



Dare to measure: Evaluation designs for industrial policy in The Netherlands

Final report of the Impact Evaluation Expert Working Group

November 2012

In Memoriam Jules Theeuwes 1944-2012

As chair of the working group, Jules Theeuwes' role in creating this report was substantial. Sadly Jules did not live to see it in its final form. Despite his illness, Jules' commitment to this report continued to the end. It is therefore our hope that Jules' ideas are echoed in this final version.

Working group members

Wolter Hassink (UU)

Bas van der Klaauw (VU)

Marco van Maasacker (Netherlands Court of Audit)¹

Wim Schaasberg (Statistics Netherlands)

Bas Straathof (CPB Netherlands Bureau for Economic Policy Analysis)

Jules Theeuwes (SEO Amsterdam Economics, chair) †

Thomas Dirkmaat (Ministry of Economic Affairs; General Economic Policy Department, secretary)

Tim Gelissen (Ministry of Economic Affairs; General Economic Policy Department, secretary)

Jeroen Heijs (Ministry of Economic Affairs; NL Agency)

Luuk Klomp (Ministry of Economic Affairs, Directorate-General for Entrepreneurship and Innovation)

The working group would like to express its gratitude to the following people for their willingness to share their knowledge and experience:

Godelieve van der Beken, René den Boer, Pieter de Bruijn, Jacqueline Coenraads, Jan Dixel, Jeroen Doorman, Anja Hezemans, Jan Hendrik Hoekstra, Edi Huster, Martijn Janmaat, Roe van Kasteren, Sander Kes, Mineke de Lange, Pieter Moerman, Aisia Okma, Marja Peperkamp, Theo Roelandt, Karen de Ruijter, Robert Schaart, Paul Silvertant, Gideon van der Staaïj, Koos Straver, Marjolijn van Valkenhoef, Suzanne Verboon, Dmitri Verhoeven and Upton van der Vliet.

Special thanks go to Piet Donselaar both for his material contribution and for his editorial work.

A final word of thanks is owed to the following people who contributed their expertise as deputies for some working group members:

Peter Kempkes (Netherlands Court of Audit), Ruurd Schoonhoven (Statistics Netherlands), Geert Thijssen (NL Agency) and Henry van der Wiel (CPB Netherlands Bureau for Economic Policy Analysis).

¹ The representative of the Netherlands Court of Audit has the role of observer in the working group, and reflects on the working group's plans based on the (development of the) recommendations made by the Netherlands Court of Audit in its investigation of innovation policy. This is an official representation of the Court of Audit. The Board of the Netherlands Court of Audit is not bound to this representation and may adopt a different opinion in future judgments.

Management summary

The government's aim with industrial policy is to improve the entrepreneurial climate in the Netherlands. This policy derives its legitimacy from the presence of market failures, such as externalities, imperfect information and coordination failure. An important consideration is that the costs of government intervention must not exceed those of market failure. The government has a fairly extensive budget for various policy measures. Furthermore, the large number of companies that come into contact with industrial policy carries an obligation to determine the effectiveness of the various policy measures with the greatest possible accuracy.

The main task of the Impact Evaluation Expert Working Group is to systematically analyse possible ways of evaluating the direct impact (output) of several of the Ministry of Economic Affairs' policy measures. The question is what proportion of the observed output is attributable to application of the policy instrument. The answer might be the volume of additional R&D that companies are persuaded to perform because of the Research & Development Deduction (RDA), or the number of newly qualified engineers and technicians at upper secondary vocational education level destined for the top sectors through the Centres for Innovative Craftsmanship (CIV). Evaluation of the effectiveness at output level can be seen as the first step in evaluating an instrument's ultimate social and macroeconomic impact: the effects on outcome level. However, the influence of the business cycle, international economic trends and other exogenous factors render measuring the outcome effects of policy instruments extraordinarily difficult.

The main problem in empirical policy evaluation is the frequently selective use of instruments, which renders the group of companies that use an instrument incomparable with those that do not. Any difference in outcomes between the two groups could just as easily be attributable to differences in company characteristics as to the policy intervention. This presents a problem, and the more so in that differences are often occasioned by unobserved characteristics of companies, which are difficult to control for. Consider the case in which efficiently managed companies form the majority of subsidy applicants. Any effect found when comparing applicants with non-applicants might have arisen because of the subsidy, but could equally originate from differences in the quality of management. The literature refers to this phenomenon as self-selection. The ideal way to resolve this problem would be to perform an experiment with random allocation, along the lines of clinical drug trials. If Randomised Control Trials (RCT) of this kind are infeasible, it may be possible to identify opportunities for natural experiments, in which existing institutions give rise to comparable groups of companies, some of which will, and others will not, have used the policy instrument. The various techniques for using a natural experiment to analyse impact are presented in Chapter 2. It is hard to measure impact without a RCT or natural experiment; the estimated impact could just as easily be explained by self-selection as the result of policy.

Survey methods may be used alongside econometric analysis in investigating the impact of policy. Surveys are frequently employed with companies, where some have and others have not participated in an intervention. As with non-experimental quantitative evaluation, self-selection is a problem with the survey method. There are also risks of strategic behaviour and a low or selective response. Nonetheless, surveys often yield a more detailed picture of the precise functioning and users' perceptions of a specific instrument. A survey is therefore advisable as a supplement and check on the quantitative methods described in this report.

The Directorate-General for Entrepreneurship and Innovation requested the Impact Evaluation Expert Working Group to develop possible evaluation designs for six policy instruments. The selected instruments include both traditional generic forms and new, more demand-driven alternatives, such as the Top Consortia for Knowledge and Innovation. From the diversity of policy instruments a picture is emerging of the impact measurement opportunities for specific types of instrument across the board of policy options. Chapter 4 describes evaluation designs for these instruments. The working group was also requested to reflect on the recent evaluation of the Promotion of Research and Development Act (WBSO) (Chapter 3). Table I lists the evaluation designs that the Expert Working Group considers most promising for impact measurement, and also gives the salient observations about the WBSO evaluation.

This exercise demonstrates that impact measurement is often feasible, by innovative and creative means, on the basis of natural experiments. However, natural experiments do not always succeed in revealing a policy instrument's impact on an entire target group of companies. Natural experiments are frequently concerned with only a subset of companies with specific characteristics within the target group or scheme, thereby limiting an estimate of the instrument's impact to part of the target group, or part of the scheme. By way of illustration, a suggestion in one of the evaluation options for Innovation Performance Contracts (IPC) was to use NL Agency's application assessments for the 2011 and 2012 tenders. Applying the regression discontinuity method the IPC applicants will be divided into two groups of companies: those with a just-sufficient score to be eligible for the IPC, and those with a just-insufficient score, which accordingly received no IPC. The other participating companies will be disregarded. The extrapolation of results to the entire population of companies is frequently not without risk.

Table I Evaluation design for each instrument

Instrument	Recommended preferred options for evaluation designs
<i>Innovation Credits (IK) of the SME+ Innovation Fund</i>	<i>Propensity score matching based on NL Agency assessment procedure</i>
<i>Research and Development Deduction (RDA)</i>	<i>Discontinuities around the maximum WBSO hours and RDA introduction</i>
<i>Certificate of Good Service (BvGD) for municipalities</i>	<i>Regression analysis based on survey output</i>
<i>Centres for Innovative Craftsmanship (CIV)</i>	<i>Comparison of ex-students and/or institutions based on difference-in-difference</i>
<i>Innovation Performance Contracts (IPC)</i>	<i>Randomized allocation (RCT) and ranking (regression discontinuity)</i>
<i>Top Consortia for Knowledge and Innovation (TKI) and TKI supplement</i>	<i>Comparison of TKI companies with other companies or comparison of TKI companies that do and do not participate in a project based on difference-in-difference</i>
<i>Reflection on WBSO evaluation</i>	<i>The dynamic panel data model attempts to adjust where possible for observed and unobserved characteristics, but the evaluation of major generic measures of this kind presents problems. Use discontinuities more in future</i>

To better clarify policy impact in the future, the Impact Evaluation Expert Working Group recommends devoting more attention in the development phase of new policy instruments to the logical substantiation of policy instruments (policy theory), and consideration of how to evaluate an instrument's impact quantitatively (evaluation design). Preferably, where possible, RCTs should be performed on the basis of (weighted) randomization. It is also advisable to work with small-scale, easily evaluated pilots before a large-scale policy roll-out. Data availability is vital for satisfactory evaluation. Measures for improving data availability include the linking of databases, active monitoring of accepted and rejected applications, and financial incentives to increase the response rate among survey subjects. It is possible moreover to combine the evaluation of instruments that have comparable policy objectives. The working group finally recommends analysing the impact of the other industrial policy instruments (with a greater financial scale), and the other Ministry of Economic Affairs policy areas, based on the approach taken in this report.

Contents

Management summary	II
Explanation and guide for the reader	V
Chapter 1 Market failure and industrial policy	1
1.1 Introduction.....	1
1.2 Market failures with generic policy.....	1
1.3 Industrial policy in practice and the top sector approach.....	6
1.4 Strengths and weaknesses of industrial policy impact measurement.....	7
Chapter 2 Statistical methods for empirical policy evaluation	11
2.1 Causality and correlation.....	11
2.2 Potential outcomes model.....	11
2.3 Randomised control trials.....	12
2.4 Natural experiments.....	13
2.5 Standard regression analysis.....	17
2.6 Survey methods.....	18
2.7 Conclusions.....	18
Chapter 3 Reflections on WBSO Evaluation: A case study	21
3.1 Introduction.....	21
3.2 The WBSO.....	21
3.3 The evaluation of the WBSO for 2006-2010.....	23
3.4 Conclusion.....	26
Chapter 4 Evaluation designs for instruments	29
4.1 Introduction.....	29
4.2 Innovation Credit (IK).....	29
4.3 Research and Development Deduction (RDA).....	36
4.4 Certificate of Good Service (BvGD).....	43
4.5 Centres for Innovative Craftsmanship (CIV).....	51
4.6 Innovation Performance Contracts (IPC).....	57
4.7 Top Consortia for Knowledge and Innovation (TKI).....	65
Chapter 5 Conclusion and recommendations	71
5.1 Conclusion.....	71
5.2 Recommendations.....	71

Explanation and guide for the reader

Introduction

The Ministry of Economic Affairs needs a clearer view of the effectiveness of its policy instruments. In response to the coalition agreement of the Rutte-Verhagen government² and reports from the Netherlands Court of Audit and CPB Netherlands Bureau for Economic Policy Analysis, the letter to Parliament '*Naar de Top: Het bedrijvenbeleid in actie(s)*' (To the Top: Industrial policy in action(s))³ expressed a clear ambition for industrial policy monitoring and impact measurement. This report gives firm starting points for refining this ambition for measuring the impact of policy instruments that fall under industrial policy.

A sound and robust analysis of policy impact starts from the statistical demonstration of a causal relationship between the instrument employed and the set objective. However, this aim is not always straightforward to achieve in practice, which prompted the Ministry of Economic Affairs to set up the 'Impact Evaluation Expert Working Group' to investigate and assess the relative merits of various analytical methods for policy instrument effectiveness.

Impact measurement is defined in this report as the quantification of an instrument's direct impact on the set objective. This report considers descriptive investigation methods that provide a more qualitative view to be a fallback option for cases in which an instrument's effectiveness is impossible to quantify. The more process-related components of policy evaluations (such as the implementation of instruments by NL Agency) are outside the scope of this report.

The Directorate-General for Entrepreneurship and Innovation requested the Impact Evaluation Expert Working Group to develop possible evaluation designs for the following six policy instruments⁴:

- 1) Innovation Credits (IK) of the SME+ Innovation Fund;
- 2) Research and Development Deduction (RDA);
- 3) Certificate of Good Service (BvGD) for municipalities;
- 4) Centres for Innovative Craftsmanship (CIV);
- 5) Innovation Performance Contracts (IPC);
- 6) Top Consortia for Knowledge and Innovation (TKI) and TKI supplement.

Most instruments are concerned with new policy, as stated in the above letter to Parliament. The selected policy instruments are representative of the broad range employed within the Directorate-General for Entrepreneurship and Innovation. The selection covers instruments with diverse budgetary burdens, and includes both traditional generic instruments (IK, RDA and IPC), new, more demand-driven instruments (CIV and TKI) and a more 'indirect' policy instrument (BvGD). The Directorate-General for Entrepreneurship and Innovation expects this selection to provide a clear picture of the strengths and weaknesses of quantitative impact measurement on a broad spectrum of policy options.

Terms of reference

The working group's principal task is to recommend the most suitable evaluation designs for ascertaining the effectiveness of the above instruments given their objective, design and operation, and the availability of data. The working group was also requested to express a view of the relative quality of the evaluation designs and rank them in order of clarity and robustness with respect to the instrument's effectiveness.

The working group was additionally requested to assess the quality of the output of ongoing evaluations and to identify points for future improvement of these evaluations. The evaluations concerned were of the WBSO and Technopartner SEED. A delay in the evaluation of the latter instrument prevented its discussion in the working group.

Finally the ministry requested the working group to examine opportunities for analysing the impact of policy on the higher policy objectives (such as economic growth and innovativeness), and the impact of the sectoral and integrated demand-driven nature of the 'top sectors' approach.

² More explicitly, the passage: 'Subsidies are provided only if the effectiveness is proven.'

³ Parliamentary Papers, session year 2010-2011, 32637, no. 15.

⁴ For recognisability reasons we decided to use the Dutch abbreviations for the instruments.

Scope of the report

The evaluation designs covered in this report are oriented primarily to clarifying the direct results achieved with an instrument, which are referred to as the first-order effects. Any impacts on higher policy objectives, which are referred to as the second and third-order effects, such as on higher employment, or economic growth, are hard to attribute to a specific instrument. The Expert Working Group reports any opportunities it observes for identifying instruments' second and third-order effects within companies and sectors, such as innovative products and productivity.

The working group refrains from commenting on the operation and economic legitimacy of policy instruments. The working group makes recommendations about evaluation designs, but performs no evaluation itself. The ministry will select the bureau that is ultimately to perform the evaluation.⁵ The working group may be requested to support and advise in the formation of monitoring committees for an evaluation. Once the evaluation has been completed, the Expert Working Group will express an opinion on the quality of the evaluation.

Method

For each policy instrument the Impact Evaluation Expert Working Group interviewed the file coordinators involved on the policy side, and the instrument implementers from NL Agency, the Dutch Municipalities Quality Institute (KING)⁶ and the National Platform Science & Technology (*Platform Bèta Techniek*). The Ministry of Economic Affairs provided the working group with factsheets and background information for each instrument, including earlier evaluations, as input for the interviews. The instruments were discussed at length in the interviews in order to build up a clear picture of the envisaged operation and execution of the instruments. On this basis the working group discussed and developed ideas for evaluation designs.

Guide to the reader

Government intervention must have legitimacy in an economic sense. Chapter 1 lists the market failures that give industrial policy its legitimacy. An evaluation of policy may use a variety of methods, a list of which is given in Chapter 2. The WBSO is a good example of an instrument that has been evaluated using quantitative methods. The working group reflects in Chapter 3 on the most recent evaluation, pointing out where room for improvement remains. Chapter 4 is the core of the report, which describes the six policy instruments identified above, lists the evaluation design candidates, and recommends the one that is most promising for ascertaining the instrument's direct impact. This chapter is more technical in nature. The technical details of the evaluation designs are intended mainly for readers who are conversant with the field. Chapter 5, finally, presents general conclusions and recommendations.

⁵ In the case of the BvGD the Ministry of Economic Affairs decides in consultation with KING. This is a shared ownership situation. In the case of CIV this is the Ministry of Education, Culture and Science in consultation with the National Platform Science & Technology, with the exception of the CIV for green education. The Ministry of Economic Affairs is responsible for this.

⁶ It should be noted that KING is not only an 'instrument implementer'. KING is co-owner and responsible for the development of the BvGD instrument in the future.

Chapter 1 Market failure and industrial policy

1.1 Introduction

This report focuses on six policy instruments within Dutch industrial policy, which is characterized by its adherence to generic framework conditions. Framework conditions are a broad concept, ranging from prudent tax and monetary policy to infrastructure, competition, access to capital, a well-educated labour force and an excellent research infrastructure. This condition-oriented policy includes generic policy instruments that are discussed in this report, such as the Promotion of Research and Development Act (WBSO), the Research and Development Deduction (RDA) and the Innovation Credit (IK). Besides improving the economic climate, these generic instruments help mitigate market failure and bring private R&D decisions in line with the socially desirable level. This chapter therefore presents market failure as a key factor in the legitimacy of government intervention: see section 1.2. Alongside this generic pillar, industrial policy also rests on the 'top sectors' policy, the main underlying idea of which is that sector-specific differences matter. A sectoral approach may then complement generic policy. Section 1.3 gives some details of the top sectors policy. Finally, section 1.4 discusses the strengths and weaknesses of industrial policy impact measurement.

1.2 Market failures with generic policy

An important element of industrial policy has to do with utilizing the positive externalities of investments in knowledge (which is also referred to as the internalization of externalities). Most of the public funds available for industrial policy are committed to this aspect. Market failure is therefore most conspicuous in innovation issues. The legitimacy of other instruments discussed in this report that are not directly oriented to innovation also stems from the existence of market or government failure. The sections that cover the instruments provide more details of this aspect. Four of the six instruments discussed in this document are oriented to encouraging innovation in the Dutch economy. For the sake of simplicity and readability, innovation is therefore taken as the starting point below.

The importance of innovation

Innovation is about the application of knowledge and ideas in new production methods and organizational forms, and about bringing new products to market. Suppliers in a competitive market economy have an interest in bringing novel, higher quality and less expensive products to market. Doing so creates a, usually temporary, lead over competitors, a greater market share, and higher than normal profits. This incentive to innovate leads not only to higher profits for the innovative entrepreneur, but also to greater consumer prosperity, in the form of better and new products, higher employment in the innovating companies, and, if the new product is placed on the international market, to more export. Innovation often leads to higher productivity, which can then be converted into lower prices, higher incomes, or more leisure time.

Over the course of time successful innovation is crucial to the increasing material and non-material prosperity of the population. The structural growth of per capita income since the start of the industrial revolution and the increase in leisure time and choice between successive generations are largely attributable to what Baumol rightly describes as 'The Free Market Innovation Machine'.⁷ Permanent and successful innovation comes about only if existing companies and new entrants are given appropriate incentives and stimuli. A competitive environment is a primary condition, albeit the relationship between intensity of competition and level of innovation is not necessarily linear.⁸ The question that then arises is whether a competitive environment leads to the optimum level of innovation, or, stated in more general terms, whether society can safely leave innovation to the free market. The answer is negative since the free market fails and generates less than the socially optimum level of innovation because of the several causes explained below. Market failure is a possible reason for the government to conduct innovation policy and to pursue a higher optimum level of innovation.

⁷ William J. Baumol, 2004. *The Free-Market Innovation Machine: Analyzing the Growth Machine of Capitalism*, Princeton University Press, 2004.

⁸ Philip Aghion, Nick Bloom, Richard Blundell, Rachel Griffith and Peter Howitt, 2005, Competition and Innovation: An inverted U relationship, *Quarterly Journal of Economics*, vol. CXX (2), pp 701-728.

Innovation does not always materialize gradually, and it is sometimes accompanied by temporary, but substantial, social costs. A widespread consequence of introducing new and better products onto the market is a loss of market share for those companies whose products are put in the shade (as in Apple's rise and Nokia's decline). Likewise, new innovative products displace other products in the market (e.g. the PC with a word processor versus the typewriter). Innovation, as Schumpeter observed, is a process of creative destruction with winners and losers. Among the winners are the investors and employees in the successful innovating companies, while the losers are the investors and employees in the companies pushed out by competition. Historically, however, the innovation process has led to a net increase in prosperity because the gains of the winners exceed the losses of the losers.

Market failures

In the *Wealth of Nations* (1776) Adam Smith, the founder of economic science, put forward the thesis that even if individuals and companies base their decisions solely on self-interest, they are led by an invisible hand that ensures that the general interest is served. Adam Smith's intuitive notion has been developed by economists through the years, to the point that it can now be demonstrated that a competitive market, under ideal conditions, with rationally acting consumers and producers, will allocate goods, services and production factors efficiently. Efficient allocation is defined as the performance of all transactions that benefit an individual or a company with no attendant disadvantage to another party.⁹ Neither too little nor too much is produced and traded, and in this sense the perfectly functioning market leads to the efficient level of social prosperity. The market fails when too much or too little is produced relative to the efficient level. Too little means that profitable transactions, with greater social benefits than cost, do not materialize; too much means that transactions with greater social cost than benefits (e.g. environmental damage) are performed nonetheless.

Market failure means that the invisible hand is not working, and that a free market economy does not lead to the efficient level of social prosperity. The government may see it as its duty – or, in other words, as a public interest – to set out to achieve the efficient level, in which case it will conduct policy with a view to correcting the market failure. Market failure as a starting point and motivation for policy is a key theme in this report.¹⁰ This, however, is not to say that market failure is the only possible starting point for government policy. Market failure is an efficiency criterion: there are too many or too few transactions compared with the social optimum. Another important criterion for government policy is the distribution of social prosperity. Under ideal conditions free market operation may lead to the ideal level of prosperity, but the resultant distribution of prosperity over households and companies may be considered unfair or unacceptable on political or ethical grounds. In that case redistribution may be an important motivation for policy. This report does not directly address redistribution policy.

Economics literature usually identifies the following forms of market failure: market power, public goods, externalities, imperfect information and coordination failure. Most of these are also relevant to a possibly inefficient level of innovation in companies and are consequently a target for innovation policy.

Market power

The typical example of market power is that of the monopolist. A rationally acting monopolist will seek to maximize monopoly gains by restricting supply, thereby allowing him to sell his product at a higher price. In a monopoly market less is traded than the socially efficient amount.

The monopolies that may arise are a cartel of companies, a legal monopoly in which the government grants a company an exclusive right to trade in certain products (e.g. the former postal service), or a natural monopoly in which the nature of the production technology means that it is always less expensive to have one provider than many (e.g. the rail network). There are many avenues for policy to correct the market failure caused by monopolies. Opposition to cartels is the responsibility of the competition authority. The regulation and deregulation of legal and natural monopolies fall under

⁹ This form of efficient allocation is referred to as Pareto efficient.

¹⁰ The 1969 presentation given by the later Nobel laureate in Economics, Kenneth Arrow, to the United States Congress Joint Economic Committee is one of the first clear accounts of this policy approach. See K. Arrow. (1969). 'The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-market Allocation' in *The Analysis and Evolution of Public Expenditure: The PPB System*. Vol. 1 U.S. Joint Economic Committee, 91st Congress, 1st Session. Washington D.C.: U.S. Government Printing Office, 59-73. In the Netherlands 'De Calculus van het Publieke Belang (The Calculus of Public Interest)', which was produced on behalf of the Ministry of Economic Affairs and the Ministry of Finance, has been influential in this approach. See C.N. Teulings, A.L. Bovenberg and H.P. van Dalen, 2003, *De Calculus van het Publieke Belang (The Calculus of Public Interest)*, Kenniscentrum voor Ordeningsvraagstukken, Ministry of Economic Affairs, The Hague.

dedicated supervisors (e.g. the Independent Post and Telecommunication Authority OPTA, and the Office of Transport Regulation¹¹).

Market power impedes innovation. The collusion of several companies in a sector to limit competition also reduces the incentive to innovate. Market power may allow incumbent companies to repel new, innovative companies, or pose an effective threat of doing so. This power may be asserted by creating overcapacity, or through predatory pricing, in which incumbent parties with 'deep pockets' maintain low (loss-making) prices just long enough to discourage potential entrants. A second way in which market power reduces innovation is through the cannibalization of a company's own product. Suppose that market power is sufficiently large to create a monopoly situation (a single provider or a cartel). If a monopolist were to put a new product on the market it would compete with its own existing product. The monopolist would make a profit on the new product at the expense of the old product, thereby reducing the incentive to innovate. The calculation is less favourable than for a new entrant unencumbered with existing products who stands only to gain from the new product.

Market power is a form of market failure that erodes the prosperity of consumers and end-users, which is why competition policy and supervision policy oppose it. The possible failure of innovation is usually already within the remit of the competition authority and the supervisors. Market failure in the form of market power requires no separate innovation policy.

Public goods

The market produces too few public goods, if any, because of the non-exclusive nature of the benefits to be gained, and the inherent lack of rivalry in their use. The non-exclusivity of benefits arises from the impossibility of preventing anyone from using a good. Non-rivalry means that use by one consumer is not at the expense of another. The classic example of a public good is a river dike. Once the dike has been built everyone living behind it is protected, and the protection of one resident is not at the expense of another. A private good, such as an apple, can be sold to one user exclusively. If that user eats the apple, it is of no value to anyone else. Consumers are able to gain from the benefits of a public good without paying for it (i.e. 'free rider' behaviour). Private producers therefore have insufficient incentive to produce public goods, even though the benefits of the public goods exceed the costs. The quantity of public goods produced by the market is inefficient. The production of public goods (such as dikes, defence, the police and the judicial system) is usually controlled by the government.

Successful innovation leads to new knowledge in the form of a new product or production process. This knowledge has all the characteristics of a public good. It is impossible in principle to exclude others from using this knowledge (non-exclusivity), and use of the knowledge by one company does not mean that the knowledge is worth less for the next company (non-rivalry). Government policy with respect to the generation of new knowledge largely follows one of two routes: fundamental research and applied, commercial research. Most fundamental research is financed entirely through government funds. The finance goes to higher education institutions and research institutes. Applied and commercial research is carried out by companies because sufficient incentive remains, certainly when commercial and applied knowledge is also protected through intellectual property rights. The use of intellectual property rights (copyright, patents and trademark rights) transforms the public good nature of the new knowledge and innovation into a private good. However, the process is imperfect, and externalities of knowledge production remain.

Externalities

An externality is a positive or negative impact of a transaction between two market parties on the consumption or production opportunities of a third party, who is neither compensated for any loss nor required to pay for any benefit.¹²

¹¹ These supervisors will be absorbed on 1 January 2013 into the Consumer and Market Authority (ACM). This new supervisor is the result of the merger of the Consumer Authority, the Independent Post and Telecommunication Authority (OPTA) and the Dutch Competition Authority (NMa).

¹² Individual decisions almost always have an effect on others. Not all externalities of this kind constitute market failure. For example, an academic graduate entering the labour market increases the graduate labour supply and will (imperceptibly) lower the pay of other graduates. A supermarket that entices customers away from a competing supermarket is another example. These kinds of market effects are sometimes referred to as 'financial externalities'. They are not considered to be market failures because they do not create a problem in terms of efficient market operation. Their effect is more one of a redistribution of income and profits (one party gains at the expense of the other, who loses). Externalities that are considered to be market failures are sometimes referred to as 'technological externalities' to stress the negative or positive impact on the consumption and production opportunities of a third party (with no compensation).

Environmental pollution is the classic example of a negative externality. For example, the fuel transaction between a car owner and an oil company leads to environmental pollution for which the rest of society suffers without compensation. The sum of the socially relevant costs and benefits of externalities is insufficiently reflected in the price of a product. The market price of a negative externality is too low, and more is produced than is socially efficient. The private costs of a positive externality are too high because the benefits to third parties are ignored and too little is produced.

