



Brussels, 28.5.2014
SWD(2014) 179 final

COMMISSION STAFF WORKING DOCUMENT

Common methodology for State aid evaluation

Table of Contents

1	Introduction	2
2	The objectives of State aid evaluation.....	3
3	The evaluation plan	5
3.1	Objectives of the aid scheme to be evaluated.....	5
3.2	The evaluation questions	5
3.3	Result indicators	6
3.4	Methods: finding an appropriate basis for comparison	6
3.5	Data collection: using the best possible sources.....	9
3.6	Timeline of the evaluation.....	10
3.7	The body conducting the evaluation: ensuring independence and expertise.....	11
3.8	Publicity: facilitating the involvement of stakeholders	12
4	Selection criteria for aid schemes to be evaluated.....	13
4.1	Large aid schemes, including those under the General Block Exemption Regulation.....	13
4.2	Novel aid schemes	14
4.3	Aid schemes affected by significant foreseen changes	14
4.4	Other aid schemes	14
Annex I:	Technical appendix on relevant methods to identify the causal impact	16
Annex II:	List of possible result indicators	34
Annex III:	Glossary	39
Annex IV:	References	40

1 Introduction

Member States provide State aid to help achieve a wide variety of policy objectives, for example, to reduce regional disparities within a country, to promote research and development and innovation activities, or to promote a high level of environmental protection.

In determining which types of aid are compatible with the common market, EU State aid rules are based on a system of ex-ante scrutiny: aid schemes¹ are approved on the basis of pre-defined assessment criteria on the assumption that, if they comply with these assessment criteria, their positive effects will outweigh any negative effects. Typically, this assessment of schemes is performed without sufficient evaluation of their actual impact on markets over time.

To date, when applying EU State aid rules, relatively limited importance has been attached to ex-post evidence on what has actually been achieved with public funds or on the impact of State aid on competition. It is however essential for decision makers both at the Member State and EU level to consider the measurable results of State aid granted in the past, and the lessons learnt. This will help to ensure that schemes financed by State aid are more effective and create less distortion in markets, and will also improve the efficiency of future schemes and, possibly, of future rules for granting State aid.

A number of countries already evaluate their subsidy measures, even if not always on a regular basis.² Similarly, EU spending (including financing from the EU Structural and Investment Funds such as the ERDF, the ESF and the EAFRD) is subject to ex-ante, ongoing and ex-post evaluation in accordance with the applicable regulations and with the guidance documents published by the Commission.³ In order to avoid duplication in the evaluations carried out by Member States, the "Concepts and Recommendations" guidance document on monitoring and evaluation clarifies that the evaluation requirements of the European Structural and Investment Funds can be fulfilled by carrying out the evaluations required by the rules for State aid.

1 Aid schemes account for the majority of all granted aid: according to the 2013 Scoreboard data, approved aid schemes represent 23 % of all aid measures and 55 % of aid amounts, and a further set of block-exempted schemes represent 63 % of all aid measures and around 32 % of aid amounts. Council Regulation No 659/1999 defines 'aid scheme' as "any act on the basis of which, without further implementing measures being required, individual aid awards may be made to undertakings defined within the act in a general and abstract manner and any act on the basis of which aid which is not linked to a specific project may be awarded to one or several undertakings for an indefinite period of time and/or for an indefinite amount".

2 For example, in several Member States, State aid evaluation reports are regularly prepared for the Court of Auditors or the Parliament.

3 The Commission guidance documents on evaluation for the 2014-20 funding period (available here: http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm) set out in detail the relevant concepts and recommendations.

The State aid modernisation initiative⁴ aims to focus the Commission's enforcement efforts on larger aid schemes that are likely to have the most significant impact on the common market. At the same time, the analysis of cases of a more local nature with minor or more limited effects on trade will be simplified, including by providing more flexibility for Member States in terms of implementing such aid measures by increasing the scope of the new General Block Exemption Regulation⁵. In order to ensure that, overall, the positive effects of State aid (in fulfilling its original objective) continue to outweigh the potential negative effects on competition and trade, and to prevent undue distortion to the market, greater simplification should be combined with greater transparency, enhanced control of compliance with State aid rules at national and European level and effective evaluation⁶.

This paper sets out a common methodology for evaluating State aid schemes. It is designed to provide guidance to public authorities involved in planning and conducting evaluations.

2 The objectives of State aid evaluation

The overall objective of State aid evaluation is to assess the relative positive and negative effects of a scheme, i.e. the public objective of the aid relative to its impact on competition and trade between Member States. State aid evaluation can explain whether and to what extent the original objectives of an aid scheme have been fulfilled (i.e. assessing the positive effects) and determine the impact of the scheme on markets and competition (i.e. possible negative effects). Evaluation therefore differs in its purpose from the two ex-post exercises currently carried out by the Commission with regard to State aid schemes – monitoring⁷ and reporting⁸.

State aid evaluation should in particular allow the direct incentive effect of the aid on the beneficiary to be assessed (i.e. whether the aid has caused the beneficiary to take a different course of action, and how significant the impact of the aid has been). It should also provide an indication of the general positive and negative effects of the aid scheme on the attainment of the desired policy objective and on competition and trade, and could examine the proportionality and appropriateness of the chosen aid instrument.

4 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, EU State Aid Modernisation (SAM), 8.5.2012, COM(2012) 209 final.

5 Commission Regulation (EU) No .../2014 of XXX declaring certain categories of aid compatible with the internal market in application of Articles 107 and 108 of the Treaty

6 See also the Council conclusions on Reform of state aid control of 13 November 2012.

7 The Commission's monitoring exercise is a periodic review of the legality of a sample of State aid measures implemented by Member States. It is designed to ensure that Member States are implementing Commission decisions correctly and are complying with the relevant legal provisions (i.e. those embodied in the General Block Exemption Regulation). The Commission also assesses compliance with the ex-ante rules and conditions among a representative sample of cases.

8 The primary objective of the annual reporting by Member States is to increase the transparency of State aid granted by Member States. It also provides a source of reliable statistics for policy-making and monitoring purposes. The data in annual reports provide information primarily in quantitative terms (for example, to show the objectives towards which State aid was directed and with what level of budget. The Commission uses Member States' reports to prepare the State aid Scoreboard.

Based on this assessment, the evaluation can confirm whether the assumptions underlying the ex-ante approval of the aid scheme are still valid and can help to improve the design of future aid schemes and rules governing State aid. It could provide the basis for adjusting future State interventions so as to improve the effectiveness and efficiency of the aid to the extent necessary to guarantee that the positive effects are sufficient to justify accepting the distortion to the market caused by the intervention. Such improvements on future schemes could range from adjustments to the design, including changes to the selection criteria and a more extensive assessment of the incentive effect, to more significant changes such as promoting the use of an alternative form of aid, redefining objectives or target beneficiaries or considering non-aid options to achieve the same policy objectives.

It is important to set an appropriate timeline for the evaluation, allowing enough time to collect sufficient evidence whilst also providing results to policy-makers as soon as possible, so that potential improvements can be introduced in due time.⁹ In view of this, State aid evaluations should normally be considered as ongoing evaluations, to be conducted while the aid scheme is still in operation, rather than as purely ex-post ones, conducted only after the implementation of the scheme is completed. Account should be taken of particular cases where the full effects of an intervention might be perceivable in a longer timeline and where the evaluation will only be able to capture and measure initial effects.

State aid evaluation should ultimately be a learning exercise for both the Commission and Member States. For this to be possible, the evaluation should meet a certain minimum standard of quality. The Commission should therefore ensure that appropriate quality control of evaluations takes place. In particular, the Commission will analyse in detail the overall reliability of the evaluation and will highlight potential shortcomings at the two crucial stages, namely the evaluation plan and the final report. Where appropriate, the Commission could seek the support of external independent experts to assist in the quality control of the evaluation.

The Commission could also organise training sessions and workshops for national administrations on methods and techniques of evaluation. Furthermore, successful experiences and best practices from Member States could be shared and used to help design more effective aid schemes in the future.

The benefits of conducting evaluations will become evident within a few years, when the first evaluation reports are ready and their findings and recommendations are made available. These will then be able to be used to improve the design of subsequent aid schemes and, possibly, rules governing State aid. In the medium to long term, evaluation could gradually lead to more fundamental changes in the general approach taken to State aid.

⁹ Some State aid guidelines refer to a normal duration of four years for evaluated aid schemes.

3 The evaluation plan

It is essential that a comprehensive plan for evaluating a State aid scheme be drafted at an early stage, in parallel with the design the scheme. Approval by the Commission of the evaluation plan is crucial to ensure equal treatment. This plan must then be rigorously implemented.

Indeed, it is generally recognised that evaluations are more effective when properly planned and prepared for in advance, in particular as this makes it easier to collect the appropriate data. Early planning is also likely to significantly reduce the resources required for the evaluation, and ultimately to improve its quality.

The evaluation plan to be notified by the Member State, according to the relevant rules, to the Commission should contain at least the following minimum elements.

3.1 Objectives of the aid scheme to be evaluated

The first stage in evaluating a scheme is to set out clearly the underlying 'intervention logic' of the aid scheme, describing the needs and problems the scheme intends to address, the target beneficiaries and investments, its general and specific objectives, and the expected impact. The main assumptions relating to external factors that might affect the scheme should also be mentioned.

3.2 The evaluation questions

The evaluation plan should define the scope of the evaluation, i.e. it should include precise questions that can be answered quantitatively and with the necessary supporting evidence. These evaluation questions should focus on the impact of the State aid scheme and can be classified according to the following three levels:

1. Direct impact of the aid on beneficiaries, e.g.:

- Has the aid had a significant effect on the course of action taken by the aid beneficiaries? (incentive effect)
- Has the aid had an effect on the situation of the beneficiaries? (For example, has its competitive position or default risk changed?)
- To what extent has the aid had the effects expected?
- Have beneficiaries been affected differently by the aid? (For example, according to their size, location or sector)

2. Indirect impact of the aid scheme, e.g.:

- Has the scheme had spill-over effects on the activity of other firms or on other geographical regions? Did the aid crowd out investment from other competitors or attract activity away from neighbouring locations?

- Has the scheme contributed to the relevant policy objective?
 - Can the scheme's aggregated effects on competition and trade be measured?
3. Proportionality and appropriateness of the aid scheme, e.g.:
- Was the aid scheme proportionate to the problem being addressed? Could the same effects have been obtained with less aid or a different form of aid? (for example, loans instead of grants)
 - Was the most effective aid instrument chosen? Would other aid instruments or types of intervention have been more appropriate for achieving the objective in question?

The evaluation should, as far as is possible, assess the impact of the aid scheme at all three levels, addressing the relevant questions in respect of the scheme's objectives. However, the direct impact of aid on the beneficiaries is typically the type of impact that can most robustly be measured. In practice, the majority of evaluation methods that have been developed are designed for assessing this type of impact. Furthermore, evaluation of the direct effects of the aid, including the incentive effect, is of paramount importance as it can provide valuable insight into the types of indirect effects and distortions to be expected. In particular, where the aid provides no incentive effect, it can be assumed that the aid is distortive, in the sense that it provides the beneficiaries in question with windfall gains.

