A2007:017

Evaluation of the Effects of Government Subsidies to Swedish Industry

Lars Bager-Sjögren, Federica Calidoni, Håkan Gadd
and A.M.M Shahiduzzaman Quoreshi

itps SWEDISH INSTITUTE FOR GROWTH POLICY STUDIES

Evaluation of the Effects of Government Subsidies to Swedish Industry

Lars Bager-Sjögren, Federica Calidoni, Håkan Gadd,
A.M.M. Shahiduzzaman Quoreshi

itps INSTITUTET FÖR TILLVÄXTPOLITISKA STUDIER

ITPS, Swedish Institute for Growth Policy Studies
Studentplan 3, SE-831 40 Östersund, Sweden
Telephone: +46 (0)63 16 66 00
Fax: +46 (0)63 16 66 01
E-mail: info@itps.se
www.itps.se
ISSN 1652-0483 (web), 1652-8486 (print)

For further information, please contact Federica Calidoni
Telephone: +46 (0)63 166 617
E-mail: federica.calidoni@itps.se

Foreword

In 2007, the Ministry of Enterprise, Energy and Communications (*Näringsdepartementet*) gave ITPS the assignment of investigating the possibility of building a model, or several models, to evaluate state aid in Sweden. The present study looks at evaluation issues from two different perspectives.

From one perspective, it presents the experiences and good practices found in other countries (Holland, Ireland and the USA) in order to shed light on the so-called *evaluation culture.* The report proposes two ways that Sweden can act to improve its evaluation effectiveness: (1) to appoint an external agency to carry out both ex-ante and ex-post evaluation, or (2) to appoint a single agency responsible for policy implementation, analysis and evaluation.

The second perspective focuses on methods at both the micro- and macro levels and, following a brief literature review of matching techniques, it sketches a possible CGE model for ITPS to develop as a general framework to evaluate the effects of policy changes at three levels: the firm level, the regional level and the national level.

ITPS would like to thank the reference group who followed the development of the project: Runar Brännlund (Umeå University), Jonas Månsson (Växjö University), Richard Murray (Swedish Agency for Public Management), Magnus Sjöström (National Institute of Economic Research) and Anders Stenberg (Swedish Institute for Social Research, *SOFI*).

The report is the result of a group work by Federica Calidoni (project leader), Lars Bager-Sjögren, Håkan Gadd, A.M.M. Shahiduzzaman Quoreshi and Johan Lundberg (CERUM, Umeå University)

Östersund, November 2007

Brita Saxton

General Director

Table of Contents

Summary **7**
1 Introduction **11**
1.1 State aid 11
2 Evaluation process and evaluation culture **15**
2.1 Evaluation in different countries 15
2.1.1 The Netherlands 17
2.1.2 Ireland 21
2.1.3 USA 25
2.2 Lessons from other countries 28
2.3 Evaluation of economic policy in Sweden 29
2.3.1 Regulatory evaluation framework 30
2.3.2 Management by objectives 30
2.3.3 What are follow-up and evaluation? 33
2.3.4 Nutek's and Vinnova's evaluation strategies 34
2.4 Challenges 38
3 Evaluation methods **41**
3.1 Micro level analysis: Non-experimental evaluation and matching estimators 41
3.2 Regional analysis and spatial effects 44
3.2.1 How can spatial econometrics be used to analyse the effects of government support for companies and regions? 46
3.2.2 Concluding discussion 50
3.3 General equilibrium model 50
3.3.1 Existing models 52
4 ITPS model **63**
5 Concluding remarks **69**
References **71**
Appendix 1 Discussion topics for interviews **75**
Appendix 2 Some programs and methods from other countries **77**
A2.1 The Netherlands 77
A2.2 Ireland 80
Appendix 3 Brief description and review of different econometric programs **82**

Summary

In 2007, the Ministry of Enterprise, Energy and Communication (*Näringsdepartementet*) appointed ITPS to investigate the possibility of building and implementing a model, or several models, to evaluate state aid in Sweden. The present study looks at evaluation issues from two different perspectives: on one side, it presents the experiences and good practices from other countries (Holland, Ireland and the USA) in order to shed light on the so-called *evaluation culture*; on the other side, it focuses on evaluation methods at both the micro- and macro levels and, following a brief literature review of matching techniques, sketches a possible general equilibrium (CGE) model that could represent a general framework to estimate policy effects at both the national and regional level.

The work is divided into two parts. The first part (Chapters 1 and 2) addresses the main issues, which are: What is evaluation? Why should a government develop an evaluation culture? What would it take to develop it?

It is clear that evaluation is a complex concept aimed at both monitoring what is done in an institution and learning from what is done. Moreover, evaluation may be the domain of a small group of specialists, inside or outside the institution, or part of the institution's general activity. In either case, the usefulness and importance of evaluation is generally agreed upon and, by consequence, the way evaluation is conceived, introduced and managed must be thought through carefully.

This poses quite a challenge for complex policies such as state aid and financial grants. This is also because evaluations are, needless to say, politically sensitive. And for the evaluation practice to be effective, institutions must be prepared to accept and to publish reports that contain negative or disturbing conclusions. This attitude can be achieved only through a deeply rooted culture of democratic evaluation at all administrative levels.

Chapter 1 describes how public intervention in the market is justified according to the theoretical literature.

Chapter 2 describes the role played by evaluation in three different countries: the Netherlands, the USA and Ireland. Through development of the *From Policy Budget to Policy Accountability* regulatory program (*VBTB*) in 1999, the Netherlands has acquired a leading role in the evaluation debate in Europe. VBTB has also led the Netherlands to rethink the evaluation function and to introduce a more systematic and structured approach towards evaluation. Moreover, the Dutch regulation states that for each new or adjusted policy objective or instrument, it should be considered whether ex-ante evaluation is useful, and each existing support measure should be evaluated, at least once every five years. The Irish experience is quite different from the Dutch one and highlights interesting evaluation features in relation to the growing attention on evaluation practice at the European Union level. Ireland, in fact, developed its government evaluation system in response to the requirements for accession to the EU, and provides a clear example of how strong external pressures, linked to the availability of significant resources, can be a key catalyst in stimulating evaluation. Finally, the experience of the United States is of interest due to the peculiar political structure and great number of actors involved, both at the federal level and at the state level, and because it offers a significant example of leadership in evaluation in its Advanced Technology Program (ATP). ATP has motivated nationally

prominent evaluators to apply new and existing methods in building an analytical and empirical basis for ATP's operations and performance.

Chapter 2 also draws conclusions on the lessons learned from the experience of other countries and from the study of the Swedish situation. The Dutch experience stresses the importance of using a good policy design that is well-thought-out as an excellent starting point for evaluation, and the importance of detailed data collection. From Ireland, we learn that programs that are limited and transparent in what they intend to achieve are easier to evaluate. Finally, the USA teaches the need to increase retrospective analyses; to incorporate both direct and indirect path analysis in cost-benefit studies; to keep monitoring and updating information; to continuously develop new evaluation techniques; to identify and address new questions that arise from evaluation; and to take greater advantage of evaluation results in the decision-making process.

The analysis of the Swedish evaluation culture emphasizes the need to deepen the understanding of the effects of different policies and to take into account the results of external experts. Given that monitoring alone is incapable of offering policy-relevant insight into policy effectiveness, the different agencies in charge of evaluation are required to move a step forward towards evaluation. This is the only way to provide legislators with the tools to design new policies and programs. Sweden needs to move towards a more complex evaluation culture, in order to trigger a learning process in which evaluations provide information and create the necessary knowledge for an improved policy design. There are two alternative paths that the Swedish government can follow to improve the effectiveness of evaluation: (1) to appoint an external agency to do both ex-ante and ex-post evaluation, or (2) to appoint a single agency responsible for policy implementation, analysis and evaluation. The focus must shift towards a new culture of evaluation while paying attention to comprehensible setting of objectives and quantitative targets, looking beyond simple monitoring approaches and improving instruments to take into account selection bias, and triggering a learning process.

The second part of the report (Chapters 3 and 4) focuses on evaluation methods and proposes a model that could be developed by ITPS in the future. The key issue in evaluation is the *identification* of a counterfactual situation and the need to answer the counterfactual dilemma of what would have happened if the policy had not been implemented.

Chapter 3 investigates how the counterfactual situation can be analysed at the micro- and macro levels, depending on the objective of the policy, the size of the effects expected, and the availability of data. By means of microanalysis the evaluator compares a so-called "experimental" group that receives the policy treatment to a "control" group that represents the universe of firms that do not receive government funds. The chapter describes the econometric methods used to construct a control group that resembles the treatment group as closely as possible, at least in terms of observed characteristics. Given that micro level analysis is not always feasible, the evaluator can study the economy at the macro level to highlight the general effects on the national (or regional) economy by means of a general- or partial equilibrium analysis, which describes two scenarios at a given point in time (the equilibrium reached by the economy without intervention and the equilibrium reached after policy implementation) and attributes the difference between them to the policy. Macro level analysis represents a valuable alternative to microanalysis in cases where data is lacking or there is a peculiar policy design.

The chapter summarizes existing models in the Swedish economy and discusses their advantages and disadvantages, in particular: EMEC (Environmental Medium Term Economic Model); FIMO (Financial Model); LUMOD (Long term dynamic simulation); KIMOD (Dynamic General Equilibrium Model of the Swedish Economy); ISMOD (Industry Structure Model); SAMGODS (Transport Analysis System); rAps (Regional Analysis and Prognosis System); RAMSES (Riksbank Aggregate Macromodel for Studies of the Economy of Sweden); *Sveriges Konkurrenskraft – att förstå och mäta nationell konkurrens*.

Chapter 3 also focuses on the role played by space and location in the study of the effects of public policies on the economy, and gives answers to the following questions: How can spatial econometrics be used in evaluation of national policies and specific regional measures? How useful can it be? Is it worthwhile to include spatial analysis in any kind of evaluation of state aid programs? How do we account for spatial effects in "traditional" econometric programs (Stata, SAS, E-views)? Is it possible? What about specific programs (GIS, GeoDa, SpaceStat)? What resources would ITPS need in order to acquire the knowledge and skills necessary to work with spatial econometrics models?

Chapter 4 describes an "ITPS model" that could be developed by ITPS in the future to study the effects of state aid at three levels: the firm level, the regional level, and the national level. The ideal way to investigate the effects at the three levels would be to start with a firm-specific analysis of every company in the economy, to group them at the regional level to investigate regional effects, and, finally, to combine the regions in order to study effects on the entire economy. Given that the sheer number of companies in the economy makes such a procedure time-consuming and technically impossible, we suggest instead the development of a flexible system, similar to RAMSES/BVAR developed by the Riksbank but on a regional level, that would enable us to investigate the effects of different policies from time to time on smaller groups of firms.

A presentation of the conclusions drawn is given in Chapter 5.