Innovation leads to positive externalities: a company's innovative efforts benefit other companies and the rest of society, without the company concerned being rewarded for these externalities. The innovating company does not factor these positive externalities into its decisions about the size of the innovative investments. From the perspective of society the company will therefore innovate too little.

Positive externalities are also referred to as spillover or knock-on effects. Government innovation policy in this connection is oriented to encouraging innovating companies through subsidies and other means to take account of these spillover effects on third parties, and therefore to innovate more than they would for private reasons alone.

Two kinds of spillover are usually recognized in this connection.¹³ The first is knowledge spillover. The knowledge development that accompanies innovation has the characteristics of a public good. As soon as a provider has put a new or better product on the market, a competitor is able to copy it immediately and put a similar product on the market. A company will then be less inclined to innovate. Putting new products on the market or introducing new production methods is always a matter of investment and experiment. If a competitor is then able to copy the ideas almost for free, the incentive to innovate will be severely curtailed.

Intellectual property rights limit the opportunities for copying and imitation, but do not entirely eliminate them. Because an innovator is never able to pluck all the fruits of his innovative effort, there is less innovation than the socially optimum level. A rather more subtle argument of knowledge spillover is that new knowledge usually builds on existing knowledge ('standing on the shoulders of giants'). For instance, the first generation of a new product or new production process may inspire competitors to produce a variant and a second generation of new products, where the profit opportunities do not benefit the initial innovator.

A second form of spillover is rent spillover. A new product is said to generate consumer surplus if it gives end-users greater prosperity than the market price that they are required to pay. Price discrimination (e.g. higher prices for 'early adopters') may allow an innovative company to skim off some of the consumer surplus. However, the innovative company is unable to collect all of the proceeds of its innovation and will therefore innovate less than the social optimum. A similar argument can be made for a new machine, for example, which lowers the production costs for the companies that purchase it, thereby increasing the producer surplus for the companies that use the machine. This is additional surplus that the innovative company can only partly skim off in the most favourable case.

Imperfect information

If one market party has more information than the other, opportunistic and suboptimum behaviour will be the likely consequence. The problem may be inadequate information about prices, quality, costs, or risks. For example, if raising quality leads to higher costs for providers, but buyers are unable to properly assess the quality, then providers have no incentive to improve quality beyond the mediocre, and consumers will see no reason to pay for it. A negative spiral then arises on the market, with only products of inferior quality being offered, and a persistently unfavourable price-quality ratio. The literature refers to this form of market failure as adverse selection.

On some markets it is not the customers but the providers who suffer from lack of information. A case in point is the possible problems if a private insurer were to introduce an unemployment insurance policy. The probability of an unemployed person finding a job depends not only on the economic cycle but also on the behaviour of the person concerned. The effect of an unemployment policy may be to reduce the urgency of the policyholder's job seeking, thus increasing the length of unemployment. The insurance premium will rise accordingly, and a private insurer might price

¹³The CPB Netherlands Bureau for Economic Policy Analysis publication 'Innovatiebeleid in Nederland: De (on)mogelijkheden van effectmeting (Innovation policy in the Netherlands: The strengths and weaknesses of impact measurement)' from 2011 identifies 'business stealing effect' as a third form of spillover. This effect is related to Schumpeter's creative destruction process. However, the creative destruction of business activity, with the accompanying bankruptcy of failing companies and growth of successful companies, is a normal consequence of market operation.

itself out of the market. This problem is known as moral hazard, which renders some risks uninsurable by private insurance companies, justifying government intervention.

Providers in the market often tackle the problem of 'adverse selection' by backing up higher quality with guarantees (e.g. a full refund if not completely satisfied) or by gradually building up a reputation, possibly linked to a brand. The problem of 'moral hazard' is tackled in the market by never providing full insurance but always leaving the insured party with some of the risk.

Market failure caused by inadequate or asymmetric information mostly exists for financing innovation projects on the capital market.¹⁴ This failure of the capital markets is caused by a lack of information about the innovative project's probability of commercial success. The applicant for a loan for an innovation project has more reliable information about the project's probability of success than the finance provider. This information asymmetry will make the loan provider reluctant to accept innovation projects. A market with asymmetric information, such as the capital market for innovative projects, can lead finance providers to demand relatively high interest rates and set other strict loan conditions, so that most loan applications will be for the poor risks (innovative projects with a low probability of success) (adverse selection). The external nature of finance may also lead to a relaxation of control over the innovation project that would not happen with entirely internal financing, thereby increasing the risk (moral hazard). Market failure because of lack of information results in a smaller capital market for innovation projects, leading to fewer opportunities for an entrepreneur with an innovative idea to attract sufficient funds from the market.

Some of the problem of asymmetric information is solved by the market itself. A loan applicant's good reputation can raise the lender's confidence. Another way of mitigating information asymmetry is for the borrower to put up collateral for the loan. However, the problem remains for start-up entrepreneurs that have yet to build up a reputation, or cannot put up sufficient collateral or cofinancing to reduce the capital provider's uncertainty.

Coordination failure

A market economy works perfectly and leads to an efficient level of prosperity only if it has a complete set of markets. In other words, there have to be $N \times T \times S$ markets, allowing all N goods and services to be traded, with the availability of futures contracts for all future periods T and insurance contracts covering all possible risks S . In practice no market economy anywhere has a complete set of markets. The number of futures markets is limited, and similarly it is not possible to insure every conceivable risk. A lack of markets is a form of market failure that renders transactions impossible that could be interesting for individuals and companies. Markets fail to materialize for several reasons. The first is the transaction costs attached to the operation of the market, such as for specifying and devising cover for all possible risks and contingencies when drawing up futures and insurance contracts. The costs of collecting and documenting all this information can be so large as to stifle all interest in contracts of this kind.

A possible consequence of transaction costs where innovation is concerned is to hamper or prevent the development of certain finance markets. For example, there might be a market for financing extremely high-risk innovation projects (through specialized venture capital markets) and for low-risk projects (easily financed with a bank loan), but no finance market for intermediate risks.

A second reason for a lack of markets is the search and information costs involved in identifying a suitable counterparty. Before they can conclude a mutually advantageous market transaction, a buyer and a seller first have to find each other. In most cases (e.g. daily grocery shopping) this poses no problem, but it is not always immediately clear where a suitable counterparty for a transaction might be found. Parties then have to search for each other. The labour market and the housing market are two examples where this aspect is relevant. Searching for a suitable counterparty takes time and money. An employer with a vacancy must place an advertisement and select and evaluate candidates. Someone putting their house on the market has to advertise the fact – possibly with the help of an estate agent – and open their home for inspection. If no suitable counterparty can be found quickly, the search costs will rise, which may be a reason to abandon the search, or to conclude the transaction with whatever non-optimum counterparty happened to be available at the time. In either case the market will have failed to achieve an efficient result.

¹⁴Note that the sometimes extremely large risks associated with investment in uncertain innovation projects are not deemed to be a form of market failure. The market translates large uncertainty into high risk premiums for financing. Successful innovation projects may also yield extremely high profits.

Parties in the innovation market, such as research institutes and university researchers, possess knowledge of possible interest to companies for development and commercial exploitation. Before they can make agreements, the two parties first have to find each other. Researchers do not know which companies might be interested in their knowledge, and companies do not know which researchers possess knowledge of possible commercial interest to them. The search costs they have to incur in finding each other can be large enough to prevent the contact ever occurring, so that beneficial innovations for society do not happen.

Market failure as a starting point

Some pitfalls have to be avoided when taking market failure as the central basic principle for innovation policy.

It can first be argued that market failure tends towards undersupply of innovation. As we have seen, spillover effects, capital market failure and coordination failure all lead to undersupply and therefore to innovation below the socially efficient level. At the same time, however, it is impossible to measure the extent of market failure, and therefore the magnitude of undersupply. The optimum scale of government intervention is likewise uncertain. The scale of innovation policy can be said to be optimum from the perspective of spillover effects if it results in all innovating companies to take the benefits of others into account alongside their own in their innovation investments. It is almost impossible to evaluate whether the scale of innovation policy is optimum: neither higher nor lower than needed to compensate precisely for spillover effects. At the very least, effective innovation policy for correcting market failure must be oriented to raising the level of innovation in the economy.¹⁵

A second point is that the correction of market failure is not the only possible starting point for innovation policy. Another possible motive for innovation policy is redistribution, with some companies receiving more innovation subsidies than others. Redistribution is a political choice, with possibly inefficient results.

Thirdly, innovation policy also generates costs (e.g. administrative costs and tax collection costs for financing innovation subsidies). It is not out of the question for these policy costs to exceed the benefits created in the form of a higher level of innovation, in which case the innovation policy will be inefficient.

Fourthly, in addition to market failure there is also government failure. The starting point in the theory of market failure is that the government pursues the general interest. Market failure means that the private market is incapable of satisfying the general interest, prompting the government to pursue corrective policy measures. However, the government does not necessarily always have the general interest in mind, and some government decisions may even cause a lower than optimum level of innovation. For example, certain forms of safety regulation may needlessly impede innovative activities. Alternatively, some policy may support private interests rather than the general interest. Incumbent companies in a given industry may lobby for the introduction of entry permits, compulsory registration and other administrative regulations, which unnecessarily block access to new, possibly innovative, entrants.

1.3 Industrial policy in practice and the top sector approach

Many companies take advantage of the opportunities presented by industrial policy. Between 1998 and 2008 some 20,000 companies were reached with subsidy and loan schemes. The Top 25 in terms of participant numbers include the 'big 9' (Philips, ASML, Shell, DSM, NXP, Unilever, Océ, KPN and AkzoNobel) plus many small companies. The major companies participate mainly in the collaborative schemes, and the small companies in the knowledge transfer schemes.¹⁶ The largest amounts are expended on the generic pillar of industrial policy. For instance, the tax-related innovation instruments¹⁷ turn over approximately € 1.8 billion. However, despite these generic instruments, several obstinate bottlenecks persist, such as unsatisfactory private R&D investment, limited knowledge valorization on the part of publicly financed research institutions and universities, a failure to coordinate economic diplomacy between the various public authorities, an excess of industry-specific regulation and red tape, and a poor match between education and the labour market.

¹⁵ Preliminary report of the Royal Netherlands Economic Association (KVS, 2004), *Innovatie in Nederland: De markt faalt en de overheid draalt* (Innovation in the Netherlands: The market fails and the government dallies), editors Bas Jacobs and Jules Theeuwes.

¹⁶ See NL Agency, *Deelnemersbedrijven aan innovatieregelingen 1998-2008* (Participation of companies in innovation schemes, 1998-2008), November 2010.

¹⁷ WBSO, RDA and Innovatiebox.

Alongside their own knowledge development, companies are relying increasingly on other companies, and perhaps even more on public and semi-public knowledge institutes, for expanding the breadth and depth of their knowledge base. Effective interaction and teamwork within companies, between companies, and between companies and knowledge institutes, is therefore vital.

The above trends demand a sectoral approach that is able to operate alongside the generic policy to eliminate the obstacles that continue to exist. The new industrial policy with its focus on nine top sectors seeks to address the market imperfections discussed above effectively, by linking together industry, knowledge institutes and the government.

However, in addressing market failure the government itself is also confronted with information asymmetry and coordination failure. The government often lacks the necessary industry-specific knowledge and information to tackle market failure effectively, which in turn can lead to government failure. Furthermore the way in which the government is organized (with multiple autonomous layers on different levels) gives rise to coordination problems, which tend to undermine efficient policy. In addition positive externalities and obstacles differ from one sector to another, necessitating a differentiated response to knowledge spillovers.

In order to deal with these problems the new industrial policy is industry-oriented, demand-driven and integrated. The integrated approach in the new industrial policy is concerned with eliminating obstacles to a favourable economic climate in a broad sense (such as finance, industry-specific regulation, innovation and human capital). Another aspect of the integrated approach is that the policy is not organized along the lines of internal government boundaries, but of economic sectors, allowing policy themes to be viewed in their totality.

What this means in practice is the creation of teams for nine top sectors, which produce industry-specific recommendations in consultation with the industries. The various pillars on which these recommendations rest are innovation and knowledge road maps, human capital agendas and internationalization agendas. The recommendations also include proposals for regulation and sustainability. This dialogue with the industries puts entrepreneurs, academics and the government in a position to identify tailored solutions that enjoy greater success in eliminating industry-specific obstacles than predigested government policy. The TKIs and the CIVs are examples of new instruments that fit within this approach and set out to correct coordination failure. However, the sector agendas are broader and comprise specific actions for industry, knowledge institutes and government. The actions are extremely diverse. For example, the logistics sector wishes to develop a synchromodal transport system to increase loading efficiency. This will require the industry itself to invest in ICT and the more frequent and intelligent combination of cargoes, the government to amend legislation, and knowledge institutes to research new contract forms, business models and complex decision-making models.

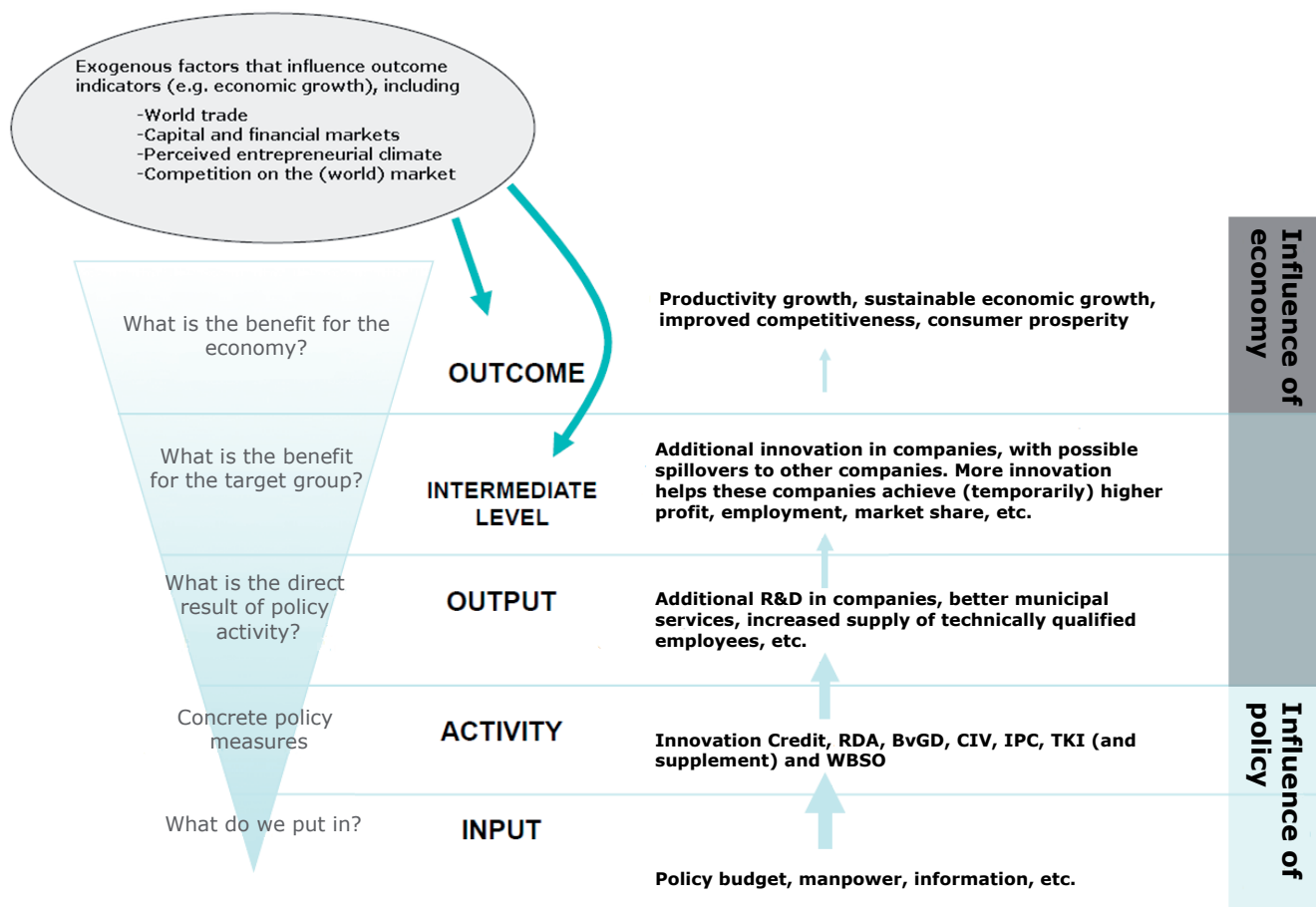
1.4 Strengths and weaknesses of industrial policy impact measurement

The Ministry of Economic Affairs uses the forms of market failure set out above to justify policy. A sound justification in policy theory is not sufficient to guarantee that the policy will actually be effective. As ever, 'the proof of the pudding is in the eating'. The amount of taxpayers' money involved in policy makes it essential to ascertain its effectiveness. The working group was duly requested, as an important first step, to establish the direct impact of policy instruments given the instruments' objectives and the availability of data. It is relevant here to identify the strengths, but especially also the weaknesses, of impact measurement.

The work of the Impact Evaluation Expert Working Group mainly focuses on measuring the output additionality of the various policy measures (see Figure 1). It is certainly more difficult, if not impossible, to measure the higher-order policy effects on, for example, economic growth or competitiveness, in view of current knowledge and methods. With each step up through the various levels (see Figure 1) the influence of the policy instrument becomes weaker. Time also imposes restrictions. It takes a relatively long time for the impact to make itself felt; additional R&D does not translate immediately into more innovation, and the innovation does not translate immediately into faster economic growth or a stronger competitive position. In the ideal situation in which a policy instrument has only one policy objective, an evaluation that finds no immediate impact may be taken to suggest that any contribution of the policy instrument to higher-order effects is unlikely.

The interpretation of impact on output level is often extended to impact on outcome level based on theoretical reasoning. However, there are risks in bridging the theoretical divide between output and outcome. After all, a theoretical correlation does not imply causality. Furthermore correlations are often overestimated, and the law of diminishing returns is frequently overlooked.

Figure 1 - Impact of industrial and other policy on various levels



An alternative is to examine outcome effects in a broader context based on academic studies. For instance, the econometric literature reports numerous studies into the outcome effect of additional R&D. These are summarized in the recent CPB Netherlands Bureau for Economic Policy Analysis study by Lanser and Van der Wiel (2011), which relies significantly on the overview article by Hall et al.¹⁸ This overview article concludes that the social return of R&D, measured as the (marginal) contribution of the R&D effort to GDP, is high but uncertain. The outcomes depend strongly on the selected interval, industries and countries, and the econometric specification used. Moreover, the private return on R&D cannot be considered constant, but is the result of a complex interaction between business strategy, competitors' strategies and a stochastic macroeconomic environment. In this respect too, caution is therefore advised.

Where possible the working group provides a glimpse of higher-level impact measurement, in particular of additional R&D spending in more innovative products (a second-order impact). This impact may also be investigated in an evaluation if sufficient data are available covering a lengthy period. However, as noted, measuring this impact remains difficult. It is easier to investigate second-order effects if a satisfactory control group is available, as in the case of a RCT (see Chapter 2).

¹⁸ B. H. Hall, J. Mairesse and P. Mohnen, 2009, Measuring the returns to R&D, Working Paper 15622, National Bureau of Economic Research, Cambridge, Massachusetts. A more recent NBER working paper also contributes to this discussion and has a list of references: Atkeson & Burstein, 2011, Aggregate implications of innovation policy, NBER WP 17493.

It is hard to evaluate the top sector approach as a whole. The monitoring system developed by the Directorate-General for Entrepreneurship and Innovation is a useful instrument for measuring the progress of the approach and for making comparisons between the various top sectors. It compels parties to justify their target level and enables comparison between top sectors. The objective trees and tables are less relevant for impact measurement, because the objectives are endogenous. They are selected by the top teams themselves, including the target value, which raises the problem of selectivity. Measuring the effectiveness of policy programmes such as the top sector approach is still in its infancy. The Organization for Economic Cooperation and Development (OECD) has set up an Expert Group on Industrial Policy in which countries will exchange their experiences with the evaluation of policy instruments, and policy programmes. This may provide additional reference points for the future.

Chapter 2 Statistical methods for empirical policy evaluation

2.1 Causality and correlation

This chapter briefly discusses the existing econometric methods for estimating the impact of policy interventions. Empirical policy evaluation focuses on how the use of policy instruments changes outcomes. It is an attempt to discover any difference in the target variable¹⁹ with and without the policy intervention, or, in other words, whether the policy intervention has a causal effect on the target variable. This is not easy to determine, because factors other than the policy intervention may also cause the target variable to change. For example, suppose the objective of a government subsidy is to cause firms to increase innovation and that the effectiveness of the intervention needs to be established. The relevant target variable might be R&D spending, the number of innovative products, revenue, profit, or the labour productivity in a company. The dependence of these target variables on many factors, such as the business cycle, innovations abroad and other possible policy changes, makes it difficult to establish a causal effect between the policy intervention and the target variable.

The major problem in empirical policy evaluation is that the application of instruments such as subsidies and loans is often non-random. In other words, the group of firms participating in a given intervention is not comparable with the group of firms that did not participate. A difference in outcomes between the two groups may therefore just as easily be attributable to differences in firms' characteristics as to the policy intervention. The correlation between participation and outcome in the case of selective use is therefore an unreliable estimate of the intervention's causal effect.

We discuss below a number of methods for dealing with this problem, in which we examine Randomised Control Trials (RCT) and natural experiments. In a RCT it is randomly determined whether companies will or will not participate in the intervention. Natural experiments do not have this random allocation with policy evaluation in mind, but rely on existing institutions to give rise to groups of comparable companies, in which some will and others will not have participated in the intervention. This chapter gives only a brief summary of existing methods. There are more extensive overviews in Angrist and Pischke (2009), Blundell and Costa Dias (2007), DiNardo and Lee (2011) and Imbens and Wooldridge (2009).²⁰

2.2 Potential outcomes model

We start with the discussion of a stylistic policy evaluation model. The potential outcomes model assumes that every company²¹ has two possible outcomes. The first outcome Y_1^* is the one that would be observed if the company participated in the intervention, whereas outcome Y_0^* would be observed if the company did not participate. The difference between the two outcomes $\Delta = Y_1^* - Y_0^*$ is the causal effect of the policy intervention. In this case the policy intervention could be the receipt of an Innovation Credit and the relevant outcome would be a company's R&D spending.

The main problem is that the causal effect Δ for a company cannot be observed. A company will either have participated or not participated in the intervention, and therefore only one of the possible outcomes Y_1^* and Y_0^* will be observed. The outcome that remains unobserved is referred to as the counterfactual.

¹⁹ Target variable is another term for dependent variable.

²⁰ Angrist, J.D. & S-S Pischke (2009), *Mostly Harmless Econometrics*, Princeton University Press. Blundell, R. & M. Costa Dias (2007), Alternative Approaches to Evaluation in Empirical Microeconomics, *Journal of Human Resources* 44, 565-640. DiNardo, J. and D.S. Lee (2011), Program evaluation and research designs, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics, Volume 4A*, Amsterdam: Elsevier Science. Imbens, G.W. & J.M. Wooldridge (2009), Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature* 47, 5-86.

²¹ In some cases the policy focus is not on companies, but, for example, on schools (CIV), municipalities (BvGD) or knowledge institutes (TKI). This chapter refers only to companies for reasons of clarity.

We then define a variable D that indicates whether the company participated in the intervention; if so D is set to 1 and otherwise to 0. The standard problem is that data frequently come from a situation in which companies choose for themselves whether they participate in an intervention, which is referred to as the selectivity problem. An example is a situation in which all companies qualify for a subsidy, but not every company chooses to apply. A similar problem arises if a subsidy provider selects only the best applications from a larger number, in which case participating in the intervention is not independent of the possible outcomes. Comparing companies that participate in the intervention with those that do not will yield a biased estimate of the intervention's impact.

2.3 Randomised Control Trials

The most unbiased method of estimating the impact of an intervention is to perform a RCT in which participation is allocated at random. Lots are drawn to determine whether a company participates in the intervention. In practice the randomized subsidy allocation will take place only after companies have submitted an application. The target group therefore consists of companies that have an interest in the subsidy. This gives rise to an experimental group²² and a control group that are mutually comparable, except in respect of participating or not in the intervention. No selectivity problem occurs in this case and the difference between the mean outcomes in the experimental group and control group are a good estimate of the intervention's impact.

There are two underlying assumptions. The first is that participation in the intervention D is independent of the possible outcomes Y_1^* and Y_0^* . This is a crucial assumption, the validity of which is guaranteed by the random allocation to the intervention. The second assumption is that the outcome for a company is independent of how many and which companies participate in the intervention. The validity of this assumption will not be patently clear, even in an experiment. The outcomes for a company may also be affected, for example by the award of an innovation subsidy to a direct competitor. If the companies in the control group also gain from the additional innovation performed by the companies in the experimental group, then the impact of the intervention will be underestimated, or vice versa. In each situation, the interaction between companies in the experimental and control groups needs to be monitored and the possible effects assessed.

RCTs are relatively uncommon with industrial policy compared with, for example, education, labour market or development policy. In these other policy areas, RCTs are ideally pilots to test in advance whether an intervention meets the expectations, which also implies that RCTs are often fairly small scale. The scale must be predetermined through power analysis²³, taking into account the expected impact of the intervention and the normal unexplained variation in outcomes.

In many cases the data are not from a RCT, and it cannot simply be assumed that participation in the intervention is independent of the possible outcomes. An option in these cases is to work through a complete economic model that specifies all the relationships between the intervention and the outcomes, or to perform a regression, controlling for all relevant differences between companies that did and did not participate in the intervention. However, both these approaches involve strong assumptions: the model specification must be correct, including the assumed causes of participation in the intervention. The data must moreover include a very considerable amount of information about the companies.

²²Experimental group is sometimes referred to as treatment group.

²³A power analysis gives an estimate of the minimum sample size needed for the evaluation to produce statistically significant outcomes.

Box 2.1 Randomised Control Trail

Comparing the mean outcomes of the experimental and control groups with data from a RCT gives an unbiased (non-parametric) estimate of the intervention's impact. The difference-in-means estimator is:

$$\Delta = \frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n (1-D_i) Y_i}{\sum_{i=1}^n (1-D_i)}$$

where Y_i is the observed outcome and D_i the independent variable that results from randomized allocation to the experimental group or a control group. The same estimate of the intervention's impact can be obtained through regression:

$$Y_i = \alpha + \Delta D_i + U_i$$

where U_i are error terms and α the constant. Because D_i is independent of the error term U_i , the parameter is a measure of the causal effect of D_i on Y_i . This regression is estimated using the least-squares method. It is usual to include other observed characteristics X_i as control variables in the regression.

$$Y_i = \alpha + \Delta D_i + \beta X_i + U_i$$

The most important reason for including the other observed characteristics is the possibility of reducing standard errors, because of the smaller variance of U_i . Adding other observed characteristics must not influence the magnitude of the intervention's estimated impact. In most cases this will have hardly any influence on the size of the estimated coefficient.