3.3 Result indicators

The evaluation questions should lead to the choice of specific result indicators that capture quantified information about results achieved by the State aid scheme. Annex II provides an indicative and non-exhaustive list of result indicators covering both the direct and indirect impact of a scheme, including the possible effects on competition and trade. The result indicators will depend on the objective of the aid being evaluated. The evaluation plan should explain why the chosen indicators are the most relevant for measuring the impact of this aid scheme.

3.4 Methods: finding an appropriate basis for comparison

State aid evaluations should be able to identify the causal impact of the scheme itself, undistorted by other variables that may have had an effect on the observed outcome, e.g. general macroeconomic conditions or firm heterogeneity (e.g. differences in firm size, firm location, financial means or management capabilities). The evaluation plan should set out the main methods that will be used in order to identify the effect of the aid, and discuss why these methods are likely to be appropriate for the scheme in question.

This causal impact is the difference between the outcome with the aid and the outcome in the absence of the aid. While the outcome with the aid is observed for firms who receive the aid, the outcome in the absence of the aid is only measured for firms who do not receive aid. By definition, we do not observe what the outcome would have been without the aid for the firms

who received the aid. To estimate the effect of the aid on aid beneficiaries, it is therefore necessary to construct this counterfactual, based on the most comparable firm(s) or control group.

The quality of this control group is crucial for the validity of the evaluation.

Firms who receive aid may well be in a different situation from firms who do not receive aid. They might, for example, face different local supply and demand conditions, have less easy access to credit or be more or less efficient. These factors may all have an impact on the performance or activity level of the firms, both when they receive aid and when they do not. Comparing the performance of beneficiaries with that of non-beneficiaries is likely to reflect this reality more than the effect of the aid itself. An evaluation of the aid scheme cannot therefore rely on a simple comparison between beneficiaries and non-beneficiaries, but must take into account the different characteristics of the two groups of firms, both those which can be observed and those which cannot.

In the case of regional aid for example, aid beneficiaries in regions where market conditions are unfavourable (i.e. where the local product, labour or capital markets are weak) typically perform worse than non-beneficiaries in more prosperous regions. This by no means reflects the effect of the aid itself, however. The relevant question is whether they performed better than they would have without the aid, not whether they performed better than non-beneficiaries in other regions.

Similarly, general industry trends must also be taken into account when identifying the effect of the aid. Even if beneficiaries of regional aid reduce their staff numbers, the aid may still have been effective. For example, when conditions within a particular industry as a whole are deteriorating and all firms are cutting jobs, aid beneficiaries might reduce employment to a lesser extent than they would have otherwise. This is illustrated in the graph below, which shows a negative trend in the amount of employment provided by firms receiving aid, both before and after the aid was granted. Nevertheless, the trend becomes less negative after the firm has received the aid. The difference in the extended trend line without aid and the line showing employment actually offered by the firm after receiving the aid isolates the positive influence of the aid.

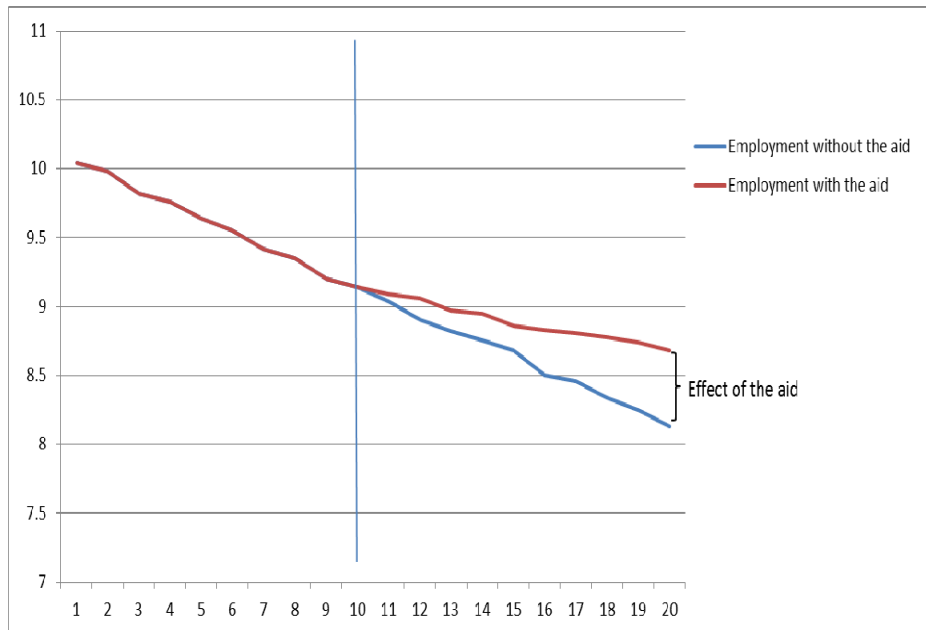


Figure 1 — positive influence of the aid where the current trend is negative

A specific problem emerges in terms of identifying a control group when non-beneficiaries have decided themselves to apply or not to apply for aid. For instance, if all firms are eligible (i.e. all firms who propose a project and apply for aid do receive some aid), then the firms who do not apply are likely to be those without projects. The firms' results may show that firms that did not receive aid performed worse in absolute and relative terms than those who did receive aid. This finding may however be entirely explained by the mere fact that the former group had no project to begin with, whereas the latter did, i.e. the management of the former group are lacking interest or creativity. It is therefore crucial that firms in the control group (firms who did not benefit from aid) are part of that group for reasons that have no influence on the measured outcomes. In particular, where firms have self-selected and voluntarily decided not to apply for aid, this condition may not be fulfilled.

Any systematic difference between State aid beneficiaries and non-beneficiaries should be properly accounted for in the design of the evaluation, in order to avoid a bias in the results (selection bias). In recent decades, several reliable methods have been developed to address this issue. The choice of method depends on the design of a particular State aid scheme and on the data available. The methods each have their limitations and are only valid when certain assumptions hold. Recognising and discussing these limitations and assumptions openly is crucial for the credibility of a study.

Randomising the process used for selecting beneficiaries is one way of making sure that the evaluation is unbiased. If aid beneficiaries are selected entirely at random, any systematic difference observed in the performance of the firms can be attributed to the aid. This method may however be difficult to implement in practice, in particular for large existing schemes. Other methods aim to use existing sources of exogenous variation in the environment in

which firms operate (i.e. variation not determined by parameters and variables in the model) to identify causality.¹⁰ Annex I to this guidance paper presents in more detail the most relevant methods, focusing on the practical aspects of their use. It discusses the way in which each method identifies causality, this being of particular importance in the context of State aid evaluations where the ex-ante design of the evaluation serves to ensure that a proper evaluation of the effects of the aid is possible.

Finally, the impact of multiple aid, either from one scheme, from several schemes or ad-hoc aid, should be controlled for. If non-beneficiaries in the given programme receive aid from other programmes, or if beneficiaries of the given programme receive additional aid from other programmes, the evaluation of the effects of the given aid scheme are likely to be distorted.

3.5 Data collection: using the best possible sources

Consistent and sufficient data must be collected on both the aid beneficiaries and the control group. Identifying the data required and obtaining access to the sources of the data forms are part of the planning of the evaluation.

Effective monitoring of the intervention and accurate collection and processing of data are crucial for ensuring the quality of the evaluation. As soon as the aid scheme is approved, a mechanism should therefore be put in place to monitor the intervention and to collect and process the appropriate data. This is likely to significantly reduce the costs of the evaluation.

Making sure that the necessary data on aid applicants and beneficiaries is collected is a crucial step in designing the evaluation plan, if the availability of this data can be made part of the eligibility conditions for aid.

With the exception of data on aid applications (including rejected applicants, when available), the data sources for aid beneficiaries and for the control group must be identical, for the data to be comparable. It is very likely that data will have to be taken from multiple sources, e.g. combining data from databases containing information about aid receipts with data from firm registries. The evaluation may need to draw on existing data sources, such as administrative data sources (e.g. the tax office, the companies register, innovation surveys and the patent office). The evaluation plan therefore needs to review the existing data sources, decide whether they provide sufficient information for the evaluation and ensure that access to them will be possible within the relevant timeframes.

Data from administrative sources, e.g. national statistical offices, is likely to be made available to evaluators only under certain conditions relating to privacy and confidentiality of business data. The conditions for access to this data must be described in the evaluation plan. Whenever necessary, the authority granting access to the data must ensure that the experts carrying out the evaluation have access to this data.

¹⁰ The most commonly used methodologies are differences-in-differences, regression discontinuity design and instrumental variables.

When data from several sources is used, it is very important that it is collected in a format that allows variables to be matched consistently. It may be necessary to find unique identifiers for observation units in each data set used. For example, firm and plant identifiers must be unique in all datasets, addresses must be collected in a format that allows geo-localisation, etc. The exact origin of the identifier may differ between Member States. It could, for instance, have a fiscal origin (e.g. a VAT number) or be directly provided by statistical institutes (e.g. SIREN and SIRET in France, the business identifier number and establishment identifier number respectively, both provided by the national institute for statistics and economic studies (INSEE)).

The evaluation of State aid could be complemented by information from surveys of aid beneficiaries and/or interviews with scheme managers. Qualitative information of this type is subjective by nature and answers may reflect the strategic interests of the beneficiaries rather than providing a genuine assessment on the effect of the aid. This risk is particularly high if the interviewee assumes that a positive testimony will improve the scheme's chances of receiving aid in the future. Nonetheless, if treated with the necessary degree of caution, information from qualitative exercises such as interviews and case studies can be a useful complementary source and can help in interpreting the results of the evaluation.

Whenever personal data will be processed in the context of the evaluations, EU law on the protection of personal data applies, in particular Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data and the national legislation implementing it as well as Regulation (EC) No 45/2001 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data.

3.6 Timeline of the evaluation

An evaluation plan should provide information on the precise timeline of the evaluation, which will be set in accordance with the approved duration of the scheme, and should include milestones, i.e. for collecting the data, carrying out the evaluation and submitting the final report. The timeline could vary according to the scheme and should therefore be discussed and agreed with the Commission on a case-by-case basis. Those involved in the management of schemes are advised to facilitate informal discussion on the content of the plan before submitting their official notification to the Commission.

In order to allow a proposed extension to an aid scheme to be assessed, the final evaluation report should be submitted to the Commission in sufficient time (e.g. six months before the scheme is scheduled to end). If no extension is envisaged the report can be submitted once the scheme has come to an end.

