Oxera comments on draft European Commission guidance on how to conduct state aid evaluation

Note prepared for DG Competition

7 February 2014

Introduction

Oxera is pleased to have the opportunity to comment on the European Commission's draft guidance paper on evaluation in the field of state aid.[1]

We understand the purpose of this paper to be to set the standard for robust ex post evaluation of substantial aid schemes, such that evaluations can be used with confidence to optimise the design of future aid schemes.

Oxera's comments address the following main questions:

- how should ex post evaluation for state aid be conducted?

- who should conduct the ex post evaluation?

- are the methodologies proposed in Annex 1 fit for purpose?

1 How should ex post evaluation for state aid be conducted?

1.1 Oxera's understanding of the aim of the guidance is to encourage both more and higher-quality ex post evaluation of state aid schemes, such that lessons can be shared about what types of aid are most effective and what is best practice in designing effective schemes.

1A Precision and sophistication versus data requirements and simplicity

1.2 The consultation document, as currently drafted, places demanding requirements for analysis in terms of data and techniques. For example, instrumental variables,

---

[1] European Commission (2013), 'Evaluation in the field of State aid: Draft Methodological Guidance Paper —concepts and recommendations', November,
http://ec.europa.eu/competition/consultations/2013_state_aid_modernisation/draft_guidance_paper_en.pdf

regression discontinuity design and difference-in-difference techniques will not be familiar to the non-technical reader, and the descriptions provided of these techniques suggest high costs in terms of data and analytical resource.

1.3   However, within this framework, much insightful analysis can be undertaken which may not be too onerous. For example, a well-designed study may be able to use a relatively simple controlled trial (an arithmetic difference-in-difference in economics terms), which could be calculated quickly and easily.

1.4   One of the biggest benefits from such appraisals is likely to be identifying the schemes that do exceptionally well, and promoting their use, and identifying those that are very ineffective, and discouraging their use. For this purpose, simpler methods of comparing key performance indicators (KPIs) between different groups may be informative in identifying particularly successful and particularly ineffective projects.

1.5   Many of the techniques suggested, while setting the benchmark for best practice in policy assessment, have substantial data requirements. In practice, researchers may be constrained by data limitations which mean that the analysis may not be as precise as would be required for publication in an academic journal. A section on what could be done with less data, and the trade-off between accuracy and more detailed data, may help to ensure that a scheme is analysed to some extent, despite the data being imperfect. As has been discussed in relation to quantifying antitrust damages, when data is more limited, simpler methods can sometimes be used, albeit with great care:

> All methods and models rely on the data and information that is available to calibrate them. When more data is available, a wider range of methods and models can in principle be used in any given situation. Several of the more complex methods and models require significant amounts of data in order for them to be used at all (eg, difference-in-differences panel data regressions), while many of the simpler approaches can be used with more limited data (eg, difference-in-differences comparisons of averages).
>
> The simpler approaches that are straightforward to understand and calculate, such as comparisons of averages, are useful when the basis for the counterfactual is of a high quality. When there are important factors that mean that the comparator may not mirror the counterfactual, these simple approaches should be employed with care since they could over- or underestimate the counterfactual variables.[2]

1.6   Some examples of best-practice studies in policy evaluation would help to demonstrate what is possible and how it can be implemented. Oxera has undertaken several assessments of government policies designed to enhance welfare, such as the impact of tax-efficient share schemes for HMRC which uses a dynamic panel, difference-in-difference approach.[3]

1.7   In summary, a balance can be struck between the technical quality of evaluation and the need to evaluate the most important state aid measures (even when data is imperfect). The Commission has made a similar point in the field of quantifying antitrust damages, stressing the intrinsic limitations of counterfactual assessment:

> It is impossible to know with certainty how a market would have exactly evolved in the absence of the infringement of Article 101 or 102 TFEU. Prices, sales volumes, and profit margins depend on a range of factors and complex interactions between market