2.4 Natural experiments

If no RCT is feasible, an attempt to estimate the causal effect of an intervention may be made using non-experimental data. The techniques discussed in sequence below are regression discontinuity, difference-in-difference, instrumental variables and propensity score matching.

The methods are based on searching for exogenous variation in participation in the intervention. This exogenous variation leads to only one of two comparable companies participating in the intervention. Comparing the outcomes of these companies allows the causal effect of the intervention to be estimated. The exogenous variation in participating in an intervention is often caused by existing institutions. A number of methods are described in greater detail below.

Regression discontinuity²⁴

The first method to be discussed is regression discontinuity. The underlying idea is the existence of a threshold that determines the eligibility of a company for an intervention. An example is the grading of companies' subsidy applications, where applications that score above 6 out of 10, say, will be accepted, and those with a lower score will be rejected. The idea of regression discontinuity is to compare applications that just qualify with those that just do not. For example, applications could be selected with scores lying in a narrow interval between 5.9 and 6.1. The applications will be from companies that are comparable in all respects other than that those just above 6 received the subsidy and those just below did not. The wider the interval assumed, the less comparable the companies will be, preventing the causal effect of the subsidy from being established.

An assumption in regression discontinuity is that companies are unable to anticipate the scoring close to the cut-off point. Companies are therefore unable to manipulate whether they fall just above or just below the cut-off point. Statistical tests are available to verify this aspect. A simple check is whether companies above and below the cut-off point have the same mean observed characteristics. An exogenously determined cut-off point should yield comparable means of the observed characteristics across the two groups. An experiment that satisfies this assumption is referred to as a local RCT. Therefore,

²⁴For a theoretical discussion of regression discontinuity see Hahn, J., P. Todd & W. van der Klaauw (2001), Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica* 69, 201-209.

the empirical methods available for RCTs are applicable. However, a significant difference is in the interpretation of any causal effect found. Because regression discontinuity is concerned with a local RCT, the causal effect found must also be interpreted locally. The estimated effect is therefore the causal effect for companies on the margin between receiving and not receiving intervention. Regression discontinuity therefore has the great disadvantage of not yielding the total effect of an intervention, and in particular it does not estimate the effect for the companies far above the cut-off point that will certainly participate in the intervention.

A more practical problem is the choice of bandwidth that determines whether companies are included in or excluded from the analysis. The companies in the above example scored between 5.9 and 6.1 for their applications. However, it can happen that few companies fall within this range, in which case the local effect of the intervention will be estimated with a small sample, with a correspondingly greater uncertainty. Widening the interval will ensure the inclusion of more companies in the empirical analysis, yielding a more precise estimate of the intervention's impact. However, this entails a risk of introducing other (unobserved) differences between the group of companies that participated in the intervention and the group that did not. If that happens, the estimated intervention impact will be biased. Use is often made of regression techniques to adjust for some of this bias (see for example Lalive, 2008).²⁵

Difference-in-difference

The second method discussed is difference-in-difference. This is a panel data method that involves tracking companies in time. Difference-in-difference is a useful method if a number of companies receive an intervention within the observation period for outcomes of companies. The idea behind difference-in-difference is that the time trends in outcomes of the companies that do and do not receive the intervention are approximately the same, which is referred to as the common trend assumption. The change in outcomes over time of the companies with the intervention is compared with the change in outcomes of companies without the intervention. If the common trend assumption is valid, then this is a measure of the intervention's causal effect.

As an example, consider a subsidy given to several companies that are tracked for some time interval around the intervention: several periods prior to the subsidy and several afterwards. Since a panel data model is involved, a dummy variable can be introduced for each company in the sample. The target variable is also regressed on a time trend, which therefore registers the trend growth rate over the entire period. The model also includes an indicator variable for the subsidized companies from the time they received the subsidy. This variable registers the causal effect of the subsidy on the target variable.

Box 2.2 Difference-in-difference

The difference-in-difference model is a standard linear panel data model:

$$Y_{it} = \alpha_i + \delta D_{it} + \beta X_{it} + \mu_t + U_{it}$$

where α_i represents company specific impact and the dummy variables μ_t model the time trend. The variables U_{it} are the error terms. The model specification may also include other explanatory variables X_{it} of company i at time t , which may improve the precision of the intervention δ estimated causal effect. Another possibility, instead of a dummy variable for participating or not in the intervention, is to include a continuous variable for the experimental group. In that case instead of the specification including whether a company participated in the intervention, it might include the amount of subsidy received. The interpretation of δ is then how the outcome is changed by one additional euro subsidy, which is sometimes referred to as the bang for the buck.²⁶

²⁵Lalive, R. (2008), How do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach, *Journal of Econometrics* 142, 785-806.

²⁶For applications of difference-in-difference see e.g. Card, D. & A.B. Krueger (1994), Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review* 84, 772-793. Dufló, E. (2001), Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment, *American Economic Review* 91, 795-813.

The advantage of difference-in-difference is that it measures the causal effect of the intervention for the entire population of participating companies. However, the common trend assumption is too strong in some cases. For example, this assumption disregards the eligibility for an intervention of companies that have performed extremely well in the past period. A control group of companies must then be identified with a trend comparable with that of the companies that received an intervention. The validity of the common trend assumption is often argued with reference to the trend in outcomes in the periods prior to the intervention.

Instrumental variables

The following method to be discussed is instrumental variables. In the instrumental variables method one or more variables are defined that influence which companies receive the intervention, but have no direct effect on the outcomes of the companies concerned. Identifying variables of this kind turns out to be extremely difficult in practice. In some cases institutions may influence the choice. There may then be a cut-off point like that in regression discontinuity, except that it does not determine with certainty which companies receive an intervention. Companies just above the cut-off point have a high probability of receiving the intervention, but not all will do so, while a far lower proportion of companies just below the cut-off point receive the intervention. This situation is also referred to as a fuzzy regression discontinuity.

The interpretation of the estimated causal effect depends on how the instrumental variable influences the probability of a company receiving an intervention. If the instrumental variable influences the probability of an intervention only around a cut-off point, as with a fuzzy regression discontinuity, the estimated effect must be interpreted locally. If the instrumental variable value is relevant for all companies, then a more general interpretation of the effect found may be given. It is important for instrumental variables to have a sufficiently large effect on receiving the intervention. If the effect is small or inaccurately measured, then the estimated causal effect may be extremely biased.

Box 2.3 Instrumental variables

An instrument is a variable Z_i that is correlated with participating or not in the intervention, and that has no direct effect on the outcome. The instrumental variables model therefore has two equations:

$$Y_i = \alpha + \delta D_i + \beta X_i + U_i$$

$$D_i = \gamma + \pi Z_i + \rho X_i + V_i$$

Instrumental variables methods may produce a biased estimate of the effect of an intervention for small samples. It is important in this regard to inspect the π , which gives the effect of Z_i on the intervention D_i . The rule of thumb is that the test statistic of an F test for significance of π must be greater than 10, which ensures that the instruments are not weak.

Among the applications of instrumental variables methods are estimating demand-supply models and dynamic panel data models. In estimating, say, a demand function, the price is instrumented with factors that do influence supply, but not demand. The general method of moments (GMM) is frequently used in dynamic panel data models

$$Y_{it} = \alpha_i + \gamma Y_{it-1} + \delta D_{it} + \beta X_{it} + \mu_t + U_{it}$$

to estimate the parameters. The idea is to use delayed endogenous variables Y_{it-2}, Y_{it-3} , etc. in estimating the parameter. A related assumption is the absence of autocorrelation in the error terms U_{it} . The risk with GMM is of using too many delayed variables as instruments, rendering the instruments weak and increasing the bias in the estimators. There are many variants of GMM, including the system estimator.²⁷

²⁷Verbeek, M. (2012), *A Guide to Modern Econometrics*, fourth edition, Wiley.

Propensity score matching

The next method to be discussed is propensity score matching. This is actually a collective term for techniques that assume that all background variables relevant to receiving the intervention and that also influence the outcomes of the company are observable.²⁸ Propensity score matching sets out to identify pairs of companies that are comparable in background variables, but where only one participated in the intervention. Comparing the outcomes of these two identical companies yields an estimate of the causal effect of the intervention. The propensity score is used to determine how similar companies are. The propensity score gives the probability of a business receiving an intervention, given the company's background characteristics. An example of the application of propensity score is a subsidy to a company based on a number of criteria. Each of the companies receives a score for these criteria. First the probability is estimated of a company receiving the subsidy, given the scores on the multiple criteria. This probability is referred to as the propensity score. The various methods are then applied to determine the effect of the propensity score on the target variable. Linking pairs of companies that have a comparable propensity score, where one received the subsidy and the other did not, automatically produces two comparable groups.

Box 2.4 Propensity score matching

The propensity score p_i is the probability given X_i of company i participating in the intervention, so that:

$$p_i = Pr(D_i = 1 | X_i)$$

Suppose N_1 companies participated in the intervention and N_0 companies did not. A general specification of the estimator for the impact of the intervention is:

$$\Delta = \frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_i - \sum_{j=1}^{N_0} w(i, j) Y_j \right)$$

For every N_1 observations in the experimental group a counterfactual is chosen that is produced by weighting the N_0 observations in the control group. The weights $w(i, j)$ express the degree of comparability of observation i in the experimental group with observation j in the control group. These weights are determined with reference to the estimated propensity scores, for which various algorithms are available. The linking of an observation in the experimental group with the observation in the control group having the closest propensity score is referred to as nearest-neighbour matching. Linking it to the mean outcome of all observations in the control group having an approximately comparable propensity score is referred to as neighbourhood matching. The comparison of observations on intervals of the propensity score is referred to as blocking. It is also possible to weight each observation in the experimental and control group with the inverse of the propensity score.

The propensity score can be used in a wide variety of ways. It can be used to identify two identical companies to be linked together, as described above. Another possibility is to link companies that receive an intervention with multiple companies without intervention, and vice versa. Other methods exist that weight each observation in accordance with the propensity score, thereby avoiding any loss of observations.

The advantage of propensity score matching is that it estimates the impact of the intervention for all companies where counterfactuals exist. Companies whose background characteristics mean that they will certainly receive an intervention are disregarded.

A disadvantage of propensity score matching is the implicit assumption of random allocation of an intervention for companies with the same propensity score. This disadvantage may be less significant if much background information is

²⁸For a general description see Becker, S.O. & A. Ichino (2002), Estimation of Average Treatment Effects Based on Propensity Scores, *Stata Journal* 4, 358-353. Caliendo, M. and S. Kopeinig (2008), Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys* 22, 31-72.

available, but the assumption that all relevant background characteristics are observed cannot be verified. The researcher will therefore have to argue the assumption's validity on the basis of institutional knowledge.

2.5 Standard regression analysis

All the methods discussed above attempt to solve the selectivity problem in employing an intervention by seeking out companies that are comparable except in respect of participating or not in the intervention. The actual application of one of the above methods is not always feasible in practice, in which case a standard linear regression model will often be used.

A regression analysis distinguishes between companies that did and did not participate in the intervention. The outcomes of the two groups can be compared by means of a regression function, while controlling for as many observable factors as possible, including an indicator variable for the company.

Box 2.5 Regression analysis

The standard linear regression model has the following form:

$$Y_i = \alpha + \delta D_i + \beta X_i + U_i$$

The coefficient δ gives the correlation between the intervention D_i and the outcome Y_i (conditional on the other observed characteristics X_i). Often a (conditional) correlation will not indicate a causal effect. For a correlation to be interpreted as actually causal, participation in the intervention D_i must not be correlated with unobserved characteristics U_i . In a non-experimental context this will often mean including many observed characteristics X_i in the regression analysis. However, the inclusion of many observed characteristics entails the risk of also including what are known as confounding variables, which are influenced by the intervention and also have an impact on the outcome. Suppose for example that an intervention is intended to increase a company's labour productivity. If we wish to estimate the intervention's impact on profit using labour productivity as an explanatory variable, then the regression will underestimate the intervention's causal effect on profit.

It is necessary to be alert to the following problems. First, companies may self-select. The data will frequently come from a situation in which companies themselves determine their participation in an intervention. An example would be where all companies qualify for a subsidy, but not all opt to apply. A second problem occurs when a subsidy provider selects the best from a number of applications, in which case participating in the intervention is not independent of the possible outcomes. In many cases comparing companies that do and do not participate in the intervention will overestimate the intervention's impact. In an overestimate, the measured value of the impact is greater than the actual causal effect.

An imperfect solution for selectivity is to develop a complete economic model specifying all the relationships between the intervention and the outcomes, or to perform a regression while controlling for all relevant differences between companies that did and did not participate in the intervention. However, both these approaches involve strong assumptions. The model specification must be correct and the data must contain a great deal of information.

The above discussion of regression analysis implicitly assumes a cross-section of companies. However, data are frequently available for several intervals, in which case panel data will be available. Panel data can be adjusted for certain forms of unobserved company characteristics. The company characteristics concerned do not vary in time (see Box 2.3).

Besides the standard fixed-effect panel data methods, which can also be used to estimate the difference-in-difference model, more dynamic models may also be analysed. A model is dynamic if outcomes from earlier periods influence the current outcome of the company (i.e. delayed endogenous variables). A dynamic panel data model can also be used to model aggregated outcomes for a given sector over time.

The presence of delayed endogenous variables makes estimating dynamic panel data models more difficult than estimating a standard panel data model. The problem is that the delayed endogenous variables on the right-hand side of the equation may be correlated with the error terms. Instrumental variables are often used in solving this problem, with the underlying assumption that error terms are not correlated over time. If they are, further delayed endogenous regressors can be used as instrumental variables. The number of possible instrumental variables will then be extremely large, and therefore GMM will often be used for estimating. GMM must be considered to be a generalization of instrumental variable methods that allows the use of many instruments. As stated in Box 2.3 there is a large risk of including too many (weak) instruments, leading to biased estimators. There are various GMM variants, including first-difference GMM²⁹ and system GMM³⁰.

2.6 Survey methods

Surveys of companies were frequently used in the past for evaluating industrial policy. These surveys include a group of companies that participated in an intervention and a group that did not. Companies are asked directly about the impact of the intervention, as well as about several characteristics and outcomes of the company. These surveys tend to be extremely extensive, and have the potential to provide a complete picture of an intervention's impact.

However, survey methods also involve several assumptions. The first is that companies answer truthfully in surveys, and consequently there are no systematic errors in reporting. Examples of the risk in this assumption are socially desirable or strategic answering. An additional problem is the often modest response rate in company surveys. The second assumption therefore is that response does not differ selectively between companies that did and did not participate in the intervention. This assumption is relevant when differences in the response rates cause certain groups to be underrepresented or overrepresented in the survey. A clear difference between an underrepresented group's behaviour and that of an overrepresented group will bias the outcomes. A third assumption is the comparability of surveyed companies that did and did not participate in the intervention. Differences in means are often calculated when analysing data from a survey, in which it is unusual to adjust for differences in background characteristics between companies. It is then all the more important to have comparable groups of companies that did and did not participate in the intervention, which can be achieved by ensuring that the choice of companies to be surveyed includes comparable companies in the two groups. Finally, a survey method also assumes no direct spillovers of the intervention between companies, which is the same assumption that was explicitly discussed under RCTs and is also necessary in all the other methods discussed above. Despite these assumptions, surveys are often able to yield important additional information about exactly how an instrument functions and the users' experiences, making this an important supplementary method alongside the statistical methods outlined above.

2.7 Conclusions

Various empirical policy evaluation methods are discussed above. The major problem in the evaluation is the frequently selective use of interventions, which hampers empirical analysis. In these cases a linear regression will often yield a biased estimate of the intervention's causal effect.

A RCT with random allocation yields the least biased estimate of an intervention's causal effect. The estimated effect in this case is that of the intervention as a whole. As a rule the design of a RCT is limited in size, which may affect the precision of estimating an intervention's impact. However, randomized allocation of interventions is rarely used in innovation policy evaluation. There are various alternative methods if no RCT is available, which differ in the effect estimated, and in the consistency and precision of the estimated impact.

Regression discontinuity bears some resemblance to a RCT, but estimates a causal effect only locally. Difference-in-difference gives a more general picture and uses panel data, and therefore also takes into account some unobserved company characteristics, subject to satisfying the common trend assumption. Instrumental variables are often hard to

²⁹ Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies* 58, 277-297.

³⁰ Blundell, R. and S. Bond (1998), Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.

identify. The propensity score matching method is nearly always possible, and often also gives a complete picture of an intervention. The main problem with propensity score matching is the extreme strength of the underlying assumption and the need for accurate data. Surveys also rely on assumptions, such as about selectivity. An additional risk with surveys is that the data are often based on the interviewees' perceptions rather than on objective, administrative records. Nonetheless surveys can reveal much information about the functioning and execution of an instrument that complements the output of the quantitative methods discussed above.

Chapter 3 Reflections on WBSO Evaluation: A case study

3.1 Introduction

The Directorate-General for Entrepreneurship and Innovation of the Ministry of Economic Affairs has substantial experience with policy evaluations. These evaluations make considerable use of surveys and other qualitative investigation methods. Moreover, the Directorate-General for Entrepreneurship and Innovation has already acquired relevant evaluation experience using more quantitative methods, notably in the evaluations of the Promotion of Research and Development Act (WBSO), which was enacted in 1994. The WBSO was evaluated in 2001 and 2002 by PricewaterhouseCoopers, Dialogic and TU Delft, and this was the first application of linear regression to quantify impact. An attempt to do so in an earlier evaluation in 1998 failed because of data limitations. In 2005 CPB Netherlands Bureau for Economic Policy Analysis examined amendments to the design of the WBSO scheme and opted for regression discontinuity. In 2006-2007 EIM and UNU-MERIT jointly evaluated the WBSO for the 2001-2005 period, using a fixed effects panel data model. The most recent WBSO evaluation for the 2006-2010 period was executed by EIM based on a dynamic panel data model that was estimated using GMM.³¹ The Impact Evaluation Expert Working Group has been requested to assess the most recent WBSO evaluation and identify possible improvements in the analysis of effectiveness.

3.2 The WBSO

Objective

The objective of the Promotion of Research and Development Act (WBSO) is to provide an incentive to Dutch companies and entrepreneurs that carry out research and development. The incentive takes the form of a tax concession to lower the salary costs of researchers.

Market failure

The motivation for introducing the WBSO in the 1990s did not refer to market failure, it then being relatively uncommon to focus on this phenomenon. Reference was made instead to the great uncertainty that accompanies investment in research and development, and the substantial investment in equipment and laboratories required, leading to lower private R&D spending. Wage costs in the Netherlands were also regarded as relatively high, which was seen as having a possible detrimental effect on the economic climate for (R&D intensive) companies in the Netherlands. A secondary objective of the WBSO was therefore to create a tax climate that would encourage foreign companies to base their R&D activities in this country.

Positive externalities may occur in the development of knowledge and technology; the social benefits are greater than the private benefits. From a social perspective, the existence of positive externalities leads to underinvestment in R&D. A tax concession for the salary costs of R&D employees (a wage cost subsidy) is then a possible incentive that lowers the cost of innovative investment (see also the discussion in Chapter 4 of the Research and Development Allowance (section 4.3)).

Operation

The WBSO is the collective name for two R&D-promoting tax facilities:

- the research and development payroll tax deduction:
 - for withholding agents with employees who perform R&D (companies with employees);
 - for withholding agents that do not run a business, if they perform R&D for the expense of a company or consortium, or a product or industry board (knowledge institutes with employees);
- the research and development income tax allowance:
 - for self-employed people who devote more than a certain number of hours to R&D in their business.

³¹ EIM, 2012, Evaluatie WBSO 2006-2010 (2006-2010 WBSO Evaluation), Zoetermeer.

The research and development payroll tax deduction for withholding agents has two bands. In 2010 there was a 50% rebate for the first €220,000 of actual wage costs for research personnel and an 18% rebate for the remainder. For self-employed taxpayers there was a fixed income tax allowance of €12,031 on taxable income, provided at least 500 hours were devoted to R&D annually. The rates in the first and second bands were reduced in 2012 to 42% and 14% respectively. The ceiling of the first band was lowered to €110,000.

The proposal for 2013 is to raise the upper limit of the first band again from €110,000 to €200,000, which would be financed by lowering the rate of the first band from 42% to 38%, and the rate of the first band for new start-ups from 60% to 50%

Starting entrepreneurs are given an additional incentive to perform R&D. A 64% rate was applicable to the amount in the first band for new start-up withholding agents. This rate has now been lowered to 60%. There was a supplementary allowance for new start-up self-employed taxpayers of €6,017 on taxable income.

To prevent major companies appropriating a disproportionate amount of the budget, the tax facility is subject to a maximum R&D salary cost limit. The ceiling in the WBSO is €14 million per calendar year.

Execution

Entrepreneurs and knowledge institutes may submit a digital application to NL Innovation up to three times a year, at least one month prior to the period in which the work will be performed. The application must be for a period between three and six months. An annual application is allowed in certain cases. There is no annual maximum number of applications for self-employed people, but they must apply before the end of September.

The process an applicant follows can be divided into eight steps:

1. identification by applicant of development or research requirement;
2. submission of application, and, if appropriate, provision of Citizen Service Numbers;
3. setup of R&D accounting system;
4. NL Innovation checks completeness of application;
5. NL Innovation performs substantive assessment of application;
6. applicant settles financial benefit in tax return;
7. if necessary notification of R&D hours to NL Innovation;
8. NL Innovation may visit or audit applicant.

Budget

The WBSO budget for 2012 is €872 million. Since 1994 the WBSO budget has been determined annually in the Tax Plan, whereby actual expenditure may deviate from the budget.

Evaluation

The WBSO was evaluated for the 2006-2010 period in 2011 and 2012. The next evaluation is due in 2016.

3.3 The evaluation of the WBSO for 2006-2010

Hypotheses

Core hypothesis³²

- 1) The WBSO has a positive impact on a company's total R&D wage expenditure.

Hypothesis in support of the core hypothesis

- 2) Companies that receive a relatively large subsidy spend relatively more on R&D salaries than companies that receive a lower rate of subsidy.

Hypotheses for assessing spillover effects

- 3) The WBSO helps increase innovation on company level.
- 4) The WBSO leads to spillovers within the same sector.
- 5) The WBSO leads to spillovers outside the sector.
- 6) The WBSO helps increase added value per worker.

The discussion below focuses mainly on the first hypothesis and the supporting core hypothesis, and therefore only examines the basic model used in the 2006-2010 WBSO evaluation.

Econometric specification and identification

To measure the impact of the WBSO, EIM examined the relationship between (the logarithm of) the total salaries of R&D personnel employed by company i (referred to below as the R&D salary bill) and the deduction granted to the company under the WBSO divided by the total salaries (the WBSO deduction ratio). This method is the standard regression analysis that was described in Section 2.5. A delayed endogenous variable is included in the regression function in view of the strong correlation of the R&D salary bill over time. The model therefore measures the relationship between changes in the total salaries and the rate of subsidy. EIM expected a positive relationship between (the growth of) the total salaries and the rate of subsidy; if the WBSO was effective, companies receiving a relatively large subsidy would spend more on research than those with a lower rate of subsidy (see EIM 2012, p. 12 of the appendix). This model therefore tests the above hypothesis in support of the core hypothesis.

In estimating the model, EIM also took into account a number of control variables and (unobserved) differences between companies that were constant over time. This was to facilitate a causal interpretation of the estimated impact. EIM used the following specification:

$$\ln Y_{it} = \alpha_i + \gamma \ln Y_{it-1} + \delta \frac{WBSO_{it-1}}{Y_{it-1}} + \beta X_i + \mu_t + U_{it}$$

Where Y is a company's R&D salary bill in a year, α_i is a fixed effect for each company, $WBSO$ is the reduction in tax attributable to the WBSO, X is a series of control variables and μ_t is a fixed effect for the years. U is the error term.

This specification differs from that used in the above WBSO evaluation for estimating the WBSO impact on the R&D salary bill. The way the WBSO as explanatory factor was expressed in that evaluation was not as a ratio based on the R&D salary bill, but as (the logarithm of) an absolute amount (EIM and UNU-MERIT 2007).³³ Moreover, no delayed endogenous variable was then included in order to preclude possible result bias in combination with fixed effects. The reason for this was the correlation between the delayed endogenous variable and error terms, which complicates estimating (see the discussion in the following section).

³²The EIM evaluation does not work with hypotheses but with survey questions. The research questions have been translated into hypotheses in order to maintain consistency with the rest of this report.

³³A different approach was taken in the previous WBSO evaluation, in which the WBSO's impact was derived from the correlation between total researchers' salaries and the user cost of R&D.

The model suggests a relatively simple association between the rate of subsidy and total salaries, whereby the hypothesized positive impact of the WBSO on R&D wage expenditure should be accepted as significantly greater than zero. There are three possible pitfalls. First the magnitude of total salaries determines the rate of subsidy: the higher the total salaries, the lower the marginal WBSO rate. The delayed endogenous variable $\ln Y_{it-1}$ adjusts for the level of total salaries, but imposes a linear functional form that is possibly too restrictive.

The second pitfall is that companies for which the WBSO is an effective stimulus will be more likely to end up in a higher WBSO band because of the growth in their research activities. The intention is for this selection effect to be adjusted through the fixed effects, but this relies on the assumption that they are constant in time.

The third pitfall is the dynamic nature of the specification. EIM formulated the hypothesis with respect to levels. However, the estimated coefficient is more comparable with the relationship between the *change* of total salaries and the rate of subsidy. This may cause problems if a company's growth potential declines as company size increases. The company size that corresponds with maximum mean growth is an empirical question. There are signs in the EIM report of more instances of fast-growing total salaries among companies in the higher WBSO bands (see Appendix III of EIM 2012, p. 57). The proportion of total salary growth attributable to the WBSO cannot be established in this way.

It is hard to solve these problems for a generic measure such as the WBSO with non-experimental data. As explained above, in the absence of a satisfactory control group additional assumptions must be made about the model specification.

EIM's basic specification is not easy to estimate because the fixed effect also influences the delayed endogenous variable. EIM used several dynamic panel data methods to adjust for this problem. The methods use instrumental variables, usually delayed dependent variables and their first differences. In principle, instrumental variables can also be used in an attempt to eliminate selection effects and other forms of endogeneity, but success depends strongly on the instrumental variables that are available. The EIM report does not discuss the problems referred to above. In addition even more assumptions are needed in estimating dynamic panel data models. EIM's use of these econometric techniques in estimating the basic specification is discussed below.

