are asked.

The danger in this is that once expressed as numbers in a data set, these answers take on an appearance of veracity which the manner of their collection cannot reasonably sustain.

4. SPECIFICATION ERROR

While recalling quantities by simple reflective memory or by cognitive reconstruction for an occasion occurring in the last year, is part of the uncertainty associated with the figures produced from the two questionnaires, there is also the challenge of specification error concerning what the respondent's mind is addressing in the recall attempt. These are respectively the "how much" and "on what" aspects of asking spend questions. The fisher survey presents respondents with a list of types of spending, giving examples of what the type refers to in most cases.

A largely "scoping" approach was used to ask the spend on each of a range of goods and services that the fisher spent on during/for the most recent trip for consumables. These were totalled and the fishers asked what percent was for their share, versus other persons.

The test of such a list of items is whether the description of each is sensitive enough to capture the fishing driven component and concurrently specific enough to capture only the fishing driven component. The three lists (consumables, "small" equipment and "big" equipment) contain items insufficiently qualified or detailed in a way that would avoid over or under inclusion of a given spend as fishing-trip driven. Misattributions are nested within the broader terms. For example, in the consumables list, would spending on habitually consumed products be legitimately included, eg, a pack of cigarettes or aspirin, an evening beer or meal, etc. Should fuel for the vehicle be excluded if nothing needed to be spent on the trip day because the vehicle was filled a day or two before according to the owners' habitual pattern? Similarly, misallocations are a common hazard across the items. A spend is asked for bigger items like holiday homes, vehicles, boats and the like where these were purchased primarily for fishing purposes. It would be rare for such items to be so uniquely purposed in the recreational sector, and the allocation of a proportion of spending to that driven by saltwater fishing would be onerous, subjective and arbitrary.

Spending recall is complex in the task of retrieving the information into memory, and where tied to multiple drivers calls for a structure of controlled questions to be asked. It would be fair to view Q10-Q13 of the fisher questionnaire as only scoping figures. These might constitute the beginnings of the creation of a subset of sequentially devised probes to isolate specifically saltwater fishing driven figures.

Faced with the need to arrive at an estimate of the spending of overseas based fishers that is driven by saltwater fishing the researchers were obliged, given their time frame and resources, to rely on overly simplistic questioning of the charter boat operators. Q4 and Q7 ask them to say what percent of their customers across the year were international visitors, and Q8 asks what percent of these in turn were in NZ for the primary purpose of saltwater fishing. Neither of these items was known to the operators in advance, obliging them to source from memory, incidental information and open impressions to give a numeric answer.

Here as in the fisher's questionnaire the researchers took a pragmatic approach to get something rather than nothing with which to work toward an estimate. While fishers and charter operators in turn might feel comfortable conversing around such estimates, the informal context lacks any check on the accuracy of what is reported. Most pertinently the conversational context is one where those reporting are themselves unaware of whether their reported figures are close to, or well off, the actual figures. Having them believe they are answering accurately is not enough.

5. RESPONDENT BEHAVIOUR

The next section deals with the statistical properties of panel samples and the risks that type of sample poses. This section concerns behaviours of internet users who may join a panel, and how these might have played through into the data of the present study.

Internet users join a research panel, or for that matter several panels, to obtain rewards for completing survey interviews. Panel operators try to prevent members from fulfilling interviews they are not eligible for by placing non-disclosing screen questions in the initial section of the request to members. For this fishing survey the panel firm placed a list of outdoor recreational activities for which respondents had to select the ones they participated in, in the last year. Only those who selected marine fishing were asked (allowed) to complete the fishing survey itself. The report of this step does not say whether a panellist could select all or a high proportion of the outdoor activities to maximise their chance of proceeding into the survey itself, thereby being able to earn the participation reward. The use of these screening items is widespread in order to contain the behaviour of avid panellists, seeking to maximise their rewards. Panellists are knowledgeable about this and have generally responded by trying to frustrate the purpose of the screen to prevent their exclusion. A proportion of the participants may have selected marine fishing among the activities they selected in the screener with the intent of improving their chance of moving into the reward earning stage.

