<u>Academic review for Ofgem</u>

Extreme Value Analysis (EVA) methodology used for cost-benefit analysis (CBA)

Dr Robert Ritz, Assistant Director, Energy Policy Research Group (EPRG),
Cambridge Judge Business School, University of Cambridge.[1]

This version: 15 May 2018 – final

Executive Summary

- The EVA methodology presented in the report is based on standard financial economics and is, in general terms, a suitable methodology for CBA of investment options involving low probability high impact events. The standard "highest NPV" criterion is used to choose between different investment options.
- The specific application in the report of the EVA methodology to the Feeder 9 case involves additional substantive assumptions (such as on probability distribution and the range of uncertainties used in stress tests) that are often based on expert judgment and currently difficult to evaluate. The modeling choices and reporting of results could preferably be more transparent.
- A suggestion is that CBA for large investments involving low probability high impact events should be based on results from several methodologies (rather than relying on a single approach such as this EVA) to facilitate comparison. From a consumer perspective, using several methodologies could help make the investment case and regulatory decision more robust.

Questions for the review

Ofgem has requested a critical review of a new CBA methodology called Extreme Value Analysis (EVA) used by National Grid Gas Transmission. This review addresses three questions posed by Ofgem:

1. Is the EVA methodology as applied to the Feeder 9 CBA suitable and appropriate as a tool to inform network investment decisions in general, and Feeder 9 in particular?
2. What are the weaknesses and strengths of this approach as applied to the Feeder 9 question? Would an alternative approach be more likely to deliver a better outcome for consumers?
3. Does the CBA appropriately take account of uncertainty about the probabilities of different events and their consequences? In particular, are the assumptions about the range and distribution of probabilities reasonable?

The review is based on three documents provided by Ofgem:

---

[1] This review is written in a personal capacity as a member of Ofgem's Academic Panel; it does not necessarily reflect the views of Ofgem or any other organization.
Contact details: Peterhouse, Cambridge CB2 1RD. E-mail: rar36@cam.ac.uk

1. *Project EVA Case Study 1 Final Report* (including appendices) (report by Business Modelling Associates, dated January 2018) – "the report"
2. *A quantified risk assessment of the underwater section of the Feeder 9 pipeline in the Humber Estuary* (report by DNV-GL, dated January 2016) – "DNV-GL report"
3. *Humber Estuary Feeder No.9 Pipeline – Independent review of Freespane development and Remediation Planning* (report by Intertek, dated January 2018) – "Intertek report"

The report uses information from the DNV-NL report but does not refer directly to this particular Intertek report.

Scope of this review
The perspective of this review is economics rather than engineering. It focuses primarily on the suitability of the EVA methodology presented in the report.

This review has not checked a number of important underlying features and assumptions of the CBA, including:

- The premise that the Feeder 9 is pivotal to Britain's Future Energy Scenario;
- The environmental/engineering case supporting the replacement of Feeder 9;
- The availability of only three types of investment: maintenance, trench and tunnel;
- The interaction of the timing of this CBA with the RIIO price control framework (including if the costs associated with the replacement of the Feeder 9 pipeline were "substantially uncertain" as of the beginning of the RIIO price control period in 2013);
- The validity of any specific numbers used in the CBA.

Summary of main conclusions from the reports
The DNV-GL report finds that a failure of the Feeder 9 pipeline would involve substantial societal costs but that these are also highly infrequent. The Intertek report finds that frond mattresses can offer short- to medium-run mitigation but that there is no long-run engineering solution (page 31); it therefore suggests removal of the Feeder 9 pipeline to safeguard gas security of supply (page III). Both of these reports include discussion of the earlier failure of the Feeder 1 pipeline.

The report presents "Extreme Value Analysis" (EVA) as a methodology to deal with the challenges of modeling low probability, high impact events. Its EVA implementation is based on Monte Carlo simulations on a decision-tree setup. The headline result from the EVA is that the Tunnel 2012 option has NPV £53m higher than that of the next-best option Trench 2012, followed by Mitigate 2016. This NPV ranking is found to be highly robust: Tunnel 2012 dominates the other investment options "across the full range of stress test scenarios" (page 8).

Comments on methodology

This review faces an overarching challenge: the report is simultaneously being used to (a) evaluate the EVA methodology in general terms, and (b) evaluate its specific application the Feeder 9 case. The report does not make explicit whether particular assumptions made in the application to Feeder 9 would also need to be made when applying its EVA model implementation to other gas network investments (for which the available data may be different).