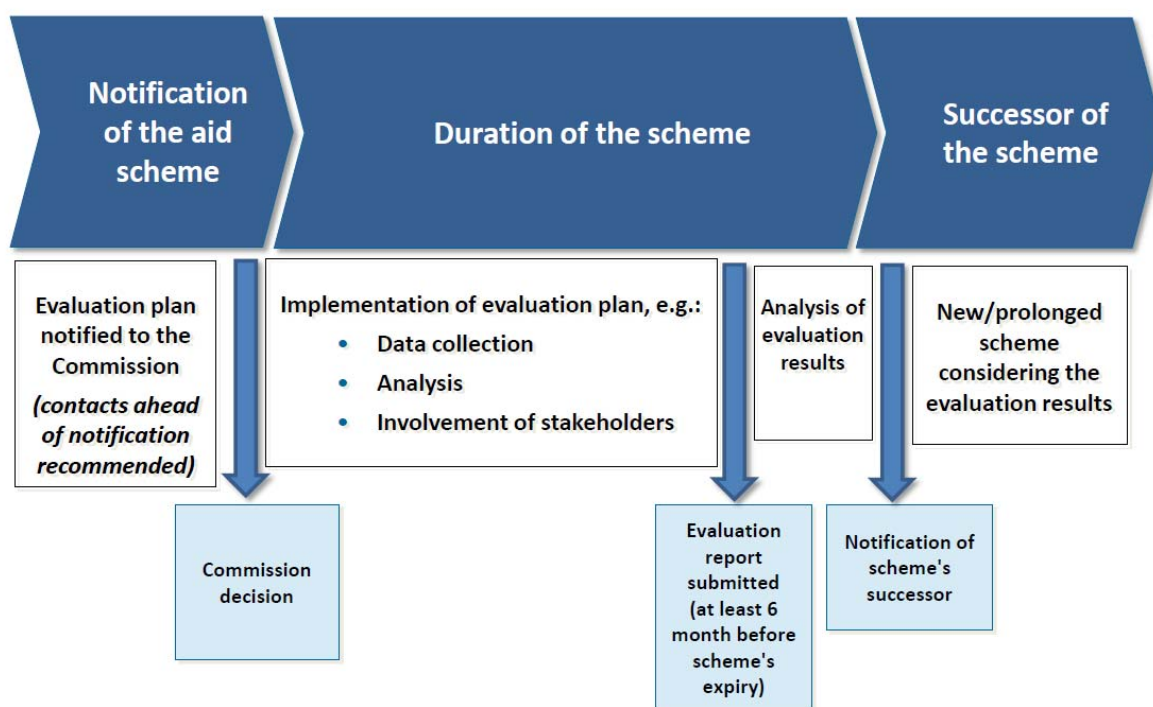


Figure 2 — overview of the evaluation process in the case of a notified scheme

3.7 The body conducting the evaluation: ensuring independence and expertise

Evaluation of the impact of State aid schemes should be objective, rigorous, impartial and transparent.¹¹ Each evaluation should be conducted on the basis of sound methodologies, by experts who have the adequate and proven experience and the methodological knowledge to carry out the exercise.

Evaluations should be carried out by a body that is at least functionally independent from the authority granting the aid, and that has the necessary and proven skills and appropriately qualified personnel to carry out such evaluations. The functional independence of the evaluator from the authority granting the aid is critical for ensuring the quality and credibility of the evaluation. This does not necessarily mean that a new body needs to be set up, nor that the evaluation needs to be outsourced to commercial evaluators. Depending on the specific organisations present in each Member State, it could be possible, for example, to make use of the independence and skills of organisations such as statistical offices, central banks, courts of auditors, public or private universities or research centres. This can be decided on a case-by-case basis for each scheme.

¹¹ See, for example, European Commission's Evaluation Standards, OECD Evaluation Norms and Standards, United Nations' Evaluation Standards and the World Bank's Independent Evaluation: Principles, Guidelines and Good Practice.

Early involvement of the body conducting the evaluation, for instance at the point of designing the scheme, is important for the success of an evaluation. It ensures that the State aid scheme will be able to be evaluated in the way proposed and guarantees that the necessary data will be collected. Whenever possible therefore, the evaluation plan should be drafted by, or at least in very close collaboration with, the designated evaluator. It should also include information, even if only of an indicative nature, on the necessary human and financial resources that will be made available for carrying out the evaluation. Information on the identity and role of each key expert involved in the evaluation and an estimate of their level of involvement are of particular relevance.

The evaluation plan should describe precisely the body conducting the evaluation or, if not yet chosen, the detailed criteria that will be used for its selection, in particular regarding independence, experience and skills. It should include existing alternatives whenever possible. Where the evaluator has not yet been selected, or has been selected but has not participated actively in the drafting of the evaluation plan, the reasons for this must be clearly stated. Even in this situation, the evaluation plan must be sufficiently detailed to allow a proper assessment of the validity of the evaluation to be made.

3.8 Publicity: facilitating the involvement of stakeholders

The evaluation should be made public. This implies that both the evaluation plan and the final evaluation report, once approved, should be given adequate publicity by being made available in the places described in the evaluation plan, for example, on a website. The Commission could also make these documents public¹².

If data used for the evaluation is personal and/or confidential, confidentiality needs to be guaranteed throughout the process of the evaluation, namely in accordance with Articles 8, 16 and 17 of the EU Charter of Fundamental Rights. Nevertheless, confidentiality does not extend to the results of the evaluation. In particular, no confidentiality clause can be included in the contract for the evaluation, apart from: 1. non-disclosure obligations applying to personal and/or confidential data; and 2. obligations to comply with general provisions of national statistical law and statistical secrecy, such as related to the presentation of the results.

The data collected during the evaluation should be made accessible for the purpose of replicating results or for further studies under conditions not more restrictive than those imposed on the body conducting the initial evaluation.

The authority granting the aid could ensure appropriate involvement of relevant stakeholders, who should be consulted at least once during the implementation of the evaluation plan. For

¹² With the exception of business secrets and other confidential information in duly justified cases (Commission communication on professional secrecy in State aid decisions, C(2003) 4582, OJ C 297, 9.12.2003, p. 6). Any publication of personal data must be done in compliance with EU law on the protection of personal data, in particular Directive 95/46/EC and the national legislation implementing it as well as Regulation (EC) No 45/2001.

example, stakeholders could be invited to discuss initial evaluation findings on the basis of an interim report. Such arrangements should be included in the evaluation plan.

4 Selection criteria for aid schemes to be evaluated

In principle, every State aid scheme is eligible for evaluation, but while evaluation is regarded as good practice, it is not required under State aid rules in all cases. State aid evaluation should remain a proportionate exercise and, in general, should be carried out for schemes that have a potentially significant impact on the internal market and may carry a risk of causing significant distortions if their implementation is not reviewed in due time. The focus in the relevant State aid guidelines is therefore on aid schemes which are: (1) large, including those under the General Block Exemption Regulation; (2) novel; or (3) face the possibility of significant (market, technological or regulatory) change in the near future that may require the assessment of the scheme to be reviewed. The individual State aid guidelines also specify other types of schemes that would benefit from evaluation.

4.1 Large aid schemes, including those under the General Block Exemption Regulation

In line with the Communication on State aid modernisation, the Commission could require the largest aid schemes to be subject to evaluation, since: (1) such schemes can impact the single market most severely if not well designed; (2) the largest efficiency gains can be made due to their high budgets; and (3) large schemes with many different types of beneficiaries can provide sufficient data for evaluation.

Certain aid schemes may still not be subject to evaluation if, despite their size, they do not entail any specific problematic aspect (e.g. routine cases, cases where a high number of beneficiaries is each receiving small amounts of aid, and cases where there is no risk of significant changes or when no serious distortions could arise).

Furthermore, the new General Block Exemption Regulation (GBER) defines large aid schemes on the basis of their budget (average annual budget exceeding EUR 150 million) and, for some categories of aid¹³ provides for their evaluation.

In order not to delay the entry into force of these large schemes, but also to ensure that they will be subject to an effective evaluation, the GBER provides for an exemption from notification for a maximum period of six months, which can be extended by the Commission upon approval of the evaluation plan¹⁴. The evaluation plan should be notified as soon as possible and at the latest within 20 working days following the scheme's entry into force.

13 Regional aid (except regional operating aid), aid for SMEs, aid for access to finance for SMEs, aid for R&D&I, aid for environmental protection (except aid in the form of reductions in environmental taxes under Directive 2003/96/EC) and aid for broadband infrastructures.

14 The Commission could also exceptionally decide that an evaluation is not necessary given the specificities of the case.

The new GBER also foresees the case of modifications or successors of these large schemes subject to evaluation, which should be notified unless the modifications are of a purely formal and administrative nature or are carried out within the framework of the EU co-financed measures.

4.2 Novel aid schemes

The definition of ‘novelty’ could vary across aid instruments and across Member States. Novelty will in principle be considered in terms of the nature of the aid scheme or the markets it is targeting, e.g. emerging markets where market developments are at a very early stage. These schemes have the potential to shape industries in a lasting and fundamental way. The scope for both benefits and distortions is therefore particularly large. Such novelty could include, for example, the introduction of a new capacity mechanism in the energy sector, aid to new types of technologies, or a novel type of support for renewable energy sources in the context of environmental aid. Evaluation of novel schemes also helps those currently designing new schemes as it allows them to take into account the latest developments on the market.

4.3 Aid schemes affected by significant foreseen changes

The possibility of significant (market, technological or regulatory) changes in the near future will be assessed on a case-by-case basis. Such significant changes could include, for example, the anticipated revision of an applicable regulation or aid to fast-moving industries where the market environment and the available technologies are developing at a rapid pace. If schemes are not adapted to the effects of these significant changes, there is a risk that public funding will not be used effectively (for example, funding may be given to a potential ‘market failure’ which will cease to exist) or that significant distortions will arise affecting new market entrants differently to incumbent companies, or creating unequal conditions for new technologies and legacy technologies. As illustrative examples, the revision of an existing regulatory framework (for example, in the electronic communication sector), the high fluctuation of input or output prices (for example, in the case of solar panels) or the launch of a new technology on the market (for example, the availability of the fourth generation mobile network for broadband services) are all cases where evaluation could be justified, in order that future schemes can take new market developments into account.

4.4 Other aid schemes

The guidelines for the different State aid fields also identify certain aid schemes where an evaluation would be particularly relevant.