1 Introduction

In 2007, the Ministry of Enterprise, Energy and Communication (*Näringsdepartementet*) assigned ITPS the task of investigating the possibility of building a model, or models, to evaluate state aid in Sweden. The present study looks at evaluation issues from two different perspectives: on one side, it presents the experiences and good practices from other countries (Holland, Ireland and the USA) in order to shed light on the so-called *evaluation culture*; on the other side, it focuses on evaluation methods at both the micro- and macro levels and, following a brief literature review of matching techniques, sketches a possible CGE model that can be used as general framework to evaluate the effects of policy changes at a regional level.

To investigate the *evaluation culture* in other countries, we decided to meet the people in charge of evaluation personally, and travelled to The Hague, Washington D.C. and Dublin to conduct a number of interviews. The interviews were semi-structured, based on a series of discussion topics presented in Appendix 1.

According to estimations by ITPS and the Swedish Agency for Economic and Regional Growth (NUTEK), in the period 2003–2005, Sweden invested between 25–30 billion SEK annually in different types of state aid to business. Included in these figures are employment subsidies, start-up grants, regional support in the form of reduced social security contributions, transport support and regional development grants, as well as aid to small businesses. One of the purposes of these types of support measures is to increase employment and thus also promote economic growth. In evaluating the effects of state aid measures, the application of relevant statistical methods and theoretical models is of great importance.

Traditionally, the core topics of interest in the framework of ITPS research have been economic growth, productivity, employment and regional issues. With this in mind, our investigation covers both the opportunities to adapt existing models and to build a new one. The rest of the analysis therefore focuses on evaluation models, with particular attention paid to regional issues.

1.1 State aid

This first paragraph gives a brief outline of a theoretical framework to justify public intervention in the market. In most cases, government intervention occurs when markets are not working optimally and there is sub-optimal allocation of resources in a market or industry, or to stimulate the country's competitiveness. In simple terms, the market may not always allocate scarce resources efficiently, in a way that achieves the highest social welfare, and may not always compete efficiently in the international market. Governments therefore justify their intervention as being in the public interest.

According to the economic literature, market failures have negative effects on the economy because an optimal allocation of resources is not attained. In other words, the social costs of producing the goods or services are not minimized, and this results in a waste of some resources. The issue of market failures and how they should be addressed is a source of contention between different schools of economic thought.

In the neoclassical perspective, if a certain result is Pareto efficient, then it is not considered a market failure, regardless of whether or not it serves the public interest. The

Keynesian and neo-Keynesian school interpret failure to automatically attain full employment of resources in terms of theories of market failure; for New Keynesian scholars in particular, the main stress in on the non-adjustment of prices and wages. Furthermore, the so-called Public Choice School and advocates of laissez-faire capitalism argue that there is no such concept as market failure. They argue that market failure does not necessarily imply that government should attempt to solve market failures, because the costs of government failure may be worse than those of the market failure it attempts to fix. Others, such as social democrats and New Deal liberals, view market failures as a common problem of any unregulated market system, and therefore argue for extensive state intervention in the economy.

However, if, on one hand, state intervention is necessary to reach Pareto efficiency, it could, on the other, be harmful to competition, something that has led to a great amount of legislation concerning the legitimacy of state aid.

According to the European Community Treaty,[1] state aid is defined as financial aid to business, which exhibits the following characteristics:

- It is granted by a Member State or through State resources.
- It favours certain undertakings or production of certain goods.
- It distorts or threatens to distort competition.
- It affects trade between Member States.

Examples of state aid that fall within the scope of the treaty are: State grants; interest rate relief; tax relief; tax credits; State guarantees or holdings; State provision of goods or services on preferential terms; direct subsidies; tax exemptions; preferential interest rates; guarantees of loans on especially favourable terms; acquisition of land or buildings either gratuitously or on favourable terms; provision of goods and services on preferential terms; indemnities against operating losses; reimbursement of costs in the event of success; State guarantees, whether direct or indirect, to credit operations preferential re-discount rates; dividend guarantees; preferential public ordering; reduction of, or exemption from, charges or taxes, including accelerated depreciation and the reduction of social contributions; deferred collection of fiscal or social contributions; assistance financed by special levies; capital transfers; certain State holdings in the capital of undertakings.

Less obvious examples where state aid might arise include: consultancy advice; advantages resulting from the activities of agencies for urban renewal; assistance to help companies invest in environmental projects; assistance to help a public enterprise prepare for privatization; legislation to protect or guarantee market share; public-private partnerships and contracts not open to competitive tendering; receipt of landfill tax credit funding.

In conclusion, there are three main requirements for state aid:

- The aid must address a well-defined market failure.
- The aid must be well-targeted: state aid must be an appropriate instrument; the aid measure must have an incentive effect and must be proportionate to the problem tackled.

[1] *http://www.europa.eu.int/comm/competition/state_aid/legislation/aid3.html - A*

- The distortions to competition and trade resulting from the aid measure must be limited enough so that, on balance, it can be declared compatible.

2 Evaluation process and evaluation culture

In the past decades, evaluation has become more and more of an independent science with roots in many disciplines, and a useful tool for understanding and implementing policy studies, performance assessments, engineering design, investment portfolios, etc. Even in the political arena evaluation practices have acquired increasing importance, and many countries have started to legislate on mandatory evaluation of different programs.

This means different countries have different views of the role of evaluation in the political debate. It is commonly understood that evaluators strive to contribute to social betterment, and this can only be achieved if evaluation findings are fed back to inform program administrators, policymakers and other stakeholders and to improve the program structure and operations.

2.1 Evaluation in different countries

The following section describes the role played by evaluation in three countries: the Netherlands, USA and Ireland. In order to collect specific information on these countries, we have travelled to The Hague, Washington D.C. and Dublin to meet some of the people in charge of evaluation. The interviews were based on a series of discussion topics, presented in Appendix 1.

The reasons behind the choice of each one of these countries are different. In the case of the Netherlands, it is interesting to recall that, since the mid-1980s, much has been done to improve the legitimacy and efficiency of the use of public funds. In the mid-1980s, political attention was focused in particular on the eligibility of government expenditure. Consequently, attention centred on auditing and control. During the 1990s, the focus shifted and questions of the effectiveness and efficiency of public policy entered the debate. There was a shift from a focus on regularity to a focus on efficiency, performance and the effectiveness of public policy. The key rationale behind this development was the desire to improve the management of government organizations: management should become more output- and, especially, more outcome-oriented. Improving government performance was seen as the key objective, by clarifying the relationship between the deployment of resources, products and services and the outcomes to be attained, and to take this as a starting point in (1) policy making, (2) policy implementation, and (3) policy evaluation. The result is a switch from purely financial accounting to policy accounting.

It is, however, mainly the development of the VBTB program[2] (*From Policy Budget to Policy Accountability*) that has given the Netherlands a leading role in the evaluation

[2] *The main conclusions of the VBTB program can be summarized as follows:*

- *Monitoring data and evaluation research should be used in a coordinated way;*
- *For any new policy development that is introduced a careful and serious assessment of the need for an ex ante evaluation should be made;*
- *The main policy objectives should be evaluated on a regular basis;*
- *The methodological quality of evaluation instruments should be ensured;*
- *Those politically and administratively responsible should be informed about the outcome of evaluation research;*
- *Within Ministries there should be a clear division of responsibilities for conducting evaluations.*

debate in Europe. VBTB has led the Netherlands to rethink the evaluation function and to the introduction of a more systematic and structured approach towards evaluation.

In the Netherlands, as in other European countries,[3] state aid program evaluations have provided valuable insights for improving existing programs and have identified learning items for new support programs. In this sense, the cost of evaluation studies can be met through gains in efficiency and effectiveness.

Moreover, in the Netherlands, the law stipulates that for each new or adjusted policy objective or instrument, it should be considered whether ex-ante evaluation is useful, and that each existing support measure should be evaluated, at least once every five years.

The Irish experience is quite different from the Dutch one and highlights interesting evaluation features in relation to the growing interest in evaluation practice at the EU level. Ireland developed its government evaluation system in response to the requirements for accession to the European Union. The system has subsequently been strengthened for internal reasons, related to the government's desire to improve the value-for-money obtained from all areas of public expenditure. This has been reflected in the government's Expenditure Review Initiative.

Ireland provides a number of lessons about success factors and impediments to developing a monitoring and evaluation system. One lesson is that strong external pressures, linked to the availability of significant resources, can be a key catalyst in stimulating evaluation. Another lesson is the merit of periodically reviewing progress in developing such a system, and reorienting the system – sometimes substantially – as a result. Ireland is continuously piloting further improvements to its evaluation system.

Finally, the experience of the United States is interesting because of the peculiar political structure and great number of actors involved at both the federal and state level.

At the federal level, the evaluation culture in the US is personified in the Government Accountability Office (GAO), in charge of studying the programs and expenditures of the federal government and advising Congress and the heads of executive agencies about ways to make government more effective and responsive. GAO evaluates federal programs, audits federal expenditures, and issues legal opinions. At the same time, each state has its own State Auditor Office, which reproduces the same kind of studies at the state level.

The second reason that justifies our interest is that, in the US, evaluation is an essential component of publicly funded R&D programs, both in support of program management and public policy. A significant example of leadership in evaluation is the Advanced Technology Program (ATP), which has engaged nationally prominent evaluators to apply new and existing methods to build an analytical and empirical basis for ATP's operations and performance.

After a brief summary of the way evaluation is legislated and performed and the roles played by the actors involved, we will focus on some interesting methods and on the main features that can guide the development of evaluation practice in Sweden.

[3] *For example, Austria, Belgium, the Czech Republic, Finland, Ireland, Norway, Portugal, the Slovak Republic and the United Kingdom.*

2.1.1 The Netherlands

As of 1999, the Dutch VBTB law prescribes that, whenever public funds are spent, the policies must be evaluated in detail in order to examine whether the resources have been put to good use.

Through a series of interviews with the actors involved, we have tried to understand how the evaluation process has been developed in the Netherlands. Given the limited time available, we chose to focus on R&D and innovation programs and held meetings and interviews at the Ministry of Economic Affairs, SenterNovem and CPB, Netherlands Bureau for Economic Policy Analysis.

Thanks to a meeting with Luuk Klomp and Arjan Wolters at the Directorate-General for Enterprise and Innovation Strategy, Research and International Affairs Department at the Ministry of Economic Affairs, we have built a picture of the current situation in the Netherlands. The existing policies can be divided into two categories: thematic programs and base instruments. Within each thematic program are a number of instruments, while the basic instruments are more or less basic policies. The individual instruments represent the lowest level that has to be evaluated, by law – at least once every five years.[4]

Since the first of January 2002 policy-evaluation and policy preparation are subject to the ministerial decree on performance measurement and evaluation (*Regeling Prestatiegegevens en Evaluatieonderzoek, RPE*). The decree poses a number of requirements on policy preparation (ex-ante evaluation), monitoring and ex-post evaluation. The requirements concern:

1. The use of evaluation instruments.
2. The obligation to consider an ex-ante evaluation when starting to think about a new instrument. Policy-makers have to consider different policy alternatives unless there are good arguments not to do so (urgency, limited financial and societal risks, sufficient information from ex-post evaluations, dictated by international agreements).
3. The frequency and extent of ex-post policy evaluations. Every instrument has to be evaluated.
4. The quality of the evaluation instruments.
5. Informing the minister, head of the department, and parliament about the outcomes of an evaluation.
6. The distribution of responsibilities within the department with regard to the implementation of this decree.