The report recognises that in their survey approach there was not only the risk of gaming of the questionnaire, but also recall and response biases. While the quality control step is not fully described or quantified, the authors assure the reader that they evaluated the raw data in great depth. This included flagging records where reported activities and expenditures appeared inconsistent. Measurement errors of the method and mischief errors of the mode are not elaborated on but it is clear that the researchers put effort into constructing and applying an edit step to deal with both. Nevertheless, there will always be unease around edit steps that rely on observation to detect them and judgement to resolve them in cleaning the data.

One finding leads this reviewer to wonder whether a proportion of panellists penetrated the researcher's best efforts to protect the integrity of the eligibility of respondents, and of their answering against a genuine background of marine fishing. Table A4 on p100 of the revised edition of the "Technical Steps" report states:

> "There is the potential for fishing activity and expenditure behaviours to exhibit a seasonal trend. In order to evaluate the presence of seasonal differences, we grouped the 12 months of the year into two season (winter and summer) and asked fishers to tell us about their most recent trip taken during the prior year in either or both of these seasons. Based on the number of fishers, our analysis does not find that fishers prefer fishing during the summer rather than winter, and thus no strong seasonal trend."

The sentence above from the revised edition of "Technical Steps" contradicts the May 25th letter from Southwick Associates to Dr Martin Cryer responding to a query on the absence of seasonal effects. The letter is prepared between the initial and revised editions and states:

> "Regarding the absence of seasonal trends as mentioned in the report, we are referring solely to the absence of a statistically different level of spending among fishers across the two seasons defined within the study. In other words, the data suggest that fishers spend the same amount, statistically speaking, per trip regardless of the time of year. This does not suggest that their activity level is the same throughout the year."

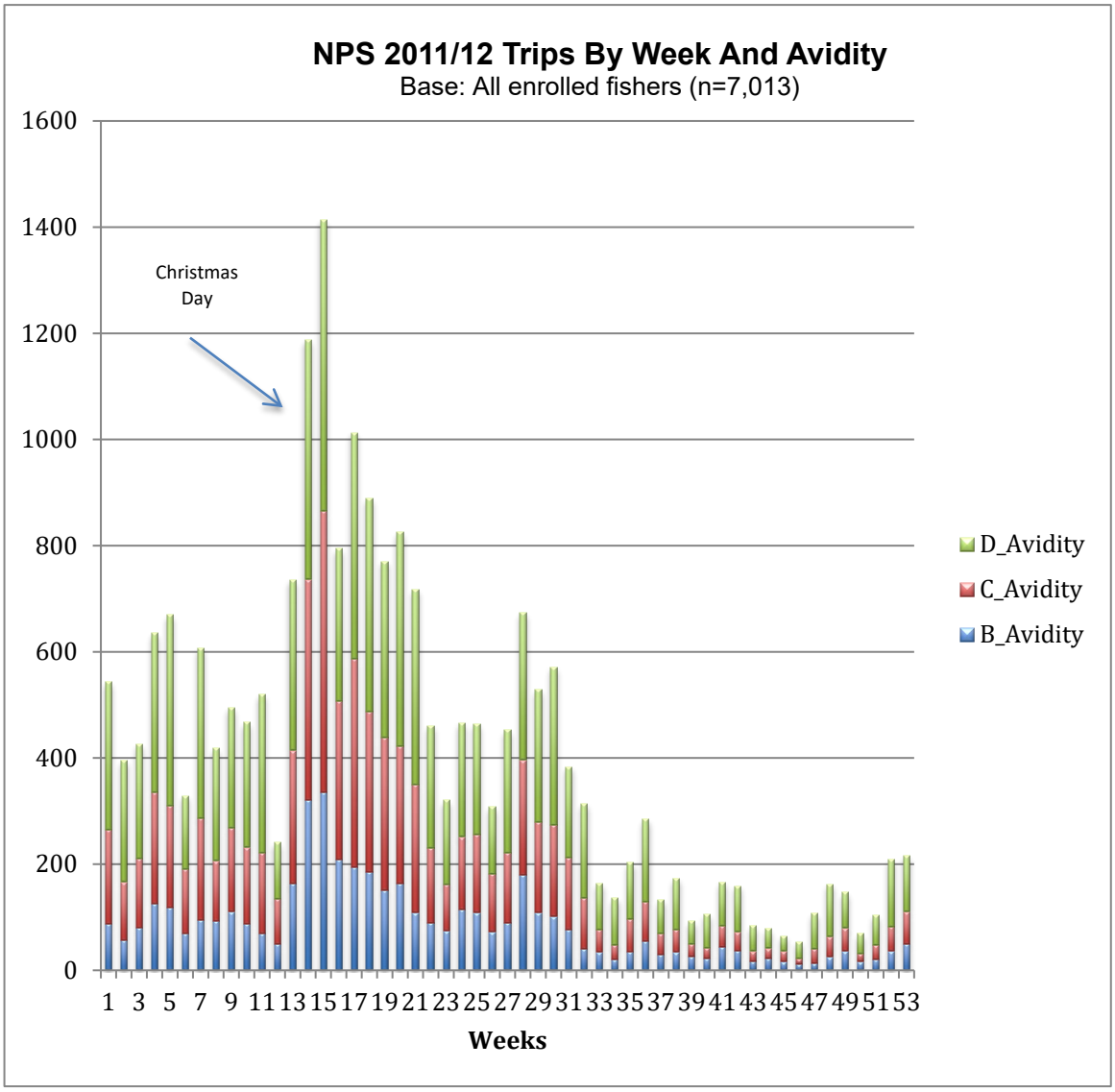
However, "activity" behaviour is indeed mentioned in the explanation of the seasonal figures (above), and it is the May letter which is incorrect.