---

[2] http://ec.europa.eu/competition/antitrust/actionsdamages/quantification_study.pdf, page xi.
[3] Oxera (2007), 'Tax-advantaged employee share schemes: analysis of productivity effects—Report 1: Productivity measured using turnover', August, HM Revenue & Customer Research Report 33, http://www.hmrc.gov.uk/research/tax-advantaged-report.pdf. Oxera (2007), 'Tax-advantaged employee share schemes: analysis of productivity effects—Report 1: Productivity measured using gross value added', August, HM Revenue & Customer Research Report 33, http://www.hmrc.gov.uk/research/tax-advantaged-report2.pdf.

participants that are not easily estimated. Estimation of the hypothetical non-infringement scenario will thus by definition rely on a number of assumptions. In practice, the unavailability or inaccessibility of data will often add to this intrinsic limitation.

For these reasons, quantification of harm in competition cases is, by its very nature, subject to considerable limits as to the degree of certainty and precision that can be expected. There cannot be a single 'true' value of the harm suffered that could be determined, but only best estimates relying on assumptions and approximations.[4]

1B   Other sorts of assessment

1.8   While the assessment methods suggested are all sensible ways of measuring the impact of policy changes, they do not necessarily identify why something was successful or unsuccessful. If a good scheme is poorly implemented, its outcome will not be as good as it could have been. For example, a Member State may have a very good idea for de-carbonisation that requires state funding. The scheme perhaps should have worked well, but the particular way it was administered led an ineffective use of resources. The economic impact assessment would suggest that the scheme was ineffective and discourage other similar schemes, whereas it should be encouraging a more effective implementation.

1.9   In some cases, there may be a role for more implementation assessments, possibly done by auditors, management consultants or social researchers (including economists), which examine whether a scheme was implemented in line with best practice, as well as assessing its actual economic impact.

1.10   On a similar theme, section 3.7.1 of the Commission's paper notes:

The design of aid schemes are based on ex-ante assumption. However, the impact of the aid on markets may differ from expectations, for instance due to imprecise or incomplete assumptions or to unforeseen changes in the market conditions.

1.11   It seems unlikely, at the stage of ex post evaluation (i.e. years after an aid scheme has been initiated), that the Member State's ex ante expectations as to the effects of state aid will be known with complete detail. But an economic impact assessment which would not control for the difference between expectations and out-turn market conditions will tend to deliver misleading results. In particular, it may condemn schemes which have failed for exogenous factors, or promote schemes which have only succeeded due to an exogenous positive shock. The empirical strategy for aid evaluation should be sensitive to these aspects, i.e. understanding *why* an aid scheme has succeeded or failed in addition to measuring the magnitude of success.

1C   Handling multiple outputs

1.12   The methods suggested in the draft guidance work well if there is a single output that the policy aims to maximise (or minimise). However, many schemes have multiple outputs, as identified in possible result indicators listed in Annex II of the guidance.

1.13   While the methods suggested can be run on multiple different outputs, this does not capture the trade-offs between different outputs. For example, is the scheme incentivising broadband coverage of one speed at the expense of coverage?

---

[4] http://ec.europa.eu/competition/consultations/2011_actions_damages/draft_guidance_paper_en.pdf. The Commission's paper draws on Oxera's study on quantifying antitrust damages: http://ec.europa.eu/competition/antitrust/actionsdamages/quantification_study.pdf

1.14   Techniques exist to handle multiple inputs and outputs, such as input–distance functions, or non-parametric approaches, such as linear programming (see the literature on data envelopment analysis, DEA).

1D   Where is a scheme most effective?

1.15   The guidance, as drafted, focuses on whether or not a scheme is effective on average. However, it is possible that schemes benefit some types or subsets of firms and not others, and can be better tailored to target only the firms where it is likely to be effective.

1.16   The question being asked is what effect additional support has on firms in helping them turn inputs (materials, capital and labour) into outputs (goods and services). There is extensive research in the efficiency literature on how a policy has affected the production function of firms. Frontier efficiency techniques, such as corrected ordinary least squares, stochastic frontier analysis and DEA, are based on the production functions. These techniques examine how firms turn inputs into outputs, and may provide valuable insights into which firms benefit the most from aid, and, in turn, help in designing better schemes. For example, Oxera's analysis for HMRC of the effectiveness of UK tax-advantaged share schemes (referred to above) demonstrated that some firms benefited significantly from the schemes while others did not, depending on the firms' characteristics.[5]

1E   Accessible presentation

1.17   A report focusing on the regression estimator and its strengths or weaknesses is less likely to change policy than a report that identifies where aid does or does not work using accessible graphics and clear language.