At present, the Netherlands is undergoing a transition phase towards a reduction in the number of instruments, higher flexibility and a better combination of instruments. The future scenario will present larger programs due to the combination of different instruments, and increased evaluation of the different mix of instruments.

Effectively, the Ministry is in charge of designing the different instruments and their combination in a thematic program, while SenterNovem is in charge of implementation of all

[4] *In 2006, the law was partially changed and the 5-year period is no longer as strict. The five years is a flexible limit considered mainly a guideline, and evaluations can be done after longer or shorter periods according to political reasons or the policy cycles. There is, however, still the obligation to evaluate each instrument.*

the instruments dealing with innovation and environmental issues. Moreover, SenterNovem maintains contact with the recipients and is responsible for monitoring and data collection for the programs they implement. And, finally, external evaluators enter the picture in order to perform the mandatory evaluation after five years of life of the instrument.

Although the existing law does not mention any obligation concerning ex-ante evaluation, such studies are considered necessary in order to weigh different options to solve the problems at hand. Ex-ante evaluation therefore takes place often, but, contrary to the ex-post studies that are always externally commissioned to consultant firms, it is done by an internal commission to justify which instrument is the best solution to the problem, normally using existing evaluations. Most of the time, in order to be able to perform ex-post evaluation, a zero-based study is conducted before implementation. The zero-base study gives a picture of the situation before the implementation and it is mainly done for the thematic programs.

The Ministry of Economic Affairs is, however, ultimately responsible for the evaluations performed and is involved throughout the process. It is the Ministry that determines the terms of reference for evaluation and the performance indicators. It is also the Ministry who is responsible for organizing the impact evaluation, even though an independent panel supervises the research carried out by a private consultant, thus making a clear distinction between the evaluation (consultant) and monitoring (organization involved in implementation – SenterNovem).

Ex-post evaluations are therefore performed by external independent private consultants, but internal checks and balances guarantee that the private consultants address the proper questions and there is an independent panel responsible for the quality and independence of the research.

Different evaluation techniques and models are used depending on the data available and the subject addressed and because:

> *Sometimes evaluations are just in time in relation to the policy making cycles, but there is a kind of tension between good evaluations and the policy cycles as the full impact of different programs are only visible far beyond the time horizon of politicians in charge.*

During our interviews, one of the most discussed topics was the definition of objectives, in particular the process of setting objectives and suitable indicators, the resources available to achieve them, and the stated expectations to attain the objectives. It is in fact the Ministry that defines the objectives and the performance indicators for both thematic programs and basic instruments, but the targets themselves are discussed with the business community and researcher in order to make them as realistic as possible. The setting of intermediate and reachable goals is done by the Ministry in cooperation with the implementation agency. The Unit Manager of the DG for Enterprise and Innovation, Luuk Klomp, states that:

> *The targets are always quite measurable and time-specific. They are not, however, always realistic. For example, the Lisbon strategy sets the target of 3 per cent R&D financed by the private sector, which has been taken as the objective in Dutch policies, but is not realistic in the Netherlands, where the amount of private R&D expenditure has been stationary at 2 per cent for 10 years. So the target is not realistic, even if it is relevant, measurable and time-specific. The problem is that although we really specify the targets everything depends on much more than just the policy and the program. If*

you don't reach the goal, it does not necessarily mean that the policy doesn't work, and it is difficult to find out what the reasons are.

Moreover:

If the goals are externally set, it is much harder to define the right amount of resources to allocate. While, if the ministry itself defines the goals, it is easier to know where you are and where you can get to. The goals set by the ministry are more realistic and the resources allocated more adequate.

Another hot topic was the availability of data for evaluation purposes, which leads to the need to use different types of methodology in order to find better answers to the questions. The data collected during the implementation process usually consists of basic information and it is only after five years, at the time the evaluation is due, that more rigorous information is collected via questionnaires.

Finally, in the Netherlands, the various ministries complete an annual Central Government Evaluation Survey (EOR) to inform the Court of Audit and other organizations about the evaluation studies planned, underway or completed in a particular year, and about their regular performance data systems. The EOR is a useful instrument that, among other things, can help ministries form an impression of the extent to which their policy is covered by evaluation studies and regular performance data systems, and informs external parties about the evaluation instruments used.

As mentioned, we have interviewed different actors involved in the process of design, implementation and evaluation of state aid policies and it is therefore relevant to point out the main outcomes of our interviews.

SenterNovem is the agency of the Dutch Ministry of Economic Affairs in charge of implementation of R&D and innovation policies. As stated on their website, their core competence is "converting government policy into reality" and, on behalf of the Dutch government, they implement policy regarding innovation, energy and climate change, and environment and spatial planning.

Here we report part of our interview with Hendrik Blanksma, Innovation Intelligence and Coordination, from the Directorate of Innovation.

In recent years, the agencies started to have more influence on the programs themselves and now parts of the economic affairs policy making have been transferred to SenterNovem to enable closer interaction between policy making and execution. Now SenterNovem is more active with monitoring and reporting on effects of the program and evaluation.

SenterNovem participates in the discussion of defining the goals and the way in which they will be measured. We use the SMART criteria and try to apply the goals in the long-term analysis. We look at the results of the activities, projects and programs, and always have to refer to the goals defined by the Ministry. When the Ministry reports to Parliament, they often speak about the results in the field, whether they are the results of the program or not. Often Parliament asks: how is the cooperation between business and universities working now compared to 10 years ago? And they state the difference. The next question would be: how much did the program contribute to the changes in the picture? But this is not always the most important question or the most important part of the report. It is often the general trend that is important for the Ministry.

We work at three levels: monitoring, registration of the effects and evaluation. The monitoring is done continuously for each program. It concerns only the companies that received the funding; the control group is not followed. Then we have to look at the effects of the individual projects of the program, and SenterNovem helps the individual projects to register their effects. Then there is an external evaluation of the program for which SenterNovem provides part of the data.

The Ministry does not clearly state the budgeting for evaluation: in some programs, the resources for evaluation are not clearly stated and neither are the resources for monitoring and reporting.

CPB (Netherlands Bureau for Economic Policy Analysis) is one of the consultants in charge of evaluation of public policy. CPB is a research institute that is independent with respect to content, but it is formally part of the central government and its funding comes from public resources from the budget of the Ministry of Economic Affairs. CPB conducts independent economic analyses to inform politicians and policy makers as well as societal organizations, the scientific community and the general public.

In our interviews, Marc van der Steeg and Björn Vroomer, researchers at CPB's Department for Growth, Structure and Knowledge Economics, represent the "voice" of evaluators and we believe it is relevant to report some of their comments on evaluation practice in the Netherlands.

CPB mainly does ex-post evaluation, but sometimes we are required to do ex-ante evaluation as well, for instance, ex-ante cost-benefit analysis when there are policy proposals to invest in certain infrastructure projects, knowledge projects, and so on.

CPB does evaluations commissioned by the Ministry as well as evaluations on its own initiative. We are very keen on doing experimental evaluation with random control, and we would like these evaluation methods to be extended in policy making and become routine.

In general, we are involved only in the last step, when it is time to evaluate the program, so we often have to state the impossibility of evaluating a certain program because the data available are not good, or we have to point out the lack of data and the drawbacks. The Ministry often requires only that we identify success or failure factors, and they are not interested in causality or additionality effects. They attribute the results to the policy even though they are not able to control for other factors, and this is a problem.

As we evaluate a lot on our initiative, we evaluate when we see an opportunity to convince the Ministry and to give them an example of how things could work. With the evaluations done on our own initiative, we try to demonstrate that if the Ministry designs the policy in a certain way it is easier for us to do a good evaluation and say a lot more about causality.

Generally the objectives are not very clear, not specific and stated too generally. We need to translate the objectives into measurable indicators.

Behavioural additionality is a new concept and we have tried to create a questionnaire and to identify some constructs that indicate the perception. It is a promising evaluation field but it is not as strong as the experimental set up; in R&D and networking studies, the behavioural additionality is a good way to measure networks and social capital, but there is the problem of socially desirable answers.

Evaluation culture: not many people are accustomed to doing evaluation and thinking about it when designing policy. In the Netherlands, especially in the education ministry, there are some people who are convinced of the usefulness of evaluation and they think about it when setting up a program. It is becoming more and more a part of the mentality.

The main problems are policy design, lack of data, lack of control groups, and lack of studies before implementing the program.

The design of the policy is crucial. Moreover, data should be collected both for the control and for the experimental group, but if you do not have a good design then you could collect a lot of data and still not manage to carry out a good evaluation.

The Ministry is required to do some kind of evaluation, not necessarily an evaluation with experimental and control groups. Keeping evaluation in mind during the set-up of the policy can be very useful but that would be the ideal world and the Netherlands is not the ideal world. The basic evaluation culture is changing, but there is still a lot to do and we are now trying to convince them that experimental evaluation is better.

The Ministry should avoid cherry-picking when the results are presented, and they should take into consideration all the effects of the policy – not only those concerning the main objectives stated.

For a more in-depth discussion on programs and methods, see Appendix 2.

2.1.2 Ireland

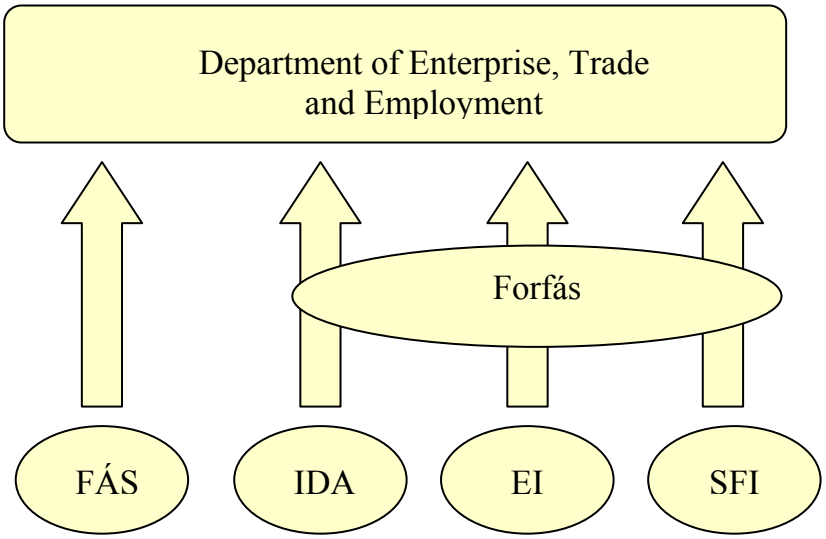
During the 1950s and 1960s, Ireland was noted for its high degree of protectionism, which it had put in place to protect its domestic industry. Ireland then became a member of the EU, which brought about major structural changes – but cold winds started to blow too, since Ireland also found itself having to open up its economy to the rest of Europe. In the mid-1980s, however, the macroeconomic situation began to improve. The government assembled all the relevant players (which was unheard of in Ireland at the time) to discuss the country's economic policy and bring about consensus on what would have to be done. Specific industrial investments and foreign direct investment (FDI) increased once Ireland had pushed through reforms, both in education and in its labour market policies.