The Ministry's 2011/12 NPS survey carried out over all 52 weeks of the year showed marked seasonal differences in the activity of marine fishing. The bar chart overleaf shows the seasonal activity.

FISHING TRIPS BY WEEK – NPS 2011/12

The following table shows the number of fishing trips in each week conducted by the NPS panel, regardless of whether there was any catch or not.

The variability of fishing conducted is shown, which naturally depends on many factors such as: season, which weeks contain holiday days other than weekends, weather, and more.



The six summer months of October to March account for over 70% of the trips taken whether any marine species were successfully harvested or not. Perhaps a scrutiny of the trips reported for winter by the Horizon panel(s) would help understand why no seasonal difference in trips is evident. Did some respondents feel they should have some entries on both the winter and summer sections of the trip reporting in order to appear authentic?

C. SURVEY STATISTICS

1. OVERVIEW

There is no evidence in the report that the Resident Fisher Survey (RFS) is a probability based sample with some proportion of non-response, handled in a standard way. Probability based sampling gives a randomization distribution on which inferences can be made, and accuracy of estimates made, free of assumptions. It provides, at the least, for non-self-weighting samples a direct way of weighting from the sample to the population of interest. Other sampling schemes require an underlying probability model and some typically subjective justification as why biases in the sample selection can be ignored, and a way to weight to the population of interest.

Of course, handling of non-response in probability based samples requires some model of non-response. A simplistic model assumes that the non-respondents are similar to respondents once we have controlled for certain variables, leading to calibration models. But we have more confidence when the response rate is high as in the NPS where it was 78%. There is no discussion of non-response and how it was dealt with in this report. Nor why we might expect that the calibration model might control for both the biases in the sampling scheme and the nonresponse.

Fred Smith (Smith, T. M. F. [1983], "On the validity of Inferences from Non-random Samples", *Journal of the Royal Statistical Society A*, 146, 394-403) gives technical conditions for drawing conclusions from a sample selected by non-random methods. He argues that the model-based approach is appropriate in situations like the following:

- there is a single client
- the results are for the use only of that client, and will not be published more widely;
- statistician and client agree that the required assumptions hold, at least approximately.