1.18   If economic impact assessments are to have a major influence on policy, the results need to make sense to the general reader. Explaining in an accessible and helpful way the intuition behind the results and the implications will help to ensure that the assessments reinforce a culture of innovative policymaking and analytical testing of the results. A tick-box list of what makes a report helpful might be useful here, as well as some examples of what are considered good, accessible studies.[6]

2   Who should conduct the ex post evaluation?

2.1   One difficulty with ex post evaluation is that the data required must be determined before the aid scheme is approved, but the analysis cannot start until the aid scheme has been in place for some time. There may be a gap of several years between specifying the data and undertaking the analysis. Ideally, therefore, a Member State would appoint the evaluator before the aid scheme is approved, in order that the empirical strategy for aid evaluation is known ex ante.

2.2   The techniques set out at Annex 1 of the Commission's paper imply that aid evaluation will normally conducted by academics or researchers with substantial experience in micro-econometrics. It may be helpful to consider how best to channel funds through existing bodies such as the ESRC.[7] Having aid evaluation results published in peer-reviewed econometric journals could improve the incentives for high-quality research. However, as discussed above, there is also a role for forensic

---

[5] http://www.hmrc.gov.uk/research/tax-advantaged-report.pdf; http://www.hmrc.gov.uk/research/tax-advantaged-report2.pdf.

[6] See, for example, the 'policy briefing' papers published by UEA: http://competitionpolicy.ac.uk/publications/working-papers-2014.

[7] http://www.esrc.ac.uk/about-esrc/what-we-do/index.aspx.

evaluation of the implementation of aid schemes: auditors, economic consultants and even management consultants could be better placed to undertake this evaluation.

2.3   Member states already conduct value-for-money assessments, such as the work of the National Audit Office in the UK and the Courts of Auditors in other member states.[8] It would be useful to consider the relationship between these studies and the evaluation of aid schemes since there may be considerable overlap, especially if a Court of Auditors is already tasked with examining matters concerning state aid.

2.4   Where the empirical strategy for evaluation will involve seeking qualitative information (e.g. interviews, case studies) as a complement to quantitative data, it may be most appropriate to consider appointing consultants to carry out the research. Oxera's study on counterfactuals to restructuring aid is one example of case studies being used to add insight and colour to the results of a complex econometric model.[9]

2.5   Likewise, if the aim of research is to influence policymakers at the highest level, the presentation of results must be non-technical and compelling. The choice of aid evaluator should be sensitive to this. In some circumstances it may be useful to consider a dual-track empirical strategy, with an academic and a consultant working alongside one another. This could ensure that results will both advance the scientific understanding of the effects of aid and be accessible to non-technical policymakers.

3   Are the methodologies proposed in Annex 1 fit for purpose?

3.1   This section of Oxera's comments focuses on the econometric techniques as proposed in Annex 1 of the Commission's draft paper.

3A   Choice of method according to circumstances

3.2   The Annex notes that 'several reliable methodologies' (p. 20, line 37) are available for valid evaluation of the causal effect of state aid. Econometric methodologies are rendered meaningful by assumptions, which may be more plausible in some situations than in others. Although across all possible situations there are undoubtedly several methodologies that can be used, in any particular case the researcher is less likely to find an abundance of possibilities.

3.3   The Annex could guide the researcher by relating the assumptions underpinning each method to the particular context in which the causal effect of state aid is being measured. It may even be possible, with good contextual knowledge, to characterise general features of the state aid effects evaluation problem, and to give the researcher a prima facie reason to look at a subset of methods from the treatment effects literature.

3.4   Implicitly, the Annex does this by devoting much of the discussion to quasi-experimental methods, and less to randomised experiments and structural models. However, it may be worth exploring whether it is possible to incorporate more comparison and contrasting of the different quasi-experimental methods, specifically in the context of state aid effects evaluation.