Ireland's economic development has been radically transformed since the mid-1980s; indeed over the past 15 years, the country's economic development is without parallel at any other time in its history. In comparison to other EU nations, Ireland's economic growth has broken all records, giving rise to the concept of the "Celtic Tiger". From the late 1980s, the abundance of transfer payments from the EU made a strong contribution to Ireland's development. Follow-up, evaluation, supervision and auditing are all important components of a system for studying and analysing the results and effects of public development initiatives. What role is played by evaluation in Irish economic and regional policies?

Information was assembled and interviews were carried out in Dublin on 24–25 April 2007. Interviews were carried out with people at the Department of Enterprise, Trade and Employment: Margaret Ryan, of the State Aid Unit, and Eamonn Culbert and Gerard Monks of the Enterprise Agencies Unit. Further interviews were carried out with people at Forfás (the National Policy and Advisory Board for Enterprise, Trade, Science, Technology and Innovation), Jaqueline Allan and Seamus Bannon, and with Gerahard O'Brien of Enterprise Ireland.

In Ireland, economic and regional policies are intimately linked with labour market and research policies and are controlled by a single, shared department – the Department of Enterprise, Trade and Employment – and its associated agencies and authorities (see Figure 1). Economic policies are primarily the responsibility of Enterprise Ireland (EI), an agency subordinate to the Department that disposes of an annual budget corresponding to 2.6 billion SEK. Other important agencies are the Industrial Development Agency (IDA), which is responsible for the promotion of direct foreign investment, and Science Foundation Ireland (SFI), whose remit is to stimulate the Irish research system. At the implementation level, there is also an authority that handles practical labour market issues for both individuals and companies, the National Training and Employment Authority (FÁS). A central authority operating under the auspices of the Department is Forfás, which supervises, investigates, analyses and evaluates policies, and advises the Department. Forfás has now been carrying out evaluations for some 15 years. Forfás is also responsible for coordinating economic policies. This is carried out by its Board in coordination with the heads of IDA and Enterprise Ireland.[5]

Figure 1 Organization of agencies and authorities.



Ireland's evaluation capacity has been affected by two separate driving forces. Ireland's membership in the EU entailed pressure from outside, while, since the end of the 1990s, the government's ambition to develop an evaluation culture for all areas covered by public spending has built up pressure within the country itself.[6]

Ireland has been a recipient of EU support since the introduction of the structural funds in 1989. These funds finance sections of a number of operative programs linked to an overall policy document (the Community Support Framework plan). In 1989, the demand for evaluations within the funding rules and regulations led the Irish government to set up two

[5] *2003. Review of Industrial performance and Policy 2003. Department of Enterprise, Trade and Employment.*
[6] *Boyle, R. 2005. Evaluation capacity development in the republic of Ireland. ECD Working Paper Series No. 14, June 2005, Washington: World Bank.*

independent evaluation units[7] and, between 1994 and 1999, a third, central evaluation unit based in the Department of Finance. During the 1999–2006 funding period, the influx of structural funds fell, since parts of Ireland could no longer be classified as Objective 1 areas. Following this change, capacity at the central evaluation unit was stepped up, while the two independent units were discontinued. At the beginning of 2007, public evaluation capacity is represented by two central evaluation units at the Finance Ministry and various agencies with an evaluation remit. One unit deals with evaluation of all operational programs in Ireland, while the other performs analyses and evaluations of state-run programs.

In 1996, basing itself on the Australian model, Ireland introduced a program[8] to examine public spending. In 1997, this was followed by an Expenditure Review Initiative (ERI) put together to examine spending by the ministries once every three years. Two main principles were to guide this work: a systematic analysis of what has been achieved by public financing, and the production of basic documentation facilitating decisions on prioritization within and between programs.[9] The purpose of the ERI was to create an evaluation strategy for all ministries and to encourage a culture of evaluation within them that focused on productivity, efficiency, relevance and prioritization of initiatives. A central steering committee (the Expenditure Review Central Steering Committee, or ERCSC) is responsible for supervising ERI monitoring.

After examining the Irish national audit, the ERCSC established a network of persons involved in the ERI system, introduced a quality assurance system for investigations and examinations, and a system for following up the results of investigations already carried out.[10]

However, the main impression is that evaluation capacity is divided when it comes to the resources available for evaluation/analysis of economic policy. With changes over time, the European Commission's demands for completed evaluations and analyses of specific economic factors as the basis for approval of new support programs have brought about increased awareness of the necessity of evaluation and monitoring for new or modified support programs. This development is also partially supported in that economic policy is reviewed at the macro level once every ten years.

State aid in Ireland is to a large extent governed by "maps" approved by the EU and by directives stipulating how the support may be used. The maps of the state aid regions were drawn up by the state aid unit at the Department of Enterprise, Trade and Employment. In addition to designated support programs, there are also certain types of support that fall under the de minimis rules.

Ever since the late 1980s, much of Irish policy has been channelled through programs. In the early 1990s, Irish economic policy was reviewed in its entirety, implementation being systematized in the form of programs. All formulation of policy takes place at the national level.

---

[7] The two units were independent of the implementing organization of the structural funds programs.

[8] The "Delivering Better Government" program.

[9] Department of Enterprise, Trade and Employment. 2004. "Expenditure review of IDA Ireland´s property programme," p. 13.

[10] Boyle R. 2005. Evaluation capacity development in the Republic of Ireland. ECD Working Paper Series No. 14, June 2005, Washington: World Bank, p. 16.

Various types of support are ingredients of most programs. In this context, programs are more than the traditional support programs that often comprise a single form of support or aid. Program thinking permeates the macro level in almost all political spheres. A National Development Plan (NDP) was drawn up for the period 2000 to 2006, and another has been launched for the period 2007 to 2013. The NDP is the primary policy document governing Ireland's development in the future. Subsidiary to the NDP is Ireland's and the EU's policy document on how Ireland is to use structural funding provided by the EU. Linked to the NDP are operational programs and associated national strategies.[11] The national strategy for economic policies specifies the national ambition for various areas in terms of goals, objectives, strategic actions and performance indicators. (See Appendix 2.)

It is Forfás's task to carry out evaluation of economic policy data in a wider context, while Enterprise Ireland focuses on evaluating programs that it itself is responsible for running. In-house evaluations are carried out primarily by Forfás; otherwise, universities, independent consulting firms and audit companies are used. In the case of the programs run by Enterprise Ireland, a program manager is responsible for implementation and evaluation and reports to Enterprise Ireland's board, which has the prime responsibility for the program and in turn reports to the Department of Enterprise, Trade and Employment.

Much of what is done at the national (ministerial) level consists of reviews of state expenditure in various areas. From the Department's perspective, this is an example of evaluation. Otherwise, reviews are referred to the equivalent of the National Audit Office in accordance with the terms of the Public Service Management Act.

A survey of some of the evaluations carried out by Forfás shows that they concentrate on analysis of program input and output. No connection is made with regional economic- or macroeconomic contexts.

Evaluations discussing deadweight are in actuality not possible, since most companies in Ireland are taking- or have taken part in some sort of support program. It is therefore more relevant to discuss partial deadweight, which addresses the effectiveness of a given means of support as measured for a number of output variables by comparison with another group.

One reason why evaluations do not use control groups to any great extent is the availability of data. Ireland does not have the same possibilities as Sweden for the collection of data. The Swedish Institute for Growth Policy Studies (ITPS) keeps a database of individuals and companies (IFDB), which enables companies to be linked with individuals and offers considerable possibilities for analysis. In Ireland there are no such possibilities. An example of this problem emerged during the interviews. One agency is planning an investigation aimed at following the progress of recently enrolled research students in order to gain an understanding of the factors affecting their graduation rate. The investigation is to be carried out following a decision by the Irish government to increase the number of those who actually present a doctoral thesis. Accumulation of data on research students requires that the universities voluntarily make such data available. In Sweden, such data are

[11] *The Statement of Strategy for 2005–2007 establishes goals and identifies strategic actions for economic and labour market policies. A separate strategy has been developed for each political field. For economic policy, a new strategy is established every three years. "Ahead of the curve – Ireland's place in the global economy" appeared in 2004, and a new strategy will be presented in summer 2007. The Strategy for Science, Technology and Innovation 2006–2013 presents goals, strategic actions and the expected effects of policies in the areas in question.*

publicly available from Statistics Sweden. In Ireland, there is no equivalent of the ITPS database, a decisive difference being the lack of legislation on the collection of statistics.

Another reason control groups are not used stems from the resources put into evaluation. Although Forfás is comparable to the ITPS, it allocates only one tenth of the resources set aside by the ITPS for evaluation each year. Such comparison gives some indication of the basic circumstances but naturally does not say anything about the actual situation since it has not been possible to accumulate the data in question.

In Boyle's view, the obstacle that must be overcome in order to develop evaluation capacity and culture in Ireland is the lack of directives ensuring that evaluation can be integrated with decision-making.

2.1.3 USA

Thanks to a very interesting interview conducted on 19 April 2007 with Bob Boerner from the National Conference of State Legislatures (NCSL) and Gary VanLandingham from Florida Legislature Office of Program Policy Analysis and Government Accountability (OPPAGA), we have built a clear picture of the organization of state aid policy in the US.

The first feature to point out is the coexistence of two overlapping levels of legislation: the federal level and the state level. Policy design, policy implementation and policy evaluation present different characteristics when considered at the federal or at the state level and, focusing here on evaluation, there is absolutely no connection between the Government Accountability Office (GAO) and the different audit offices in the individual states except for the fact that, for financial audits, states must follow the guidelines of the Yellow Book. The individual states are completely separate from the federal government, about half of which perform evaluation following the Yellow Book; other states follow the evaluation standards of the American Evaluation Association (AEA) or good research practices. Contrary to the federal level, where evaluation is guided by the GAO's Yellow Book, there are no requirements about guidelines at the state level.

Moreover, like OPPAGA, state audit offices do not have any power in terms of policy design or implementation; they are only in charge of evaluation and they can advise the legislature. The level of communication with the other actors involved can be very different according to the agencies a state has to deal with. The monitoring and data collection process is done inside the agencies in charge of implementation, and the quality of the data depends very much on the agencies' work.

The second interesting element is the autonomy and singularity of the different federal programs and the different degree of importance that evaluation has in each program depending on the policy design.