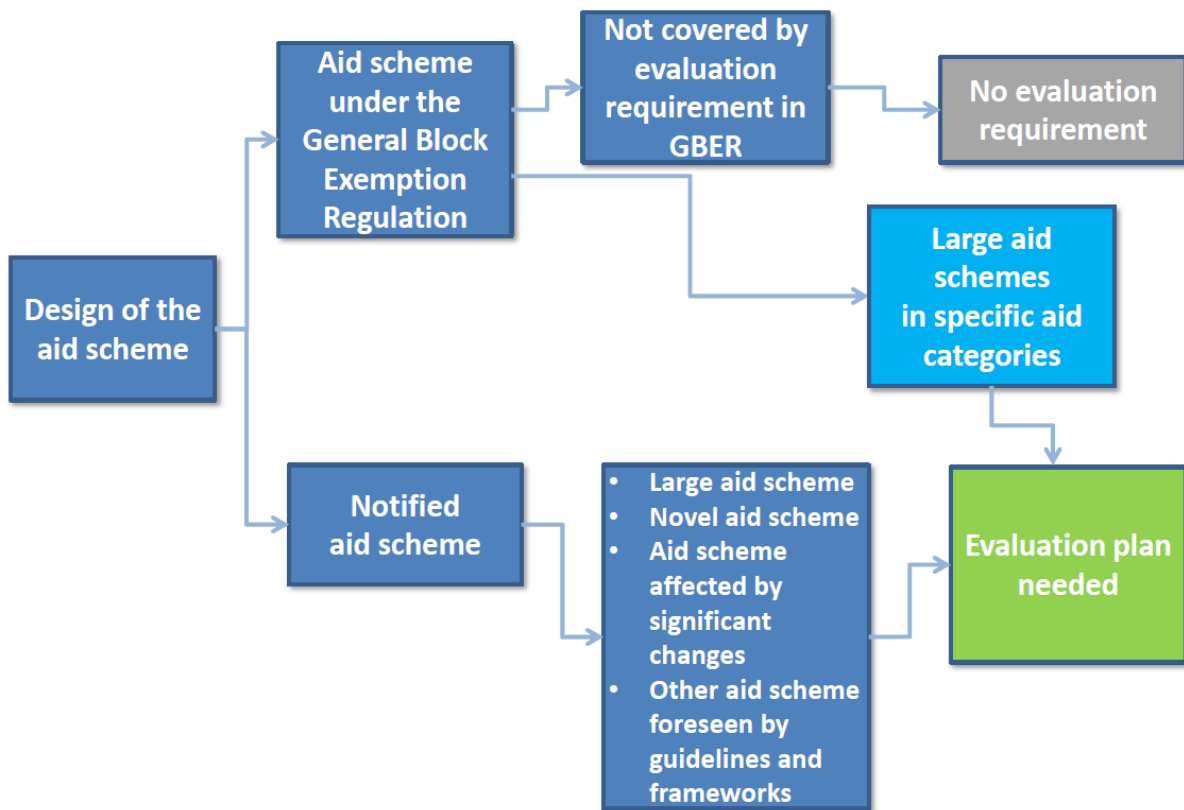


Figure 3 — selection of aid schemes for evaluation purposes

Annex I: Technical appendix on relevant methods to identify the causal impact

A State aid scheme can have impact at very different levels. It is normally expected to have a direct effect at the level of the beneficiary. Understanding the magnitude of this effect is crucial to assess the level of efficiency and effectiveness of a public measure. However, since aid is directed towards firms who interact in markets or regions which compete to attract economic activity, State aid also normally has indirect effects. These effects could for instance be spill-over effects on other firms (e.g. positive spill-overs from R&D or the crowding out of investment by other competing firms) or displacement effects (e.g. shifts in economic activity from one region to another). These indirect effects are the basis for both the potential harm and the benefits stemming from State intervention in the economy. Therefore, evaluating public measures requires assessing the magnitude of these indirect effects as well.

Measuring the direct and indirect effects of a policy normally requires the use of different tools. The last decades have seen an important development of methodologies and techniques intended at assessing the direct effect of policies on its beneficiaries. These techniques are presented in greater detail later in this section. Unfortunately, it is only in rare circumstances that these techniques will also allow assessing the indirect effects of the aid scheme on firms or regions. The evaluation of the indirect effects of the State aid scheme usually requires other types of evidence than what is used for assessing the direct effects on the recipients and interpretation normally relies more on economic theory and modelling. It is more difficult to provide precise guidance on this type of exercise as it has to be tailor made to the possible and expected positive and negative effects of the policy. Therefore, this evaluation has to be carried out after a careful and rigorous analysis of the most credible possible indirect effects of the policy. Based on this analysis, evaluators can derive measures based on micro data from non-aid beneficiaries, in particular in the same region, cluster or industry, as well as in neighbouring regions. This should form the core of the assessment of the indirect effects of the State aid scheme. If necessary, this can be complemented by more macroeconomic data and, most importantly, carefully chosen case studies.

The evaluation of direct effects is a necessary and crucial first step. However, a rigorously performed assessment of the indirect effects of the aid serves as an important piece of evidence in the assessment of the broader effects of the scheme. If the absence of additional investment by aid beneficiaries is, broadly, indicative of failure of the policy, even a positive effect is not sufficient to conclude a policy has fulfilled its objectives. In particular, if it turns out that the direct impact of the aid on the beneficiaries is very small or even non-existent, the scheme is very likely to be considered as not fulfilling its goal, unless very convincing arguments can be made about the existence of large and beneficial indirect effects. The contrary is also true: even if the evaluation finds that positive direct effects for the aid, the question remains whether there may be negative indirect effects that offset or even outweigh these.

Moreover, it is not always easy to clearly separate direct and indirect effects. A firm might have invested more (alleged direct effect) because its own investment has crowded out investment by competing firms (interacting indirect effect). A firm might also invest more because it expects spill-overs and investments by other firms. Moreover, it might be the aid itself or simply the granting of the aid which could have either effect. The likely presence, direction and expected magnitude of indirect effects should be discussed in detail in the evaluation of the direct effects. The economic theory that links the indirect effects to the aid should be explicitly stated and additional information that may serve as evidence supporting this theory should form an integral part of the evaluation.¹⁵

Causal Inference

The causal impact of aid is the difference between the outcome with the aid and the outcome in the absence of the aid. The outcome in the presence of the aid is observed for firms who receive the aid. However, the outcome in the absence of the aid is only measured for firms who do not receive aid. By definition, we do not observe what the outcome would have been without the aid for the firms who received the aid. To estimate the effect of the aid on aid beneficiaries, it is thus necessary to construct this *counterfactual*, i.e. to establish a reasonable scenario capturing what would have likely happened to the recipients of aid had they not received it. This requires finding a *control group*, i.e. a group of firms which should be as similar as possible to the group of firms that received the aid in all respects except for the aid itself.

The quality of the control group is crucial for the validity of the evaluation. Firms who receive aid typically differ in their characteristics from those who do not receive aid. They might for instance be active in a poorer area with less market potential, be more credit constrained, be more or less efficient, have a project to carry out or not, etc. Hence, naively comparing beneficiaries with non-beneficiaries is likely to reflect this reality more than the effect of the policy itself.

Making sure that this systematic difference between State aid beneficiaries and the non-beneficiaries (the so-called selection effect) does not bias the results is the core issue to carry out a valid evaluation. Several reliable methods have been developed in the last decades to address this issue. The choice of the method depends on the policy to be evaluated and on the available data. In addition, each of the methods has limitations and is only valid under a certain number of assumptions. The credibility of a study can be increased by explicitly identifying and discussing these limitations. This technical annex presents the most relevant

15 Although this document focuses on the direct effects of aid, the fact that the aid may have indirect effects does impose some analytical challenges on the assessment of direct effects, and special care has to be taken to the effects of market interactions.

methods, focusing on the most practical aspects and stressing the importance of a good identification strategy.¹⁶

A. Randomised experiments

The identification of a proper control group is key to obtaining good (i.e. unbiased) estimates of the effect of the policy. The most favourable case is when there is no selection effect because beneficiaries were selected randomly.¹⁷ Then, there is no systematic difference between beneficiaries and non-beneficiaries apart from the aid and the difference in the outcomes can be attributed to the policy.

However, random selection of aid beneficiaries is sometimes criticised for being at odds with the aim of many schemes to select the best possible aid beneficiaries on the basis of objective criteria. Still, in certain circumstances it might be possible to introduce elements of randomness in the eligibility or in the incentives to participate of beneficiaries. One example is setting a fixed budget for the given scheme. If the applicants' demand for support exceeds the budget and they are fairly equal in their characteristics, then one may try to establish randomness in treatment. Another example is randomly exposing potential recipients of aid to different levels of information about the scheme.

Pilot projects provide further opportunities for random allocation of aid. In case of innovative policies it might be advisable to evaluate a smaller scale pilot first. This pilot could have a smaller size and beneficiaries may more easily be chosen randomly. Another alternative would be to ramp-up a scheme, for instance to make eligible 25 % of randomly selected firms the first year to respectively 50, 75 and 100 % the second, third and fourth year (or alternatively, to advertise the scheme to a larger and larger audience). For a new policy, a period of ramp-up is in many cases an administrative necessity.

These ideas may be better suited for the implementation of totally new schemes or a large variation of existing schemes. It is probably fairly difficult to randomise eligibility (directly or indirectly) for the continuation of an existing scheme. However, this does not mean that random experiments cannot be used for parts of their evaluation. In particular, it is still possible to randomly select beneficiaries for potentially more efficient, more targeted and/or less distortive variants of the scheme. For instance, in the case of a grant scheme, it may be possible to randomly propose a newly introduced loan scheme instead.

16 This annex offers a quick and non-technical presentation of the econometric methods for policy evaluation. This presentation takes many elements from Givord (2010), other very good presentations can be found in Imbens and Wooldridge (2009) and Angrist and Pischke (2008).

17 Randomised experiments have for instance been the only acceptable methodology for the assessment of the effects of drugs and medical treatments for decades.

B. Quasi-experimental methods

Even though random experiments are the best possible way to evaluate the effect of policies, it is not always possible to implement them. Other methods have been developed to evaluate the effects of a policy from an ex-post perspective. They share the aim to use exogenous variations of the environment in which firms operate, to create situations very close to experiments (so-called natural or quasi-experiments).

It is generally a challenge for ex-post assessment to identify natural or quasi-experiments. However, a careful analysis of the design of the policy can entail an analysis of the existence of sufficient exogenous variation. If necessary, the initial setup can be adjusted to introduce more elements to allow identification of the effects of the policy.

Controlling for observable differences

As explained above, there normally exist significant differences between aid beneficiaries and non-aid beneficiaries. It is then necessary to account for these differences when comparing the outcomes between the two groups of aid and non-aid beneficiaries.

Many of the differences in characteristics are typically observable. The most common way to take these differences into account is to use linear regression. Linear regression seeks to control for the influence of observed characteristics on the outcomes. It assumes a linear relationship between the outcome, for instance the investment in R&D, and other characteristics of the firm, for instance the sector, age, size etc., including the granting of the aid. It is possible to see linear regression as a linear approximation of more complicated relationships.¹⁸ Linear regressions can be seen as general purpose techniques and are used in many different evaluation contexts.

An alternative to linear regression is to use matching techniques. Matching techniques aim at pairing each beneficiary with another firm that ‘looks’ very similar but did not receive aid. The observables used for matching can be firm characteristics or the estimated probability to receive aid (propensity score matching). Matching can be a useful way to control for observables in the context of a valid empirical strategy.