1. Is the EVA methodology as applied to the Feeder 9 CBA suitable and appropriate as a tool to inform network investment decisions in general, and Feeder 9 in particular?

i. The EVA methodology presented is a variation on standard financial-economics: for each investment choice, probabilities are assigned to different states of the world, cash flows are calculated, discounted back into an NPV—and then the different investment choices are compared to find that with the highest NPV. This is augmented with Monte Carlo simulations and a decision-tree analysis to yield “extreme risk weighted analysis of Net Present Value” (p. 10); the Monte Carlo simulations are based on “fat tailed” probability distributions that incorporate a wide range of uncertainties with guided by expert inputs. Finally, the report presents further stress test scenarios with additional discrete changes to other input parameters.

ii. The presence of such low probability high impact events, which are hard to assess based on historical experience, may in general argue for using several CBA methodologies rather than relying on a single approach—especially for large important infrastructure decisions. The report presents results only from EVA methodology so no comparison with results from another methodology is currently possible. From a consumer perspective, using several methodologies this could help make the investment case more robust. The obvious comparator would be a traditional CBA methodology; given their otherwise similar approaches, this would give a clear picture of the differential impact of using the present EVA methodology. Another potential comparator is "least worst regret” analysis. Finally, in a different but related context, the literature on climate-change policy has developed a range of modeling approaches to deal with low probability high impact (“catastrophic”) events[2]; some of these techniques might lend themselves to application within CBAs of energy infrastructure investments.

iii. In a sense, the modeling of low probability high impact events is simultaneously the methodology’s strength and its weakness: small

[2] See e.g. Nicholas Stern et al. (2006). Stern Review on the Economics of Climate Change. HM Treasury & Cambridge University Press – as well as the subsequent critique by Martin Weitzman (2011). Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change. *Review of Environmental Economics and Policy* 5 (2), pp. 275-292.

differences in probabilities across different options end up having huge NPV impacts—which are currently difficult to cross-check against any other source. An example is Figure 37 (page 70): tiny differences in probabilities such as the "average annual frequency", even with identical per-unit impacts, account for the bulk of difference in NPVs. This adds to the case for using several different methodologies to test the investment case.

iv. The welfare standard underlying the EVA methodology would benefit from further clarification. The Executive Summary states that the modeling incorporates "probability weighted consequences of failure of Feeder 9 valued from a safety, environmental, social, commercial, and reputational perspective" while the results are presented in terms of NPV and consumer bills. How are these reputational risks measured (probability, impact)? By what mechanism are they passed through to consumer bills? These issues did not appear to be discussed.

v. It would be useful to have a clearer presentation of which assumptions are important in driving the conclusions. A number of sensitivity analyses are presented throughout Section 6; these initially stem from the Monte Carlo simulations and then from a set of further stress test scenarios (which were presumably done in conjunction with further Monte Carlo runs). Despite this variety of analyses presented, it would still be useful to have results, for example, on *simultaneously* varying several input parameters in the stress tests.

vi. The EVA implementation around rare events involves input expert input/judgment on probability distributions from National Grid, i.e., the regulated company itself (e.g. pages 25-26). For obvious reasons, it would be valuable to know: (i) if these expert inputs are consistent with data that has been used in past CBAs or in other Ofgem-related regulation, and (ii) if other experts from outside the regulated company (perhaps from another country) would be in a position to provide similar inputs.

vii. The report does not discuss in much detail any knock-on effects for the rest of the GB gas network; for example, is any risk shifted to other nearby pipelines under any of the options considered? Or does closing the Feeder pipeline push up the capacity utilization of other pipelines or infrastructure elsewhere, with consumer repercussions? Perhaps the answer to these questions is "no" but it would still be valuable to have this confirmed as well as whether the methodology can, in principle, address such network considerations.