A very interesting example for our purposes is the Advanced Technology Program (ATP) of the National Institute of Standards and Technology (NIST). ATP is a federal program targeted to help industry invest in longer-term, high-risk research with payoffs far beyond private profit. By sharing the cost with companies, ATP accelerates the development of early-stage, innovative technologies, helping industry raise its competitive potential.

Critical evaluation of the ATP's impact on the economy is an important part of the program. To measure the long-term effects of R&D funds on the economy, the ATP has established economic analysis procedures that are pushing the state of the art in evaluating the long-term outcomes of an R&D investment.

Moreover, it is important to keep in mind that as of 1993 the Government Performance and Results Act (GPRA) hold federal agencies accountable for using resources wisely and achieving program results.

GPRA requires agencies to develop plans for what they intend to accomplish, measure how well they are doing, make appropriate decisions based on the information they have gathered, and communicate information about their performance to Congress and to the public. GPRA requires agencies to develop:

- A five-year Strategic Plan, which includes a mission statement and sets out long-term goals and objectives;
- Annual Performance Plans, which provide annual performance commitments toward achieving the goals and objectives presented in the Strategic Plan;
- Annual Performance Reports, which evaluate an agency's progress toward achieving performance commitments.

Agencies are therefore required to plan their goals and objectives, ensure that resources are available to carry out these plans, measure and assess progress and link resources actually used to results achieved, and report on progress achieved and impacts on future efforts.

The GPRA therefore proposes a new framework for evaluation in an effort to focus the attention of government on the outcomes of its programs.

As Susan E. Cozzens[12] points out:

> *The GPRA clashes with the old framework … in several ways. The first mismatch is between evaluation at the input end of the research process and evaluation of outcomes. […] A second clash is the culture of autonomy versus the culture of planning, which has been reflected in the struggle over qualification of outcome goals. […]The third new element that GPRA has introduced is stakeholder consultation.*
>
> *In sum, GPRA has the potential to make research agencies, program managers and researchers better strategic thinkers, with a clearer sense of how their activities pay off for the public.*

The Advanced Technology Program (ATP)

The Advanced Technology Program describes itself as "bridging the gap between the research lab and the market place." It was created to fund government-industry partnerships to support the development of new technologies with potential application across the American economy.

Since the beginning of the program NIST has stressed the need for assessment and evaluation of the outcomes and the desire for objective analysis of the goals, operations and results of partnership programs. The economic impacts of ATP projects can be measure in

---

[12] *Cozzens S. 2003. Frameworks for Evaluating R&T Policies in the United States. In Philip Shapira and Stefan Kuhlmann (Eds.), Learning from Science and Technology Policy Evaluation - Experiences from the United States and Europe. Edward Elgar, Cheltenham, UK and Northampton, USA.*

various ways – such as productivity gains, new businesses created, benefits in employment and increases in GDP – including both private returns to the company involved in the project and spillover effects.

When evaluating the effects of the ATP program, NIST has been very cautious in tracing the program's impacts along direct and indirect paths. The direct path follows the awardees and includes private returns to the particular companies directly involved in the ATP-funded projects and spillover effects to their customers. The indirect path involves the take-up of the knowledge generated by the project by others, outside the projects, which have not directly contributed to the investment cost. The impact along the direct and indirect paths combined represents what the economist call the "social return" of the project.

To track these impacts, the ATP uses different tools, such as:

- Statistical profiling of applicants, projects, participants and technologies;
- Progress tracking of all projects and participants;
- Status reports for all completed projects;
- Detailed microeconomic case studies of selected projects and programs;
- Econometric and statistical studies of innovation, productivity and portfolio impacts;
- Macroeconomic analysis for selected projects;
- Special issue studies;
- Development and testing of new assessment models and tools.

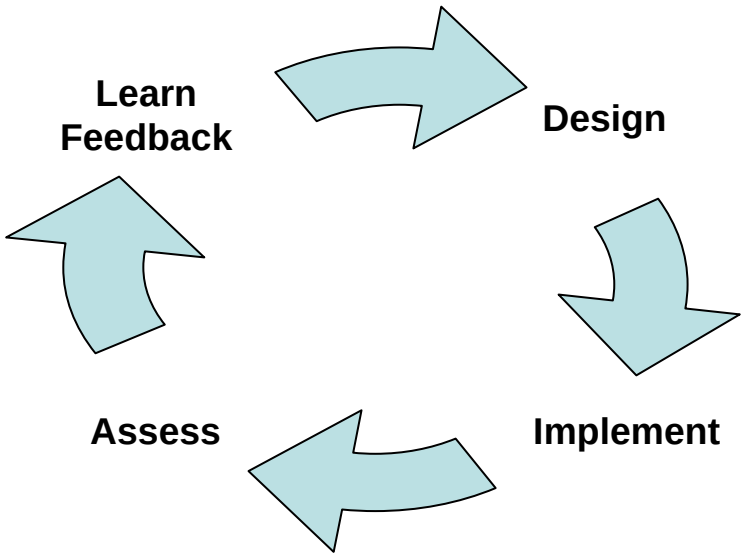
As Irwin Feller from Pennsylvania State University states:

> *ATP has gone beyond the efforts of other programs that have sought to measure direct benefits, by trying to measure indirect or spillover effects. Measuring these impacts is a difficult task. ATP assessment techniques are at the "state of the art" and in many ways have advanced the "state of the art".*

An evaluation effort for the program was put into place for two reasons: (1) as a management tool, to meet program goals and to improve program effectiveness; (2) to meet the many external requests for ATP program results. Evaluation is a powerful tool when it is integrated into program management and to maximize effectiveness it has to follow four steps:

- Design
- Implementation
- Assessment
- Learning and feedback

Figure 2 Program management's cycle.



Source: www.atp.nist.gov

However, as shown in Figure 2, this is not simply a linear sequence, but a cycle in which assessment, lessons learned and feedback would be reflected in appropriate modifications to program design, implementation, etc. Program evaluation is therefore fully integrated in the program's dynamic structure and contributes to continuous program improvements.

Rosalie Ruegg at NIST asserts that:

> *There are some basic principles to follow in setting up an evaluation program. The basic principle is to measure against the mission. Another important principle is to link, in a systematic way the program's activities to its mission: the output to the activities and the shorter and longer run outcomes to the outputs. This is called "developing an evaluation logic model."*

2.2 Lessons from other countries

It is clear that the previous description of evaluation practices in other countries triggers a deeper discussion. At which level should the evaluation be done? How important is ex-ante evaluation? What is a reasonable time period for evaluation? What is the current state of evaluation in Sweden? Is it enough to monitor the results?

The Dutch experience offers several conclusions concerning the evaluation element of the policy process.

1. A policy design that is well-thought-out offers a good starting point to obtain convincing evidence of the causal link between the policy instrument and its output.
2. Detailed data collection from both the firms that are able to benefit from the policy and those that are not is essential. Without information on a control group of non-users, it is very difficult to find convincing evidence of effectiveness. And without

clear and specific questions and concepts, the response rates from the surveyed SMEs will be low and the responses difficult to interpret.

3. Effects that come into play over the longer term can only be observed over the longer term.
4. Certain policy instruments may also affect the behaviour of non-recipient firms and this should be taken into account to avoid over- or underestimation of the effectiveness of different policies.

In Ireland, program thinking has been systematically implemented in several policy areas. From an evaluation standpoint, programs have both advantages and disadvantages. A program comprising many different areas and actions is generally more difficult to evaluate. An advantage, however, is that programs are limited and transparent in what they intend to achieve, which makes evaluation easier. Moreover, Ireland's system of "high level objectives" and a number of macro-indicators that are systematically followed up and reviewed is more beneficial to assessments of accountability. In most evaluations, the aim has been not to evaluate effects but to evaluate means of improving the support program. To this end, peer reviews have been used to acquire a deeper understanding of how support is used. The role of evaluation as a learning factor is not without significance.

Finally, the most interesting lesson learned from the USA can be easily summarized in the words of two of the interviewed people. Using Gary VanLandingham's words (OPPAGA):

> *…the legislator can only pay attention to a certain number of issues at any time, so, if you evaluate everything, the legislator will not care and the government will not take into account most of the results…. We have to ask the legislator what he cares about in a specific program so that we know which questions we have to answer and what the objectives of the evaluation are. Generally, the legislator does not set the objectives in a very specific way because the programs are typically designed as a political compromise.*

In Kathleen McTigue's words (NIST-ATP Program):

> *We can provide some recommendations: increase retrospective analyses; incorporate both direct and indirect path analysis in cost-benefit studies; continuous monitoring and update of information; continuous development of evaluation techniques; identify and address new questions that arise from evaluation; analyse both failures and successes; use an effective mix of internal and external evaluation studies; and, finally, take greater advantage of evaluation results in decision-making processes.*

2.3 Evaluation of economic policy in Sweden

Does Sweden approach evaluation in the same way as the Netherlands, Ireland and the USA? In this section, we will look at the formal Swedish regulations pertaining to evaluation and examples of steps taken by two government agencies to improve evaluation in their own particular fields.

2.3.1 Regulatory evaluation framework

The first two subsections of the State Budget Act (1996:1059) are formulated as follows:

> *§ 1 All State operations and activities shall be carried out with a view to achieving a high degree of effectiveness while ensuring good economic management.*
>
> *For the purposes of this Act, State operations and activities are defined as any operations and activities managed by the Government, the courts and the administrative authorities subordinate to the Government.*
>
> *§ 2 The Government shall inform the Riksdag (Parliament) of its intended goals and objectives and of the results achieved in various areas of operation.*

The 1996 Budget Act brought sweeping changes to public administration. Among other things, the Ministry of Finance implemented a project on financial management instruments for central government (*Verktyg för ekonomisk styrning inom staten*, or *VESTA*), which in 2000 published a ministerial communication on the effectiveness and transparency of economic management (*Ekonomisk styrning - effektivitet och transparens*, 2000:68) as part of its development program.[13]

A report discussing the further implications of the Budget Act with respect to evaluating and developing the budget process (*Utvärdering och vidareutveckling av budgetprocessen*, SOU 2000:61) states that subsection 2 shall be understood as requiring the Government to inform Riksdag of its intended objectives and the results it has achieved.

> *The Budget Act requires the Government to inform the Riksdag of its intended objectives and the results it has achieved in its various operational fields. According to the budget act bill, the aim of this provision is to establish the significance of management by objectives and results in all government operations and activities. The intention of management by results is stated to be both to make clear to recipients of government funding that they must state what it is they aim to achieve and to afford the various parliamentary committees greater scope for follow-up and evaluation.* (SOU 2000:61, Subsection 2, page 35)

The Riksdag Act was amended in the following year (Swedish Code of Statutes 1974:153, s. 4, ss. 18), such that, as of 2001:

> *In considering the matters at hand, the committees shall be expected to follow up and evaluate decisions of the Riksdag in the areas indicated in ss. 4–6 and supplementary clauses as pertinent to each individual committee. (SFS 2003:180)*

In 2002, to support Riksdag and its committees, the State National Audit Office (*Riksrevisionen*) was formed. This is one of Rikdag's instruments of control. However, the creation of the new State National Audit Office, following closure of the previous National Audit Office (*Riksrevisionsverket*), also saw a decrease in Government's evaluation resources.