Estimation methods

The dynamic panel data model has a delayed endogenous variable on the right-hand side, implying that the standard fixed effects estimation methods cannot be used. It is usual in cases of this kind to take first differences, i.e.:

$$\ln Y_{it} - \ln Y_{it-1} = \gamma (\ln Y_{it-1} - \ln Y_{it-2}) + \delta \left(\frac{WBSO_{it-1}}{Y_{it-1}} - \frac{WBSO_{it-2}}{Y_{it-2}} \right) + \beta (X_{it} - X_{it-1}) + \mu_t - \mu_{t-1} + (U_{it} - U_{it-1})$$

The great problem in estimating models of this kind is correlation between the regressor $(\ln Y_{it-1} - \ln Y_{it-2})$ and the error term $(U_{it} - U_{it-1})$. The same could be true of the regressor $(\frac{WBSO_{it-1}}{Y_{it-1}} - \frac{WBSO_{it-2}}{Y_{it-2}})$, and should be handled in the same way. The usual approach is to identify instruments, which are variables that are correlated with the endogenous regressors, but uncorrelated with the error term. The instruments for dynamic panel data models are often sought in further delayed endogenous variables. Possible instruments therefore are $\ln Y_{it-2}$, $\ln Y_{it-3}$, etc. Where only one delayed variable is used as an instrument, in this case $\ln Y_{it-2}$, it is referred to as the Anderson-Hsiao estimator. GMM is an efficient parameter estimation method where there are more instrumental variables than regressors. This method is referred to in the WBSO evaluation as GMM-dif.

This GMM approach involves two underlying assumptions. The first is that the instruments $\ln Y_{it-2}$, $\ln Y_{it-3}$, etc. are exogenous, which means that they are not correlated with the error term $(U_{it} - U_{it-1})$. This assumption is valid if there is no autocorrelation in the error terms U_{it} , as can be verified after estimating the model. The second assumption is that the instruments are relevant, which means that they are correlated with the endogenous regressor $(\ln Y_{it-1} - \ln Y_{it-2})$. This correlation must be large enough to avoid excessive estimator bias because of what is known as the weak instruments problem. Instruments may be deemed weak for a variety of reasons. The first is if γ is close to 1, and the second is that the fixed effects α_i have a relatively large influence on the outcomes compared with the error terms U_{it} . The problem of weak instruments is exacerbated by the inclusion of a large number of instrumental variables. There are a variety of tests for the sufficient strength of instruments, such as F tests for instruments in the first-stage regression, and the Gragg-Donald test.

If γ equals 1, the dynamic panel data model is not stationary and estimation of the model with the above procedure is hampered by the instrumental variables becoming irrelevant. In this case the system GMM estimation method will often be used. This method does not use the model in first differences, but uses the model in levels, and it is the instruments that are specified in first differences. This method is referred to in the WBSO evaluation as GMM-sys.

The WBSO evaluation used the maximum number of instruments for both GMM-dif and GMM-sys. The inclusion of many instruments increases the risk of biased estimates because of the weak instruments problem. However it is impossible to tell whether this problem occurs in this case because the usual tests for weak instruments are not reported. Likewise, no tests of the autocorrelation in the error terms are included in the report.

The estimates produced by the Anderson-Hsiao method are close to those of GMM-dif, which is encouraging and implies a certain degree of robustness in the results. However, the relatively large differences between the GMM-sys estimating results and those of the other two estimation methods may point to possible problems with the GMM-sys estimator, but no tests were reported. The right choice was probably made in the report in the interpretation of the estimating results by assuming those obtained with GMM-dif and Anderson-Hsiao. Nonetheless without the above-mentioned specification tests the validity of the results remains hard to assess.

The next section gives an interpretation in economics terms of the results, based on an assumption of a correctly estimated WBSO impact on R&D wage expenditure, despite the remaining uncertainty.

Economic interpretation of the measured impact

The most significant variable in the EIM report in clarifying the WBSO's contribution to research and development is the 'bang for the buck' (BFTB) criterion. Bang for the buck measures the additional R&D salary expenditure³⁴ caused by a one euro tax reduction.³⁵ The EIM report analyses the first-order impact of the WBSO in this way. Since the BFTB was also calculated in other evaluations, different estimation methods can be compared readily with BFTB.

The mean BFTB for the 2006-2010 period was estimated at between 1.55 and 1.99 euros, with the most probable estimate being 1.77. For the 2001-2005 period the mean BFTB was estimated at 1.99. The difference may be attributable to both the improved econometric approach and a greatly improved database, but also to the 'law of diminishing returns', in that the marginal impact decreases as the regulation expands.

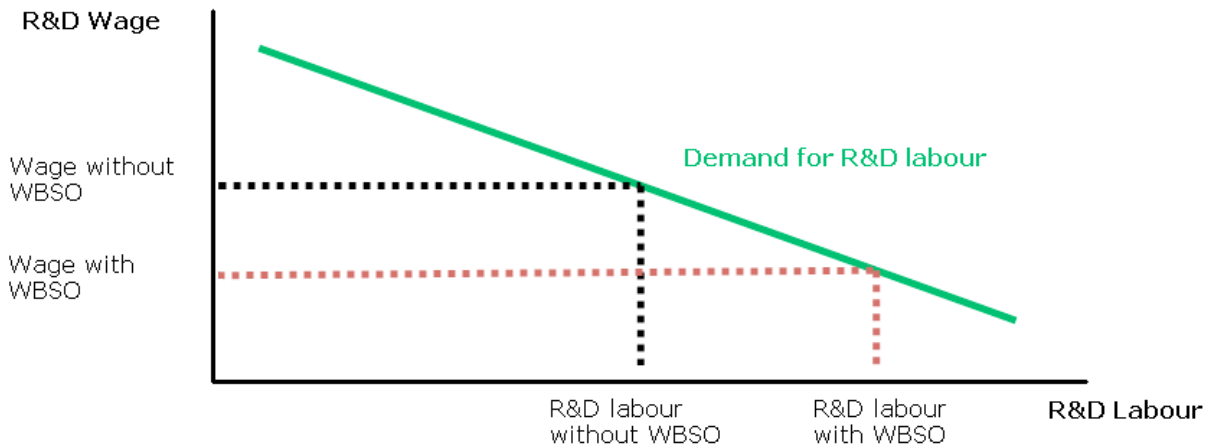
How should the mean BFTB be interpreted? The report (p. 62) gives a specimen calculation for 2007, giving after rounding total gross R&D salary costs of €3.4 billion, of which 60%, or approximately €2 billion, was supported by the WBSO. In 2007 companies received a total tax advantage of €0.46 billion because of the WBSO in the form of an R&D rebate awarded by NL Agency, or a payroll tax deduction. On average the tax advantage is therefore 23 per cent (0.46 divided by 2.0 billion). For a mean BFTB of 1.77 this means, according to the report, total additional R&D salaries of $€0.46 \times 1.77 = €0.8$ billion.

The report subsequently claims that the difference of €1.2 billion between the WBSO salary bill (€2 billion) and the additional R&D salary attributable to the WBSO (€0.8 billion) equals the deadweight loss, which is defined as the 'salary supported by the WBSO that would also have occurred in the absence of the WBSO'. From an economic theory perspective this is an unconventional interpretation of the 'deadweight loss' concept. Economics literature associates this concept with a loss of production attributable to supply restrictions imposed by a monopolist or a company with monopoly power. The definition of 'deadweight loss' is sometimes extended to cover a subsidy provided for activities that would also have taken place without the subsidy. This more extensive definition makes 'deadweight loss' the opposite of additionality. The definition of 'deadweight loss' used in the report is confusing. It would perhaps have been clearer to have referred to 'windfall gain'.

³⁴In its report EIM incorrectly referred in this connection to the impact on additional R&D spending, which is a broader concept than additional R&D salary cost expenditure alone. However, in estimating the model (Report appendices, p. 11, Equation 1) the parameter that underlies the BFTB variables is related to the R&D salary costs (variable *rdw*) and not to the total R&D costs.

³⁵The formula for this can be found in Appendix 2 of the econometric appendix to the main report of the 2006-2010 WBSO evaluation.

Figure 2: Relationship between demand for R&D supply and wage



In essence the WBSO is a wage cost subsidy, which, despite being calculated as a reduction of the salary bill rather than as a proportion of the hourly wage, nonetheless ultimately leads to a lower hourly wage for R&D work. Figure 2 depicts the demand for R&D labour as a decreasing function of the hourly wage. The WBSO deduction causes the hourly wage to decrease from ‘wage without WBSO’ to ‘wage with WBSO’, with an associated rise in demand for WBSO work from ‘employment without WBSO’ to ‘employment with WBSO’.

Underlying the report’s estimate of the BFTB is an estimate of the demand elasticity for R&D work. The report (p. 66) refers to the connection between elasticity and the BFTB. However, the elasticity described there is not the R&D labour demand elasticity. The elasticity described in the report is a measure of the relative effect of a 1% increase in WBSO deduction on the R&D salary bill. This is an increase of 0.38%.

Suppose that the 1% increase in WBSO deduction is interpreted as a 0.23% reduction in remuneration per worker.³⁶ The effect on the R&D salary bill would then be a 0.38% increase. The relative increase in the R&D salary bill (+0.38% in this case) is equal for a given R&D salary (before the WBSO deduction) to the relative increase in R&D employment. If this interpretation is correct, then the implicit estimate of demand elasticity that underlies the BFTB is 1.65 (= 0.38/0.23), which in the empirical literature is a relatively high elasticity, and more so in that specialized employees are involved.

Second and third-order effects

The evaluation also attempts to test Hypotheses 3 to 6, inclusive, by measuring the WBSO’s impact on innovation (second-order effect) and on business performance (third-order effect), both within and outside the sector concerned. An attempt is made to identify these effects by means of elasticities (second-order effect) and a production function model (third-order effect). However, it is hard to draw a conclusion from this since the problems discussed above in measuring the first-order effects are more pronounced for higher order effects. Even more caution is therefore called for in interpreting these effects.

3.4 Conclusion

It was opted in the WBSO evaluation to give a prominent role to quantitative analysis. However, the evaluation of a generic measure used on a large scale is no simple matter. The chosen dynamic panel data model is more commonly used in the evaluation of measures of this kind, but certainly has its limitations. The lack of clarity about the source of the variation in the use of the WBSO necessitates various assumptions in the model, in both the specification and the estimates. The empirical analysis takes various awkward econometric problems into account. Several specification tests needed for a proper assessment of the analysis are absent. The causal interpretation of the effect found continues to be problematic

³⁶We follow here EIM’s assumption of completely elastic availability levels of researchers. This assumption is not very realistic in the short term for R&D employees.

because the instrumental variables do not necessarily correlate with the WBSO deduction ratio while also being independent of the R&D salary bill.

In order to be in a better position in future to make statements about the effectiveness of the WBSO the working group recommends the following.

- Give more attention in future evaluations to discontinuities. The last evaluation also examined discontinuities, but the large number of simultaneous changes that were implemented at the time made it hard to differentiate between the various effects. In future a difference-in-difference approach may allow analysis of the impact of the extension of the band to €200,000 that is scheduled for 2013. The proposed shift from tax schemes for innovation to the TKI supplement that was expressed in the coalition agreement 'Bruggen slaan (Building bridges)' may lead to more changes in the future. If used skilfully, this approach may offer interesting evaluation opportunities. For example, a higher ceiling would not be considered amenable to a difference-in-difference analysis because of the restricted number of observations. The present evaluation also examined the effect of changes in the system of bands, but because the wider bands were introduced mainly as a crisis measure, difference-in-difference did not yield robust results.
- Future evaluations could gain in strength by more clearly establishing the relationship between survey output and econometric results.
- Avoiding an overestimate of the WBSO's impact requires accurate knowledge of whether the companies that have an R&D declaration also actually perform R&D. For WBSO applications submitted in or after January 2013, applicants are obliged after year-end to declare the number of R&D hours actually expended (which is consistent with the RDA). WBSO applicants are currently required to declare these hours only if they are fewer than 90% of the originally allotted number. There will therefore be a clearer view of this aspect in a future evaluation.
- In respect of the second-order effect, the impact of the change in the bands on innovative activities could be examined.

Chapter 4: Evaluation designs for instruments

4.1 Introduction

This chapter is the core of this report and presents the Expert Working Group's recommended evaluation designs for each of the six designated policy instruments. Each evaluation design is assessed on its merits. Before discussing the evaluation designs associated with an instrument, an explanation is given of the objective, the justification in terms of market or government failure, and the instrument's operation, implementation and budget. The selection comprises instruments with diverse budgets and includes traditional generic instruments (Innovation Credit, Research & Development Deduction, Innovation Performance Contracts), alongside new, more demand-driven instruments (Centres for Innovative Craftsmanship and Top Consortia for Knowledge and Innovation), and a more 'indirect' policy instrument (Certificate of Good Service for municipalities). This chapter therefore addresses a broad spectrum of policy options, building up a picture across the board of what to expect, and more in particular of what not to expect, of impact measurement for each type of instrument. The descriptions of evaluation methods in this chapter reflect the comprehensive account of the evaluation techniques given in Chapter 2. Unlike the previous chapters, which set out to guide the reader as much as possible, this chapter assumes a higher level of technical knowledge and goes into the econometric material in greater depth. The aim nonetheless has been to present the evaluation designs as intuitively and clearly as possible. The econometric equations develop the intuitive ideas and may present more of an intellectual challenge to advanced readers, while also providing points of reference for further development in future evaluations.

4.2 Innovation Credit (IK)

Objective

The Innovation Credit (IK) scheme is aimed at increasing private R&D spending oriented to technical and/or clinical high-risk development projects for new products, services and processes.³⁷

Market failure

Based on the economic theory of prosperity there are two reasons for the government to provide Innovation Credit. The first is that the social return of an innovation project may be greater than the private return, in particular because of knowledge spillover. The second is to use Innovation Credit to mitigate capital market failures where the government is better informed about an innovation project than private investors. This situation may arise if entrepreneurs are more willing to divulge information with NL Agency than with private investors,³⁸ and if NL Agency has more specific technological and innovation-related expertise than private investors. NL Agency may then be better able to assess the risks of a specific innovation project.

Operation

The market for funding innovation projects may be segmented along the lines of risk and return.³⁹ At one end are high-risk, high-return projects, which obtain funds on the venture capital market. Low-risk, low-return projects are usually financed through bank loans. The Innovation Credit scheme targets the middle segment, which comprises projects that are too high-risk for banks, but offer insufficient return to interest venture capitalists. These projects therefore encounter difficulty in obtaining private funding.

Unlike other policy instruments, the Innovation Credit scheme applies the direct benefit principle in the event of technical success. On average 70-80% of Innovation Credits are repaid. The return on the successes is too low to cover losses on the failed projects. The reasons are: 1) the charge for a loan under the Innovation Credit scheme is close to market interest rates in the lower risk segment (bank loans), despite the greater risk involved, and 2) unlike shareholders' equity, the government does not share in the profit of successful projects, but does bear the losses of the loan portion of projects that fail technically. The implicit subsidy element of the Innovation Credit scheme accrues to the venture capitalist or entrepreneur.

³⁷The projects involved are in the experimental phase and are new for the Netherlands.

³⁸This is referred to in the economic literature as the hold-up problem.

³⁹The term 'return' as used here always refers to expected return.

The interest rate for the Innovation Credit scheme is linked to Euribor. The base is 4% for technical projects of companies that have sufficient cash flow, rising to 7% if cash flow is uncertain. Clinical projects are charged 7% interest, rising to 10% if the cash flow is uncertain. Cofinancing through the Innovation Credit scheme may be up to 35% of the eligible expenses of the types of project given above. An Innovation Credit applicant must have prospects of financing, and will therefore also have to approach other, private, lenders.

Implementation

The advisers and experts on the NL Agency's credits team assess applications on six criteria: entrepreneurship/business, business case/earning potential, novelty/innovative content, technical risk, project plan and financing. A rating is given for each criterion, on a scale ranging from -- to ++. The opinion of the credit committee (i.e. an internal committee with an external entrepreneur) is then solicited. This independent assessment counts as a second opinion. The credit committee gives a qualified opinion on the application, stating its reasons. The opinions of NL Agency's 'bureau committee' and the credit committee differ in approximately 20% of cases. The credit committee's recommendation is not followed in every case. The ultimate go/no go decision is made by the NL Agency's Innovation Credit manager on behalf of the ministry. The average application period for an Innovation Credit is three to four months (with a maximum of four months), which is no longer than for other, private, lenders. For more than 50% of Innovation Credit applications the project encompasses the entire company.

The assessment is not influenced by the amount of credit requested. The average amount requested is €1.5-2 million. Approximately 80 applications are submitted each year, of which 20 are withdrawn by the applicant after receiving feedback on the plans from NL Agency's advisers. Of the remaining 60, 30 ultimately receive an Innovation Credit. The budget was almost doubled in 2012, and the number of applications will probably increase in response to the policy amendment to scrap and partly convert subsidies into fiscal measures and loans, plus the broadening of the Innovation Credit target group to include large companies with effect from 1 January 2012.

Budget

€95 million was available for the Innovation Credit scheme in 2012. This includes administration costs, which were approximately €3.7 million in 2012. Unutilized budget flows back to the fund and through to subsequent budget years, which helps avoid premature closure of the instrument because of insufficient budget, but without having to issue too much, which is a drawback of open-ended schemes. The fund is thus able to adjust its size flexibly if necessary within the total budgetary scope of the SME+ Innovation Fund. Income, such as the repayment of granted credits, also flows back to the fund and is therefore available for new funding. This uncertain income is not included in the National Budget, but does contribute to the revolving nature of the fund and the underlying pillars such as the Innovation Credit scheme.

Evaluation period

The Innovation Credit scheme is to be evaluated in 2012 and 2013. The working group has produced a preliminary recommendation for this evaluation. A research bureau was recently engaged to perform the evaluation.

Hypotheses

Core hypothesis

- 1) Companies spend more on R&D if they receive an Innovation Credit.

Hypothesis in support of core hypothesis

- 2) Projects that are supported by the Innovation Credit scheme materialize faster than rejected projects.

Hypothesis for assessing spillover effects of application procedure

- 3) The NL Agency application process improves the prospects of a loan from private lenders for companies that are rejected for the Innovation Credit scheme.

Hypothesis for assessing social benefits

- 4) The social benefit of successful Innovation Credit applications exceeds the social cost of the scheme's implicit subsidy element.

Indicators

Various indicators are used in testing the above hypotheses.

- Whether or not the project goes ahead, including the scores for the underlying criteria and the credit committee's recommendation. The Innovation Credit scheme should be targeted at projects that would otherwise not be executed. Influences on project pace and size are also important. (Hypotheses 1 and 2)
- Cofinancing; in answer to the question: what influence did granting or not granting the Innovation Credit have on cofinancing by other investors? (Hypothesis 3)
- Project success, both technical and commercial aspects. The technical success or otherwise of the project becomes known sometime during the term of the Innovation Credit, and repayment of the credit depends on success. It is possible to measure the degree of technical success relatively early in some cases based on the outcome of a patent office search report⁴⁰, and in the long term based on the patents obtained. The sale and resale of patents says something about commercial success. Technical success can also be derived from the NL Agency assessment of whether the credit is to be repaid, which is a far more direct measurement. (Hypothesis 4)
- A measure of commercial success is the growth or contraction of the companies in terms of revenue, investments, profit, follow-on projects and change in employee numbers. Other possible points to examine are internationalization of the company, or a measure of market share. (Hypothesis 4)
- Spillovers may be investigated separately, within the same sector and in others. The question then is how quantifiable this aspect is for the projects supported with an Innovation Credit. The results for spillover effects are more readily acquired if the effects of the total R&D spending within the same sector and in others are investigated on individual business performance (following the same approach as in the WBSO evaluation). (Hypothesis 4)

Much of the necessary data for an evaluation can be obtained from NL Agency, including information about all companies ever to have applied for an Innovation Credit. Because the Innovation Credit scheme is concerned with specific projects, investigation will initially focus on the projects that were the subject of an Innovation Credit application. The group of companies with a successful Innovation Credit application is relatively easy to track. It can be derived from NL Agency project information whether a project was a technical success, and the portion of the credit that has been or is to be repaid. These companies can also simply be requested for information about the magnitude of R&D spending, the commercial success of the supported projects, the financing from other investors and the economic performance of the company in general.

Companies whose Innovation Credit application was rejected can also be interviewed, for example to ascertain whether any alternative financing was obtained. The feasibility of obtaining sufficient data about these companies is questionable.⁴¹ In the case of rejected applications where the project is the same as the company, WBSO data may possibly be used to ascertain whether finance was successfully obtained. If a company is present in the WBSO data it may be assumed to have been successful in arranging financing. Along the same lines WBSO data could possibly also be used to track non-start-up companies with a rejected Innovation Credit application, provided the application is 'large' relative to the total company size. If an Innovation Credit would increase an applicant's current total assets by at least 100%, it would be reasonable to assume that the performance of the entire company is closely linked to the success of the innovation project. This assumption is invalid for large companies, for which the success of an innovation project cannot be derived from the annual figures. The idea therefore is that WBSO data provide an indication of the growth of a company's R&D spending. The absence of growth following a rejected application leads to the assumption that no alternative financial backer was found.

Various data about companies (including revenue and employee numbers) can be obtained from Statistics Netherlands, which requires the linking of NL Agency data with Statistics Netherlands data (General Business Register, ABR) and with data from the tax authorities. The linking is through Chamber of Commerce file numbers. A weakness of the above is that data from Statistics Netherlands surveys is extremely incomplete for relatively small companies. WBSO data from NL Agency is an important source of information about R&D-spending, in particular of relatively small companies, as a supplement to R&D data from Statistics Netherlands. WBSO data normally cover all companies with substantial R&D activities (including companies with fewer than ten active employees, which are not observed in the Statistics Netherlands R&D survey). These data are also needed in assessing Hypothesis 4. Statistics Netherlands and WBSO data do not help

⁴⁰ Acquiring a European patent can take many years. The 'search report' gives an early indication of an application's chance of success.

⁴¹ Provided WBSO is received for the same project, of the same size. Account must also be taken of the approximately 70% of Innovation Credit applicants that can be retrieved from the WBSO database.

answer the question as to whether projects materialize faster because of the Innovation Credit scheme. A supplementary survey will probably still be needed to answer this question. Patent data are available by linking Chamber of Commerce data with Netherlands Patent Office data.

Table 2 List of indicators for Innovation Credit evaluation design

Hypotheses	Indicators		Data sources	
	<i>Dependent</i>		<i>Independent</i>	
1) Companies spend more on R&D if they receive an Innovation Credit.	<ul style="list-style-type: none"> • Go/no go of project • Project size 		<ul style="list-style-type: none"> • Dummy for Innovation Credit grant 	<ul style="list-style-type: none"> • NL Agency • Statistics Netherlands
2) Projects that are supported by the Innovation Credit scheme materialize faster than rejected projects.	<ul style="list-style-type: none"> • Project pace 		<ul style="list-style-type: none"> • Dummy for Innovation Credit grant 	<ul style="list-style-type: none"> • NL Agency • Survey
3) The NL Agency application process improves the prospects of a loan from private lenders for companies that are rejected for the Innovation Credit scheme.	<ul style="list-style-type: none"> • Available cofinancing 		<ul style="list-style-type: none"> • Dummy for Innovation Credit grant 	<ul style="list-style-type: none"> • NL Agency • Statistics Netherlands
4) The social benefit of successful Innovation Credit applications exceeds the social cost of the scheme's implicit subsidy element.	<ul style="list-style-type: none"> • NL Agency assessment of technical success • Repayment of credit • Patents • Growth of companies in terms of revenue, investments, follow-on projects, employee numbers, internationalization, etc. • R&D within the sector as a measure of spillover 		<ul style="list-style-type: none"> • Dummy for Innovation Credit grant • Amount of Innovation Credit granted 	<ul style="list-style-type: none"> • NL Agency • WBSO data • Netherlands Patent Office • Statistics Netherlands
Control variables [not exhaustive]	<ul style="list-style-type: none"> • Scores on NL Agency assessment criteria and the credit committee's recommendation • Project information • Company characteristics 		<ul style="list-style-type: none"> • NL Agency • Statistics Netherlands 	

Evaluation design

An effective demonstration of the Innovation Credit scheme's primary impact (Hypothesis 1) requires the identification of comparable groups, where one group received the Innovation Credit and the other did not. Ideally the comparable groups would emerge from randomized allocation of Innovation Credit to eligible companies (see the discussion in Chapter 2). The absence of a randomized experiment for the Innovation Credit scheme necessitates the use of other evaluation methods (see Chapter 2).

Option 1: Propensity score matching

Experimental and control groups

Companies that apply for an Innovation Credit fall into three groups: those with no chance of an Innovation Credit, those almost certain to obtain an Innovation Credit; and those in a 'grey area' where the outcome of the assessment procedure could be either positive or negative. The non-random nature of Innovation Credit allocation precludes the use of a simple comparison of accepted and rejected applications to test the hypotheses.

The hypotheses (with the exception of Hypothesis 3) are concerned with all accepted Innovation Credit applications and the evaluation should focus on companies that are almost certain to obtain a credit and those in the grey area. The evaluation problem is the lack of counterfactuals for applicants that are assured of receiving the credit, therefore theoretically ruling out an empirical evaluation for this group based on comparison with a control group. The comparable groups must be drawn from the grey area, but the interpretation must take into account the exclusion of some accepted Innovation Credits. If there is a relationship between the certainty of application acceptance and the (positive) impact of the Innovation Credit, then comparison in the grey area will yield an underestimate of the actual effectiveness. This means that if evidence of a certain hypothesis is found in the grey area, that hypothesis is probably also valid for all granted Innovation Credits.

Identification strategy

Much information is available about the process of granting applications. NL Agency assesses applications on six criteria and the credit committee also expresses its opinion. In addition, other characteristics of the application and the applicant are known. Two steps are involved in identifying an Innovation Credit's impact. First the probability of acceptance is determined for each application based on the observed characteristics (the propensity score).⁴² In step two the impact of the Innovation Credit is determined by comparing accepted projects with rejected projects that have (approximately) the same probability of being granted.

Econometric specification and methods

Having determined the propensity score for each observation, various methods are available to establish the impact of the Innovation Credit on the target variables (see Chapter 2). Working with weights is possibly preferable to the one-on-one matching of observations with and without Innovation Credit, because it leads to less data loss. The weight given to an observation will reflect the similarity of the company in the control group with the company that received the Innovation Credit. The limited sample size is probably the greatest obstacle in the empirical evaluation. The large proportion of Innovation Credits that go to start-ups means that company age is an important variable for matching. The methods may be applied by means of the Stata software package. A brief explanation of the technique is given in Box 2.4. The weights are determined by weighting the target variable with the inverse of the estimated propensity score (the allocation probability of accepted applications and probability of rejection for the non-accepted applications).

The above is possible only if the effect of accepting an Innovation Credit application or otherwise is the relevant choice. It is more difficult to directly estimate the benefit of one euro of credit, which is known as the bang for the buck, and depends not only the selectivity in allocation but also on the selectivity in the size of the application. However, it is possible to obtain an aggregated idea of the bang for the buck by dividing the causal effect of an allocation (e.g. on commercial success, profit, or revenue) by the mean size of the credit. It must also be taken into account that on average most of the credit is repaid, because that is also part of a project's benefit.

⁴²In principle the NL Agency score sheet should be sufficient because it weighs all relevant information for an application's chance of success. Moreover it would probably be difficult to include even more information based on a maximum of eighty applications. Nonetheless it would be useful to consider other characteristics as well. You should actually have concerns about the 'propensity score' if other characteristics besides the information from the NL Agency score sheet also influence an application's chance of success.