2. What are the weaknesses and strengths of this approach as applied to the Feeder 9 question? Would an alternative approach be more likely to deliver a better outcome for consumers?

i. The decision tree analysis is seemingly applied to decisions that have partly already taken place in the past (pages 13-14). For example, the preparatory investment for the Tunnel 2012 option has already happened; this leads to the unusual situation that the apparent "status quo" of mitigation on Feeder 9 is penalized for creating a partially stranded investment for the tunnel. Would the Tunnel 2012 option remain preferred had the preparatory investment not taken place? How was the decision on the preparatory investment taken; was a CBA conducted?

ii. The report notes that its analysis conforms to "the principles of real option pricing" (page 8). In economics, real options analysis is usually associated with the option value of *delaying* a decision until a later date—by which time new information on costs or benefits will be available to further inform the decision. The report only appears to consider a single decision "now" between different options (mitigate, trench, tunnel); there does not appear to be an option such as "wait for another 3 years (i.e., mitigate), and then decide whether to go ahead with tunnel". Perhaps such delay is infeasible but it would be valuable to clarify whether such other options do exist, and how they were modeled.

iii. The report mentions "no environmental costs were considered in terms of Bored Tunnel because it was concluded that it has a minimum impact on habitat" (page 12). This is a strong assumption that will presumably favour the tunnel option in the CBA. More sensitivity discussion/modeling of this aspect would clearly be useful.

iv. A different perspective on the results could be very useful. In particular, under which assumptions does the Tunnel 2012 option become *inferior* to the "mitigate" and "trench" options? Understanding these scenarios, even if they may turn out to be very implausible or unrealistic, could further strengthen the claim that Tunnel 2012 is preferable. The report comes closest to providing such a perspective in Table 38. In particular, at a higher cost of capital of 6.0% the tunnel option is only marginally superior to trench; the result looks like it would flip for a cost of capital above 6.5%. Does the report assign zero probability to such a scenario? If yes, on what basis? Given the very long lifetime of these assets, the interest-rate environment and cost of capital might by the 2020s and beyond be very different from the recent past.

v. Table 15 and 16 give the estimated impact of outage of Feeder 9 on the wholesale gas price (pages 41-42). Tunnel 2020 incurs a zero price impact from 2020 onwards while all the other options involve large price increases all the way out to 2068. Presumably this is an important driver of the NPV advantage of Tunnel 2020—though the report does not appear to confirm to this explicitly. It was unclear if and how this wholesale price impact features later on in the report. Specifically, in Table 38, there are stress tests with respect to "Wholesale cost of gas price increase reduced by 50 (and 90) percent". How do these relate to Table 15 and 16; do they incorporate changes to the price impact of a Feeder 9 outage? If yes, it seems surprising that this does not have a stronger impact on the NPVs and ranking of the options—so it would be valuable to have this further clarified.

vi. The report shows a calculation that appears to cast increased GHG emssions in monetary terms (page 44). The value of £59/tCO_2 employed for this calculation appears to be the social cost of carbon; should this value not vary over time as a function of when the GHG leakage occurs?

3. *Does the CBA appropriately take account of uncertainty about the probabilities of different events and their consequences? In particular, are the assumptions about the range and distribution of probabilities reasonable?*

i. The report lays out two probability distributions, triangular and uniform, which are used to capture the "fat tails" of extreme events (pages 25-26). The triangular distribution has the advantage that, in addition to the two extremes, it also includes an explicit "best guess" (i.e., the mode). In this sense, it is preferable to employ the triangular distribution where possible. It would be good to have an explanation of why which distribution was used for a particular source of uncertainty; what prevents a "best guess" from being obtained for some of the sources of uncertainty?

ii. When the uniform distribution is used, the maximum is often assumed to be 100 times as large as the minimum (e.g. Tables 2 and 10). In such cases, the distribution specification thus appears to be fully determined by the choice of a *single* number. Also, this wide span makes the uniform distribution look increasingly like a "diffuse prior", i.e., there is no way of placing greater weight on any particular value. More explanation would again be useful; varying the multiple of 100, perhaps again relying on expert judgment, would be a way of obtaining further robustness.

iii. The report does not appear to discuss the issue of *correlation* between different sources of uncertainty. The implicit assumption in the Monte Carlo simulations therefore seems to be that the underlying risk factors are independent of one another. This also seems to be the case for the stress test scenarios presented in Section 6. If so, this is a major assumption that deserves more justification, specifically of why it is reasonable to assume that particular sources of uncertainty are uncorrelated. If the modeling does allow for non-zero correlations, this should be reported more explicitly in the results.

iv. The treatment of discounting is said to follow the guidelines set out in the Green Book, with a social discount rate of 3.5%. However there is no detail on the how the cost of capital (WACC) is estimated. Very late in the report, a sensitivity analysis for cost of capital over the range 4.8–6.0% is presented (page 76) but this does not show the baseline assumption. Do any of the numerous uncertainties modeled, some of which are directly faced by National Grid, enter into the WACC calculation (as systematic risk)?