2.3.2 Management by objectives

The relationship between Government, its authorities and agencies, and Riksdag is regulated not only by the Budget Act and Riksdag Act but also by the Government Agencies Ordinance (1995:1322), the Public Administration Act (1986:223) and the Swedish Consti-

[13] *See also government communication Skr 1000/01:151.*

tution. From an evaluation standpoint, it is the Government Agencies Ordinance that is of the greatest interest, as this is the document controlling how the activities of government authorities and agencies are organized.

Subsection 7 of the Government Agencies Ordinance states that it is the duty of directors of government agencies to:

1) be properly economical with government funds;

and to:

3) continuously follow up and examine the activities of the agency and the consequences of the statutory regulations and special decisions pertaining to the agency's activities and to take any steps that may be deemed necessary.

The concept of effectiveness is incorporated in the board's responsibility to ensure that the work of the agency is carried out in an efficient manner. The concept of objectives is also present, although only in the requirement that employees be familiar with the objectives of the agency's operations.

Subsection 31 of the Ordinance specifies the information that must be included in the document (documentation) drawn up for every decision by a government agency. From this, it is apparent that, under Swedish law, such documents need contain only:

- The date;
- The substance of the decision itself;
- Who made the decision;
- Who presented the matter;
- Who took part in processing the matter but not in the actual decision.

In addition to the above legislation, each individual agency is governed by instructions (in the form of an SFS), annual Government Regulations, and special assignments.

In the case of special assignments, an agency's operations are controlled with regard both to the resources consumed during the operation – financial control, and to performance and effects – results management. Together, the two approaches are known as "economic control". This, essentially, is the tool used to maintain control of government finances to achieve the distribution of resources decided by the government's political priorities and to be able to report back to Riksdag as required by the Budget Act. Competing distribution ambitions and a tight budget mean that the demand for effectiveness in the use of national funds is of no small importance.

The above-mentioned ministerial communication (Ds 2000:68) suggests that the Riksdag has demanded that the government's objectives can be measured and followed up and that it must be meaningful to set performance against cost. The so-called SMART criteria have been proposed as a suitable norm for the drawing up of policy objectives. These criteria are given in the table below.

Table 1 SMART criteria for policy objectives.

Specific	It must be clearly stated exactly what is to be achieved.
Measurable	It must be possible to follow up results using result indicators, key ratios or the like.
Accepted	They must be accepted and perceived as relevant by those who will be implementing the operation in question.
Realistic	They must be possible to achieve.
Timetable	The time by which the objectives shall have been achieved must be clearly indicated.

Source: DS 2000:63, page 54[14]

The said communication also notes that the SMART criteria can be followed more extensively at lower levels of government, i.e., below political area levels such as operational field and operational specialty level. This proposal is fully in line with developments now taking place all over the world along with demands for increased transparency and clarity in public administration. In Sweden, however, the proposal has not been enacted in law, although the formulation of the annual government regulations and the work carried out within the agencies themselves (see below) are developing in accordance with the objectives of the SMART criteria.

One of the problems inherent in management by objectives is the relationship between an objective and the resources earmarked to achieve it, either partially or in full. It is up to the elected government to implement its policies and define its objectives accordingly. The problem is that these objectives can be formulated at different levels. Some objectives are more visionary and can be linked to follow-up indicators only with difficulty. Previously, Swedish IT policies aimed at "becoming the first to offer an information society for everyone," however, without any part of that ambition having been made clear in the bill adopted by Riksdag.

A second problem lies in the difficulty of linking visionary objectives with the limited resources available to the agencies. The Swedish solution has been to hold yearly talks between the government offices and the agencies in so-called "objectives and results dialogues" (as discussed in the above-mentioned SOU 2000:61).

Yet another problem of management by objectives is that the more concrete the objectives set up by the government, the more the work needed to follow them up, which thereby tends to be distributed as follows: the agencies must interpret the government's objectives, translate these into concrete measures within their own areas of operation, obtain consensus from Government, and, finally, report on the results they have achieved. Government, for its part, must assess the "interpretations" submitted by the agencies and determine whether these appear to offer a successful means of achieving its (Government's) objectives.

A suitable tool for facilitating adaptation of SMART criteria to operations and political fields is what is known as *operational logic*. The National Financial Management Authority has described the implications of this in a report (ESV 2001:16). Operational logic is one of several names used to refer to a method of analysing the link between selected measures and the objectives those measures seek to achieve. Operational logic analysis aims at identifying hypothetical courses of events and the underlying suppositions for selected measures; it elucidates the probability of a certain outcome and the extent to

[14] *There is also an English version of the SMART criteria that has been recommended by the European Commission since 2002. See reference in "SMART Innovation – A practical guide to evaluation innovation programmes."*

which the measure implemented has played a role in producing it.[15] Operational logic is discussed in greater detail in the following sections.

2.3.3 What are follow-up and evaluation?

In the absence of markets, public administration needs a tool for its effectiveness audit. Follow-up and evaluation are generally the tools that are used. However, there is no common standard stating exactly what follow-up and evaluation mean in practice. To put it simply, we might say that follow-up and evaluation can be related to results produced by public measures taken within a certain period of time. In DS 2000:68, follow-up and evaluation are stated as being set in relation to predefined objectives, meaning that the results achieved are compared with a set of objectives established in advance.[16]

By follow-up and evaluation, we mean quite simply that we describe and analyse how the results or lack thereof of public measures may best be understood. The purpose of follow-up and evaluation has long been to provide documentation enabling us to decide whether the public measure should be continued, perhaps with some slight adjustment of level, or discontinued.

Effect and effectiveness are key terms that must also be considered in this context. Effectiveness can only be related to some form of reference norm. Unless this norm is defined, we cannot talk about effectiveness. A public measure can only be said to have had an effect if it can be shown to have achieved something that would not otherwise have been achieved within the period during which it was intended to produce a result. Effectiveness in the sense of cost-efficiency means that public measures must be produced at the lowest possible cost.

Of all the huge amount of literature on evaluation, Vedung[17] has become the international standard.[18]

Vedung defines evaluation as:

> *Careful ex post facto assessment of outcome, final performance and administration in public programmes.* (Vedung 1997, p. 33)

Vedung views evaluation as comprising two main components: *qualified follow-up* and *effect measurement* (ibid., p. 32).

The EU divides evaluation into *ex-ante*, *mid-term* and *ex-post* stages. Documents on the suitability of a public measure must be produced before implementation (*ex-ante*), during implementation (*mid-term*), and after a measure has been concluded (*ex-post*). This break-down makes for a practical approach to the evaluation procedure.

---

[15] The concept goes under a variety of names, such as program theory, events chain and change theory. The concept of "balanced scorecards", used in managerial models, is a close parallel.

[16] For a discussion of standards, see www.fteval.at for examples of what evaluation should contain. However, there is a conflict in the evaluation world between "qualitative" and "quantitative" methodology (not method), which also Swedish authorities encounter in inquiries for evaluation tenders.

[17] Vedung, E. 1997. Utvärdering i politik och förvaltning (Evaluation in Public Policy and Public Administration), Studentlitteratur.

[18] Two current reports ordered by EU DG Enterprise are "Review of methodologies to measure the effectiveness of state aid to SMEs" by Mosselman, Prince and Kemp, and "SMART Innovation – A practical guide to evaluation innovation programmes."

Vedung's terminology differs from that of the EU in two respects. One is that Vedung stresses that evaluation is something that takes place after a public measure has been carried out/initiated, preferring to replace the term *ex-ante evaluation* with the concept of *previewing* or *impact analysis*. And, secondly, Vedung considers qualified follow-up to be more than just follow-up of individual result variables; instead, it also includes operational logic, which we will be discussing in greater detail below. *Effect measurement* focuses in turn on the causal relationships between public outcomes/results and desirable consequences.

It is not possible to determine which terminology is *a priori* most accurate. In purely pragmatic terms, it would be quite reasonable to adapt oneself to EU conventions while striving to develop the quality content of each of the evaluation stages, regardless of whether they are styled ex ante, mid-term or ex post.

Two concepts we have already touched upon, and which are of significance throughout, are *operational logic* and *effect*.

2.3.4 Nutek's and Vinnova's evaluation strategies

Evaluation strategies in economic and innovation policies have changed over the past few years. Vinnova (Swedish Governmental Agency for Innovation Systems) is expected to report the effects of its (and its predecessors') programs. Nutek (Swedish Agency for Economic and Regional Growth) has restructured its organization partly to afford the management greater influence over the design of new measures.

Both Nutek and Vinnova have recently produced written descriptions of how they envisage the development of follow-up and evaluation methods at their agencies in the future. According to ITPS, both documents represent a step in the right direction in enabling us to determine more systematically whether the economic and innovation policy measures for which these agencies are responsible actually have an effect on the national economy.[19]

The following paragraphs should be seen as a description of follow-up ambitions in Sweden for comparison with similar descriptions by other countries named in this document and should not be regarded as an evaluation of the strategies in themselves.

Nutek

Nutek's *Nytta* discusses the SMART criteria described above and emphasizes the importance of well-formulated objectives. It also takes up the concept of *change theory*, which is the same thing as the concept of *operational logic* named above. Nutek also describes differences between the activity, results and effect of measures, and states that indicators for these should be identified as reflecting quality demands in selected indicators. *Nytta* points out the significance of formulating a follow-up and evaluation plan in which all the various stages can be properly documented.

[19] *Nutek 2006. Nytta – Nutek's system for follow-up and evaluation.*
Vinnova 2007. Vinnova's overall strategy for follow-up evaluation and effect analysis, Dnr 2006-04079.
Vinnova has previously published an outline for a strategy in Winqvist (2003), "Strategies behind Vinnova's evaluation policy," Fteval no. 19, and an outline for a follow-up system for research sections of innovation policies in Winqvist (2006), "Approach to a national system for monitoring university research in Sweden," Fteval no. 27.

In organizational terms, Nutek has changed since the 1990s, when analytical programs were run in parallel with operational activities (which, in Nutek's new terminology, are known as "supply processes"). Analysis – comprising analysis, new developments and evaluation – is now part of the agency's management process and thus stands above operational activities. The only purpose of these latter is to perform certain specific assignments and develop new measures, and only together with colleagues from the management process.

Nytta gives examples of the desired procedure, summarized in Table 2, below.

Table 2 Nutek's organization for planning and follow-up of new measures.