3.5   A concise summary comparison of different quasi-experimental methods could be useful to allow the researcher to appreciate the bigger picture with less effort. Particular attention could be given to the question of whether the differences between the methods, together with contextual knowledge (of state aid), could inform the choice of method within the quasi-experimental class.

---

[8] http://www.courtofaudit.nl/english/Publications/Topics/European_Union/The_role_of_audit_institutions.
[9] Oxera (2010), 'Should aid be granted to firms in difficulty? A study on counterfactual scenarios to restructuring state aid', http://www.oxera.com/Oxera/media/Oxera/Restructuring-state-aid.pdf?ext=.pdf

3.6   The Annex does provide a good description of the fundamental problem in causal inference.[10] In a simple comparison of firms that received aid with those that did not, the researcher would have to assume that there were no systematic differences between the firms that received the aid and those that did not. In other words, the probability that a firm received aid must not depend on factors that also determine the outcomes being measured.

3.7   In experimental settings, this possibility may be ruled out by assigning aid entirely independently of any factors that determine outcomes in the absence and presence of aid.

3.8   In quasi-experimental settings, such random assignment is not feasible. It would help the reader if a high-level overview were provided of the quasi-experimental class of methods.

3.9   It seems that the causal effect of aid may be estimated consistently in an econometric setting, in one of two ways:

   i) the probability of receiving aid is explicitly taken into account in the calculation either directly (regression and matching designs and sharp regression discontinuity designs), or indirectly (regression designs with instrumental variables and fuzzy regression discontinuity designs);[11]

   ii) actual (not constructed) control groups are used under the assumption of a common unobservable time trend.

3.10   This dichotomy should be brought out clearly. This because, in case i), cross-sectional data is sufficient, while, in case ii), repeated observations (in time) are required.

3.11   All of the quasi-experimental methods reviewed in the Annex are well described in widely available textbooks. Therefore the Annex could perhaps add more value by focusing on areas where contextual knowledge (concerning state aid, theoretical and empirical) can inform the choice of econometric techniques.

3.12   It could be worth exploring whether those sections where contextual knowledge is not used could either be furnished with it, or shortened by making references to primary literature or a textbook treatment.

3B   Technical accuracy of description of methods

3.13   This section of Oxera's comments takes the content and emphasis of the Annex as given, and comments on specific technical aspects.

   Regression versus matching approach

3.14   The distinction between regression and matching seems too sharp in light of the underlying econometric theory. In practice, the researcher does face a choice between these methods and their implementation in software requires a clear distinction to be drawn. However, in terms of the identification strategy, the methods are very similar, and this is not brought out sufficiently in the Annex. The following sentence, for example, creates some scope for misunderstanding.

---

[10] In this note, it is assumed that the causal variable of interest is a dummy variable determining whether state aid was granted or not.

[11] For a helpful explanation of the 'anatomy' of regression and matching, see, for example, Angrist and Pischke (2008, section 3.3.1). Angrist, J. and Pischke J. (2008), 'Mostly Harmless Econometrics: An Empiricist's Companion', Princeton University Press.

> In particular, in the case of matching, comparing the outcomes between a beneficiary and its matched 'twin' without aid, allows avoiding the selection effect only if the granting of the aid is unrelated to unobserved variables that also influence the outcome. (p. 23, lines 3–5)

3.15   The absence of unobserved variables that affect the outcome of interest, and are also related to whether aid is granted, is *also* a necessary assumption in regression analysis, if the researcher is to uncover the causal effect of aid.

3.16   One way to make clearer the differences and similarities of regression and matching would be to cite Angrist and Pischke (2008, section 3.3.1), especially as this text is already referenced in footnote 19.

3.17   Both regression and matching techniques rely on two assumptions: i) conditional independence, as already noted in the Annex; and ii) some overlap between the aid status of firms, for similar values of the observables that influence both the granting of aid and the outcome of interest. That is, unlike in sharp regression discontinuity designs, for any particular values of the observables, there can be expected to be both firms that were granted aid and those that were not.