Sensitivities

The number of observations in the grey area needed for the execution of a worthwhile impact measurement is difficult to state in advance. This depends on the variation in the data, for two reasons. First there must be sufficient overlap in the distribution of propensity scores of accepted and rejected applications. In other words, the larger the grey area, the more precise and complete the estimation method. Second, the level of risk of the projects is significant. High-risk projects require a larger sample than low-risk projects in order to provide a reliable estimate of a mean outcome (e.g. project completion, or R&D spending). The modest volume of Innovation Credit applications in the grey area, the number of which cannot be anticipated clearly, makes the success of the evaluation method described above uncertain in advance.

It was noted above that the propensity score method illuminates the impact of the Innovation Credit scheme only in the grey area. For the sake of completeness it is also necessary to examine projects that are certain to be accepted. The actual size of this group must first be established. It may be possible to say something about the effectiveness in this area by examining the interaction of the Innovation Credit impact with the propensity score. If projects with a higher propensity score are more effective, then the propensity score method will evidently produce an underestimate of the total effectiveness of the Innovation Credit scheme. However, the success of this approach depends on the sample size, and interpretation continues to involve extrapolation outside the area of observation. It must therefore be noted that complete representativeness can be achieved only through an experiment with (partial or conditional) randomized allocation.

Option 2: Panel regression

Experimental and control groups

This option involves estimating the correlation between Innovation Credit allocation and subsequent outcomes. The companies fall into three groups: companies allocated an Innovation Credit, companies whose application for an Innovation Credit was rejected, and companies that did not apply.

Identification strategy

The large amount of panel data in Statistics Netherlands' possession allows for panel regression in which the independent variables are both a dummy for companies with an application and a dummy for companies allocated an Innovation Credit. However, it must be noted that an application for, or allocation of, the Innovation Credit probably does not satisfy the condition of strict exogeneity. Companies apply for an Innovation Credit because they are in a phase of development, growth, or both.

Econometric specification and methods

Simple fixed effects panel data techniques can give a biased picture because panel dropout is probably selective. It is therefore advisable to examine the time of introduction of the Innovation Credit and to consider a kind of difference-in-difference model. It is unclear what constitutes a good control group with the same time trend. The following difference-in-difference function could be estimated:

$$Y_{it} = \beta_0 + \beta_1 IK_{it} + \beta_2 A_{it} + \beta_3 X_{it} + \eta_i + \theta_t + U_{it}$$

In this equation Y represents company i 's R&D spending, the β s represent the coefficients of the estimators, IK is an indicator of whether the company received an Innovation Credit, A is an indicator of whether a company submitted an application, and X is a control variable vector. Furthermore an adjustment is needed for unobserved characteristics of companies and time effects ('fixed effects' for companies, η , and time, θ), and U is the error term.

Sensitivities

The greatest problem with this approach is that a causal interpretation for correlations is not always possible, but requires adjustment for all selectivity in the Innovation Credit allocation process (see Chapter 2). This selectivity can be handled by adjusting the regression model for heterogeneity, while being alert to the addition of what are known as confounders. Confounders are variables that absorb some of the effect that the researcher is interested in. In addition, the limited sample size may mean the grey area has too few observations for a sound evaluation.

Option 3: Survey method

Soliciting information from Innovation Credit recipients and rejected applicants about the instrument's importance (conditional questions). This resembles the previously executed TOK evaluation.⁴³ For example: what influence did the credit have on progress, size and speed of project execution, and what were the economic results? Control questions can be used to guard against opportunistic answering.

Conclusions and recommendation

The first evaluation option differs greatly in quality from the second and third options. The first option uses more information about applications and applicants in forming a comparable control group than is the case in the two other options, giving a more reliable view of the effectiveness of the Innovation Credit scheme. The practical feasibility of the first evaluation will be unclear in advance. The total number of applicants is small, and the size of the evaluation's grey area unclear (and it is impossible to state in advance how large it should be). A second disadvantage is that the hypotheses are tested only for applications in the grey area, which can also pose a problem if the grey area is small. If so, the interpretation of the effect is limited.

Hypothesis 1 can be tested with the propensity score method (Option 1). Hypothesis 2 can be tested by means of panel regression (Option 2). For Hypothesis 3, the working group advises the use of a survey with conditional questions (Option 3), in view of the extremely high costs of identifying a non-user control group. Hypothesis 4 cannot be tested by simple means, and all spillover effects of the Innovation Credit scheme must be examined. However, there will currently be only limited long-term effects and spillover effects of the Innovation Credit scheme, because it started only in 2008.

The impossibility of establishing the size of the grey area in advance, and the inability of the propensity score method to test all hypotheses, mean that it makes sense to use both other evaluation options alongside the propensity score method. Doing so will give the most complete picture possible, and may also clarify the pros and cons of the different evaluation designs, which will be of use in determining how to perform future evaluations. The Innovation Credit scheme started in 2008, and serious far-reaching analyses will become possible only in a subsequent evaluation. The first evaluation must therefore be used as an opportunity to learn about the best kind of evaluation design for the Innovation Credit scheme.

⁴³Bureau Bartels (2000). Evaluatie van de TOK-regeling (Evaluation of the TOK scheme), Amersfoort.

4.3 Research and Development Deduction (RDA)

Objective

The objective of the Research and Development Deduction (RDA) is to increase private R&D investment. Alongside the objective of increasing private R&D investment, a secondary RDA objective (as with the WBSO and Innovatiebox) is to create a tax climate that is attractive to foreign companies opting to base their R&D activities in this country. This secondary objective is not discussed in this report.

Market failure

The legitimacy of the RDA derives from companies' spending less than the socially desirable amount on R&D, because they do not factor knowledge spillovers of possible benefit to other companies into their own R&D spending decisions. The result is a suboptimum level of investment in R&D (see the discussion in Chapter 1). The financial expenses for R&D salaries were already lowered by the WBSO. The RDA lowers the other R&D costs (e.g. investment in equipment, consumables), which gives a more balanced tax treatment of R&D labour and capital, and causes less distortion of the correct choice of capital-employment ratio.

Operation

The WBSO and the RDA are closely interwoven. Eligibility for the RDA depends on the simultaneous submission of a WBSO application. A single application form is available for the two schemes. The reason for making this link was to minimize the scheme's administrative burden and administration costs.

There are two RDA regimes.

- Deduction based on actual costs and expenditure⁴⁴: An additional corporate income tax (VpB) or income tax (IB) deduction for 40% of R&D costs and expenditure (based on 2012). For a corporate income tax rate of 25% the net benefit is therefore 10% of the R&D costs and expenditure that fall under the RDA.
- A flat-rate deduction: applicants may deduct 40% of €15 per WBSO hour from profit, to a maximum of 150 WBSO hours per month, i.e. 1800 WBSO hours a year (which is approximately 1 FTE). The €15 is based on the ratio between labour costs and other costs for R&D in small companies, and is intended to compensate for these other costs. Since the total amount that may be deducted is determined by the number of WBSO hours, the RDA in the flat-rate regime actually acts to lower the tax costs of R&D labour. Until the 1800 hours is reached, the more R&D labour a company employs, the more deduction the company receives. (€15 RDA per WBSO hour, based on 2012, results in €15 x 40% = €6 deduction from taxable profit per WBSO hour). The flat-rate regime has an escape mechanism. Companies that fall under the flat-rate regime by virtue of the number of hours, but have more than €50,000 of costs and expenditure, are eligible for deduction based on actual costs. Having once opted to use the escape mechanism, a company can no longer revert to the flat-rate regime in the same year.

The main reason to introduce the flat-rate regime was to reduce the administrative burden for (small) companies with relatively small-scale R&D and to limit the administration expenses. It is estimated that the budgetary burden of the group in the flat-rate regime in 2012 is approximately €12 million (5% of the budget), accounting for approximately 60% of applications.

Implementation

Within three months following year-end an applicant must notify the actual hours and (if applicable, for the actual-costs scheme applicant) costs and expenditure. Where necessary an adjustment may be applied based on actual figures relative to allocation. Only downward adjustment is allowed (for either regime); no increase is permitted in the event of higher actual figures.

⁴⁴Not all costs are deductible. For instance, the costs of outsourced research are excluded. For a complete list of deductible and non-deductible costs and expenses, refer to the Research and Development Deduction (RDA) manual, NL Agency (2012).

The RDA is consistent with normal loss accounting, whereby a maximum of thirteen years' carry-forward or carry-back is allowed.

Budget

The RDA budget for budget year 2012 is €250 million. The intention is for the RDA budget to rise in the next few years to €375 million in 2013 and then in 2014 reaching the maximum of €500 million that is envisaged for 2014 and beyond. If the budget increases in 2013 and 2014 go ahead, and depending on use in 2012, the current 40% deduction level will be increased. The above will apply to both regimes.

The RDA and WBSO are subject to the t-1/t+1 budgetary rule, whereby overruns or underruns in t-1 are compensated in the budget in year t+1 (which is established in year t). The compensation may be through parameter adjustments with an unchanged budget, or through additional budget.

Evaluation period

The RDA started on 1 January 2012, with applications accepted from 1 May 2012, both new and retroactive with effect from 1 January 2012. The first RDA evaluation is scheduled for 2016.

Hypotheses

Core hypothesis

- 1) The RDA has a positive impact on a company's total R&D spending before deduction of RDA: the RDA lowers the total price of R&D activities, encouraging companies to spend more on R&D (there is both an 'income effect' and a 'substitution effect').

Hypotheses in support of core hypothesis

- 2) The RDA based on actual costs alters the ratio of R&D salaries to other R&D costs in favour of the latter through a change of relative prices (there is a 'substitution effect' within the R&D spending). In determining the substitution effect, adjustment is required for any change in the ratio of R&D salaries to other R&D costs as a consequence of larger-scale R&D activities (larger companies are often more capital intensive).
- 3) The flat-rate RDA alters the ratio of R&D salaries to other R&D costs in favour of the former through a change of relative prices (there is a 'substitution effect' within the R&D spending).

Hypotheses regarding second-order and third-order effects

- 4) Companies' additional R&D spending attributable to the RDA (first-order effect) has a positive impact on innovativeness (second-order effect) and business performance (third-order effect).
- 5) Knowledge spillovers mean that alongside private returns for the companies that perform the additional R&D there are also externalities for other companies. The social (economic) return of the RDA (and WBSO) is therefore greater than the private return.

Evaluation design

An effective demonstration of the RDA's primary impact (Hypothesis 1) requires the identification of comparable groups, where one group received RDA and the other did not. Ideally the comparable groups would emerge from randomized allocation of RDA to eligible companies (see the discussion in Chapter 2). The absence of a randomized experiment for RDA necessitates the use of other evaluation methods (see Chapter 2).

Table 3 List of indicators for RDA evaluation design

<i>Hypotheses</i>	<i>Indicators</i>		<i>Data sources⁴⁴</i>
	<i>Dependent</i>	<i>Independent</i>	
1) <i>The RDA has a positive impact on a company's total R&D spending.</i>	<ul style="list-style-type: none"> • <i>Expenditure on R&D</i> 	<ul style="list-style-type: none"> • <i>Dummy for use of RDA</i> 	<ul style="list-style-type: none"> • <i>NL Agency</i> • <i>Statistics Netherlands</i>
2) <i>RDA based on actual costs has a 'substitution effect' within R&D spending.</i>	<ul style="list-style-type: none"> • <i>Number of WBSO hours</i> 	<ul style="list-style-type: none"> • <i>Dummy for use of RDA</i> • <i>Dummy for actual-costs component</i> 	<ul style="list-style-type: none"> • <i>NL Agency</i>
3) <i>The flat-rate RDA has a 'substitution effect' within R&D spending.</i>	<ul style="list-style-type: none"> • <i>Number of WBSO hours</i> 	<ul style="list-style-type: none"> • <i>Dummy for use of RDA</i> • <i>Dummy for flat-rate component</i> 	<ul style="list-style-type: none"> • <i>NL Agency</i>
4) <i>RDA has a positive impact on innovativeness and business performance.</i>	<ul style="list-style-type: none"> • <i>(Labour) productivity</i> • <i>New products & processes</i> • <i>Patents</i> • <i>Growth of companies</i> 	<ul style="list-style-type: none"> • <i>Dummy for use of RDA</i> 	<ul style="list-style-type: none"> • <i>NL Agency/WBSO</i> • <i>NL Agency/Netherlands Patent Office</i> • <i>Statistics Netherlands</i>
5) <i>The social (economic) return of the RDA (and WBSO) is greater than the private return.</i>	<ul style="list-style-type: none"> • <i>Innovation per sector</i> 	<ul style="list-style-type: none"> • <i>Use of RDA per sector</i> 	<ul style="list-style-type: none"> • <i>NL Agency</i> • <i>Statistics Netherlands</i>
<i>Control variables [not exhaustive]</i>	<ul style="list-style-type: none"> • <i>Project information</i> • <i>Company characteristics</i> • <i>Company-specific variables that relate to the supply of and demand for R&D services⁴⁵</i> 		<ul style="list-style-type: none"> • <i>NL Agency</i> • <i>Statistics Netherlands</i>

Option 1: Phased asymmetric extension of RDA

Experimental and control groups

The RDA is currently being rolled out. The RDA budget between 2012 and 2014 will probably be expanded from €250 million to €500 million. In general this is an ideal phase to introduce exogenous variation in RDA participation, with a view to supporting reliable impact measurement at a later stage. The RDA's budgetary scope, and the budgetary burden ratio between actual-costs and flat-rate applications of approximately 1:20, mean that it makes sense to increase the actual-costs component in the first instance. The group of companies that use the flat-rate RDA would then form the control group. The next increase will apply to both the flat-rate and actual-costs component, where the budgetary burden imposed by the flat-rate component will double relative to 2012.⁴⁶ Both groups will then be an experimental group.

⁴⁵NL Agency has access to all application data of WBSO and RDA applicants. For the evaluation these data could be linked to Statistics Netherlands data and tax data. It emerged from the WBSO evaluation that it was possible to link approximately 95% of the companies in the WBSO database with Statistics Netherlands data. Use can also be made of information from NL Agency's company visits. It is possible to categorize data by top sector. Statistics Netherlands has recently produced a standard for this through custom work and a baseline measurement.

⁴⁶In policy terms it is more justifiable to withhold the flat-rate group's increase for a year, because the RDA lowers the price of R&D labour for that group, while the scheme is actually intended to make the other R&D input less expensive.

Identification strategy

With an asymmetric roll-out the RDA's impact can be determined by comparing trends in indicators for the experimental group with those of the control group. Any perceptible difference in trends during the asymmetric roll-out phase would point to the differences being caused by the RDA extension. The investigation then follows a difference-in-difference design. Since the distinction between the experimental and control groups was made with evaluation in mind, this choice may be assumed to be exogenous. Extending the instrument in this way enables relatively unbiased estimating of the instrument's effect. Otherwise it must be noted that this approach measures the effect of the increase in the roll-out phase. Under the assumption of diminishing returns of the tax instrument, this approach gives a lower limit for the total effect.

Hypotheses 1 and 2 can be tested with the first asymmetric extension. According to Hypothesis 1 the increase of the actual-costs RDA should lead to higher R&D spending in the experimental group than in the control group. According to Hypothesis 2 the non-wage component of R&D spending should be higher in the experimental group than in the control group. Hypothesis 3 can be tested with the second asymmetric extension alongside Hypotheses 1 and 2, because of the simultaneous increase in the flat-rate RDA. In the second extension the increase for the flat-rate group is relatively higher than for the actual-costs group.

If a company's R&D efforts are found to increase, then the outcome measures to be examined to test Hypothesis 4 could be labour productivity, the number of new products (and/or processes), patents and company growth, either in terms of employees, revenue, or profit. Testing Hypothesis 5 requires investigation of a given sector's innovativeness. Sectors that make much and little use of the flat-rate component will then have to be compared.

Econometric specification and methods

The proposed approach can be estimated with a standard linear model in accordance with the difference-in-difference method (see the discussion in Chapter 2). The dependent variable is (the annual change in) an output or outcome indicator, and the independent variables are dummies for treatment and any control variables. Difference-in-difference is sometimes linked to (propensity score) matching, in order to obtain double-robust estimators of a treatment effect. This approach may be included as a robustness analysis.

Sensitivities

Companies that are eligible for the flat-rate RDA have the alternative option of actual-costs RDA if the company's costs are €50,000 or more a year. The composition of the control group therefore depends on the size of the increase of the actual-costs RDA. Companies that opt for the actual-costs RDA despite being eligible for the flat-rate RDA could possibly be excluded from the sample.

Companies that are aware of a planned increase in the flat-rate RDA the following year that does not affect the actual-costs RDA may act in anticipation, which may lead to an underestimate of the RDA's effectiveness. If a future increase in the flat-rate RDA is unknown in advance, then companies might possibly respond differently.

Option 2: Discontinuities around the maximum WBSO hours and RDA introduction

Experimental and control groups

Companies expending more than the 1800 WBSO hours receive the policy intervention that the RDA is concerned with: the promotion of R&D by stimulating R&D spending other than on salaries. Companies expending fewer than this number of hours will not receive the policy intervention, but will receive a decision granting a flat-rate RDA. The group of companies that use the actual-costs RDA can then be compared with the group that uses the flat-rate RDA. Both groups are experimental. The groups will become better comparable by disregarding companies that expend many fewer or many more than the 1800 WBSO hours.

Identification strategy

The discontinuity around the maximum WBSO hours can be used in identifying the RDA's impact on the proportion of salary costs in the total R&D costs. If companies with flat-rate RDA have a higher proportion of salary costs than their actual-costs counterparts, then either or both Hypothesis 2 and Hypothesis 3 may apply. Hypotheses 2 and 3 are therefore tested together: no distinction is made between them.

In addition a distinction can be made between companies based on the proportion of salary costs in their R&D spending, which may be included as an additional continuous variable in the regression discontinuity model. According to Hypotheses 2 and 3 flat-rate companies will increase their proportion of salary costs in line with increasing R&D spending because of the RDA, while actual-costs companies conversely will reduce their proportion of salary costs as their R&D spending increases. The hypotheses are therefore supported if the correlation between the proportion of salary costs and R&D spending is greater for flat-rate companies than actual-costs companies. This approach therefore tests the truth of Hypotheses 2 and 3 simultaneously, but an affirmative answer is unable to identify which of the individual hypotheses is true.

Finally, use may be made of the fact that the RDA was introduced only recently. On the introduction of the RDA a distinction may be made between flat-rate and actual-costs companies. Both groups are experimental. The varying extent to which companies used innovation subsidies that were withdrawn shortly beforehand means that the introduction of the RDA will have different effects on the two regime groups. One option therefore is to group companies not only on the basis of RDA regime, but also on starting position. As with Option 1 it is possible to use the difference-in-difference method, but in this case around the time of RDA introduction. It is important to map out clearly the financial incentives that existed prior to the RDA and how they influenced the various companies. Which hypotheses should be tested depends on the prior situation.

These additions may be analysed either separately or in combination. Combined testing requires a difference-in-difference model that includes the regression discontinuity.

Econometric specification and methods

All the above analyses may be performed in a standard linear difference-in-difference model as discussed in Chapter 2. The dependent variable is an output or outcome indicator and the independent variables are a dummy for the treatment and any control variables. Dummy variables must be included to indicate the regime the companies are in, and it must also necessary be ascertained which financial incentives were applied prior to the RDA.

The robustness of the estimated impact can be checked by varying the bandwidth around 1800 WBSO hours used in building the experimental and control groups. A larger bandwidth will reduce the comparability of the control group and the experimental group, but increase the number of observations.

Sensitivities

The proposed method tests only how the proportion of R&D salaries responds to the RDA. Without additional assumptions it is impossible to state how the RDA influences total RDA spending. The above method therefore yields only a partial evaluation.

Companies that are eligible for the flat-rate RDA may opt instead for the actual-costs RDA if the costs or expenditure exceed €50,000. The composition of the control group therefore depends on the size of the increase in the actual-costs RDA. To obtain an unbiased picture, companies that use the actual-costs RDA despite being eligible for flat-rate RDA are disregarded. Unlike the situation in Option 1, there is no control group, but only groups that differ in treatment. An unbiased measurement of the RDA's impact is therefore unlikely. The subsidies that companies used prior to the RDA were fairly numerous and diverse. It will therefore be difficult to establish on company level what the treatment was.

Option 3: Demand and supply system (traditional regression analysis with time series)

Experimental and control groups

N/A.

Identification strategy

This option determines whether there is a correlation between a lower price of R&D attributable to the WBSO/RDA and companies' R&D spending. Because R&D costs are market outcomes, an endogeneity problem arises: a negative correlation between R&D prices and R&D spending may be attributable to a greater supply of R&D production factors, or to declining demand for R&D production factors. An instrument will therefore have to be found that does relate to the demand, but not to the supply (and vice versa). This is a familiar problem in estimating demand and supply systems. Mohnen and Lokshin (2010) estimated a demand and supply system for the WBSO.

Econometric specification and methods

Two functions are used, one for the demand for R&D services and one for the supply of R&D services in year t :⁴⁷

$$demand_t = \alpha_v + \delta_v (1 - \tau_t) price_t + \beta_v X_t + \gamma_v Z_t^v + U_t^v$$

$$supply_t = \alpha_a + \delta_a price_t + \beta_a X_t + \gamma_a Z_t^a + U_t^a$$

Both the demand and supply depend on price and on observed factors X . For parties with a demand for R&D services, the RDA lowers the price by a fraction τ .⁴⁸ The phased introduction of the RDA causes this fraction to change over the years. The major problem in estimating demand-supply models of this kind is that the price is not exogenous. The price is determined on the market where supply and demand are equal. Therefore the observed equilibrium price in a given year is a function of the demand and supply function in that year. Estimating the above model functions therefore requires the use of instrumental variables Z , each of which either affects only the demand or only the supply.

The parameters are the most important policy parameters, and show how supply and demand respond to price. However, the RDA changes firstly the demand, which increases. If demand exceeds supply the price will increase until supply and demand are again equal. This means that the impact of the RDA on R&D services (in absolute value) is less than δ_v . After estimating the model it must be simulated in order to determine the new equilibrium after RDA introduction.

Time series elements are often also included when estimating this kind of demand-supply model. Seemingly unrelated regression, for example, may be used, which is a standard time series method for estimating demand-supply models with aggregated data.

Sensitivities

It is often difficult to find variables Z that influence either supply or demand alone. In this case too, there are no obvious candidates for these variables. Without these instrumental variables it is impossible to solve the endogeneity problem of the price of R&D services.

If linear equations are assumed, this approach will give a good indication of the marginal impact of changes in the prices of R&D production factors. It is not always possible to estimate a non-linear system satisfactorily.

Option 4: Compare the Netherlands with other countries

Experimental and control groups

There are two experimental groups: flat-rate and actual-costs Dutch companies. The control group comprises comparable companies abroad.

Identification strategy

It would appear logical to choose Germany, where nothing resembling RDA exists. This too should use a difference-in-difference model. However, the comparison is difficult in that it requires insight into both the Dutch and the German institutions.⁴⁹ In addition, linked information is required (albeit Statistics Netherlands has already created the links). The advantage of this method is that it illuminates the effectiveness of the entire instrument rather than just the extension, or the difference between the two regimes. Its disadvantage is the implicit assumption of the same (conditional) trend in the Netherlands and Germany.

Econometric specification and methods

The proposed approach can be estimated with a standard linear model in accordance with the difference-in-difference method. The dependent variable is (the annual change in) an output indicator and the independent variables are a

⁴⁷R&D services may include both R&D labour and other R&D spending.

⁴⁸This is a simplification. Unlike the WBSO, the RDA is not a direct input subsidy, but operates in conjunction with the tax on profits. If a company makes no profit, the RDA benefit is therefore zero.

⁴⁹Germany, for example, has no federal WBSO, but operates an intensive innovation policy through the states.

dummy for treatment and any control variables. Difference-in-difference is sometimes linked to (propensity score) matching, in order to obtain double-robust estimators of a treatment effect. This approach may be included as a robustness analysis.

Sensitivities

Identification: The disadvantage is the implicit assumption of the same (conditional) trend in the Netherlands and Germany.

Data: This approach relies on data from an appreciable number of different sources. For instance, the definitions and survey methods used by the Federal Statistical Office of Germany are different from those of Statistics Netherlands. This problem is limited if Eurostat standards can be used.

Option 5: Survey method

Experimental and control groups

N/A.

Identification strategy

RDA applicants may be asked directly to give their opinion about the RDA and changes in their behaviour because of the RDA. 'Are you a new research project started by the RDA?' 'Have you made new investments through the RDA?' An example of a survey of this kind can be found in Section 4.7 of EIM's 2006-2010 WBSO Evaluation (2012).

There may be questions about the past, the present and expectations for the future. The survey will include a question about the forecast profit of companies in the sector concerned, specifying the period.

Econometric specification and methods

The outcomes of a survey depend in part on how questions are phrased. A pilot study with open interviews is needed to validate the questionnaire. The consistency of answering must be tested by repeating important questions in a slightly different form. 'Markers' must be used to standardize the outcomes. (For example, an employee might answer differently from a director.)

Sensitivities

Outcomes: Answers do not always give an accurate picture of the real world. Socially desirable answering in particular may distort the outcome.

Conclusions and recommendation

The recent decisions about the 2013 Tax Plan do not include an asymmetric RDA roll-out. Since the RDA will probably acquire its final form in 2014, the opportunity for asymmetric roll-out to obtain exogenous variation is limited. It would nonetheless be advisable to bear this option in mind in the event of further extension of the RDA.

Option 2 would be preferable to Option 3, Option 4 and Option 5. The discussion of Option 2 refers to various possibilities for adjusting for selectivity under certain conditions. This is more difficult with a survey, as discussed in Option 5. Making a comparison between countries, as discussed in Option 4, is likewise unreliable at times, because countries often differ in more respects than the introduction of new policy alone. It is therefore uncertain that the same trend occurs in different countries. Estimating a demand-supply model as mentioned in Option 3 uses instrumental variables. It is unclear in this case which variables would be good instruments, and the lack of valid instruments would cause biased estimators for the RDA's causal impact. A limitation in Option 2 is that the research focuses only on relatively small RDA users close to the limit of 1800 WBSO hours. Another limitation of this option is that it investigates only the impact of the RDA on the portion of salary costs within R&D spending, but not the strength of the RDA's stimulus to companies' total R&D spending.

4.4 Certificate of Good Service (BvGD)

Objective

The objective of the Certificate of Good Service (BvGD) for municipalities is to raise entrepreneurs' level of satisfaction with municipal transactional services, which include all (simple and complex) services arising from individual contact with a municipality. To achieve this objective the instrument aims to improve both objective and perceived quality. Objective quality may be improved, among other things, by complying with set application and remedy periods, reducing response times and providing up-to-date municipal information. Perceived quality is closely related to the provision of information and the involvement of entrepreneurs in improving objective quality. Perceived quality may be increased by actively involving entrepreneurs in improvement plans and actively providing information to entrepreneurs.

Government failure

Unlike the other instruments examined in this report, the BvGD is not concerned with market failure, but with government failure. Incurring excessive municipal costs, such as for licence applications, administrative burden, incorrect or inaccessible information, when starting an activity that benefits society could reduce the availability of certain goods and services that entrepreneurs can provide, which will impede the efficient allocation of goods and services.