Phase	Activity	When	By whom
Planning	Objective formulation	All inputs	New development process
	Change theory	All inputs	
	Define indicators	All inputs	
	Establish follow-up and evaluation plan	All inputs	Established by colleagues in the supply process in consultation with colleagues in evaluation
	Reference measurement	On identified need	
	Advance assessment (ex ante)	Large, complex inputs should have an independent ex ante	
Follow-up		All inputs	Performed by colleagues in the supply process in consultation with colleagues in sub-process evaluation
Effect evaluation		On identified need	Performed by evaluation in consultation with colleagues in the supply process
Evaluation		On identified need	Performed by evaluation

Source: Nutek (2006, p. 21)

Nutek uses a schematic example to show how a fictive program is concretized in outcome objectives, result objectives and effect objectives, and gives a description of a change theory.

Vinnova

Vinnova's focus on innovation policy, in the form of both support for pure research in certain fields (needs-oriented research), intended above all as support for skills centres, and support awarded for stimulating increased research in small companies, presents special challenges to identification of results and effects.

Vinnova's report is similar in content to Nutek's. Like Nutek, they place information collected from follow-up and evaluation in a system context. The information gathered is intended to describe the contribution of all the agency's measures and how they interact. It is emphasized that, since many of Vinnova's measures aim for effects that can only be demonstrated in the long term, it is necessary to be able to collect information for up to 20 years. This report, too, draws attention to the significance of the fact that:

> *…the planning of the operation also embodie[s] very clear effects logic, i.e. a review of the effects to which the operation is to give rise in the long term and what the road there looks like, or the mechanisms that can be supposed to be of significance for the outcome. (Vinnova DNR2006-04079, page 14)*

The authors also stress the importance of ensuring that a measure's effects logic is based on reasonable assumptions as to the nature of the relationship between measure and outcome. Vinnova devotes more space than Nutek to a discussion of developments over time. One diagram, for example, shows how national economic effects accumulated over time can only be expected to get off the ground several years after termination of the project.

Unlike Nutek, Vinnova's evaluation strategy does not indicate any organizational changes in the agency itself that might serve to promote implementation of the evaluation strategy. The report notes that the organization is "decentralized", by which is meant that the responsibility for follow-up and evaluation of concrete measures lies with the department, unit or officer responsible for implementation.

To help promote follow-up and evaluation, Vinnova recommends instead the development of a norm (uniform work procedure, Vinnova DNR2006-04079, p. 31) for handling programs, application calls and policy development through so-called "pilots". In addition, Vinnova is of the opinion that support, in the form of a functioning IT system enabling flexible input, processing and analysis of project data, will contribute towards more effective follow-up of individual projects along with comparisons between different program inputs. Systematic data capture in all R&D programs is a measure that they would like to see introduced as soon as possible. Support for the departments is also available organizationally from Vinnova's own evaluation expertise, above all the Department for Strategic Development. The report also notes that there is a special group, the QA group, charged with supporting departments in the formulation of program documents, with regard to predicted results and effects, and with the formulation of application calls. The report states that "a clear-cut organization for the development of Vinnova's effects analyses is under construction" (Vinnova DNR2006-04079, p. 21).

Vinnova participates in several international networks, including Taftie, with a view to developing its skills in evaluation and analysis.

Nutek's and Vinnova's work on clarifying the operational logic of the measures they initiate is welcome. Although Sweden has probably been no worse than other countries in inspecting its public measures, the move from the previous view of analysis being a supporting activity to the notion that "analysis is one of the agency's strategic tools" is a change that brings with it the potential for moving towards more effective programs. It is naturally too early as yet to judge how the agencies may develop in the future, but, based on these documents, the ITPS considers that discussion should be initiated in the following areas at least.

Deepen our understanding of what an effect may imply

The first area for discussion is that agency managements show insight into the fact that the "results" achieved by an agency are not identical with the "effect" of the agency on the national economy. The authorities must strive to develop their ability to distinguish between what can be called effects and what may be styled results. For Nutek, the choice of the concept of "effect evaluation" is significant, although they have allowed it to be

used for the support recipients' own evaluations of the support they have received from the authority.[20] This procedure is commonly adopted when evaluations are commissioned from agencies or authorities. However, it provides insufficient proof of the presence and impact of effect and has met criticism in the evaluation literature.[21] It is possible that Nutek believes that more causally valid methods will be used under their "evaluation" concept, but this remains to be seen. However, a dilemma arises in that Nutek notes that, for government assignments, "effect evaluation" is to be carried out two years after the end of the program. The question then arises as to what "effect" each of the parties is talking about. One way of supplementing subjective measurements, says Nutek, would be to introduce measurement of the starting point, although no such measurement is referred to in the effect evaluation of the product development program.

Above, we saw that, in this context, the concept of "effect" expresses a relationship with a specified norm or, in other words, the counterfactual situation. Just like all other counterfactual statements, "effect" conclusions are conditional on measurements of the starting point and how well the relevant starting point has been specified.

Vinnova's report makes mention of the concept used by the EU for effects, "additionality", but does not refer to the distinctions made between "input", "behaviour" and "output". Of these, output is the most desirable but at the same time the most difficult, which has sparked an interest in studies of its effects on behaviour and input strategies.

Include risks in operational logic

In the view of ITPS, program design and outcome depend on whether the agency succeeds in identifying relevant change logic. A dilemma of both Vinnova's and Nutek's reports is that neither takes up risks and events that lie beyond the agency's control but may affect what the agency itself intends to affect. It is probably realistic to assume that there are influential factors that can take the edge off the measure in question or drown the "effect" in noise, e.g., a strong upswing in the business cycle that brings with it greater potential for financing previously risky projects. If descriptions of operational logic are to stimulate the development of relevant measures, factors such as these must be identified.[22]

Discuss populations

Measures making up part of the national economic and innovation policy, especially those taking the form of financial support for companies, are relatively small in comparison to the overall investment volume of the business sector. How is it believed that they will make any difference? Agencies and other public players, such as ALMI, Innovationsbron and Industrifonden should relate their measures to relevant populations. In the above-mentioned effect evaluation of the product development program, it is stated that in 2005 Nutek distributed 60 million SEK among 322 companies. Nowhere in the report does it state how many companies there were in the group of companies legitimately entitled to seek support. Nor is it stated how many of these have received the information.

---

[20] Nutek 2006. Effect evaluation of the product development programme, file no. 70-2006-4350.

[21] Storey, "Six steps to heaven".
Duignan, P. 2007. Seven possible outcome evaluation designs,
http://www.parkerduignan.com/oiiwa/model/

[22] See ESV 2001. Verksamhetslogik (Operational logic), ESV 2001:16. See also ESV 2006:7, Måluppfyllelseanalys (Target fulfilment analysis), p. 27 ff.

Any discussion of operational logic must reasonably incorporate a discussion of the potential for increasing the scope of the program, assuming that it is successful. However, this presupposes an analysis of representativity, i.e., the makeup of the group of companies included in the program.

Invite outside expertise to participate in the discussion

The second of ITPS's views pertains to public access to documentation. Neither Nutek's nor Vinnova's report indicates the extent to which they intend to make documentation available for discussions with outside stakeholders. Here, it would be of interest above all to discuss causal links and any hidden assumptions before launching the measure. In the view of ITPS, one of the agencies' strong points is that they make use of the customary reference group procedures here as well. Vinnova, for example, recently used an extensive reference group in producing a proposal for boosting the innovation potential of small and medium-sized companies. When this strategy becomes operational, the operational logic derived from it should be discussed with outside experts.

2.4 Challenges

The need for a common framework and the potential for mutual learning through the comparison of practices in different countries have inspired the OECD to develop guidelines that allow international comparability. According to this framework, programs should be evaluated with regard to their efficiency in a number of areas, including:

- Appropriateness: Does the program address an important objective?
- Superiority: Is the program more effective than other policies, programs or instruments that would achieve the same goal?
- Systemic efficiency: How does the program interact with other programs and to what extent does its efficiency depend on conditions created by other government actions?
- Own efficiency: Is the program cost-effective?
- Adaptive efficiency: To what extent have results from evaluations been fed back into policy design and implementation? Does policy design ensure a degree of flexibility in responding to unpredictable changes?

In addition, OECD stresses that the evaluation methodology should be clearly stated, evaluation schemes should include cost-benefit frameworks, measurable criteria should be used to determine success where possible, the counterfactual scenario should be used as a way of describing what would have happened in the absence of a program or policy, and the results of evaluations should be fed back into the program design in order to improve it. According to OECD's reports, this last point has been the weakest aspect of evaluation practices in many OECD countries.

When discussing the existing challenges for evaluation in Sweden, we have to keep in mind the OECD framework and focus on the actors involved and the current practice of Swedish evaluation for both ex-ante and ex-post evaluations.

The OECD framework of evaluation is particularly interesting for ex-ante studies, in particular with regard to appropriateness and superiority. The main goals of an ex-ante evaluation are:

1. To describe a problem, most likely due to market imperfection, policy imperfection or system imperfection.
2. Investigate alternative ways to solve the problem and suggest a program logic, which allows solving the problem step by step.

Moreover, the OECD framework emphasizes the impossibility of conducting an evaluation in the absence of clearly specified objectives. Governments should set objectives, indicating which, if there is more than one, takes priority. Once the objectives are set, then quantitative targets need to be specified. Only then can evaluation take place.

David Storey (2000) provides an analytical framework to classify a variety of types of analysis. As shown in Table 3, Storey's framework identifies six approaches, where steps 1–3 are viewed as monitoring and steps 4–6 as evaluation.

Table 3 The six steps.

MONITORING	
Step 1	Take up of schemes
Step 2	Recipients opinion
Step 3	Recipients view of the difference made by the program
EVALUATION	
Step 4	Comparison of the performance of recipients with non-recipient firms
Step 5	Comparison with *matched* firms
Step 6	Taking account of selection bias

Source: Storey (2000)

Given this simple framework, it is now interesting to ask ourselves the following question: Are Swedish agencies (such as ITPS, NUTEK and Vinnova) performing *evaluation* as defined in Storey's steps 4–6? What Nutek defines as "effects evaluation" (*effektvärdering*) is rather subjective evaluation in which recipient companies are asked (by means of questionnaires and interviews) to judge the effectiveness of a given policy. This is common practice, not only in Sweden but in the majority of the OECD countries, in evaluation performed by both governmental agencies and external evaluators. But can we really call this evaluation?

Proceeding in this way certainly has its advantages, especially in terms of cost. The monitoring procedures that identify the characteristics and nature of the recipients of a scheme, as well the collection of recipients' opinions, are cheap and easy procedures, they inform about the subjective judgment of the recipient firms, and they are common practice in many countries. However, monitoring by means of questionnaires and interviews gives no indication about policy effectiveness. Moreover, questionnaires and interviews tell nothing about satisfying objectives or about additionality. They represent the surviving firms only in a partial way and it is likely that the answers provided are socially desirable answers.

Monitoring alone is therefore incapable of offering policy-relevant insight into policy effectiveness, and the different agencies in charge of evaluation need to move a step forward. This is the only way to provide the legislators with the tools to design new policies and programs.