3.18   Both regression and matching achieve their purpose of controlling for observables by aggregating the local measures of the effect of aid, based on firms that are similar in ways other than aid status. The difference between regression and matching is in the weights used in this aggregation.

3.19   In the context of difference-in-differences estimation, regression and matching are described in a way that suggests that they are closely related. For reasons given above, the sentence...

> The two methods are two different ways to take observable differences into consideration but there is no fundamental difference in terms of identification of the causal effect of the policy. (p. 25, lines 20–22)

...in essence also applies to other quasi-experimental designs.

Description of matching methods

3.20   The description of matching methods could be refined, to make clear it is the model which is doing the matching rather than the analyst's judgement. Most techniques, such as propensity score matching, use some form of econometric technique, such as logit, to predict the probability of treatment (ie. that they receive aid). It would be helpful to identify the specific different techniques within this family, such as propensity score matching, and also consider non-parametric methods.

Description of instrumental variable methods and endogeneity issues

3.21   The discussion of instrumental variables would benefit from mentioning which tests can be done to test the appropriateness of instruments. While it is true that no test of the validity of instruments exists, specification tests can be applied to assess the plausibility that some of the assumptions related to the instruments are valid, if the rest are assumed to be valid without testing. If this test is failed it is an indicator that the assumptions surrounding the instruments may not be valid.

3.22   It is not entirely clear what is meant by:

> Since benefitting from aid can be seen as an endogenous explanatory variable of the performance of a firm in a linear regression context… (p. 27, lines 2–3)

3.23   Presumably, in this context, the endogeneity problem relates to the dummy variable describing whether aid is given or not (if this is the causal variable of interest, which it

must be if the causal effect of interest is the effect of state aid). The endogeneity of the aid status may in turn be explained by some model through which the extent to which a firm benefits from aid partially determines the probability that it is granted aid; here, benefits to the firm from state aid should be thought of as an omitted variable.

Regression discontinuity designs

3.24   In view of the earlier distinction that the Annex makes between regression/matching and instrumental variables methods, it would be helpful to also explain the difference between sharp and fuzzy designs.

3.25   In a sharp design where the explanators predict whether the firm receives aid or not, we either observe only firms that either receive aid or only firms that do not receive aid. This means that a greater degree of extrapolation is necessary than in the case of fuzzy measures, and the choice of functional forms we specify matters more for satisfactory measurement of the causal effect than it does in regression/matching. Consequently, sharp discontinuity designs usually concentrate on firms that are just above the threshold for receiving the aid do or are just below, the rationale being that these firms will be otherwise similar to one another, except for becoming the recipient of aid. In the case of fuzzy designs, the discontinuity in the probability of receiving state aid is used as an instrument for receiving state aid.

3.26   It is not clear whether sharp or fuzzy designs were intended in the Annex and some clarification around this would be helpful. Giving an intuitive explanation of the different types of discontinuity designs and referring to literature, eg, Imbens and Lemieux (2008), or Angrist and Pischke (2008, Ch. 6), might assist the reader.

Structural models

3.27   The discussion of structural estimation could be revised. It is not entirely clear that a full section on structural estimation is helpful, especially as no references are provided for a reader wishing to look into this more. A researcher wishing to use this framework must build a complete economic model of the effect of state aid and estimate a statistical model, the parameters of which can be related to the parameters of the economic model. If particular values of parameters of the statistical model correspond to a unique value of the 'state aid effect' parameter of the economic model, the causal effect is identified, and can be estimated consistently, provided that the statistical model adequately describes the observed variation in the data. In practice, this is likely to require the use of IV methods. In principle, this class of method provides the opportunity to model any situation, but can be demanding in terms of data requirements.

3.28   Some discussion of the potential benefits of the structural approach would be helpful. As it stands, the description merely conveys the enormity of the task, stating:

> It is impossible to provide general guidance on structural estimation… (p. 31, line 30)

and on the next line (which seems a little contradictory):

> Nevertheless, the general guidance provided before still applies. (p. 31, line 31)

3.29   Yet, researchers sometimes use structural models, and the reader could be interested in why this is the case.