Operation

The BvGD instrument consists of two phases. Phase 1 is an audit of various entrepreneurs' files based on ten products selected by the municipality. The audit comprises three subaudits: of files, customers and issues. The file audit assesses files of the selected products based on quantifiable facts. The customer audit solicits information about specific experiences of companies with the services provided by the municipality. The issue audit forms a general opinion about a given subject for selected products based on information retrieved from the municipality. The entire audit embodies a set of ten standards⁵⁰ and the output is expressed in the form of a spider diagram. When the audit is complete an improvement plan is drawn up for the municipality based on the scores obtained. A municipality receives the BvGD, which is valid for two years, upon completion of the audit and the drafting of an improvement plan. The improvement plan therefore does not have to have been executed before the award.

Phase 2 is the execution of the improvement plan that was drawn up in Phase 1. A period of two years is referred to, but this is not binding. Whether and how thoroughly Phase 2 is carried out may therefore differ from one municipality to another.

A voucher scheme has been in operation since 2009 to promote use of the BvGD, through which the Ministry of Economic Affairs gives financial support to municipalities, provincial governments and district water boards for the engagement of an independent consultant to assist in the introduction. The scheme involves two vouchers on the basis of 50% minimum cofinancing by the municipality. The first voucher is for the execution of Phase 1 and has a maximum value of €5,000. A €10,000 voucher is available for Phase 2. A very small number of municipalities perform the audit and execute the improvement plan outside the voucher scheme. The voucher scheme funds go directly to the consultancies.

Implementation

The top ten licences by number of applications will normally be selected for the audit, but the ultimate choice is up to the municipalities. The licences chosen for the audit may be problem files or smoothly running files. The consultancy checks files on a sample basis with reference to fixed criteria, which consultancies tend to interpret differently in practice. For large municipalities the sample consists of approximately 200 files, and for small municipalities approximately 55.

Since early 2012 the executing organization and co-owner of the BvGD has been the Dutch Municipalities Quality Institute (KING).

⁵⁰These ten standards are: 1.) Compliance with application periods, 2.) Remedy periods for missed deadlines, 3.) Completeness of requests and applications, 4.) Technical knowledge and expertise, 5.) Perceived supervision, 6.) Accessibility of the municipality, 7.) Up-to-date municipal information, 8.) Customer satisfaction, 9.) Sound decisions, 10.) Administrative burden. This standards framework was drawn up jointly with the Confederation of Netherlands Industry and Employers (VNO-NCW), the Dutch Federation of Small and Medium-Sized Enterprises and the Association of Netherlands Municipalities (VNG).

Budget

Only the voucher scheme as part of the BvGD imposes a budgetary burden. The funds available for the scheme were €800,000 in 2012. The term of the scheme has been extended twice (in 2010 and 2011), but it will be phased out in 2012 (after which municipalities will have to bear the full costs of Phase 1 and Phase 2 themselves). Applications will still be accepted until the end of 2012, after which the scheme will be run down until the end of 2013. After 2013 the national government will no longer contribute financially. The €800,000 is to cover the entire term of the extended voucher scheme (i.e. from 30-12-2011 to 31-12-2013). When the available budget is exhausted the scheme can no longer be used.

Evaluation period

Evaluation was originally scheduled for this year (2012). For reasons of costs and response-related benefits, the evaluation was postponed for one year to allow use of the output of the 'SME-friendliest municipality of the Netherlands' survey. The evaluation covers the 2009-2012 period. The actual evaluation will now take place in 2013.

Hypotheses

Core hypothesis

- 1) The BvGD leads to greater entrepreneur satisfaction with municipal services.

Hypotheses in support of core hypothesis

- 2) The BvGD leads to better information provision to entrepreneurs about municipal services.
- 3) The BvGD involves entrepreneurs more closely with municipal services.
- 4) The BvGD improves the objective quality of municipal services.

Optional hypothesis

- 5) The BvGD has a reputation effect, which persuades municipalities with a poor entrepreneur satisfaction score to improve their performance.

Indicators

The BvGD involves ten standards, see Appendix. Three standards are based on the opinion (score) given by entrepreneurs, and two of these are reflected at least in part in the 'SME-friendliest municipality of the Netherlands' survey.

- Standard 4. Expertise of official (Professional and capable): this may link up with the score given to 'Expertise' (under Question 4 of the 'SME-friendliest municipality of the Netherlands').
- Standard 8. Customer satisfaction (Customer orientation and commitment): this may link up with the score given to 'Sympathetic to entrepreneurs' (under Question 4 of the 'SME-friendliest municipality of the Netherlands').

The working group advises the inclusion of additional questions in the 'SME-friendliest municipality of the Netherlands' survey, in order to obtain a broader picture.⁵¹ Standard 5 and Standard 7 in particular are amenable to investigation.

- Standard 5. Perception of supervision (Limited regulatory burden): e.g. new subquestion/questions in Question 4 of the 'SME-friendliest municipality of the Netherlands': score for 'Supervision of compliance with rules'.
- Standard 7. Up-to-date municipal information: e.g. new subquestion/questions in Question 4 of the 'SME-friendliest municipality of the Netherlands': score for 'Up-to-date website' (also links up with the score for 'Communication').

For the evaluation designs discussed below, it is necessary to control thoroughly for the heterogeneity between municipalities, which calls for the inclusion of background variables for each municipality in the empirical analyses. Some examples of variables to be included are: degree of urbanization, size of municipality, mean per capita income of municipality, proportion of SMEs/large companies, number of non-public sector jobs, geographical location, number of entrepreneurs per 1000 residents, and employment rate.

Statistics Netherlands has many financial figures for municipalities. For example, Statline lists fees and taxes in a variety of categories for each municipality. In principle Statistics Netherlands can supply more detailed figures than those on Statline for analyses of this kind.

⁵¹ A disadvantage is that this survey does not encompass an assessment of large companies.

Table 4 List of indicators for BvGD evaluation design

Hypotheses	Indicators		Data sources
	Dependent	Independent	
1) <i>The BvGD leads to greater entrepreneur satisfaction with municipal services.</i>	<ul style="list-style-type: none"> • Entrepreneur satisfaction by subscore on BvGD Standards 2 - 4 • [optional] Entrepreneur satisfaction on all ten BvGD standards 	<ul style="list-style-type: none"> • Dummy BvGD Yes/No 	<ul style="list-style-type: none"> • SME-friendliest municipality of the Netherlands' survey • [optional] Supplementary survey of entrepreneurs
2) <i>The BvGD leads to better information provision to entrepreneurs about municipal services.</i>	<ul style="list-style-type: none"> • Extent of information provision to entrepreneurs about municipal services 	<ul style="list-style-type: none"> • Dummy BvGD Yes/No 	<ul style="list-style-type: none"> • Supplementary survey of municipalities
3) <i>The BvGD involves entrepreneurs more closely with municipal services.</i>	<ul style="list-style-type: none"> • Extent of entrepreneur involvement with municipal services 	<ul style="list-style-type: none"> • Dummy BvGD Yes/No 	<ul style="list-style-type: none"> • Supplementary survey of municipalities
4) <i>The BvGD improves the objective quality of municipal services.</i>	<ul style="list-style-type: none"> • Performance on ten BvGD standards 	<ul style="list-style-type: none"> • Dummy BvGD Yes/No 	<ul style="list-style-type: none"> • Supplementary survey of municipalities
5) <i>The BvGD has a reputation effect, which persuades municipalities with a poor entrepreneur satisfaction score to improve their performance.</i>	<ul style="list-style-type: none"> • Entrepreneur satisfaction 	<ul style="list-style-type: none"> • Dummy BvGD Yes/No • Dummy poorly scoring municipalities Yes/No 	<ul style="list-style-type: none"> • 'SME-friendliest municipality of the Netherlands' survey • Interadministration administrative burden baseline measurement 2009 • Information from BvGD data
Control variables <i>[not exhaustive]</i>	<ul style="list-style-type: none"> • Degree of urbanization • Size of municipality • Mean per capita income of municipality • Proportion of SMEs/large companies • Number of non-public sector jobs • Geographical location, • Number of entrepreneurs per 1000 residents • Employment rate 		<ul style="list-style-type: none"> • Statistics Netherlands

Evaluation design

The non-random way in which the group of municipalities that participate in the BvGD is formed implies that there is no natural control group of municipalities with which to test the above hypotheses. The ideal evaluation design in the form of a randomized experiment cannot therefore be applied. Participating municipalities differ greatly from their non-participating counterparts, as is evident from the fact that 30 of the 36 largest municipalities participate, while the participation rate among smaller municipalities is much lower. It is therefore extremely difficult to identify comparable municipalities with and without BvGD, which hampers quantitative evaluation. A second limiting factor is the relatively modest budget available for evaluation. Several suggestions for the evaluation that take into account the limited budget are given below.

Option 1 Panel regression

Experimental and control groups

N/A.

Identification strategy

Answering the core hypothesis requires entrepreneurs' view of the municipal services to be explored. Interviewing entrepreneurs is expensive and a large-scale survey of entrepreneurs will not fall within the available budget, so that it is necessary to link with existing data sources. The most logical candidate is the 'SME-friendliest municipality of the Netherlands' survey⁵². This survey was first conducted in 2008, and for a second time in 2010/11. Preparation for the third survey to be conducted in the first half of 2013 is about to start. A disadvantage of this survey is that it yields no information about how larger companies view municipalities' services. The survey moreover does not cover all the BvGD standards. The first three hypotheses can be covered in part with these data, but this is impossible for Hypothesis 4 about objective quality.

Therefore a supplementary survey of municipalities is needed alongside 'SME-friendliest municipality of the Netherlands' to determine what municipalities have actually done (Hypotheses 2, 3 and 4). The survey must ask municipalities about the objective quality of services and whether they involved entrepreneurs or provided them with information, distinguishing between the time of the survey and the situation three years previously. These municipalities are supposed to answer the same questions twice, once for the current period and again for the situation three years previously, enabling the creation of a retrospective panel.⁵³

The costs of the supplementary survey of 100 municipalities is estimated at €40,000⁵⁴ assuming face-to-face interviews at the municipalities. A telephone, postal, or Internet survey would be far less expensive, but would introduce the risk of a lower response rate. There would also be less opportunity to go in depth into the underlying reasons and choices made (because of limited or no interaction).

Econometric specification and methods

With the exception of the additional questions in the 'SME-friendliest municipality of the Netherlands' survey of 2012-2013, this survey offers an opportunity to create a panel. The supplementary survey under municipalities also provides observations at two points in time.⁵⁵ For various reasons a panel can increase the robustness of the results, in particular because of the possibility of adjusting panel data for municipality-specific effects. This adjustment can be applied if enough municipalities start the BvGD application process within the panel observation period. If so, two intervals are

⁵²The 'SME-friendliest municipality of the Netherlands' questionnaire is administered by Lexnova. This bureau also performs the measurements for BvGD at municipalities. The Ministry of Economic Affairs must be alert to possible conflicts of interest when deciding to engage Lexnova in an evaluation.

⁵³Ideally municipalities and companies would be questioned at three points in time (1. baseline measurement, 2. BvGD received, 3. improvement plan executed (or not), or companies actively involved with or informed about the municipal services). Since results are not recorded in Phase 2, then alongside municipalities without BvGD and companies in that municipality you would also need to question municipalities with BvGD (and companies) again in Phase 2.

⁵⁴This is calculated as follows: 100 municipalities times a half day per interview (including travel time and time for reporting) times a rate for a researcher of € 800 (excluding VAT) a day.

⁵⁵The estimates based on a retrospective panel are less robust than those from a standard panel, but a retrospective panel offers more analysis facilities than a cross-section, at no additional cost.

sufficient to adjust for municipality-specific effects (see the discussion of difference-in-difference methods in Chapter 2).⁵⁶ The model to be estimated is a linear panel data model with the following specification:

$$\text{Outcome}_{it} = \alpha_i + \beta \text{BvGD}_{it} + \gamma X_{it} + \theta_t + \xi_{it}$$

The parameter in which we are interested is β , which represents the effect on the outcome variable of having a Certificate of Good Service. A separate model is estimated for each outcome variable, thereby revealing the influence of having a Certificate of Good Service on each outcome for a municipality in the ‘SME-friendliest municipality of the Netherlands’ survey and the survey of municipalities. The BvGD_{it} variable indicates whether municipality i had a Certificate of Good Service at time t . Also included are municipality-specific effects α_i , as well as calendar time effects θ_t , and adjustment is made for other observed characteristics of municipalities that are included in X_{it} . The linear panel data model must be estimated with fixed effects methods (see Chapter 2). If the panel data covers two periods, then the model reduces to the standard difference-in-difference model as discussed in Chapter 2.

In order to test Hypothesis 1, the two to four standards obtainable from the ‘SME-friendliest municipality of the Netherlands’ survey can be used as outcome variables. Hypothesis 2 is tested in the same way as the previous one, except that the dependent variable is the extent of information provision to entrepreneurs about municipal services. To test Hypothesis 3 the dependent variable is the extent of involvement of entrepreneurs with the municipal services. Hypothesis 4 is concerned with objective quality, mapped out by means of the ten standards (or at least some of them). For some standards this is a relatively easy process (e.g. throughput times, numbers of complaints and the compliance rate with application periods) because they can be obtained from a survey of municipalities. The situation is more difficult for other standards (e.g. customer satisfaction) and information can be obtained only by means of a survey of entrepreneurs.

Hypothesis 5 is concerned with poorly scoring municipalities. This hypothesis asserts that BvGD applications have a different effect for municipalities that score relatively poorly. A new variable that identifies poorly scoring municipalities is constructed to test this hypothesis. Poorly scoring municipalities are those in the lowest quartile of the quality distribution. The new variable is a dummy (poorly scoring municipalities=1, other municipalities=0). The regression model is then extended to:

$$\text{Outcome}_{it} = \alpha_i + \beta \text{BvGD}_{it} + \delta \text{BvGD}_{it} \times \text{Poor}_{it-1} + \lambda \text{Poor}_{it-1} + \gamma X_{it} + \theta_t + \xi_{it}$$

A positive (and significant) estimated interaction effect δ suggests that the BvGD indeed leads poorly scoring municipalities to greatly improve their outcomes. The receipt of a BvGD at time t then has an additional positive effect if performance was poor at $t-1$.

Reputation effects may also occur, which may be of interest to the evaluation. Reputation effects may be relevant if many municipalities within a certain region participate in BvGD. A municipality that does not participate in the BvGD in a region of this kind may come under pressure to do so, or to improve its services outside this scheme. This effect can be investigated by adding a variable to the set of background variables X that measures the proportion of BvGD-participant municipalities in the neighbourhood. This assumes that municipalities compare themselves principally with their local peers, and reputation therefore has a mainly local effect. The analysis can also seek evidence for a reputation effect of BvGD over time, whereby in later years municipalities improve their performance irrespective of their participation in the scheme.

A great risk involved in a panel data method is the relationship between the timing of a BvGD application and outcomes in previous periods (the economics literature refers to this as the Ashenfelter dip). Municipalities might decide to apply for BvGD at a time when entrepreneurs perceive their services to be good, or conversely when their services are seen as poor. In the first case the BvGD’s impact will be overestimated, and in the second there will be an underestimate. The bias is attributable to the difference in trend between municipalities that do and not apply. Investigating this phenomenon requires a study of how the timing of an application relates to earlier outcomes of the services in the municipalities.

⁵⁶This is therefore not concerned with panel data in a pure sense, but data obtained from answers to questions set at the same time regarding two points in time.

If the data cover only two periods, formally speaking the possible presence of an Ashenfelter dip cannot be investigated, but some insight can be obtained nonetheless from the following regression:

$$BvGD_{it} = \alpha + \beta Outcome_{it-1} + \gamma X_{i,t-1} + \xi_i$$

If it is significant, it may be a sign that the decision to apply for BvGD depends on earlier outcomes. However, this is not necessarily the case, because it is also possible for a BvGD application to be related to the municipality-specific effect. If the panel data cover at least three periods it will be possible to perform a more reliable test using the following regression:

$$BvGD_{it} = \alpha + \beta (Outcome_{it-1} - Outcome_{it-2}) + \gamma X_{i,t-1} + \xi_i$$

If it is now significant, it is an indication that the decision to apply for BvGD is related to the trend in the outcomes. This situation is problematic in that it is impossible unequivocally to attribute a change in the outcomes of a participating municipality to the BvGD. If municipalities with improving outcomes are more likely to apply for BvGD, then the BvGD impact will be overestimated, and conversely if mainly municipalities with deteriorating performance apply for BvGD, then an underestimate will result.

Sensitivities

As noted above, municipalities that participate in the BvGD differ from those that do not. In order to obtain a sense of the robustness of the results, it is necessary to account for self-selection as completely as possible. Other methods can be devised alongside the analyses described above to explain BvGD participation through delayed outcomes. First, the time sequence of municipalities' BvGD applications can be examined. The underlying idea is that the first municipalities to apply for BvGD are those that are already alert to the business climate, and that therefore necessarily score better than other municipalities. If so, the BvGD impact will be overestimated. In order to formalize this idea further, the econometric regression analysis could include an interaction term between dummy BvGD and the timing of application, alongside the dummy for BvGD. In the absence of self-selection effects, the interaction term will have no statistically significant effect on the dependent variable (the performance variable).

Second, use can be made of the 'Interadministration administrative burden baseline measurement 2009'⁵⁷. Although the administrative burden score relates to just one of the ten standards, this baseline measurement does give an indication of municipalities that might possibly score better. This variable should also be added to the regression function. In the absence of selection effects this variable will have no effect on the dependent variable.

The robustness of the panel data method can be improved by preselecting municipalities for the panel data analysis based on observed characteristics. The guiding principle is that municipalities exhibit the same trend in the quality of services before the introduction of the BvGD. This is ultimately the most important assumption of the analysis. This therefore means that municipalities that applied for BvGD based on the variable $(Outcome_{it-1} - Outcome_{it-2})$ must be linked with municipalities that did not.

Option 2 Survey method

At the instrument session the instrument managers expressed a wish for the current year to be evaluated. If awaiting the results of the 'SME-friendliest municipality of the Netherlands' survey is not an option, a resort will have to be made to surveying companies in municipalities both with and without BvGD. However, in the working group's opinion this would require a substantial increase in the available budget. Assuming a sample of fifty municipalities with and fifty municipalities without BvGD, and ten companies in each municipality, then 1000 companies would be involved in the survey. This would require a multiple of the budget for supplementary surveys of municipalities. A sufficiently high response rate among companies can be obtained only in a telephone survey, which is an expensive method.

⁵⁷ <http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2009/06/11/brief-aan-de-tweede-kamer-over-de-nulmeting-administratieve-lasten-interbestuurlijk.html>.

An advantage of a supplementary survey is that it also sheds light on the standards of the BvGD that are not included in the ‘SME-friendliest municipality of the Netherlands’ survey. For supplementary surveys of entrepreneurs it is also recommended to enquire about the quality of the services both now and three years previously, with a view to forming a retrospective panel. In that case all the empirical analyses discussed above may be applied.

This method allows municipalities to be presented with conditional questions. For example, ‘to what extent would you have improved your municipality’s services for entrepreneurs without applying for a BvGD?’ Conversely, entrepreneurs can be asked how much their municipality’s services improved compared with before it received a BvGD. The sample would have to be small in view of the limited budget.

Conclusion and recommendation

The selectivity involved in municipalities’ participation in BvGD makes it impossible to assemble a satisfactory control group to test the hypotheses. The ideal evaluation design of a randomized experiment is infeasible, so that additional assumptions must be made in establishing causality between instrument and observed result. A frequently used fallback option is the survey method with conditional questions (Option 2). However, this option carries the risk of socially desirable answering, which is detrimental to the reliability of the outcomes.

The working group therefore recommends a compromise of performing an econometric analysis of survey results (Option 1). The analysis requires the collection of as many retrospective data as possible in order to control for existing differences between municipalities. The extent of and reason for self-selection must also be examined. The output of this analysis may give an impression of the effectiveness of the BvGD, but the reliability of the results depends on the nature of the self-selection of BvGD applicants. This method allows the testing of all five hypotheses, subject to a supplementary survey of municipalities alongside the ‘SME-friendliest municipality of the Netherlands’ survey.

The limited budget currently available for the evaluation conflicts with the need for evaluation in the short term. This limits the opportunities for gaining insight into the instrument’s effect. The working group recommends increasing the budget for the evaluation if there is a need to gain an impression of the functioning of the BvGD sooner, or if there is a need to evaluate the BvGD on all ten standards. If supplementary surveying is opted for, Options 1 and 2 may be performed in parallel.

Appendix Standards for Certificate of Good Service

I Periods	Minimum standard	Plus standard
1. Compliance with application periods > Clear handling times	100% of applications granted within the legal application period	70% of applications granted faster than the legal application period
2. Remedy period for missed deadline > Anticipation of missed deadline	60% of missed deadlines are remedied within 5 working days	80% of missed deadlines are remedied within 5 working days
3. Completeness of requests and applications > Admissibility of first application.	85% of applications are admissible on first submission	95% of applications are admissible on first submission

II Professionalism	Minimum standard	Plus standard
4. Expertise of official > Professional and capable	7	9
5. Perception of supervision > Limited regulatory burden	7	9

III Entrepreneur orientation	Minimum standard	Plus standard
6. Speed of response > Time between confirmation of receipt and the substantive answer from the municipality	Mean response time is 3 days	Mean response time is 1 day
7. Up-to-date municipal information	90% of information on the website and in documentation is up-to-date	100% of information on the website and in documentation is up-to-date
8. Customer satisfaction > Customer orientation and commitment	7	9

IV Reliability	Minimum standard	Plus standard
9. Sound decisions	10% of objection and appeal procedures are well founded	5% of objection and appeal procedures are well founded
10. Administrative burden for entrepreneurs > Limited administrative burden	25% lower than national mean administrative burden	35% lower than national mean administrative burden

4.5 Centres for Innovative Craftsmanship (CIV)

Objective

The aim with Centres for Innovative Craftsmanship (CIV) is to increase the number of newly qualified engineers and technicians at upper secondary vocational education level in order to satisfy demand from industry in the top sectors, to stimulate greater involvement of companies in practical assignments, and to increase the number of enrolments of top industry sector employees on CIV courses. The underlying idea is to reduce the gap between supply and demand on the labour market for engineers and technicians.

Market failure

Educational institutions are currently given insufficient incentive to provide educational programmes that match the needs of companies, leading to a suboptimum allocation of workers and a suboptimum supply of (refresher) educational programmes. The lack of incentive for a satisfactory match between educational programmes and jobs may be attributable to information asymmetry between educational institutions and companies. CIV serves to reduce the information asymmetry between educational institutions and companies by informing both educational institutions and students at an earlier stage about changes in the demand for labour, thereby improving labour market performance and spending government funds more efficiently.

Operation

CIV focuses on new forms of education and on raising the quality of mainstream education through the active involvement of industry in the curriculums, such as in work placements, practical assignments and education in the workplace. CIVs are created as public-private partnerships (PPPs)⁵⁸ and therefore reside outside school and industry. They are usually linked to educational institutions.

The educational institutions draw up business plans with the active involvement of a network of top sector companies in the region. The best proposals for national government financial support are selected in a tender. Subsidies are provided on the basis of cofinancing. The national government requires at least 50% cofinancing of the investment costs for a CIV. The national government puts up €2 million over a five-year period (€400,000 a year). A CIV has to be able to continue independently after this five-year period. The other half must come from the educational institution (25%) and affiliated companies (25%). There is therefore a total of at least €4 million available for a CIV over a five-year period. The size of the group of companies that invests in a CIV (the primary partners) is usually between six and twenty.

With sufficient willingness of companies and local authorities to provide finance, upper secondary vocational education institutions without a CIV subsidy may nonetheless attempt to achieve their business case by attracting finance from elsewhere, and the more so if the case is assessed as good.

Implementation

Each top sector is eligible for one CIV. Funds earmarked for one top sector can be diverted to another if the original applicant submits plans of insufficient quality on two separate occasions.

Incoming proposals in the form of business cases are selected and ranked on the basis of a pitch before a selection committee. Proposals will be selected in order of ranking in the event of insufficient budget to honour all proposals. The assessment observes the following criteria.

- Constraints: the centre establishes a profile on and within the top sector, there is equal commitment from relevant institutions and other partners (borne out by cost-benefit analyses), cofinancing is arranged, and the centre has access to scientific knowledge and input.
- Core criteria: ambition, partners, feasibility and sustainability. Ultimately the plan's ambition is weighed against its clarity of definition and feasibility.

The selection process involves two rounds, where an expert committee first assesses compliance with the constraints and then proceeds to assess the business plan. Non-compliance with any of the constraints automatically entails an unsatisfactory assessment of the application.

⁵⁸For example in the form of a foundation (stichting), private limited company (BV) or a (cooperative) association.

Applications that are compliant with all constraints are assessed on each of the four core criteria on a ten-point scale (giving a maximum of 40 points). Applications scoring between 0 and 3 points on any of the core criteria are automatically assessed as unsatisfactory.

The National Platform Science & Technology puts out the tender, grants the successful proposals and monitors them from the time of granting by tracking performance agreements, and through audit committee visits, joint workshops and cooperation and exchange events.

The subsidy's continuity depends on interim results as monitored after two years on the basis of performance contracts agreed with the National Platform Science & Technology (PBT). In the event of unsatisfactory results after two years the subsidy will be withdrawn for the remaining three years.

Budget

For the centres on upper secondary vocational level the Ministry of Education, Culture and Science has set aside a budget of €11 million for six CIVs. The Ministry of Economic Affairs also provides a budget for two centres in the green sectors, but these funds have their own procedure and process separate from the other CIVs. For this reason, the evaluation designs below show only the CIVs that are assessed and supervised by PBT.

Evaluation period

The CIV evaluation is scheduled for 2017. The evaluation will cover the 2012-2017 period.

Hypotheses

Core hypothesis

- 1) CIV leads to more newly qualified engineers and technicians at upper secondary vocational educational level destined for the top sectors (quantitative effect).

Hypothesis in support of core hypothesis

- 2) Engineers and technicians at upper secondary vocational education level who gained a qualification through a CIV find work more quickly after their course than other engineers and technicians at that level who did not do so (qualitative effect).
- 3) Engineers and technicians at upper secondary vocational education level who gained a qualification through a CIV earn higher salaries after their course than other engineers and technicians at that level who did not do so (qualitative effect).

The CIV subsidy may prompt other upper secondary vocational education institutions with no CIV to embark on similar initiatives. The measurement of this effect requires the testing of Hypotheses 1, 2 and 3 also for project plans that were assessed as promising in the tender but were not honoured (Hypotheses 4, 5, 6 & 7):

Hypotheses about the CIV's boosting effect

- 4) Project plans that are assessed as promising lead to more newly qualified engineers and technicians at upper secondary vocational educational level destined for the top sectors (quantitative effect).
- 5) Engineers and technicians at upper secondary vocational education level who gained a qualification from an institution whose project plan was assessed as promising find work more quickly after their course than other engineers and technicians at that level who did not do so (qualitative effect).
- 6) Engineers and technicians at upper secondary vocational education level who gained a qualification from an institution whose project plan was assessed as promising earn higher salaries after their course than other engineers and technicians at that level who did not do so (qualitative effect).