His summary is, "*... in the public sector ... there is no simple well-defined user and it is reasonable to ask that the sampling method used should have wide acceptability. Random sampling methods provide that wide acceptability.*"

Nearly a decade later Deville (Deville, Jean-Claude [1991], "A Theory of Quota Surveys", *Survey Methodology*, 17, 163-181), takes a fresh and detailed look at the forms of modelling which underlie quota sampling. His conclusions are similar: "*In a survey, the use of any speculative model represents methodological risk-taking. This may be perfectly reasonable if the users are aware of it, and if they have ratified the speculations leading to the specification of the model .... Official statisticians, on the other hand, are responsible for generating data that can be used by the entire society; ... Official statistics should not tolerate any uncontrollable bias in its products. It should carry out sample surveys using probabilistic methods.*"

More recently Gray (Gray, Alistair [2014] "Analysis of *the Travel diaries mobile trial* run by TNS for the Ministry of Transport", *Consulting Report for MoT*) noted that:

- Existing nonprobability based panels such as *SmileCity* show both socio-economic characteristics and travel behaviour which is not controlled for by balancing on the usual demographic characteristics (sex, age, ethnicity, region, etc).
- Recruiting respondents from previous probability based surveys also seems to lead to a panel with different travel behaviour from a probability based sample.

A similar situation could hold with the Horizon panel on marine fishing (see the earlier summer/winter anomaly). At the least there ought to be a discussion about the quality and characteristics of the panel and if possible some modelling of the propensity to belong to the panel. Such propensity scores might substitute for selection probabilities of a probability based sample. See, eg, Chapter 11 in Bethlehem, Jelke and Biffignandi, Silvia [2012] *Handbook of Web Surveys*, Wiley.

In the next sections I point out some other problems with the process in producing the estimates.

2. USE OF THE NPS: WHICH POPULATION?

Data from the NPS can be used to make estimates for several populations which are nested. Do we include or exclude:

- customary fishing;
- personal allowance from Commercial Catch;
- people who are on a charter boat trip;
- people who release everything;
- people who catch and keep something.

Do we include or exclude these people from the drop-in survey (People whose stated avidity was A, but who were asked six monthly whether they had fished.)

From the numbers in Table A1 in the Appendix A it appears that all these groups are included from both the main survey and the drop-in survey. Although as my estimate is 593328 fishers and 2657061 days and this is not quite the numbers in the table, I may be wrong about this. For example, they may be deleting personal allowance from Commercial catch since it may be already accounted for in the GDP subgroup for the Agriculture, Forestry and Fishing Industry. Similarly, whether customary fishing should be valued is a moot point. This question is analogous to whether unpaid work should be valued. A discussion of this would be helpful.

In any case, I think that the full definition of population the researchers are considering should be mentioned somewhere. In fact, it seems they are at times including the drop-in survey and at other times excluding it. For example, in Table A2 using both surveys I estimate 443346 boat based fishers and 2181585 days. On the other hand, maybe it is a miscalculation as, if I estimate using the main survey only I get 327724 fishers and 1978807 days.

Also Table A5 comparing the NFS with the NPS is somewhat confusing. I interpret Avidity in this table as some form of Stated Avidity as defined by the screening question. Presumably the 2% are people who engage in marine fishing but turned out not to have done any in the last year when asked further questions. So it is not strictly comparable to the Avidity A in the NFS for two reasons. People who had a Stated Avidity of B, C, or D who did not fish during the survey period were excluded. Also of the estimated 3034840 people of Stated Avidity A, the drop-in survey estimated 117348 people fished (with actual Avidity B) during the year. This is about 20% of the population of fishers.

The numbers in Table A5 for the NPS are mostly what I estimate if I include fishers who fished only outside the survey period and so should not be included. For example, I get the numbers by avidity to within rounding to 1. I get the male figure but the figure I get for female is 147233. I get most age group figures except I get 64606 for 15-24, 106137 for 45-54. I get 292445 for Upper North Island, 97057 for Lower North Island, 88632 for the South Island and 478144 for the total based on Regional Council boundaries. Note the total in Table A5 for Regions is actually the total for people who fished only in the survey period in the main survey.

Finally, the NPS excluded residents living in the Chatham Islands and so has no fishing recorded in FMA 4. This should be noted somewhere. Did the RFS have any fishers fishing there?

3. WEIGHTING OF THE RESIDENT FISHER SURVEY

The marginal distributions of the NPS sample across avidity, sex, age groups, and broad regions as presented in Table A5 are used to weight the expenditure data, eg, in Table A7, using an iterative proportional fitting (raking ratio) algorithm. That is the observed marginal in the RFS is calibrated to the observed population marginal in the NPS.