However, both simple linear regression and matching have some intrinsic limitations. Both are only valid under the so-called conditional independence assumption. This condition requires that, once the impact of the observable characteristics has been taken into account, the outcome is independent of the observable characteristics. In practice, this normally requires that every variable that impacts both the outcome and the selection is observable (and is taken into account with the proper functional form). If this is not the case, the mere fact that a firm participates reflects certain (unobserved) characteristics of the firm that also drive its performance. Both linear regression and matching will fail to provide a valid evaluation. For instance, if a firm has a ‘promising project’, this both affects the likelihood that it will apply

¹⁸ Moreover, it is possible to interact characteristics (for instance sales and sector) and to introduce functions of these characteristics (for instance squares of variables).

for aid (and get aid) and the likelihood that the firm is successful in growing a business. Not taking this into account will bias the results.

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. In reality, this assumption will rarely be fulfilled. Measuring all the variables that have an impact on the fact of applying or getting aid is rarely possible. Implementing matching techniques moreover requires that firms who get aid are very similar in their observable characteristics to those not getting it. If the matched firms are truly similar in every observable aspect, the reason why some firms received aid and some others did not are, by definition, unobserved. The justification to the validity of a matching-based evaluation or a simpler classical linear regression can thus not be the mere existence of a very complete dataset with many observed characteristics.

On the contrary, the potential justification for the use of matching or simpler linear regression in evaluation relies on the fact that these unobserved reasons that explain eligibility or attribution of aid have no direct or indirect influence on the outcomes (once controlled for the observables). For an evaluation based on simple matching or linear regression to be valid, one would need to be confident that the set of firms who did not receive aid has been exogenously determined. This requires that once the observables are controlled for, there remains no unobserved factor explaining eligibility or attribution of aid that would also directly or indirectly influence the outcomes. In general, matching firms that are equally eligible for aid will not fulfil this latter criterion. For instance, if all firms are eligible, firms who get investment aid are much more likely to have a project than firms who did not get aid (as they would also have applied and been granted aid otherwise). Overall, firms with a project are more likely to grow in terms of sales or employment, but this is not related to aid and matching on observables is not able to disentangle the two (unless we measure the existence of a comparable investment project).

In many situations, the conditional independence assumption is bound to fail. It may therefore be necessary to implement different techniques than mere linear regression or matching to account for the existence of unobserved selection into the treatment.

The remainder of this section presents the most common methodologies used to assess policy impact in this context in more detail, i.e. Differences-in-Differences, Regression Discontinuity Design (RDD), Instrumental Variables (IV). These methodologies derive their validity from different assumptions and the best choice is normally driven by the context of the policy and the availability of data. This presentation sets out the merits and weaknesses of each particular technique. With the noticeable exception of randomised controlled treatments (‘RCT’) presented above, there exists no technique superior to all the other ones in every aspect. The choice of a particular technique has to be guided by a careful analysis of the context of the measure and the available data.

It is worth stressing here that it is not the use of a specific econometric technique that allows identifying the effects of a policy; it is the exogeneity of the control group and hence the quality of the counterfactual. The quality of the evaluation study will therefore crucially depend on how convincingly the researcher can establish the exogeneity of the control group. In cases where residual biases might remain, it is essential to discuss these biases in detail, including their sources and the directions and likely magnitude of their effects on the results.

a. Difference in Difference

Rationale and identification

As explained earlier, a simple comparison between beneficiaries and even a well-chosen group of non-beneficiaries is unlikely to lead to a valid evaluation. The reason for this is that it is not possible to exclude the existence of unobserved differences between the two groups, leading to a persistent difference in outcomes even in the absence of the aid. Moreover, simply comparing the outcomes before and after the aid for beneficiaries is also likely to lead to a spurious evaluation. It does not allow disentangling the effects of the aid from the effects of other factors that also affect the outcome of the two groups, for instance the general economic trend, changes in the regulatory environment or increasing labour cost.

However, combining the two approaches might allow assessing the causal effect of the aid: this is the Difference-in-Difference approach. The general idea is to consider the difference in outcome between firms over time. Pre-existing differences would be attributed to other factors than the State aid. Only the change in these differences (the ‘Difference-in-Difference’) would be attributed to the aid. In other words, the method compares the difference in the performance between beneficiaries and control group *before* the aid as well as *after* the aid and then attributes the change in the difference to the aid. The method works if, over time, both the beneficiaries and the control group are affected by the other factors that also affect performance in the same way. It can then be concluded that the aid is the only relevant factor that explains the observed change in performance of beneficiaries relative to the control group.

The crucial assumption is that the differences between beneficiaries and control group are stable over time and that both groups are affected identically by common shocks (deviations from the mean) during the period. This assumption can fail in practice. For instance, if beneficiaries are the more vulnerable firms, they are likely to be more affected by economic downturns and general business climate. Therefore, the control group has equally to be made of vulnerable firms. Overall, the choice of the control group is the key for the validity of the method. Identification does not lie in the use of differences-in-differences, which is the mere technical implementation, but in the proper choice of control group.

Special care in the construction of the control group is needed if non-beneficiaries decided themselves not to apply for aid. Applying or not for aid can be expected to be related to the returns of getting the aid. Therefore, there are reasons to believe that the anticipated outcomes of firms who do not apply for aid (in terms of employment, productivity, sales, etc.) differ

from the expected outcomes for beneficiaries. For instance, if all firms who apply for aid get some aid, the only eligible firms who do not apply are those without a project (assuming the cost to apply is low). These firms are not only likely to perform worse in absolute terms but also comparatively worse as time passes, while better firms implement projects and grow. Employment, productivity or sales cannot be expected to remain parallel and double differentiation does not, in general, solve the problem.

Therefore, firms in the control group who did not benefit from aid need to have been selected for reasons that have no influence on the measured outcomes. They cannot have self-selected and voluntarily decided not to participate. The most convincing setup is when non-participation is related to non-eligibility that is the consequence of a natural experiment. In this case, non-eligibility is unlikely to be due to unobserved factors that also have an influence on the outcomes. Control groups could for instance be firms located in regions no longer eligible for aid (if this eligibility is not related to their own performance but rather to an exogenous event).

Implementation

From a technical point of view, difference-in-difference methods can be implemented either within a linear regression model or with matching. In the former case, the control group is chosen independently of the observable characteristics and therefore overall comparable to the whole group of the aid beneficiaries. Then, observable differences are taken into account in a classical linear regression. In the second case, the control group is made of firms that are individually comparable to each aided-firm in the sample based on observable factors. The outcome for each firm is compared to the outcome of its most comparable firm(s) and the results are aggregated. 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.

Depending on the circumstances, it may be worthwhile to compare the variations of outcomes of the beneficiaries and the control group before the aid. If the outcomes systematically start diverging already before the aid has actually been granted, it is likely that the control group and the group of the beneficiaries are diverging for reasons unrelated to the aid and the method does not give a valid estimate of the causal effect of the aid. This does not constitute a rigorous test of the validity of the assumption: such a test does not exist. However, this is at least a useful first sanity check.

Additional methods and robustness tests can be used when there are several potential control groups exist which are a priori valid. The first and most natural robustness check is to implement several difference-in-difference estimators and to compare the results. In addition, it is also possible to use these different control groups to build a more reliable estimate. Imagine a scheme targeted at SMEs in a particular region. Two potential control groups are the non-SME firms in this region or SMEs in an adjacent region. None of these firms voluntarily decided not to apply for aid, they were simply not eligible. Nevertheless, neither of these control groups is perfect: larger firms in the same region are likely to be affected

differently by general economic trends while SMEs in an adjacent region might be subject to different regional shocks. Instead of choosing between these two possible difference-in-difference estimators, it is possible to combine them and implement a triple difference estimator (DDD): starting from the ‘classical’ difference-in-difference between SMEs and non-SMEs in the concerned region, one can subtract the same difference-in-difference from the adjacent region to cancel the variation in outcomes between SMEs and non-SMEs in the region with aid¹⁹. Alternatively, one could systematically try to build a synthetic control group, made of SMEs from several adjacent regions and non-SMEs from the same region in order better replicate the pattern of the outcome for the beneficiaries before the aid (see Abadie, Diamond and Hainmuller, 2010 for details).

Inference

In addition to a careful design and choice of control group, the issue of inference has to be specifically addressed. The notion of inference in this context refers to the question of whether the effects that have been estimated are really significant. Statistical significance is a different issue from economic significance. The second one refers to the magnitude of the estimated effects compared to the other relevant parameters from an economic theory point of view. As explained earlier, economic significance is crucial. However, this discussion is in principle only relevant when, from a statistical point of view, the effects are estimated precisely enough, i.e. one can exclude that there is no effect at all.

There are reasons to believe that a straightforward inference under standard assumptions (such as the homoscedasticity assumption and the assumption of no autocorrelation) is likely to overestimate the statistical significance of the effects.²⁰

The first problem is related to clustering of data. If the control group as well as the group of the beneficiaries are each very homogenous (even if distinct from each other), all firms in each group are likely to be affected by similar deviations from the mean (shocks). In statistical terms, this means that the error term has a common component. If the variance of this common component is large compared to variation in outcomes observed for individual firms, the inference will be biased. With two periods and two groups, the problem can be particularly severe and borderline to an identification issue: it is impossible to separate the effect of the shocks shared within each group from the effect of the policy. The problem does not need to be as severe if the groups are not so homogeneous. However, it is always necessary to reflect on the presence of common shocks for homogeneous subsets of the groups. For instance if demand is local, it will normally be necessary to correct for the clustered structure of the error term at the level of localities. The same could apply to industries or sectors.

19 Consider the example of a regional SME scheme to create new employment. Imagine that at the end of the scheme it appears that SMEs in the region have performed 20% better than large enterprises in that region in terms of job creation. If in a comparable adjacent region (where no aid was given) SMEs also performed better than large enterprises (say 15% better), the impact of the aid may be estimated at roughly 5%.

20 This issue has been emphasised largely in the context of difference-in-difference technique, but the same problems can emerge with the other techniques covered by this paper.

The second problem emerges when panel data are used. Error terms of most firm level data like employment, productivity and investment are normally auto-correlated. This means that deviations from the mean in one period are likely to persist in the next period. Ignoring this issue leads to overestimating the precision of the estimation of the effects and to rejecting, more often than one should, the hypothesis that the policy has had no effect. This problem can be severe, as shown in Bertrand, Duflo, and Mullainathan (2004).

b. Instrumental Variables

Rationale and identification

Instrumental variables ('IV') is a classical method to deal with endogeneity of explanatory variables. Since benefiting from aid can be seen as an endogenous explanatory variable of the performance of a firm in a linear regression context, it is natural to use instrumental variables to evaluate the effect of aid.

A variable is endogenous when it is correlated with an unobserved element, which also determines the outcome. For example, imagine that one tries to identify the effect of State grant on firms' employment by regressing employment on programme participation and other observables. Let us imagine that the aid programme targets underperforming firms who are likely to face difficult local market conditions. Market conditions are not observable by the evaluator and hence cannot be controlled for directly. However, when this variable is left unaccounted for, the effect of the grant is likely to be underestimated by the evaluator due to the endogeneity of programme participation. Whether the firm faces favourable or difficult market conditions has an impact on both programme participation and on employment, i.e. programme participation is correlated with the error term explaining employment. The impact of market conditions on programme participation means that it is impossible to attribute the entire correlation between programme participation and employment to the causal impact of aid.