Optional hypotheses

- 7) CIVs and/or institutions whose project plan was assessed as promising provide more refresher courses for engineers and technicians.

Data availability

A link must be made between Ministry of Education, Culture and Science/Education Executive Agency data, and policy data (payroll tax declaration). This linking is possible through the Statistics Netherlands Social Statistics Database.

Table 5 List of indicators for CIV evaluation design

Hypotheses	Indicators		Data sources
	Dependent	Independent	
1) CIV leads to more newly qualified engineers and technicians at upper secondary vocational educational level destined for the top sectors (quantitative effect).	<ul style="list-style-type: none"> Ex-student is employed in a top sector 	<ul style="list-style-type: none"> Indicator CIV Yes/No 	<ul style="list-style-type: none"> Ministry of Education, Culture and Science/Education Executive Agency Social Statistics Database
2) Engineers and technicians at upper secondary vocational education level who gained a qualification through a CIV find work more quickly after their course than other engineers and technicians at that level who did not do so (qualitative effect).	<ul style="list-style-type: none"> Elapsed time between certificate award and entering employment 	<ul style="list-style-type: none"> Indicator CIV Yes/No 	<ul style="list-style-type: none"> Ministry of Education, Culture and Science/Education Executive Agency Social Statistics Database
3) Engineers and technicians at upper secondary vocational education level who gained a qualification through a CIV earn higher salaries after their course than other engineers and technicians at that level who did not do so (qualitative effect).	<ul style="list-style-type: none"> Starting salary of newly qualified student 	<ul style="list-style-type: none"> Indicator CIV Yes/No 	<ul style="list-style-type: none"> Ministry of Education, Culture and Science/Education Executive Agency Social Statistics Database
4) Project plans that are assessed as promising lead to more newly qualified engineers and technicians at upper secondary vocational educational level destined for the top sectors (quantitative effect).	<ul style="list-style-type: none"> Ex-student is employed in a top sector 	<ul style="list-style-type: none"> Indicator CIV Yes/No Indicator CIV rejected 	<ul style="list-style-type: none"> Ministry of Education, Culture and Science/Education Executive Agency Social Statistics Database
5) Engineers and technicians at upper secondary vocational education level who gained a qualification from an institution whose project plan was assessed as promising find work more quickly after their course than other engineers and technicians at that level who did not do so (qualitative effect).	<ul style="list-style-type: none"> Elapsed time between certificate award and entering employment 	<ul style="list-style-type: none"> Indicator CIV Yes/No Indicator CIV rejected 	<ul style="list-style-type: none"> Ministry of Education, Culture and Science/Education Executive Agency Social Statistics Database
6) Engineers and technicians at upper secondary vocational education level who gained a qualification from an institution whose project plan was assessed as promising earn higher salaries after their course than other engineers and technicians at that level who did not do so (qualitative effect).	<ul style="list-style-type: none"> Starting salary of newly qualified student 	<ul style="list-style-type: none"> Indicator CIV Yes/No Indicator CIV rejected 	<ul style="list-style-type: none"> Ministry of Education, Culture and Science/Education Executive Agency Social Statistics Database
7) CIVs and/or institutions whose project plan was assessed as promising provide more refresher courses for engineers and technicians.	<ul style="list-style-type: none"> Indicator for refresher courses 	<ul style="list-style-type: none"> Indicator CIV Yes/No Indicator CIV rejected 	<ul style="list-style-type: none"> survey
Control variables [not exhaustive]	<ul style="list-style-type: none"> Indicator variables for programme Indicator variables for Regional Training Centre Observable characteristics of ex-student 		<ul style="list-style-type: none"> Statistics Netherlands Ministry of Education, Culture and Science/Education Executive Agency

Data are available for participants/successful students of upper secondary vocational education institutions, by institution, course subject, etc. for the 2005-2010 period. Various job and social benefit databases support the investigation on person level of the career progress of ex-students on the labour market. The job data concerned are largely based on the payroll tax declaration. Along with a job, the Dutch standard industrial classification (SBI)⁵⁹ and the size category of the employer, for example, may also be included. A person's self-employed status can also be derived from the Social Statistics Database. The files concerned have data about the duration of the job/social benefit/self-employment and salary, but not about the nature of the work (in particular whether it is consistent with the course of study – which is mainly a point for attention if the employer is a large company with a different core activity (standard industrial classification) but where the person concerned is nonetheless employed in the field in which he or she qualified).

The registration numbers (BRINs) of the educational institutions and the Chamber of Commerce file numbers of the companies must be provided to enable the linking of Ministry of Education, Culture and Science/Education Executive Agency data with the Social Statistics Database. Companies can be classified in top sectors through the Dutch standard industrial classification (SBI). Educational programmes can be classified in top sectors through the Dutch standard educational classification (SOI).⁶⁰

Indicators

Important variables on the level of ex-students in experimental and control group(s) are:

- certificate award date and commencement date of first job;
- data about career path: successive job durations, unemployment hours;
- data about job income over time;
- sector in which successive jobs are found.

These variables are probably available from the Social Statistics Database. Hypotheses 1, 2, 3, 4, 5 and 6 can be tested on the basis of a comparison of the results for the various groups (given sufficient numbers). The long-term career paths of the CIV and the Regional Training Centre non-CIV ex-students can also be tracked in the Social Statistics Database, and it is possible to investigate the scale of the differences in labour mobility between top sectors and other (technical) sectors. The size of the spillover effect of the CIV technical educational programmes may then be investigated for the entire technical labour market. Information on the level of the Regional Training Centres is needed for testing Hypothesis 7, probably in a separate investigation.

Evaluation design

Two complementary evaluation designs are needed for testing the hypotheses. The first design (Option 1) can be used to test the core hypothesis. This design compares educational programme performance. The objective of the second design (Option 2) is to verify the supporting hypotheses, and involves comparing ex-students with each other.

Option 1 Comparison of educational programmes

Experimental and control groups

The statistical analysis in the first option is performed on the level of the educational programme. A distinction is made between educational programmes with and without a CIV. Furthermore, the performance of an educational programme before and after the introduction of a CIV is compared. This option is intended primarily for testing the core hypothesis and Hypothesis 4, but can also be used for the other hypotheses.

Identification strategy

The impact of a CIV can be determined by investigating whether a change in the performance trend of the educational programmes occurred after its introduction. The performance trend of the educational programmes without CIV serves as reference. This approach is an application of the difference-in-difference method discussed in Chapter 2.

⁵⁹The standard industrial classification (SBI) is a hierarchy of economic activities. Statistics Netherlands uses the standard industrial classification to classify business units according to their main activity. See Statistics Netherlands for additional information.

⁶⁰The Dutch Standard Educational Classification (SOI) categorizes educational programmes according to level and subject. The SOI was developed for use in statistics and research and for administrative purposes in the Netherlands. See Statistics Netherlands for additional information.

Econometric specification and methods

The dependent variable that measures the result of the educational programme (for the core hypothesis this is the proportion of students on an educational programme that find work in a top sector) must be regressed on an indicator of whether the educational programme j in region r has a CIV at time t . The estimated coefficient of the indicator is a measure of the causal effect of CIV on the dependent variable.

An adjustment may be made in three dimensions for (unobserved) effects that may bias the estimate of the CIV's influence. The first possible adjustment is for changes over time that affect all educational programmes. Then, if there are enough observations, (unobserved) properties of educational programmes that do not change over time can be controlled for. A third dimension is the region in which a programme is given. The regression function is then:

$$Y_{jrt} = \beta_0 + \beta_1 CIV_{jrt} + \eta_j + \rho_r + \theta_t + u_{jrt}$$

In this equation Y is the performance of the programme, CIV indicates the presence of a CIV and there are fixed effects for educational programme η , region ρ and time θ . The β is the coefficient to be estimated and u is the residue.

Sensitivities

It is interesting to perform a further comparison with the other groups. For instance, it is useful to extend the analysis with information about educational programmes for which the CIV was applied for, but had yet to be allocated. It is also possible to examine educational programmes without CIV given at a Regional Training Centre that does have a CIV for a different programme, to determine any differences within a Regional Training Centre between educational programmes with and without CIV.

Option 2 Comparison of ex-students

Experimental and control groups

For testing the supporting hypotheses (Hypotheses 2 and 3) the instrument can be evaluated on ex-student level. Three primary groups can be identified: 1) ex-students of educational programmes for which an institution received a subsidy; 2) ex-students of educational programmes for which a subsidy was applied for but not allocated, but for which alternative financing was obtained; 3) ex-students of educational programmes for which a subsidy was applied for but not allocated, and for which no alternative financing was obtained. This option is also applicable to Hypotheses 5 and 6.

Identification strategy

A comparison is made between a group of individual students who attended a CIV and individual students who attended comparable educational programmes that were not at a CIV to determine any differences in later labour market performance between comparable groups of ex-students from the two educational programmes. There are various options for the comparison groups: a) ex-students from the same course subject at a different Regional Training Centre; b) ex-students of earlier cohorts of the same educational programme (in the period before CIV started); c) ex-students of other educational programmes within the Regional Training Centre that did not obtain a CIV.

Econometric specification and methods

Two econometric strategies may be used. In the first strategy the dependent variable is the labour market performance of ex-student i of educational programme j at Regional Training Centre s with t being the time of entering the labour market. This variable can be regressed on an indicator for completing a programme through a CIV. Because most students complete only one programme, this does not constitute a panel analysis but a cross-section analysis. Account must be taken of (unobserved) differences between educational programmes, Regional Training Centres and student cohorts. The regional scope of Regional Training Centres means that no explicit account need be taken of regions in this approach. Furthermore, as many observable characteristics of the ex-student as possible must be included. The regression function is then:

$$Y_{ijst} = \beta_0 + \beta_1 CIV_{ijst} + \gamma X_{ijst} + \eta_j + \sigma_s + \theta_t + u_{ijst}$$

In this equation Y is a performance indicator for the ex-student, CIV is an indicator of whether the student was at a CIV, X is a series of control variables of student characteristics and there are fixed effects for educational programme η , Regional Training Centre σ and time θ . The β and γ are the coefficients to be estimated and u is the residue.

The second strategy estimates an econometric duration model in which the dependent variable is based on the interval between the time of completing the programme and the time of first entering employment. The regressors are identical to those for the first strategy. Some students will already have a job when they complete the programme, so that adjustment for censoring is required in estimating a duration model. The proportion of ex-students with a job immediately after completing a programme can also act as a dependent variable in a probit or logit regression. This regression could also be combined with the duration model.

Sensitivities

A problem with this identification strategy is the impossibility of controlling for unobserved differences between ex-students, because they are each linked with only one educational programme. In other words, the regression function cannot include fixed effects for the individual. For this reason it is important to adjust as completely as possible for the individuals' observable background characteristics.

Option 3 Survey of ex-students, educational institutions and companies

Students who complete an educational programme at a CIV and those who do so at a mainstream upper secondary vocational school may both be asked how quickly they found employment and about their perceptions of how well their qualification matches the labour market. Upper secondary vocational schools can be asked whether they would have undertaken similar actions to improve the match with industry in the absence of a CIV subsidy. Companies can be asked whether CIVs help them find better trained employees.

Conclusions and recommendation

The hypotheses about the CIV's impact can be tested on the level of educational programmes and on the level of ex-students. Educational programmes with a CIV and their ex-students can be compared with their non-CIV counterparts. The educational programmes for which a CIV grant application was rejected can be categorized further into those for which alternative financing was or was not found.

Option 1 can be used to test the core hypothesis ('CIV leads to more newly qualified engineers and technicians at upper secondary vocational educational level destined for the top sectors') and Hypothesis 4 ('Project plans that are assessed as promising lead to more newly qualified engineers and technicians at upper secondary vocational educational level destined for the top sectors'). Option 2, testing on educational level, is more accurate than Option 1 because better account can be taken of differences between students. The ability to observe the number of students in an educational programme before any certificates are awarded means that the first strategy (timing of intake) has the advantage of determining CIV effectiveness earlier than with the second strategy (duration model). The option to be preferred depends – for Hypotheses 2, 3, 5 and 6 – on the time of evaluation, among other factors. A supplementary survey of Regional Training Centres will be needed for testing Hypothesis 7.

4.6 Innovation Performance Contracts (IPC)

Objective

The objective of the Innovation Performance Contracts (IPC) scheme is to encourage SMEs to collaborate more on long-term innovation.

Market failure

The SMEs that the IPC targets are too often caught up in what is currently 'all the rage', are insufficiently externally oriented, and lack the necessary organizational capacity and information (who can I trust as a development partner?) to enter into innovation alliances for joint risk and account. Both coordination failure and information asymmetry are therefore present. Regarding innovative projects there are also potential knowledge spillover effects when projects with a low private return would not materialize without the incentive of IPC.

Operation

IPC targets the 'appliers' and 'followers'⁶¹ SME segment, which is a group that engages in innovation only occasionally, if at all. In order to persuade this group to collaborate on innovation nonetheless, the IPC scheme offers a subsidy with no repayment obligation, but with compulsory cofinancing. The idea behind this subsidy is the overriding importance of teamwork for innovation.

IPC has three components, which in order of deployed funds per component are: 1) IPC projects; 2) funds for exploring industry-transcending collaboration; 3) funds for exploring international collaboration.

For IPC projects, NL Agency puts out one or two tenders a year, inviting groups of between ten and twenty companies (with or without a knowledge institute) to submit joint proposals for two-year innovation projects. A leading party (frequently a trade association) supervises the entrepreneurs and applies for the subsidy. Syntens may also be requested to provide support, which is free of charge. Companies within the IPC application are interconnected by themes (collaborating companies from around the entire country) or on a regional basis (companies from the same region) and there are indirect links between the subprojects. Participants work together in smaller teams within an IPC. Each entrepreneur is required to collaborate with at least one other company in the IPC. IPC projects are concerned with product or production process innovation that is new for the entrepreneur. Besides industrial innovation, IPC is also open to service innovation.

The subsidy covers 40% of the project expenses incurred, to a maximum of €25,000 per entrepreneur. The leading party receives an allowance of €3,000 per entrepreneur. The subsidy can be used for project expenses such as materials, wage costs for work performed on the IPC project, hiring externals (from outside the IPC consortium), and so on. Internal operating expenses, such as travel expenses, accommodation and management overhead, are ineligible.

Components 2 and 3 are concerned with exploring collaboration opportunities, both nationally and internationally. A leading party investigates opportunities for potential alliances of SMEs with one or more companies or public knowledge institutes within one or more top sectors, or investigates the opportunities for collaborating with a similar organization or trade association abroad, on joint research of benefit to the entire industry. The limited budgetary burden of these two IPC components has led the working group to dwell in particular on Part 1 in the evaluation design.

Implementation

The description below refers only to IPC projects. A 'first-come, first-served' approach is applied to the budget for the other two IPC components.

Current practice

In the current (changed) situation NL Agency ranks applications for IPC projects according to three criteria: innovation, collaboration (including with knowledge institutes), and match with a top sector. The weighting of the factors is 40/30/30 per cent respectively. NL Agency then accepts the highest ranking applications to the extent permitted by the budget for the tender. The actual score for the criteria is subordinate to the ranking. There is no threshold in the form of a minimum score. Funds are distributed in order of ranking for as long as they are available. The more compliant with a criterion, the

⁶¹ However, this is not an explicit condition, 'front runners' are not necessarily excluded.

higher the score. The minimum requirement is 20% collaboration.⁶² The criterion for innovation is 'new for the company'. If it is 'completely new', the score on this criterion will be higher. Approximately two-thirds of the fifty applications submitted by 800 companies were accepted for the first tranche in 2012. The value of applications submitted in response to the tender in 2011 was €45 million against a budget of €15 million.

Development of design

The IPC selection procedure has changed several times over the years. The IPC budget was initially distributed on a first-come, first-served basis (2008 and 2009). In 2008 the budget was completely exhausted, and in 2009 and 2010 the large number of applications necessitated a randomized selection to avoid a budget overrun from day one. Later a generic tender was applied, initially with seven criteria (2011 and early 2012), and later in accordance with the methodology described above. The seven criteria have been compressed into two criteria, and the new criterion 'link with top sector' added. The original criteria are therefore still taken into consideration. For example, 'degree of teamwork' and 'collaboration with a public knowledge institute' were initially separate criteria, but have now been combined in the criterion 'quality of collaboration'.

In 2013 part of the IPC budget will be used for a multiyear SME reinforcement plan in the top sectors. The legal basis for this scheme is the new SME Top Sectors regulation, which is additional to the TKI supplement, which TKI companies apply for. More than €7 million continues to be available for the IPC scheme. These remaining IPC funds are made available for IPC projects through a tender (one each year). Ranking is based on the 'innovation' and 'collaboration' criteria. The 'link with a top sector' criterion has been removed, because SMEs in the top sectors can now apply for the new SME Top Sectors scheme. Exploration of industry-transcending collaboration will also be removed from the IPC in 2013.

Box 4.6.1 Recommended future IPC design

The working group was asked to recommend a future design for the instrument. The optimum from the perspective of impact measurement is randomized selection, because having a randomized group simplifies robust impact measurement. However, a tender is preferable from a policy perspective. In order to bridge the gap between the two interests the working group recommends the following selection process in future.

- Organize a tender to produce two groups: the rejected group and the potentially accepted group. Apply randomization within the potentially accepted group for the purpose of impact measurement.
- Weighted randomization may be chosen in order to reduce the 'unfairness' of randomization and preserve the incentive to deliver sound plans, in which the proposals assessed as best in the tender have a greater chance of success. The weighting probabilities in the randomization then form the propensity score.

Budget

The budgetary burden of IPC in 2012 is €41.2 million. In 2013 the budget will decrease to a little more than €7 million, because of the €15 million earmarked for the new SME Top Sectors scheme. Past budgets have always been exhausted.

Evaluation period

The IPC scheme will be evaluated in 2015.

Box 4.6.2 The working group's additional observations on IPC evaluation options

The design of the IPC scheme has changed frequently, which creates opportunities from the point of view of evaluation. For example, it is possible to ascertain which distribution mechanism works best. Tenders have both advantages and disadvantages. On the one hand they compel parties to submit better proposals, while on the other hand they lead to an expensive implementation process. While randomization is inexpensive in execution, it gives no incentive to submit high quality proposals. Supplementary investigation using data for the entire IPC term would allow analysis of the added value of having an IPC application assessed by NL Agency.

⁶²Every IPC participant draws up their own budget for their IPC project. At least 20% of this budget (in hours or money) must be reserved for collaboration with other companies.

Hypotheses

The hypotheses distinguish between accepted and rejected IPC applications in order to differentiate between the effect of the received subsidy and that of the application procedure.

Core hypotheses

- 1a) A granted IPC application leads to more and longer-term collaboration between the companies involved compared with companies whose IPC application was rejected.
- 1b) A rejected IPC application leads to more and longer-term collaboration between the companies involved compared with companies that did not apply for IPC.

Hypotheses regarding second-order and third-order effects

- 2a) A granted IPC application leads to more innovation on the part of the companies involved compared with companies whose IPC application was rejected.
- 2b) A rejected IPC application leads to more innovation between the companies involved compared with companies that did not apply for IPC.
- 3a) Companies involved in a granted IPC application grow faster and have higher productivity than companies whose IPC application was rejected.
- 3b) Companies involved in a rejected IPC application grow faster and have higher productivity than companies that did not apply for IPC.

Hypothesis for testing the selection procedure

- 4) A tender leads to higher effectiveness of the IPC scheme than randomization (more promising projects because of the competition element).

Optional hypothesis for testing the effect of exploration

- 5) IPC applications by means of exploration score better than applications without this exploration.

Indicators

1. Indicators for collaboration between companies. These indicators may have to be the subject of a survey.
2. (Labour) productivity per company
3. R&D spending per company per year
4. Number of WBSO hours per company per year
5. Indicators for innovation per company: new products, processes or patent applications
6. Other company characteristics: revenue, number of employees, size capital, age, etc.
7. Top sector by company

Option 1: (Weighted) randomization

Experimental and control groups

There was randomized allocation of IPCs in 2009 and 2010. With randomization the companies that are drawn go into the experimental group, and the others into the control group. If the randomization is weighted, then observations are weighted with the probability of selection. A regression analysis could be performed to verify the randomness of the allocation. This option is applicable to Hypotheses 1a, 2a and 3a.

Identification strategy

The effect of IPCs on the target variables follows directly from the difference between the (weighted) mean of the experimental group and the (weighted) mean of the control group.

Econometric specification and methods

The randomized allocation to the experimental and control groups means that an initial comparison of the means of the target variable of the two groups already gives a satisfactory estimate of the effect of IPCs on target variables. In terms of regression analysis this means a regression of the target variable on an indicator for IPC, where conditioning is possible on company characteristics such as revenue, number of employees and age. However, these variables should have no effect,

Table 6 List of indicators for IPC evaluation design

Hypotheses	Indicators		Data sources
	<i>Dependent</i>	<i>Independent</i>	
1a) A granted IPC application leads to more and longer-term collaboration between the companies involved compared with companies whose IPC application was rejected.	<ul style="list-style-type: none"> Degree of collaboration 	<ul style="list-style-type: none"> Dummy IPC Yes/No 	<ul style="list-style-type: none"> Survey NL Agency
1b) A rejected IPC application leads to more and longer-term collaboration between the companies involved compared with companies that did not apply for IPC.	<ul style="list-style-type: none"> Degree of collaboration 	<ul style="list-style-type: none"> Dummy IPC Yes/No Dummy rejected IPC Yes/No 	<ul style="list-style-type: none"> Survey NL Agency
2a) A granted IPC application leads to more innovation on the part of the companies involved compared with companies whose IPC application was rejected.	<ul style="list-style-type: none"> New products & processes Patents R&D spending by company 	<ul style="list-style-type: none"> Dummy IPC Yes/No 	<ul style="list-style-type: none"> NL Agency WBSO data Netherlands Patent Office Statistics Netherlands
2b) A rejected IPC application leads to more innovation between the companies involved compared with companies that did not apply for IPC.	<ul style="list-style-type: none"> New products & processes Patents R&D spending by company 	<ul style="list-style-type: none"> Dummy IPC Yes/No Dummy rejected IPC Yes/No 	<ul style="list-style-type: none"> NL Agency WBSO data Netherlands Patent Office Statistics Netherlands
3a) Companies involved in a granted IPC application grow faster and have higher productivity than companies whose IPC application was rejected.	<ul style="list-style-type: none"> (Labour) productivity Growth of companies 	<ul style="list-style-type: none"> Dummy IPC Yes/No 	<ul style="list-style-type: none"> NL Agency WBSO data Netherlands Patent Office Statistics Netherlands
3b) Companies involved in a rejected IPC application grow faster and have higher productivity than companies that did not apply for IPC.	<ul style="list-style-type: none"> (Labour) productivity Growth of companies 	<ul style="list-style-type: none"> Dummy IPC Yes/No Dummy rejected IPC Yes/No 	<ul style="list-style-type: none"> NL Agency WBSO data Netherlands Patent Office Statistics Netherlands
4) A tender leads to higher effectiveness of the IPC scheme than randomization (more promising projects because of the competition element).	<ul style="list-style-type: none"> Degree of promise of project 	<ul style="list-style-type: none"> Dummy tender Yes/No 	<ul style="list-style-type: none"> NL Agency
Control variables	<ul style="list-style-type: none"> WBSO hours by company Company characteristics (revenue, number of employees, age of company, etc.) 		<ul style="list-style-type: none"> NL Agency Statistics Netherlands

because the randomized allocation of companies to the experimental and control groups means they are strongly comparable. OLS can be used by regressing the target variable on an indicator for IPC and all other background characteristics.

This could possibly be specified as a difference-in-difference function, in which information prior to the IPC allocation can be used to arrive at a trend comparison of the control group with the intervention group. However, the randomization means that this is not strictly necessary.

Sensitivities

Anticipation: an announced unweighted randomization removes the incentive for companies to refrain from submitting poor proposals. A weighted randomization will reduce this anticipation effect.

Option 2: Allocation in order of application

Experimental and control groups

In 2008 IPCs were allocated in the order in which applications were submitted until the budget was exhausted. In this case the experimental group comprises companies that were allocated IPC. The control group consists of rejected companies. Both the experimental and control groups can be limited to applications made around the date on which the budget ran out. This option is applicable to Hypotheses 1a, 2a and 3a.

Identification strategy

Identification is based on the idea that the two groups of companies are strongly comparable, except that some companies were allocated an IPC and other comparable companies were not. It is possible for companies to have certain expectations about the date on which the budget will become exhausted, as a result of which applications submitted early in the year may differ systematically from those submitted later in the year. Early applications might possibly be of higher quality than later applications because they are produced by an efficient group of companies. There are other conceivable reasons for differences between early and later applications.

In order to limit companies' expectations about the closing date, the experimental group and the control group could be restricted to applications close to the closing date. This method is described in Chapter 2 as regression discontinuity.

Econometric specification and methods

Comparing the mean of the experimental group with the mean of the control group already gives a satisfactory estimate of the effect of IPCs on target variables. There is a trade-off to consider in the selection of the experimental and control groups: the effect of expectations diminishes the closer the sample is to the closing date, but the number of observations, and consequently the accuracy of the estimated effect, decrease accordingly. It is possible to condition on company characteristics such as revenue, number of employees and age, but these variables should actually have no effect because the companies should be strongly comparable. If these variables do have some effect, then it is questionable whether regression discontinuity is actually a suitable estimation method for the specific problem.

Sensitivities

Selection: strong expectations about the closing date among at least some of the IPC applicants may lead to overestimating or underestimating the effect of IPC. Furthermore, the companies before and after allocation may be strongly heterogeneous. If the companies before the closing date are more (less) productive, then this will lead to an overestimate (underestimate) of the effect of IPC.

Option 3: Ranking

Experimental and control groups

NL Agency assessed the IPC applications in 2011 and 2012. The experimental group consists of accepted applications with a relatively low (but satisfactory) assessment, and the control group consists of rejected applications with a relatively good (but unsatisfactory) assessment. This option is applicable to Hypotheses 1a, 2a and 3a.

Identification strategy

Accepted and rejected applications are compared with each other in order to estimate the effect of IPCs on target variables. The NL Agency assessment ensures that accepted applications are of higher quality than rejected applications. In order to minimize the difference in quality between the experimental and control groups the sample is limited to applications assessed as just satisfactory or just unsatisfactory. This method was described in Chapter 2 as regression discontinuity, with local impact measurement.

An alternative identification strategy is to use the information about allocation and to apply a propensity score method. Various allocation criteria can be used for 2011 and 2012 to estimate the propensity score. However, it is unclear in this case whether the criteria are sufficiently distinct to estimate the propensities with the required accuracy.