As mentioned in the previous section different populations seemed to have been used to produce those marginals. If we use a consistent population, say that including all people in the bullet points in the previous section, in both the main and drop-in survey then it turns out that the marginal distributions are close to what Southwick uses with the exception of sex and 25-34 age group. See Table 1.

The question is whether these, applying the consistent marginal for sex, age and region would produce estimates which would be different but have differences within sample error. I suspect the differences would be in within sample error since the cvs based on eg, data in Table A6 which range from 0.04 to 0.23. Hence although I do not think the marginals used are correct, using them would make no material difference.

Table 1: Comparison of Marginal Distributions

	RFS uses	NPS data
avidity		
A		
B	0.3632099	0.3642474
C	0.3681122	0.3683839
D	0.2686779	0.2673686
sex		
male	0.7507642	0.6801463
female	0.2492358	0.3198537
age group		
15-24	0.1351397	0.1377938
25-34	0.1754126	0.1685587
35-44	0.2160979	0.2185787
45-54	0.2218622	0.2250516
55-64	0.1589547	0.1585404
65-74	0.0729040	0.0706761
75+	0.0196289	0.0208007
region		
Upper NI	0.6116046	0.6112593
Lower NI	0.2030382	0.2051337
SI	0.1853573	0.1836070

However, avidity is problematic. Reading the RFS questionnaire avidity is Actual Avidity as recalled, whereas the NPS data used by the NFS is actual fishers who gave a Stated Avidity. There is a high misclassification rate.

Table 2: Misclassification Between Stated Avidity and Actual Avidity (row percentages)

		Actual Avidity			
		A	B	C	D
Stated Avidity	A	95.6%	4.1%	0.4%	0.0%
	B	59.9%	30.7%	7.9%	1.5%
	C	35.0%	36.2%	21.3%	7.5%
	D	21.2%	23.0%	27.4%	28.3%

If we look at the marginal distribution of Actual Avidity in the NPS it is:

B	C	D
0.6318511	0.2458414	0.1223075

This is substantially different both from the Stated Avidity marginal in the NPS and the observed marginal in the NFS ignoring the Avidity A respondents.

B	C	D
0.38522238	0.42539455	0.18938307

I would expect that using the Actual Avidity marginals from the NPS would alter the estimates substantially.

4. ACCOUNTING FOR SAMPLING VARIABILITY

The revised version of the paper provides some sampling errors based on replicated methods although what method is not stated. This will give a measure of the internal variability of the sample but because the sample is based on a panel and not a probability based sample it is possible that the sampling errors underrepresent the mean square error (variance plus bias squared).

Usually when calibrating a sample to population totals it is expected that the population total is known exactly, or at least the error in the population marginal is very much less than that of the sample marginal, say at least an order of magnitude less in the cv. But this is probably not the case here.

One approach would have been to replicate over both the NPS marginal (using the jackknife weights) and over the NFS sample. I suspect that this would inflate the sample error or cv. However, the following simple sensitivity analysis indicates that might be overkill.

To give an idea of the effect, suppose I have an estimate of the number of Avidity B male fishers aged 45-54 living in the Upper North Island using the NPS. Call that X. Suppose I measure the average annual expenditure on trip related spending per day for that group in the NPS. Call that Y. Then since the samples are independent so are X and Y. Then $cv(XY)^2$, which is the total expenditure by fishers in that group is:

$$\frac{V(XY)}{E(XY)^2} = \frac{V(XY)}{E(X)^2E(Y)^2} = \frac{V(Y)}{E(Y)^2} + \frac{V(X)}{E(X)^2} + \frac{V(X)}{E(X)^2}\frac{V(Y)}{E(Y)^2} = cv(X)^2 + cv(Y)^2 + cv(X)^2cv(Y)^2$$

Now the cv for X is .11 and the cv of Y is probably .52 based on that in table A11 allowing a sample size of this group being about 2.5% of the sample. So the overall $cv(XY)$ is about .53. In this case the inflation is probably ignorable given that the estimates of the cv for X and Y will have error as well. But if looked at average expenditure for Upper North Island fishers then the cv for X would be about .03, and that for Y about .12 and the combined cv would be about .124, so again ignorable.