However, there also exist other factors explaining programme participation but not employment. For instance, as in Criscuolo et al (2012), geographical location may determine the total amount of money available for the programme in the region. Moreover the list of regions covered by the programme changes over time. If the programme budget for a given region changed over time for external reasons (e.g. average EU's GDP per capita dropped), this has an effect on the programme participation but not on firm's local market conditions. The change in employment that is related to the exogenous change in the programme coverage is not related to local market conditions. By focusing on this 'part' of programme participation variable it is possible to isolate the true impact of the participation on firm's employment without interference of local market conditions. This is the logic of instrumental variables.

For the evaluation of State aid, an instrumental variable is a variable that can explain the fact of receiving the aid but has no direct impact on the other unobserved determinants of the outcome that has to be measured. Instrumental variables then allow focusing on the

participation in the scheme without interference from the selection effects. For illustrative purposes, one can see the logic of instrumental variable as follows.²¹ In a first step, programme participation is regressed on all the exogenous variables, including the instrumental variables. In a second step, the participation variable (the variable indicating whether the aid was received) is replaced with the participation as predicted in the first step: this expected participation is not correlated with the unobserved element that also determines the outcome.

Issues with weak instruments

An instrumental variable is a variable that can explain the fact of receiving the aid but has no direct impact on the other unobserved determinants of the outcome that has to be measured. This simple and classical definition, however hides a number of practical difficulties. There exist tests aimed at checking for the consistency of instruments when more instruments are used than what is strictly necessary to identify a model. However, there exists no test of the validity of instruments. The main focus of a study using instrumental variables generally is to explain why each individual instrument can be assumed to be uncorrelated with the unobserved determinants of the performance of the firms, would it be employment, productivity, sales, investment, etc. Such explanations, based both on economic arguments and factual elements, are necessary to assess the validity of the evaluation. However, they are not sufficient, especially when several instruments are used.

The discussion of the quality of instrumental variables should include the issue of weak instruments, i.e. instruments weakly correlated with the outcome variable. When instrumental variables are poorly correlated with the endogenous variable, estimates are likely to be imprecise. One might be tempted to add more instrumental variables in that situation. It is well known that by instrumenting with a large enough number of variables, it is possible to recover enough of the initial variable to get statistically significant results. At the same time, the two stage least squares naturally gets closer and closer to the biased ordinary least squares estimate.²² The potential for such bias should be explicitly addressed in any evaluation using the IV method. In particular, the issue of the credibility, not only of their individual, but also of their joined exogeneity has to be addressed.

A special case arises when the endogenous variable is assumed to be auto-correlated. If the source of endogeneity is assumed to be solely contemporaneous, it is then possible to use past values as instrumental variables. However, one would then have to reflect on the exact validity of this approach. For instance, if explanatory variables are auto-correlated, this could also be the case of the measured outcome. Then, the lagged variables are also endogenous. More generally, if the autocorrelation of the explanatory variables is very large, exogeneity

21 In practice, two stage least squares are implemented in one step for well-known inference reasons.

22 A very interesting practical discussion about the biases created by weak instruments can be found in Bound, Jaeger and Baker's (1995) discussion of the statistical biases in Angrist and Krueger (1991). Moreover, instrumental variable estimates are biased at finite distance. Therefore, even with sufficiently large datasets to ensure apparent statistical significance, non-asymptotic biases can still be important.

assumptions might fail. If it is small, one could resort to using many lags (and potentially future values) and would risk falling in the pitfall of using many weak instruments described before. Overall, instrumenting by past values could be a valid strategy but it should be used with caution.

Generally speaking, to avoid the problems described earlier, it is highly advisable to only use a small number of convincing instruments. It is then, however also necessary to show that the instruments are good predictors of the endogenous explanatory variable.²³

Variations of two step estimations: Heckman (1979) selection model

When the endogeneous variable is a treatment variable (participation dummy), the first regression of the two stage least squares can be seen as a linear probability model of the probability to be treated. This linear probability model is a linear approximation. However, in some cases, the probability to be treated, even restricting to eligible firms, might be low. Then, linear approximations might be too coarse to effectively approximate this probability to be treated and to focus on the tails of the distribution, which are precisely the matter of interest. There are several ways to deal with this issue. They all rely on replacing the linear probability model of the probability to be treated by a non-linear function.²⁴

A classical approach is to treat the evaluation problem in the context of a selection model (Heckman, 1979). This approach treats the selection effect as an omitted variable problem in the linear regression of the outcome on the observables and the participation.²⁵ Several variants of this methodology exist, for example estimating the whole model by maximum likelihood, or instrumenting the granting of aid by the predicted value of the selection equation.²⁶

23 This can take the form of computing the Fisher statistics of the first step regression. The higher this Fisher, the less likely it is that instruments are weak. Stock, Wright and Yogo (2002) propose a formal test. For one instrument, it is for instance necessary that the Fisher statistics of the first step regression is larger than 10.

24 This section provides a very brief description of the selection models in this section. For a more complete presentation, the reader is referred to the seminal paper of Heckman (1979) and, mainly, Wooldridge (2002), chapter 17.

25 This omitted variable is the difference in conditional expectation of the outcome for the selected sample (here the aid-beneficiaries). Under certain assumptions on the selection process of the aid beneficiaries (for instance a probit or logit model), this difference can be formally derived (the inverse Mills ratio) and is a function of the selection parameters. Then, the effect of the policy can be identified by adding the omitted variable to the regression. The selection parameters are unknown, but consistent parameters can be recovered in a first step estimation of the selection process. This leads to the estimation procedure sometimes referred to as 'Heckit'. It first requires recovering the parameters of interest for the selection of the aid beneficiaries, for instance a probit or logit specification. Then, a consistent estimator of the effect of the policy can be recovered by adding the estimated inverse Mills ratio to the linear regression. Statistical software packages normally have a feature to perform this Heckman estimation.

26 For the presentation of all these methods, readers can for instance refer to, Wooldridge (2002), chapter 17.

However, it is crucial to reflect on the identification and in particular on the choice of variables. It is not satisfactory to use the same variables in both steps of the estimation, even if the results are sufficiently precise.²⁷

It is only reliable to estimate a selection model with a so-called exclusion variable. An excluded variable is a variable that explains selection of the aid beneficiary but not the outcome. It is not sufficient to remove one variable from the main equation to add it to the list of explanatory variables of the selection equation. On the contrary, this exclusion variable has to explain the selection but have no impact on the outcome one is trying to explain. It is in substance very close to a valid instrumental variable. The choice of such a variable cannot be driven by convenience; it has to come from economic theory, institutional structure and/or experience.

c. Regression Discontinuity Design

Regression discontinuity design (RDD) is the latest addition to the evaluation toolbox.²⁸ It has known a large success in the academic community in the last decade, mostly due to its simplicity. This method exploits the existence of a variable which has a discontinuous impact on the probability to be affected by a policy. In the context of State aid schemes, several types of discontinuities can be useful. The first one is geographical borders: the eligibility of schemes can be linked to precise administrative borders, like localities, NUTS regions, etc. The second one comes from conditions imposed on the firms which benefit from a scheme, in particular in terms of age and size.

Let us consider an example. Imagine that projects presented by firms are rated by points (out of 100) and firms who get at least 70 points get aid while the others get no aid. A firm who scores 71 has a marginally better project than a firm who scores 69. However, the consequence of this marginal difference is dramatic: one gets some aid, while the second gets no aid at all. Comparing the outcomes for these two firms is thus very indicative of the causal effect of the aid.

Formally, the RDD requires that the probability to receive aid is discontinuous, while all the other variables are continuous.²⁹ The technical implementation can be very close to this of

27 When the selection equation is non-linear, the inverse Mills ratio is not collinear to the other explanatory variables, even when the first equation includes only a subset of these explanatory variables. Then, in theory, the model is already identified. In this case, the inverse Mills ratio very often does not show enough variation, which leads to very imprecise estimates. However, especially with large samples, the estimation could still lead to significant results. Nevertheless, when all the variables of the selection model are also in the main equation, the model is solely identified due to the non-linearities of a particular parametric form.

28 A formal and complete description of RDD can be found in Imbens et Lemieux (2008).

29 Formally, there are two different regression discontinuity designs: the sharp and the fuzzy design. In the sharp design, which is implicitly the one described here, all firms, and only them, are treated above a certain threshold. In the fuzzy design, the discontinuity is less drastic: there is a discontinuity of the probability to be treated, but this does not change from 0 to 1. In absolute term, as far as state aid schemes based on eligibility conditions are concerned, it is only if one considers the treatment to be the eligibility that the

instrumental variable, using a threshold crossing dummy as instrument. However, there are two main differences. The first one is that RDD relies on weaker assumptions. In particular, we do not a priori require the independence of the instrument. For instance, in the case of scoring, firms with better project might apply more than firms with bad projects. The only requirement is that around the threshold the probability to apply should not be discontinuous. The second difference is that the estimates are built only on firms very close to both sides of the threshold. Weaker assumptions thus come at a cost: RDD estimates are even more local than estimates by instrumental variables generally are. If the effects of the aid differ for firms further away from the threshold, the RDD estimates are not a correct estimate of the effect on all aid beneficiaries.

The locality of these estimates can be of concern if one would expect large discrepancies of effects away from the threshold. Moreover, individual companies on the other side of the border could be very significantly affected by the policy. This could for instance be the case if displacement effects are important. Then, the use of RDD at the geographical border is not a good empirical strategy. Last, the strength of the RDD is to focus on a narrow bandwidth around the discontinuity. If the bandwidth is large, the impact of the other characteristics cannot be assumed to be constant. This issue is normally not solved by controlling for the observables, which assumes a particular functional form.

Graphical inspection of the data can provide comfort as regards the reliability of the assumption underlying RDD. In particular, it is very important to control for three things. The first one is that there indeed is a discontinuity on the granting of aid at the threshold. The second one is that the outcomes to be measured have a discontinuity at the same moment and no other discontinuity of the same kind anywhere else. Third, it is also necessary to check that there exists no discontinuity in the other parameters correlated with the outcome, including the propensity to apply for aid.

At last, discontinuities might be created deliberately in order to allow an evaluation of the scheme. In particular, ramp-up of policies could be used to create discontinuities and help the identification of the effects of a policy.

C. Structural estimation

In some instances, it is possible to go a step further and confront a theoretical model, for instance of firms' investments, with the data in order to recover the key parameters of interest. This approach is qualitatively different from those presented before. Structural estimation uses a completely specified theoretical model of firm behaviour. Estimation then allows recovering parameters determining firm's behaviour. This allows an evaluation at the closest of the determinants of the individual behaviour of firms and enables to carry out simulations about the efficiency of other tools. However, structural estimation is generally more demanding in terms of resources and data as well as in terms of assumptions.

design is sharp. Otherwise, if the treatment is to receive aid, we are in a fuzzy design. On the contrary, when the allocation is based on a scoring, we only consider firms who apply and the design is sharp.