Econometric specification and methods

There is a trade-off to consider in the selection of the experimental and control groups with regression discontinuity: the effect of quality differences diminishes the closer the observation is to the minimum assessment (a score of 5.5), but the number of observations, and consequently the accuracy of the estimated effect, decrease accordingly. There must be no substantial difference in the other observed variables between the accepted and rejected companies either side of the selection discontinuity. Conditioning on revenue, number of employees, age, top sector and other non-target variables may be performed to further limit the differences in characteristics between applications in the experimental and control groups. However, a disadvantage of this conditioning is that inclusion of the variables reduces the precision of the estimates (the t values). These variables should be included only if they have a statistically significant effect. NL Agency's final assessment of an application could be binary, which is to say either 'satisfactory' or 'unsatisfactory', in which case an additional step will be required to select applications that were just allocated or just not allocated. It may be possible to reconstruct a 'score' on the basis of subcriteria and their weights.

With the propensity score the probability of allocation (relative to a rejected application) can be estimated with reference to scores on subcriteria.

Sensitivities

It must be verified that enough observations are available for comparable companies around the minimum assessment.

Option 4: Meta-option selection methods: Dummies for selection methods

Experimental and control groups

Hypothesis 4 asks whether a tender is more effective than randomization. Investigating this hypothesis involves comparing the effectiveness of IPCs between 2009-2010 and 2011-2012. Randomization was used in 2009-2010 whereas allocation by ranking was used in 2011-2012. The group consists of the allocated or rejected applications. In principle the propensities are determined as described for Option 3. Subsequently the estimated propensities are gathered into a database of all applications for 2009-2010 and 2011-2012 together.

Identification strategy

Variation of the propensities over the two periods is estimated by including an indicator variable for periods in the propensity score, allowing an assessment of the pattern over the periods. The underlying idea is to compare companies over the years, whereby companies allocated through randomization are compared with companies allocated through a tender. Companies rejected through randomization can also be compared with those rejected through a tender.

Econometric specification and methods

The propensity scores are estimated in the first round. Then in the second round the propensity scores are compared by including an indicator variable for application round (randomization/tender).

Sensitivities

The possibility of determining propensities depends on the availability of comparable scores on subcriteria in the various application rounds (with randomization and tenders).

Option 5: Meta-option selection methods: Differentiation of selection methods

Experimental and control groups

This option involves comparing the selected applications for 2011 and 2012 while disregarding the rejected applications. The underlying idea is that an IPC that is selected in a less strict selection round might not be selected in a strict selection round, and therefore the selected companies are assessed according to degree of strictness. This option is an extension of Hypothesis 4.

Identification strategy

A propensity score is applied for each cut-off point. The estimated propensity scores are pooled for the various application rounds. The various rounds are ranked by degree of strictness. Applications above the cut-off point of the stricter selection are then compared with the applications of the somewhat less strict selection (which were not selected in the stricter selection and were therefore below the cut-off point for the stricter selection).

Econometric specification and methods

Propensity scores must be given by comparing a selected company for 2012 with a selected company from 2011, thereby determining a propensity score based on the information from selected companies. The propensities for 2012 are then compared with those for 2011. An absence of differences would mean that a stricter selection did not lead to a better result of the dependent variable.

Sensitivities

Companies may decide against submitting an application if it is known that the selection has been made stricter in a specific application round.

Option 6: Survey method

Identification strategy

IPC applicants can be asked directly for their opinion about the IPC and changes in their behaviour because of the IPC. All hypotheses except Hypothesis 4 can be tested with a survey. Surveys would appear to be important for testing Hypotheses 1b, 2b and 3b in particular, for which companies whose IPC application was rejected are compared with companies that did not apply for IPC.

Questions may be included about past, present and future expectations. The survey will include a question about the forecast profit of companies in the sector concerned, specifying the period.

Econometric specification and methods

The outcomes of a survey depend in part on how questions are phrased. A pilot study with open interviews is needed to validate the questionnaire. The consistency of answering must be tested by repeating important questions in a slightly different form. 'Markers' must be used to standardize the outcomes. For example, an employee might answer differently from a director.

Sensitivities

Outcomes: Answers do not always give an accurate picture of the real world. Socially desirable answering in particular may distort the outcome.

Conclusions and recommendation

Option 1 is the preferable strategy because it is based on randomization. It also gives an estimate for the entire population as opposed to a local estimate. The options based on the propensity score method also appear to be promising, partly because of the possibility of assessing trends over years. The discontinuity method yields only a local estimate. For each of Hypotheses 1, 2 and 3 two groups of companies can be compared. The first comparison is of companies whose IPC application was granted with companies whose application was rejected (Hypotheses 1a, 2a and 3a). The second comparison is of companies whose IPC application was rejected with companies that did not apply for IPC (Hypotheses 1b, 2b and 3b). The data material for the first comparison would appear to be more readily available. The second comparison would appear to require a supplementary survey of companies that did not apply for IPC.

For the purpose of future instrument design, the working group recommends the following selection process.

- Organize a tender to produce two groups: the rejected group and the accepted group. Apply randomization within the accepted group for the purpose of impact measurement.
- Weighted randomization may be chosen in order to reduce the 'unfairness' of randomization and preserve the incentive to deliver sound plans, in which the proposals assessed as best in the tender have a greater chance of success. The weighting probabilities in the randomization then form the propensity score.

4.7 Top Consortia for Knowledge and Innovation (TKI)

Objective

The objective of a TKI is to improve the match between public R&D investment and innovation efforts in the top sectors. Greater cohesion in the knowledge chain helps promote private R&D, ultimately leading to more new (innovative) products and services. The underlying idea is that a TKI helps raise the return on public research funds because tighter control of demand in the knowledge system leads to more commercial and social application of research output. The TKI target group consists of companies and knowledge institutes (research and professional universities, TNO, the Major Technological Institutes (GTIs), and the Royal Netherlands Academy of Arts and Sciences (KNAW) institutes) with the possible addition of public authorities that share in the costs.

Market failure

The TKI's legitimacy derives primarily from the existence of information asymmetry. There is often a lack of information in the knowledge-skills-finance chain about who is in the best position to perform research and who would be best placed to market the research output. A TKI is intended to bridge this information gap. Secondly, a TKI addresses coordination problems. Locating research within a TKI prevents the duplication of research activities, leading to more efficient allocation of funds. Finally, the TKI supplement promotes collaboration on precompetitive research. This collaboration may be hindered by the presence of positive externalities in the form of knowledge spillovers.

Operation

A TKI is a multiyear partnership of private and public parties that collaborate on a programme of research and innovation in the fields set down in top sector innovation contracts. A TKI involves at least three companies and three publicly financed research organizations. Companies in a TKI make a financial commitment to a research programme. Each top sector has one or more TKIs. The nineteen TKIs established in the nine top sectors as at 1 September 2012 will serve as an umbrella for the various partnerships. A TKI is tailored to a top sector, so that TKIs may differ from one top sector to another. TKIs are open to anyone who wishes to join.

The essence of the collaboration between companies and public knowledge institutes in a TKI is the performance of research and innovation activities for joint expense and risk. A project involves at least one company and one knowledge institute. This applies throughout the innovation chain from fundamental research to experimental development, always in precompetitive collaboration between private and public parties. TKIs allow companies to pool some of the risk in investing in precompetitive R&D through an ex ante contribution and clear intellectual property arrangements in partnership with a knowledge institute. The research is performed by research and professional universities, technological institutes and companies (including private research institutions). A TKI may call on the Netherlands Organization for Scientific Research (NWO) and Netherlands Organization for Applied Scientific Research (TNO) in formulating research objectives. A TKI is responsible for the management, network creation and knowledge sharing associated with the research and innovation activities.

TKI supplement

A financial incentive known as the TKI supplement exists to encourage companies to engage in public-private partnerships and place the joint activities in a TKI. Only TKIs within the scope of the Budget Act for the year concerned are eligible for a TKI supplement. The TKI itself, rather than the individual companies and knowledge institutes, applies for the supplement. Accordingly the supplement does not go to individual companies but to the TKI. The TKI supplement must then be used for a new project, or for the project for which the private cash contribution was granted, at the TKI's discretion. A TKI uses the supplement together with the other contributions for items in the programme of activities throughout the chain, such as for PhD research at universities, applied research, or valorization. Costs may include items such as personnel costs, apparatus, building costs, costs of contract research and additional general overheads, provided they are attributable to the research programme. The base amount is determined by the annual cash contributions of companies to the TKI research programme.

Implementation

The minimum size of a TKI research programme is €5 million per year. The minimum quantitative contribution of private parties within a TKI (cash and in kind) is 40% for 2015 (2014: 35%, 2013: 30%). A TKI distributes the intellectual property of public-private research projects in accordance with the private contribution to a research project on a pro rata basis. A party that wishes to acquire an exclusive right may buy out other companies.

A TKI may apply to NL Agency for a TKI supplement. Only the amount of the private cash contribution for partnership projects of a TKI programme at research organizations determines the size of the supplement. A company receives no RDA or WBSO in respect of the private cash contribution to a research organization (the research being outsourced), whereas a company is eligible for RDA and WBSO for internally executed R&D. TKI supplement applications are always accepted subject to compliance with the conditions of the R&D&I support framework. The TKI supplement is an open-ended scheme, which means that if the private contribution is higher than foreseen the rate for the following year can be adjusted downward.

Budget

From 2013 €90 million is available for the TKI supplement.⁶³ The supplement rate is 25%. For the first €20,000 of cash contribution per company per year to a TKI research programme the supplement rate is 40%, to encourage many, usually relatively small, companies to join the research programme. Exhaustion of the budget would require approximately €360 million of private contributions.

Evaluation period

An interim evaluation of the TKIs is scheduled for 2015, and an initial evaluation of the TKI supplement is planned for the same year.

Hypotheses

Core hypotheses

- 1) A TKI increases public-private partnership in the knowledge chain.

Hypotheses in support of core hypothesis

- 2) A TKI improves the match between public R&D investment and the innovation efforts in the top sectors.
- 3) A TKI leads to more private R&D.

Hypothesis regarding second-order effect

- 4) The channelling of public research funds through a TKI leads to more commercial and social application of research output.

Hypothesis for testing of TKI supplement

- 5) The TKI supplement encourages private R&D investment in the TKIs.

Data availability

The TKI itself or NL Agency are important sources of data about companies' and institutions' expenditure within a TKI project. In principle the TKI must also be able to provide data about patents and other intellectual property arising from TKI projects.

Statistics Netherlands has several databases of companies that are available for TKI evaluation. Applicant data can be linked via Chamber of Commerce file numbers with the General Business Register (ABR) and then with other Statistics Netherlands resources. Much experience of this has been gained regarding WBSO applications, for example. The Production Statistics (PS) are a source of data about revenue, productivity, employment, and so on. A limitation is the sample nature of the PS for smaller companies. The same is true of the Company Finance Statistics (SFO).

The Statistics Netherlands innovation survey (CIS) and R&D survey are sources of indicators of R&D and innovation activity. These also have the limitation of being samples. The surveys are performed every two years and the microdata are released only 18 months later. The sample includes 0% of the companies with fewer than 10 employees, approximately 10% of the companies with 10-50 employees, approximately 20% of the companies with 50-100 employees and 100% of the companies with more than 100 employees. NL Agency WBSO data can be used for a broad coverage of R&D data over companies (in the coming years NL Agency RDA data will also be available).

⁶³According to the Rutte-Samson coalition agreement another € 110 million will be added. At the time of writing further details were unknown. This extension may provide further reference points for the construction of an evaluation design.

Table 7 List of indicators for TKI evaluation design

<i>Hypotheses</i>	<i>Indicators</i>		<i>Data sources</i>
	<i>Dependent</i>	<i>Independent</i>	
1) A TKI increases public-private partnership in the knowledge chain.	<ul style="list-style-type: none"> • More PPP publications • More PPP patents 	<ul style="list-style-type: none"> • Indicator TKI Yes/No 	<ul style="list-style-type: none"> • TKI • PATSTAT • CWTS • NL Agency/Netherlands Patent Office
2) A TKI improves the match between public R&D investment and the innovation efforts in the top sectors.	<ul style="list-style-type: none"> • No quantitative indicator available 	<ul style="list-style-type: none"> • No quantitative indicator available 	<ul style="list-style-type: none"> • Survey
3) A TKI leads to more private R&D.	<ul style="list-style-type: none"> • Private R&D investment 	<ul style="list-style-type: none"> • Indicator TKI Yes/No 	<ul style="list-style-type: none"> • TKI • NL Agency/WBSO • PATSTAT
4) The channelling of public research funds through a TKI leads to more commercial and social application of research output.	<ul style="list-style-type: none"> • More patents for knowledge institutes 	<ul style="list-style-type: none"> • Indicator TKI Yes/No 	<ul style="list-style-type: none"> • TKI • NL Agency/Netherlands Patent Office • PATSTAT
5) The TKI supplement encourages private R&D investment in the TKIs.	<ul style="list-style-type: none"> • Private R&D investment 	<ul style="list-style-type: none"> • Indicator TKI Yes/No Dummy 	<ul style="list-style-type: none"> • TKI • NL Agency/WBSO • Statistics Netherlands
<i>Control variables [not exhaustive]</i>	<ul style="list-style-type: none"> • Company characteristics • Characteristics of knowledge institutes 		<ul style="list-style-type: none"> • PS/ Company Finance Statistics • CIS

NL Agency and Statistics Netherlands have access to the patent applications made by Dutch companies in the Netherlands and within the EU. The records largely cover all patents in PATSTAT, but unlike PATSTAT the information is linked with the General Business Register (ABR).

Data about knowledge institutes are less readily available. CPB Netherlands Bureau for Economic Policy Analysis has a database of academic staff from the Association of Universities in the Netherlands with supplementary data from the individual universities. The approximate period covered is 2002-2008. Permission must be obtained from the Association of Universities in the Netherlands to reuse the data.

There is a Netherlands Organization for Scientific Research (NWO) database of Renewal Impulse subsidy applicants, which CPB Netherlands Bureau for Economic Policy Analysis uses in anonymized form within Statistics Netherlands' secure microdata environment.

CWTS has scientific publications by name. These publications are difficult to link with Statistics Netherlands data, because Statistics Netherlands mainly links on gender/date of birth/address. The Statistics Netherlands Municipal Personal Records Database (GBA) data contains no names.

Patent databases (such as PATSTAT) contain names of both the inventor and the institutions and/or companies applying for a patent. Except for companies these data are currently insufficiently standardized and therefore cannot yet be linked with other data.

Option 1 Comparison of TKI companies with other companies

Experimental and control groups

For each TKI, companies can be compared with similar companies in the same economic sector that do not participate in a TKI. Both groups of companies must be tracked from the period prior to the TKI. This option is applicable to Hypotheses 1 and 3.

Identification strategy

The time trend of the TKI companies must be compared with that of the non-TKI companies. This resembles a difference-in-difference approach, with the assumption that both groups of companies exhibited the same time trend in the period prior to the TKI.

Econometric specification and methods

The database has information about the period prior to the introduction of TKI. The dependent variable is regressed on an indicator, TKI, which is equal to 1 from the time a company joins a TKI. The effect of TKI can be determined with the coefficient of this indicator. Time adjustment is also required, with a separate indicator variable included for each year. Account is taken of differences between companies by including an indicator variable for each company, which are also referred to as the fixed effects. The regression function is then:

$$Y_{it} = \beta_0 + \beta_1 TKI_{it} + \eta_i + \theta_t + u_{it}$$

In this equation Y is the performance of company i at time t , the β s are the coefficients to be estimated and u is the error term.

Sensitivities

Companies and knowledge institutes that form a TKI exhibit self-selection. The assumption of a common trend could therefore sometimes be incorrect. It can also be investigated whether the observed impact intensifies over time, because the proportion of the contribution of the private parties within a TKI is expected to increase (40% from 2015).

Option 2 Comparison of TKI companies participating and not participating in project

Experimental and control groups

The next step is to zoom in on information of companies that participate in a TKI. Several projects may be performed within a TKI, and not all companies and knowledge institutes necessarily participate in all of them. An option is to compare the companies within a TKI that participate in a specific project with companies in the TKI that do not. This option is applicable to Hypotheses 1, 3 and 5.

Identification strategy

Companies that participate in a specific project are compared with companies that do not. The comparison is based on a difference in time trend between the period prior to a specific project and the period after starting the project.

Econometric specification and methods

A difference-in-difference specification, in which the dependent variable is regressed on indicator variables set to 1 from the time a company joins a specific project. This variable is 1 for all companies that participate in this specific project, with each project being given a separate indicator variable. Time adjustment is also required, with a separate indicator variable included for each year. Account is taken of differences between companies by including an indicator variable for each company, which is also referred to as the fixed effects. The regression function is then:

$$Y_{it} = \beta_0 + \sum_j \beta_j PROJECT_{ijt} + \eta_i + \theta_t + u_{it}$$

In this equation Y is the performance of company i at time t , the β s are the coefficients to be estimated and u is the error term.

It can also be investigated whether the observed impact intensifies over time, because the proportion of the contribution of the private parties within a TKI is expected to increase (40% from 2015). In this case therefore the #s vary in time. For testing Hypothesis 5 an indicator may be added for use by the project of the TKI supplement.

Sensitivities

Companies and knowledge institutes that participate in a TKI project exhibit self-selection. The assumption of a common trend could therefore sometimes be incorrect. Companies are allowed to participate in one or more projects.

Box 4.7.1 Variation in supplement rates

If the supplement rates change in future an evaluation might possibly use the resultant variation over the years. If the threshold values for the supplement rate change, companies just above and just below a new threshold could be compared. A problem with this empirical strategy is that companies have some influence over whether they fall above or below a threshold. This variation is therefore not entirely exogenous.

Option 3 Comparison of knowledge institutes

It is also possible to zoom in on the knowledge institutes. In principle the same specifications can be used as described above for Options 1 and 2. As dependent variable the knowledge institute's publication record or number of patents is used. This option is applicable to Hypothesis 4.

Experimental and control groups

Participating and non-participating academic institutions are compared with respect to publications and the receipt of indirect funding and funds for contract research. The comparison is based on academic production in the period prior to the TKI relative to that in the TKI.

Identification strategy

Researchers who take part in a specific project are compared with companies that do not. The comparison is based on a difference in time trend between the period prior to a specific project and the period after starting the project.

Econometric specification and methods

In principle the same identification strategy is used as with Option 2. A difference-in-difference specification, where the dependent variable is regressed on indicator variables that are set to 1 from the time a researcher joins a specific project. This variable is 1 for all researchers who participate in this specific project, with each project being given a separate indicator variable. Time adjustment is also required, with a separate indicator variable included for each year. Account is taken of differences between companies by including an indicator variable for each company, which are also referred to as the fixed effects. The regression function is then:

$$Y_{it} = \beta_0 + \sum_j \beta_j \text{PROJECT}_{ijt} + \eta_i + \theta_t + u_{it}$$

Sensitivities

A complication with this option is the substantial delay of several years associated with publications, so that this option does not seem to be promising.

Option 4 Survey of companies and knowledge institutes

In this option arbitrary companies within a TKI are compared with companies that are not in a TKI. The companies are questioned as to how much the proceeds of innovation and the use of public research results has increased relative to the period before the TKI application, and to what extent the change is attributable to the TKI.

This option is applicable to all hypotheses. A survey of this kind can also be performed with knowledge institutes, in which it is possible to zoom in on publications and the receipt of indirect funding and funds for contract research.

Conclusions

Each top sector has one or more TKIs. The proceeds of a TKI may go to the participating companies and the knowledge institutes. In order to test the core hypothesis the companies must be compared with a group of companies that do not participate in a TKI, where TKI allocation should be randomized. The same design could also apply to funds for the knowledge institutes. However randomized allocation is infeasible, so that alternative estimation methods are proposed. The planned design of a TKI would appear to rule out different clusters of companies and knowledge institutes within a top sector competing for participation in a TKI. Consequently it is impossible to compare a TKI with a different TKI within the top sector.

Of all the options presented, Options 1 and 2 would appear to have the best prospects of shedding light on the effect of TKIs on innovation output. Both options are based on a difference-in-difference specification. Option 3, being oriented to knowledge institutes, would appear to be less promising because of the substantial delays in publications. However, these options do not answer Hypothesis 2 about the match between public R&D investment and the innovation efforts in the top sectors. A survey may be considered as a fallback option to provide insight into all hypotheses, without estimating causal effects.

Chapter 5 Conclusion and recommendations

5.1 Conclusion

The calls for greater clarity surrounding the impact of policy are becoming ever louder. In response, an increasing number of ministries are actively engaged in evidence-based policy. The efforts of the Ministry of Economic Affairs are not isolated. For instance, the Ministry of Education, Culture and Science embarked in 2010 on a similar process to the one covered in this report. A natural experimental design proved feasible for approximately two-thirds of the fifty policy interventions in this programme. Of the fifty projects more than ten have now actually been put out to tender. The effectiveness of some interventions has been demonstrated, which is encouraging.

The Directorate-General for Entrepreneurship and Innovation has considerable experience with policy evaluation. The investigation methods applied in the past vary from surveys to randomized experiments. Examples of evaluations that map out the immediate quantitative impact of policy are the Innovation Vouchers and the WBSO. The Directorate-General for Entrepreneurship and Innovation wishes to further improve the existing evaluation practice to ensure clear accountability and the most effective possible spending of public funds. This report makes several suggestions for achieving this.

The main task of the Impact Evaluation Expert Working Group is to provide a picture of the multiple evaluation options for the six policy instruments covered in this report. The working group aimed to be as innovative as possible in the sessions held to discuss the various instruments. In the working group's opinion this report presents a series of evaluation designs that have the potential to quantify the immediate impact of the six policy instruments. It must be noted that since the design and operation of the instruments were taken as given by the working group, the strengths of the assumptions that underlie the various evaluation designs vary considerably. The working group therefore recommends the designs for each instrument (see Chapter 4) that they consider to be the most promising for evaluation.

The Expert Working Group was also requested to give an opinion about the most recent WBSO evaluation. The large scale and generic nature of the WBSO instrument hamper the identification of a comparable control group. The WBSO evaluation used a dynamic panel data model that attempts to adjust as well as possible for both observed and unobserved company characteristics, but the complete success of this approach is never guaranteed. There is therefore room for doubt as to whether the estimated effect is actually the causal effect of the WBSO. In general a better research strategy is to look for policy variation that ensures the existence of comparable groups of companies that did and did not participate in an instrument. This is a difficult strategy to apply to a generic policy instrument such as the WBSO. This emphasizes the importance of giving far more attention to future quantitative evaluation options from as early as the instrument development phase.

Measurement of the indirect impact of policy, which is referred to as the second and third-order effects, remains a tricky issue, firstly because of the many factors that influence higher policy objectives, such as economic growth and employment. These factors hamper the isolation of an instrument's contribution from all other effects. Time is another important factor. The observance of second and third-order effects requires more time to have elapsed than in measuring the direct impact. This condition is not always satisfied at the time of evaluation. Finally, it is important to have sufficient data available, and for it to be feasible to link databases.

5.2 Recommendations

The working group has given interim recommendations for the Innovation Credit and Certificate of Good Service in view of the imminent evaluations of these instruments. The working group sees the Directorate-General for Entrepreneurship and Innovation's energetic response to the recommendations as a sign that the ministry is taking impact measurement seriously. Hence, it pleases the working group to provide more specific recommendations in order to further improve the visibility of the impact of policy in the future.

1. Ex post evaluation starts ex ante with a clear analysis. What is the social problem? Should the government attempt to solve this problem? If so, which policy instrument would be the most appropriate? If the nature of the problem or the possible contribution of a policy instrument to a solution are unclear, evaluation will be difficult. Chapter 1 of this report explains why social problems may arise with innovation, and why the government may therefore have a role to play. *Develop policy theory ex ante for each policy instrument, defining one measurable objective for each instrument, also addressing the economic legitimacy of the government's role.*
2. Future evaluation often receives too little attention in the design of policy instruments, leading to difficulty in performing an effective policy evaluation. For example, satisfactorily comparable groups can no longer be found, or only on a local scale around an instrument's assessment threshold, which determines companies' eligibility for an intervention. In the latter case robust impact analysis will be possible for only some of the users. Extrapolation of local impact to the entire user group is not without risk, and may lead to false conclusions. Therefore resort must often be made to evaluation methods that rely on strong assumptions that are impossible to verify completely. This rules out firm statements about the impact of policy. *The working group recommends setting up an evaluation design as early as the policy instrument development phase, thereby indicating how the instrument's impact is to be evaluated in future.*
3. As stated in Chapter 2 the most unbiased estimate of an intervention's causal effect is through a RCT. An additional advantage of an experiment of this kind is that the estimated impact relates to the entire user group. RCTs rely on relatively few assumptions, thereby allowing for the clear identification of second and third-order effects. The working group is aware that random allocation is not always possible or desirable. In weighted randomization, subsidy or credit applications are first assessed on quality. Randomization is then performed within the group that falls above a certain threshold, thereby making the randomization more equitable and increasing the support for this kind of measure. *To gain an even clearer understanding of the effectiveness of instruments the working group recommends using RCTs based on weighted randomization where possible in the introduction of new policy instruments.*
4. Time is needed to measure the impact of policy. Even if a good evaluation design is available, sufficient time must pass to allow observations to be recorded. The General Administrative Law Act requires subsidies to be evaluated every five years. Fewer years will generally be available for observation after subtracting an instrument's start-up period and allowing for evaluation to start early in order to produce evaluation results promptly at the end of the five years. Furthermore, an instrument's objective may be incompatible with mapping out all effects within the five-year period. Whereas for most innovation instruments a first-order effect may be expected within the five-year period, such as an increase in private R&D investment, this period is simply too short for second and third-order effects to emerge (new products, economic growth). *Before instruments can be rolled out on a large scale it is advisable to consider a small-scale pilot for experimental purposes. A realistic evaluation time must then be chosen, in order to enable analysis of the immediate impact at least.*
5. Empirical policy evaluations provide more insight when the available data is of high quality. The working group's recommendations for data collection are consequently:
 - *to link databases from NL Agency (such as Innovation Credit and WBSO), Statistics Netherlands and the Tax and Customs Administration where possible, and to provide maximum access to the results on the website www.volginnovatie.nl, which is in line with the Open Data Policy and will contribute to better policy accounting transparency;*
 - *for NL Agency to actively track the progress of submitted proposals and projects, for rejected applications as well as granted applications;*
 - *to survey both successful and unsuccessful applicants in evaluations. Without data about the control group of non-users convincing evidence of effectiveness is hard to find;*
 - *to work with a positive financial incentive to enhance the response of (rejected) applicants (e.g. a small monetary reward for returning a completed survey).*
6. In this report the Impact Evaluation Expert Working Group gives recommendations about only six policy instruments. There are many more instruments in the ministry's industrial policy and other policy areas, the impact of which remains unclear. *The working group recommends analysing the impact of the other (major) policy instruments of both the ministry's industrial policy (such as the Innovatiebox) and other policy areas, based on the approach followed in this report.*
7. Some policy instruments have similar objectives. An example is the WBSO and RDA. *The working group recommends where possible the joint evaluation of instruments that have a comparable policy objective.*

The Impact Evaluation Expert Working Group is aware that the implementation of the above recommendations demands serious effort from the Ministry of Economic Affairs. This effort must be weighed against the scale of policy instruments. However, policy instruments of larger financial scale⁶⁴ call for thorough evaluation. In the end, a ministry must be able to justify the effective use of taxpayers' money.

⁶⁴€ 5 million or more.