I think something like the above analysis ought to be in the report so that the reader understands that accounting for the inexact calibration marginals is probably unnecessary.

5. SOME MINOR CONCERNS

- **Per trip table**: This appeared as Table A6 in the revised version of the report but not in the earlier version. It is a per trip estimate. It looks as if it is produced from Table A7 by using the average number of days per trip, about 3.5. But surely, eg, for personal vehicle use, a trip of several days would have a travel cost getting to the fishing launch place, which would be different from the travel cost being at the fishing launch place. I think an explanation of how the numbers are estimated is necessary.
- **Increase in fishing based on ANZ**: On page 13 of the revised report there is a discussion of participation in fishing as recorded in the Active New Zealand surveys. In particular that there is a 10% growth from 2007/2008 to 2013/2014. But as the sampling error is about 0.7% on each of the estimates (18.8% and 19.7% respectively) the difference is well within sampling error. Actually a more interesting question is the change in ethnicity due to migration, in particular, the increasing number of Asians who take finfish and shellfish.
- **Use of 2 as multiplier for 95% ci**: This could be reasonable for an approximation for the .975 quantile of the standard normal distribution 1.96, which is the appropriate thing to use assuming the replicate estimates are approximately normally distributed as stated on page 103 of the revised report. On the other hand it may be the value of a t distribution with 60 degrees of freedom, which would be sensible if, eg, 60 replicate estimates had been calculated. Either way a footnote explaining it would be helpful.
- **International Visitor Survey**: A less brief discussion of this would be helpful. For example, response rates by country, sample errors on the estimates, etc.

D. ECONOMETRIC EXPANSIONS

1. CONTEXT

We focus on the economics aspects of the methodology, looking at how it accords with generally accepted practice, highlighting any particular issues, and assessing the robustness of the main quantitative conclusions.

2. METHODOLOGY

The methodology employed to calculate the direct, indirect and induced economic effects conforms with generally accepted practice. Indeed, it is a good example as the authors take care to adjust the survey spending estimates in ways that some studies incorrectly omit. In particular, before calculating multiplier effects the spending figures are adjusted for direct imports and GST. Furthermore, the allocation of spending to industries distinguishes between industries that supply margins and industries that actually produce the goods, thereby converting the data from purchaser prices into producer prices, which is required before multiplier can be properly applied.

Nevertheless a few points are worth noting:

a. Ideally basic prices, which adjust producer prices for all indirect taxes should be used. The adjustment for GST goes most of the way to doing this, but there seems to be no adjustment for excise duty. For most commodities this is irrelevant, but for fuel, alcohol and tobacco it can be significant. The effect on the estimated indirect and induced effects is unlikely to be large, although the bias will be unambiguously upward.

b. However, acting in the opposite direction, GST which is (correctly) subtracted before the application of multipliers, could be added back to the final value added estimate. Excise tax could be treated the same way.

c. The assumptions regarding spending and its regional allocation, trip purpose/attribution etc all seem generally reasonable.

d. While the methodology is good, the report could do with a clearer statement in a more prominent position (than the discussion on p115-116 of the revised March 2016 report) that, with the possible exception of international tourists whose main reason for coming to New Zealand is to fish, the methodology does not purport to measure the NET economic contribution of recreational fishing. That is, it is not claiming for example that 8000 people would be unemployed if recreational fishing did not exist, or that NZ's GDP would be lower by $638m[1]. The authors may understand this, but the report as it stands make it too easy for the less informed reader to misinterpret the results.

[1] In some places in the report the total value added is given as $634m.

3. ROBUSTNESS

The main components of the total Contribution to GDP (value added) of recreational fishing of $638m are shown in Table 1.