It is impossible to provide precise guidance on structural estimation as the identification, estimation and inference has to be derived on a case by case basis. Nevertheless, the general guidance provided before still applies. First, it is necessary that the theoretical model matches the key stylised facts of the market. Second, the issues of unobserved characteristics and selection have to be explicitly and properly addressed.

D. Additional methodological remarks

Heterogeneity of treatment effects

The previous sections focused on the estimation of the average treatment effect on the treated. The very name suggests that the effect of the aid varies between beneficiaries. This heterogeneity may have many roots and many consequences. The first consequence might be that, if aid is very effective for some firms but much less for others, the average effect might be statistically insignificant. This absence of statistically significant effect does not mean that the aid has no effect for any firms. From a policy perspective, the average performance of a scheme is a very interesting first indicator. However, trying to understand the determinants of this heterogeneity is as important for the design of better schemes. It allows focusing directly on firms where the aid is the most effective and least distortive.

Thereby, whenever possible, the effect of the aid should be estimated for different types of firms, such as small firms vs large firms, young firms vs old firms, innovative firms, credit constrained firms, etc.³⁰

Distortions on the non-aided firms

Evaluating the impact of the scheme on non-participants, either directly or indirectly, is very informative for the evaluation of State aid. State aid may be distorting markets via effects on the non-beneficiaries, for example by knowledge spill overs from beneficiaries or by the reduction in relative competitiveness vis-à-vis beneficiaries, etc.

Moreover, the effects on the non-aided firms or locations can have an effect on the validity of the evaluation. For example, a part of the effect of regional aid could materialise by opportunities at the border: firms historically located on the ‘wrong’ side of the border moving their location just on the other side. Then, an RDD at the border would mostly capture this displacement effect and would risk overestimating the real aggregate effect of the policy. In such a situation, another empirical strategy has to be used (for example it may be useful to check the robustness of the evaluation on wider regions).

30 Another approach would be to systematically estimate different treatment effects for firms in different points of the conditional distribution. There is a growing body of literature estimating such quantile treatment effects, starting from Abadie, Angrist and Imbens (2002). This is a very useful tool to understand the intrinsic nature of the heterogeneity of treatment. However, it is less useful from a strict policy point of view, unless it is possible to directly target different firms depending on their position in the conditional distribution.

E. Data

Having access to appropriate microeconomic data that enables conducting impact evaluation is crucial. These data have to be consistent between beneficiaries and non-beneficiaries. Therefore, they need to have the same source, with the natural exception of information on the aid itself. The data should be from accessible at the most refined level although in some cases some form of aggregation at a later stage may be necessary.

Data capturing the result indicators of both the treatment as well as the control group are necessary, including the time at which the outcome is measured. Furthermore, as much data as possible on factors potentially influencing outcomes and the entities' decision to participate in the aid programme are necessary. This data is used to 'control for' differences between the treatment and control groups. For example, on the firm level such data may include location, size and demographics, as well as production inputs used.

The most natural source of data is of administrative origin, such as fiscal balance sheet data , or national surveys. These sources provide information on the location and activity of firms, and sometimes of individual plants. They normally allow to track investment and sales by activity as well as to compute financial ratios. Large national or community surveys, such as Community Innovation Surveys are also of interest. They cover a large and representative sample and provide very complementary information on specific topics. Last, merged employer-employees datasets are also a relevant source of information. They normally allow relating labour characteristics to each plant location. This can be crucial when the geographical dimension of labour is a matter of interest.

Apart from indicators on results and recipient characteristics, data about the aid and the aid granting process is necessary. This information would usually come from the aid granting authority itself. This includes data on the amount and timing of granting of the aid to beneficiaries. However, general data on the process of attribution of the aid is also particularly helpful. Data on rejected applicants is important, especially if the granting of the aid is made using a scoring mechanism.³¹

Access to such confidential data is normally regulated. Securing timely access to these data for the whole of the scientific team performing the evaluation is therefore crucial. Moreover, these administrative sources are normally accessible at a delay. It is important to take into account data availability when designing the evaluation plan.F.

31 Having data on rejected applications is particularly valuable for studies pursuing a regression discontinuity approach.

F. Examples

Example 1 (Regional aid): Criscuolo et al. (2012)³² have evaluated the Regional Selective Assistance ('RSA') scheme in the UK between 1986 and 2004. In this period, RSA provided discretionary grants to firms in disadvantaged areas. It was the main business support scheme in the UK. The scope for aid given under the RSA was governed by the Regional Aid Guidelines, in particular the maps of eligible regions ('regional aid maps'). These maps have evolved over time. By and large, the criterion for eligibility for a region is the relative position of the region in terms of GDP per capita or unemployment. Thereby, the status of a region can change either because it had developed over time or because the average EU per capita GDP changed (for instance when new member states joined the EU in 1995). Moreover, the indicators used to determine eligibility also change over time. Therefore, part of the change of eligibility of the firms does not depend on the situation of the firms themselves, but rather on events occurring outside of the UK or on changes in administrative rules. By focusing on this part of the changes in eligibility and assessing how these changes resulted in changes in investment activity, employment and productivity, Criscuolo et al. (2012) are able to convincingly identify the impact of the aid.³³

Example 2 (Enterprise support): Martini and Bondonio (2012)³⁴ have examined two cases of enterprise support — an investment grant available throughout Italy (Law 488) and various SME schemes in the region of Piemonte. The first evaluation is particularly interesting. It compares the firms who saw their aid application approved (i.e. the aid-beneficiaries) with comparable firms who saw their aid application rejected as the budget that was available for the aid had reached its limit. The use of rejected applicants in the evaluation is particularly useful to avoid the selection bias which typically arises if one were to just compare applicants with non-applicants. This group of firms had passed the first quality check, which means that they had a credible investment project. Therefore, they shared with the aid beneficiaries the same ambition to invest in a credible project. However, because of budgetary limits (rationing), they did not receive aid. The difference in performance between (just) successful applicants and (closely) rejected applicants provided a reliable estimate of the effect of aid.

Example 3 (Loan guarantees): Lelarge, Sraer and Thesmar (2010) evaluate the effects of a loan guarantee programme in France. The 'SOFARIS' programme provides insurance to lenders against borrowers' risk of default through guarantees. Borrowers pay an insurance

32 Criscuolo, C., R. Martin, H. Overman and J. Van Reenen, 2012. 'The causal effects of an industrial policy,' CEPR Discussion Papers 8818, C.E.P.R. Discussion Papers.

33 Technically, Criscuolo et al. (2012) are using an instrumental variable approach, as presented before in this technical appendix.

34 Report for DG REGIO. A. Martini, D. Bondonio: 'Counterfactual impact evaluation of cohesion policy: impact and cost effectiveness of investment subsidies in Italy' (2012).

premium, but this premium is subsidised. Lelarge, Sraer and Thesmar (2010) explicitly describe the nature of the selection effects. First, firms with more profitable projects are more likely to accept to pay the fee associated to the guarantee. Second, programme managers are likely to select socially desirable projects which might not otherwise get access to private funding. Overall, firms self-select into the programme and selection also occurs at the granting phase. This is likely to affect the results of naïve evaluations, based for instance on classical linear regressions or comparisons with the most comparable firm.³⁵ However, the factual and institutional context of the programme provides a source of identification of the effects of the policy. The programme was set up in the late 1980s and was initially restricted to firms active in the manufacturing and business services industries. In 1995, the public endowment of the programme was increased and new industries (construction, retail and wholesale trade, transportation, hotels and restaurants and personal services) became eligible. Lelarge, Sraer and Thesmar (2010) compare the newly eligible firms to the previously eligible firms to assess the effect of the programme on various indicators, like debt, employment, capital growth, financial expenses and probability of bankruptcy. Firms in these two groups are likely to differ. However, firms should be affected by similar macroeconomic shocks and therefore, their differences should not change over time, except for the expected effects of the policy itself.³⁶

Example 4 (Creative Credit): Bakhshi et al.,³⁷ use a randomised control trial experiment to assess the effect of an innovative business support scheme. The pilot study, which began in Manchester in 2009, was structured so that vouchers, or ‘Creative Credits’, would be randomly allocated to small and medium-sized businesses applying to invest in creative projects such as developing websites, video production and creative marketing campaigns, to see if they had a real effect on innovation. Creative Credits created genuinely new relationships between SMEs and creative businesses, with the award of a Creative Credit increasing the likelihood that firms would undertake an innovation project with a creative business they had not previously worked with by at least 84 per cent. The research found that the firms who were awarded Creative Credits enjoyed a short-term boost in their innovation and sales growth in the six months following completion of their creative projects. However, the positive effects were not sustained, and after 12 months there was no longer a statistically significant difference between the groups that received the credits and those that didn’t. The

35 This is an instance where matching techniques, here one-to-one nearer neighbour matching, is not a better way to solve selection problems than ordinary least squares. As explained earlier in this technical appendix, matching techniques are not, in general, a way to solve the issue of selection effects in the absence of natural experiments.

36 In practice, the authors implement a Heckman selection model with an exclusion variable at firm level and a classical IV strategy at sector level. See before in this technical annex for more details on these methodologies.

37 Report for Nesta, Creative Credits, a randomised controlled industrial policy experiment, Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley and N. Rathbone, June 2013, available at http://www.nesta.org.uk/sites/default/files/creative_credits.pdf.

report argues that these results would have remained hidden using the normal evaluation methods used by government, and calls for RCTs to be used more widely when evaluating policies to support business growth.

Example 5 (R&D&I support): Einiö (2013) has studied the impacts of Tekes's R&D subsidies on R&D investment, employment, and productivity in the period 2000-2006. Tekes is a national innovation agency responsible for the major part of R&D support in Finland. The study exploits regional variation in potentially awardable Tekes R&D support budget that arises from the higher ERDF funding in parts of the Northern and Eastern Finland (Objective 1 areas). These areas were initially determined in the accession negotiations of Finland in 1995 and were based on the no more than 8 persons per square kilometre population density rule. As a result of the relatively larger R&D support budget, the likelihood of receiving the support was higher in Objective 1 areas as compared to other parts of the country. This induced regional variation in treatment with a substantially larger fraction of companies being supported in the Objective 1 region. Because the regional allocation was based on the pre-determined population density rule based on the 1993 densities (and not on expected future levels of R&D investment or economic performance, for example), the study controls for the 1993 population density which effectively addresses concerns about regional selection. In practice, the treatment effects are estimated with an instrumental variables approach where an indicator for Objective 1 region is used as an instrument for programme entry. This approach identifies the impact of the support among those companies that entered the support scheme as a result of higher funding in the Objective 1 area. Validity of the setup is confirmed by showing that pre-programme trends between companies that entered the programme and the control group were not different. Einiö (2013) find positive impacts on R&D investment, employment and sales among the participants who were granted an R&D subsidy as a result of additional aggregate R&D-support funding in their region. While there are no instantaneous impacts on productivity, the study provides evidence of long-term productivity gains.