- Line A, number of fishers: The number of resident fishers is derived from earlier research by NRB, while the number of foreign fishers is derived from MBIE's International Visitor Survey. The 109,000 visitors comprise 21,000 people who come to New Zealand primarily for fishing and 81,000 people who partake in fishing, but do not visit primarily for that purpose.
- Line B, mean spend: The average spend amounts reflect separate calculations for domestic and foreign visitors (primary and other), whether or not a boat is used, and whether the expenditure is of a capital type (such as equipment) or consumable (such as fishing bait).
- Line C, total spend, is the product of lines A and B.
- Line D, gross output: Gross output is derived by subtracting direct imports and GST from total spending, and reallocating margins to appropriate industries. This is required before economic multipliers can be applied.
- Line E, value added, also known as Contribution to GDP, subtracts intermediate inputs from gross output, leaving the return to labour and capital (and some indirect taxes). This avoids double counting as intermediate inputs are simply the output of another industry.
- Line F, the value added Type II multiplier. The Type II multiplier captures the indirect effects on industries that supply inputs to recreational fishers (such as the food processing industry supplying food to hotels) and the induced effect of people who work in supplying industries (both directly and indirectly) spending their wages and salaries on general consumer goods and services. As in line B, the multipliers are calculated at a more disaggregated level than shown in the table.
- Line G, contribution to GDP is the product of lines E and F.
- Lines H-I, direct employment multiplied by Type II employment multipliers, yielding total (direct plus indirect plus induced) employment.

Table 1: Contribution to GDP of Recreational Fishing

			New Zealand Fishers	Foreign Fishers	Total
A	Number of fishers	'000	595	109	704
B	Mean spend	$	1441	815	1344
C	Total spend	$m	857	89	946
D	Gross output	$m	616	71	687
E	Value added	$m	262	31	293
F	Type II multiplier		2.17	2.22	2.18
G	Contribution to GDP (total value added)	$m	570	68	638
H	Direct employment	FTE	4083	715	4798
I	Type II multiplier		1.73	1.50	1.69
J	Total employment	FTE	7048	1076	8124

The point of setting out the results in the above manner is that the effects of changing any of the main component are easily seen. In particular, we might consider the error margins attached to:

a. The number of fishers.
b. Their mean spend.
c. The economic multipliers.

The first two items are estimated from various surveys. These are discussed in the first section by NRB.

4. OVERVIEW

The economic multipliers come from an independent source (Insight Economics) and are therefore readily verifiable. They are based on input-output tables which are typically a few years out of date. These tables show the flow of goods and services between industries and sectors in the economy; for example how much fabric is required to produce a shirt or how much households spend annually on electricity.
The application of multipliers also involves a number of implicit assumptions such as:

- There are no changes in relative prices – recreational fishing does not change the price of commercially caught fish for example, nor hotel room rates.
- Average relationships apply at the margin – for example if one charter boat can accommodate 500 fishers over a 12-month period, 1000 fishers would require two charter boats.

Accordingly it is prudent to accept a typical error margin on multipliers of ±15%. They are not stochastically determined so there is no statistical standard error around a multiplier for any given industry. However the multipliers in Table 1 are weighted averages of many individual industry multipliers – probably at least 40. If the error margins are independent, a 95% confidence interval for the multiplied effects would be about ±8.5% or ±29m.

There is also an error margin (standard error) in the underlying base figures. For the value added estimate of $293m the standard error seems to be about 15% (based on appendices in the revised technical report) or about ±$44m.

Given independence of the base expenditure estimate and the industry multipliers, the standard error for the total contribution to GDP of $638m is about ±100m.

The full report contains much more detail than the summary numbers in Table 1, notably disaggregation by region of fishing activity and by fish species. While from an economics perspective the disaggregation seems reasonable, the results should be recognised as having larger error margins than the top-line numbers.

The (expansion) weights used to adjust the survey spending estimates are a function of avidity (keenness to fish), sex, age, and region of residence; the assumption being that each of these factors might affect how much a fisher is likely to spend. These factors are additional to allowing for whether fishing is from a boat or from land, as this clearly affects the sort of equipment that fishers would buy.

The obvious factor missing here is income albeit that some influence of income may be captured in the other factors, notably age. Household income was requested in the survey, but perhaps it was poorly answered or perhaps it was found to be unhelpful. There is no discussion of it in the Technical report (unless I have missed it).