Annex II: List of possible result indicators

It should be noted that below is an indicative list for illustration purposes only. The actual result indicators should be set in accordance with the objective of the aid scheme and that of the evaluation.

Direct impact of the aid at the level of beneficiaries

	RESULT DIMENSION	RESULT INDICATORS
Regional aid	Positive impacts	Private investment matching public support Employment increase in the supported enterprises
Research, development and innovation aid	Additional RDI activity	Private investment matching public support Additional RDI expenditure undertaken by supported companies Number of new researchers employed in supported companies Number of new patents registered Number of enterprises supported to introduce new to the markets
Environmental aid	Positive environmental impacts	Reduced CO2 emissions of the beneficiary firms Additional capacity of renewable energy production Reduction of the share of waste landfilled or incinerated, Number of contaminated sites cleaned
	Early adoption of environmental standards	Percentage of companies reaching new environmental standards at least X months/years before they come into force [as minimum 1 year has been required and higher aid intensities have been allowed if earlier than 3 years]

Energy (infrastructure) aid	Reduced energy consumption	Number of households with improved energy consumption classification Decrease of annual primary energy consumption of public buildings Number of additional energy users connected to smart grids
	Renewable energy support	Production share of energy from RES
Risk finance	Positive impacts	Returns achieved in the fund; Leverage of private investments Number of firms receiving risk capital
	Picking losers:	Poor average performance of investee firms due to deficient commercial management/insufficient private participation
	Lack of sufficient degree of diversification	Too small/regionally constrained funds with limited return prospects that remain unattractive for private investors
Broadband aid	Increased broadband coverage	Additional household coverage with at least 30 Mbps broadband connection Additional household coverage or take up with at least 100 Mbps broadband connection
	Efficiency	Investments costs/aid per connecting a household (homes passed) Number of households signing up to new services
Rescue and restructuring	Positive impacts	Maintenance of employment and activity at firm-specific and regional level Changes in market share and productivity of aided firms
Aviation	Positive impacts	Number of air carriers using the airport; Private investment matching public support; Increase in regional productivity and/or gross value added (GVA))

	Negative effects	Duplication of lossmaking infrastructure or air routes; Deterioration of traffic of existing infrastructure (e. g. other airports in the catchment area or other means of transport)
--	------------------	---

Indirect impact of the aid scheme

	RESULT DIMENSION	RESULT INDICATORS
Possible positive effects	Macro-economic gains	Employment increase Increase in productivity and/or gross value added (GVA)
	Diversification of the regional economy	Number of industries under different NACE codes
	Increased cooperation between private and public	Number of enterprises cooperating with research institutions
	Positive externality / spill-over effects	Number of indirect beneficiaries (e.g. number of third parties accessing the facility) Changes in employment or activity in other firms and regions (aviation) Number of inhabitants with improved transportation means in the catchment area;
Possible negative effects on competition and trade	Sectoral bias	Aid was predominantly granted to one industry in a multi-sectoral scheme
	Bias towards loss-making firms or firms with low productivity (prevention of exit)	Proportion of high vs low productivity firms
	Bias towards incumbents	Proportion of old vs young firms
	Reinforce the market power	Change in market power of a dominant player
	Location effect	Relocation from a poorer region to a more developed one
	For security of supply	Locking-in in high-carbon energy sources, Assess whether the concerns in terms of black-outs are real and continue to exist; Foreclosure of national electricity markets
	For energy infrastructure:	Foreclosure of national electricity markets, reinforce the market power of an incumbent
	Rescue and restructuring	Changes in employment or activity in other firms and regions Changes in market share and productivity of aided firms

	Aviation	Duplication of lossmaking infrastructure or air routes; Deterioration of traffic of existing infrastructure (e. g. other airports in the catchment area or other means of transport)
--	----------	---

Annex III: Glossary

Baseline	The value of the indicator before the policy intervention at stake is undertaken.
Control group	Counterfactual analysis requires finding the most comparable firm(s) or control group, i.e. a group of firms which should be as similar as possible to the group of firms that received the aid — except that they have not benefitted from that aid.
Counterfactual	To estimate the effect of the aid on aid beneficiaries, it is necessary to construct a ‘counterfactual’, i.e. to establish a reasonable scenario capturing what would have likely happened to the aid beneficiaries if they had not received it.
Evaluation	The systematic collection and analysis of information about programmes and projects, their purpose and delivery; it derives knowledge on their impact as a basis for judgments. Evaluations are used to improve effectiveness and inform decisions about current and future programming.
Impact	The change that can be credibly attributed to an intervention. Same as ‘effect’ of intervention or ‘contribution to change’.
Indicator	A variable that provides quantitative or qualitative information on a phenomenon. It normally includes a value and a measurement unit.
Method	Methods are families of evaluation techniques and tools that fulfil different purposes. They usually consist of procedures and protocols that ensure systemisation and consistency in the way evaluations are undertaken. Methods may focus on the collection or analysis of information and data; may be quantitative or qualitative; and may attempt to describe, explain, predict or inform action. The choice of methods follows from the nature of the intervention, the evaluation questions being asked and the mode of enquiry — causal, exploratory, normative etc.
Result	The specific dimension of the well-being of people that motivates policy action, i.e. that is expected to be modified by the interventions designed and implemented by a policy. Examples are: the mobility in an area; the competence in a given sector of activity.
Result indicator	An indicator describing a specific aspect of a result, a feature which can be measured. Examples are: the time needed to travel from W to Y at an average speed, as an aspect of mobility; the results of tests in a given topic, as an aspect of competence; the share of firms denied credit at any interest rate, as an aspect of banks’ rationing.

Annex IV: References

Abadie, A., J. Angrist and G. W. Imbens (2002), 'Instrumental Variables Estimates of the Effect of Subsidised Training on the Quantiles of Trainee Earnings,' *Econometrica*, 70(1), 91–117.

Abadie, A., A. Diamond and J. Hainmueller (2007), 'Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Programme,' *Journal of the American Statistical Association*, June 2010, Vol. 105, No 490,

Angrist, J. and A. Krueger (1991), 'Does Compulsory School Attendance Affect Schooling and Earnings,' *Quarterly Journal of Economics*, 106.

Angrist, J. and J. Pischke (2008), "Mostly Harmless Econometrics: An Empiricist's Companion", Princeton University Press.

Angrist, J. D., and J. Pischke (2010), 'The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics.' *Journal of Economic Perspectives*, 24(2): 3-30.

Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley and N. Rathbone (2013), "Creative credits, a randomized controlled industrial policy experiment", Report for Nesta, available at http://www.nesta.org.uk/sites/default/files/creative_credits.pdf.

Bertrand, M., E. Duflo and S. Mullainathan (2004), 'How much should we trust differences-in-differences estimates?', *The Quarterly Journal of Economics*, 119, 249–275.

Bound, J., D. Jaeger and R. Baker (1995), 'Problems with Instrumental Variable Estimation When the Correlation Between the Instruments and the Endogenous is weak,' *Journal of the American Statistical Association*, 90(430), 443–450.

Criscuolo, C, R. Martin, H. Overman and J. Van Reenen (2012), 'The causal effects of an industrial policy,' CEPR Discussion Papers 8818, C.E.P.R. Discussion Papers.

Duflo, E., R. Glennerster & M. Kremer (2007), 'Using Randomisation in Development Economics Research: A Toolkit,' CEPR Discussion Papers 6059, C.E.P.R. Discussion Papers.

Duflo, E., and M. Kremer (2005), 'Use of Randomisation in the Evaluation of Development Effectiveness,' in *Evaluating Development Effectiveness*, ed. by O. Feinstein, G. K. Ingram, and G. K. Pitman. New Brunswick, New Jersey and London, U.K.: Transaction Publishers, vol. 7, pp. 205{232}.

Einiö, Elias (2013), 'R&D Subsidies and Company Performance: Evidence from Geographic Variation in Government Funding Based on the ERDF Population-Density Rule', *The Review of Economics and Statistics* (forthcoming).

European Commission's Evaluation Standards. Available at: http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/standards_c_2002_5267_final_en.pdf.

Garicano, L. C. Lelarge and J. Van Reenen, (2012), 'Firm Size Distortions and the Productivity Distribution: Evidence from France,' CEP Discussion Papers dp1128, Centre for Economic Performance, LSE.

Givord, P. (2010), « Méthodes économétriques pour l'évaluation de politiques publiques », WPD3E n° G2010-08.

Givord, P., Rathelot, R. and P. Sillard (2013), 'Place-based tax exemptions and displacement effects: An evaluation of the Zones Franches Urbaines programme', *Regional Science and Urban Economics*, Volume 43, Issue 1, January 2013, Pages 151-163

Heckman, J. J. (1979), 'Sample Selection Bias as a Specification Error,' *Econometrica* 47, 153–161.

Imbens, G. and J. Wooldridge (2009), *Recent Developments in the Econometrics of Programme Evaluation*, *Journal of Economic Literature*, 47:1, 5-86

Imbens, G. W., et T. Lemieux (2008), 'Regression discontinuity designs: A guide to practice,' *Journal of Econometrics*, 142(2), 615–635.

C. Lelarge, D. Sraer and D. Thesmar (2010), 'Entrepreneurship and Credit Constraints: Evidence from a French Loan Guarantee Programme,' NBER Chapters, in: *International Differences in Entrepreneurship*, pages 243-273, National Bureau of Economic Research, Inc.

Keane, M. P. (2010), 'A Structural Perspective on the Experimentalist School.' *Journal of Economic Perspectives*, 24(2): 47-58.

Martini, A. and D. Bondonio (2012), 'Counterfactual impact evaluation of cohesion policy: impact and cost effectiveness of investment subsidies in Italy', Report for European Commission, DG Regio.

Nederlandse Rijksoverheid (2012), 'Durf te meten', Eindrapport Expertwerkgroep Effectmeting, available at <http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2012/11/23/durf-te-meten-eindrapport-expertwerkgroep-effectmeting.html>

Nevo, A. and M. D. Whinston, (2010), 'Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference.' *Journal of Economic Perspectives*, 24(2): 69-82.

OECD Evaluation Norms and Standards. Available at: <http://www.oecd.org/dac/evaluation/dcdndep/41612905.pdf>.

Sims, C. A. (2010), 'But Economics Is Not an Experimental Science.' *Journal of Economic Perspectives*, 24(2): 59-68.

Stock, J., J. Wright and M. Yogo (2002): 'A Survey of Weak Instruments and Weak Identification in Generalised Method of Moments,' *Journal of Business and Economic Statistics*, 20(4), 518–29.

United Nations Evaluation Group (2005), 'Standards for Evaluation in the UN System' Available at: http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=22.

Wooldridge, J. (2002), "Econometric Analysis of Cross Section and Panel Data", Cambridge: MIT Press.

World Bank (2003), "Independent Evaluation: Principles, Guidelines and Good Practice". Available at: <http://siteresources.worldbank.org/INTDGF/Resources/Evaluation&LearningNote.pdf>.