Steps 4 and 5 of Storey's classification are not immune from criticism. Comparison of the recipients with typical firms and matched firms can present problems given that the assisted firms are not typical, that perfect matching can be very difficult, and that there is the likelihood of sample selection bias.

Notwithstanding these problems and the need to invest in the use of statistical techniques to take into account selection bias, it is clear that matching methods are more likely to yield reliable results on the causal effects of a given policy. Along ATP program's line, Sweden needs to move towards a more complex evaluation culture in order to trigger a learning process in which evaluations provide information and create the necessary knowledge for an improved policy design.

There are two alternative paths that the Swedish government can follow to improve the effectiveness of evaluation.

1. The first alternative would be to appoint an external agency to do both ex- ante and ex-post evaluation. Such an agency would be responsible for data collection but would not be in charge of policy implementation. The actors involved would therefore be three: the ministry in charge of policy design, an agency (such as Nutek or Vinnova) in charge of policy implementation, and an agency in charge of policy evaluation (such as ITPS). By assigning different responsibilities (implementation and evaluation) to the two agencies, the ministry could increase the independence and reliability of evaluation results and reduce the possible bias of the evaluator, but at the same time would face higher administrative costs due to the existence of two actors.
2. The second alternative would be to appoint a single agency responsible for policy implementation, analysis and evaluation. The actors involved would therefore be only two, the ministry and the agency, and everything would be done *in-house,* along the lines of Nutek's Nytta.[23] This alternative has the advantage of reducing administrative costs by creating or appointing a single agency, but at the same time poses a greater risk of bias, especially when the agency performs ex-ante evaluation, in order to increase their budget.

These are two alternative paths that Sweden can follow. Regardless of the way chosen, it is clear that the focus has to shift towards a new culture of evaluation that focuses on: comprehensible setting of objectives and quantitative targets; looking beyond simple monitoring approaches and improving instruments to take into account selection bias; and triggering a learning process.

To achieve this kind of culture therefore requires a slow change in behaviour on the part of all the actors involved. It is only when the objectives are clearly stated, the data systematically collected and the roles well-defined, that it is possible to concentrate on determining the most appropriate methods for carrying out evaluations.

Given this, the next chapter will focus on different methods, at the micro- and macro levels, that can be used to answer the question: What would have happened if the policy had not been implemented?

[23] *Nutek 2006. Nytta – Nuteks system för uppföljning och utvärdering.*

3 Evaluation methods

The key issue in evaluation is *identification.* We therefore need assumptions in order to make a causal inference that one variable affects another. In particular, we must focus on the best way to identify a counterfactual situation and try to answer to the counterfactual dilemma: What would have happened if the policy had not been implemented?

Such a counterfactual situation can be analysed at both the micro- and macro levels, depending on the objective of the policy, the size of the effects expected and the availability of data.

In the case of state aid, microanalysis is considered the best way to find the counterfactual situation because it allows comparison of a so-called "experimental" group that receives the policy treatment with a "control" group representing the universe of firms that do not receive government funds. Econometric methods are used to construct a control group that resembles the treatment group as closely as possible, at least in terms of observed characteristics.

However, randomization of program participants into control and experimental groups is often not feasible in field settings, and the researcher's desire to evaluate a program with a rigorous experimental design is often incompatible with the objective of serving the expressed needs of the program participants. In such cases, an analysis at the macro level, which highlights the general effects on the national (or regional) economy, can be a valuable alternative. By means of a general or partial equilibrium analysis, we are able to describe two scenarios at a given point in time, the equilibrium reached by the economy without intervention and the equilibrium reached after policy implementation, and thus attribute the difference between them to the policy. The analysis at the macro level presents two main drawbacks: it is useful only when studying the effects of programs that represent a big share of public expenditure, and it does not provide any information about causality relations. It does, however, represent a valuable alternative to microanalysis in cases where data is lacking or where the policy design is peculiar.

3.1 Micro level analysis: Non-experimental evaluation and matching estimators

The key issue in non-experimental evaluation is the *identification* of a counterfactual situation and the need to answer the counterfactual dilemma: What would have happened if the policy had not been implemented?

In experimental studies, participants from a large group of eligible applicants are randomly assigned to treatment and, through comparison of the treated group and the control group; we are able to obtain an unbiased estimate of the average treatment effect. However, this is not the case in non-experimental studies, because the various treatment groups are likely to differ from each other in a non-random way. It is therefore necessary to construct a comparison group as similar as possible to the experimental control group.

A variety of econometric methods are used to construct a control group that resembles the treatment group as closely as possible, at least in terms of observed characteristics: randomized experiments; quasi-experimental designs; matching on individual characteristics that distinguish treatment and control groups; and propensity score matching. The first of these two methods are rarely used in studying economic effects of public policies

for different reasons: the first, randomized experiments, are costly and time-consuming and are not feasible when variables cannot be manipulated; while the second, quasi-experimental designs, are criticized mainly for their substantial selection bias.

The problem of selection bias in impact evaluation relates to the fact that program participants differ from non-participants in characteristics that cannot be observed by the evaluator and that affect both the decision to participate in the program and its outcome.

If, by means of randomization, the evaluator can ensure that the treatment- and control groups are statistically equivalent, in quasi-experimental and non-experimental designs, econometric techniques are used to model the participation and outcome processes and arrive at an unbiased estimate of program impact. One method suggested for solving this problem is matching.

Matching methods have been developed and extensively used in the statistics and medical literature (Rubin 1978; Rosenbaum and Rubin 1983, 1985; Rubin and Thomas 1992), but are fairly new to economics and policy evaluation. Matching can be considered a "correction strategy" that corrects for selection bias in making estimates and consists of pairing individuals from various treatment groups who are similar in terms of their observable characteristics.

The fact that matching estimators do not require specifying the functional form of the outcome equation and are therefore not susceptible to misspecification bias along that dimension makes them quite valuable in evaluation studies. Traditional matching estimators pair each program participant with a single matched non-participant, but in recent years a variety of estimators that pair program participants with multiple non-participants and use weighted averaging to construct the matched outcome have been constructed.

In the words of Petra Todd (2006),[24] we:

> *…assume that there are two potential outcomes, denoted (Y_0, Y_1), that represent the states of being without and with treatment. An individual can only be in one state at a time, so only one of the outcomes is observed. The outcome that is not observed is termed a counterfactual outcome.*

The treatment impact for an individual is the difference between the outcome with treatment and the outcome without treatment, and it is not directly observable. The evaluation problem is due to missing data, because if we were able to observe both Y_0 and Y_1 simultaneously we could construct such a difference for everyone. Therefore, assessing the impact of a program requires making an inference about what outcomes would have been observed in the no-program state, given that we are only able to observe the total outcome on both the treated group and the untreated one.

The evaluator is thus required to use evaluation parameters of different kinds. Because of missing data problems, it is most common for evaluators to concentrate on the mean impact of treatment on the treated group and try to answer the question: How much did persons participating in the program benefit from it, on average, compared to what would have happened to them otherwise?

Matching estimators therefore assume the existence of a set of observed characteristics (Z), such that outcomes are independent of program participation conditional on Z or, in the terminology of Rosenbaum and Rubin (1983), treatment assignment is *strictly ignorable*

[24] *Todd, P. 2006. Matching Estimators.* athena.sas.upenn.edu/~petra/papers/mpalgrave2.pdf

given the set of observable characteristics. Moreover, it is required that, for all Z, there is a positive probability of either participating or not participating in the program, and the distribution of the matching variables Z should not be affected by whether the treatment is received or not.

Todd (2006) highlights that matching estimators can be difficult to implement when the set of conditioning variables is large, and when Z are discrete, small-cell problems may arise. Moreover, if Z are continuous, the convergence rates could be slow due to the so-called *curse of dimensionality*. Rosenbaum and Rubin (1983) address the dimensionality problem by stating that when Y_0 outcomes are independent of program participation conditional on Z they are also independent of participation conditional on probability of participation. Hence, when matching on Z is valid, matching on the propensity score P(Z) is also valid. This is one of the reasons why most of the literature on matching focuses on propensity scores matching methods.

In *propensity score methods,* the evaluator estimates the propensity score by means of a parametric logit or probit model or a semi-parametric estimator[25] and then, in a second stage, individuals are matched on the basis of their propensity to participate.

Other limitations of non-experimental methods compared to experimental methods are: non-experimental methods do not guarantee that the support for the comparison group equals the support for the treated group; they are likely to combine two different datasets for participants and non-participants, often collected using different methods.

The literature therefore suggests alternative matching estimators such as: Nearest Neighbour Matching; Interval Matching; Kernel and Local Linear Matching; and Difference-in-Difference Matching.[26]

If propensity score matching and multivariate regression methods control for selection on observables, it is by means of instrumental variables methods that the evaluator is able to control for selection on unobservables. *Instrumental variables* is a technique that identifies a factor that determines receipt of a project, but which does not influence outcomes of interest. This factor is then used to simulate who would have been in the treatment group and who would have been in the control group had receipt of the project been based on that factor. The difference in outcomes between these simulated treatment and control groups is then the impact of the project.

[25] *The propensity score can not be estimated with a non-parametric model otherwise the problem of dimensionality would reappear.*

[26] *The difference in a given outcome between recipients of the project (the treatment group) and a comparison or control group is computed before the project is implemented. This difference is called the "first difference". The difference in outcomes between treatment and control groups is computed again some time after the project is implemented, and this is called the "second difference". In the difference-in-difference technique, the impact of the project is the second difference less the first difference. The logic is that the impact of the project is the difference in outcomes for treatment and control groups after the project is implemented, net of any pre-existing differences in outcomes between treatment and control groups that pre-date the project.*

3.2 Regional analysis and spatial effects*

Traditionally, empirical analyses of the effects of various types of government support based on economic theory have focused on how such support has affected developments in a particular company or region for which support was paid out. Since it is reasonable to suppose that it is not only government support that affects developments within a company or region, other potentially important explanatory factors are often included in the analysis in an effort to test their influence or maintain that influence at a constant level. Examples include the availability of human capital, raw materials and natural resources, investments, local public services, local tax rates, demographic factors, etc. Time dynamics – the fact that it often takes time before the effect of a given measure can be perceived – have also received a fair amount of attention in both the applied empirical literature and the more method-oriented econometric literature. Here, the development of so-called "dynamic panel" data methods has made an important contribution (see, for example, Holtz-Eakin *et al.*, [1988], Arrelano and Bond [1991], and Ahn and Schmidt [1995]; see also Baltagi [2001] for a general survey of the field).

Although it would be reasonable to proceed from the view that developments in a given company or region are to a large extent affected by factors present within the company or region itself, it is also reasonable to assume that developments in a given company or region affect developments in other nearby companies or regions. There are several reasons for this. For example, it is reasonable to suppose that government support for a company in a particular region may also affect the company's subcontractors and/or retailers in other nearby regions. In addition, support may also impact the company's competitors and thus have an overall negative effect on, say, employment in that particular region or regions close by. Some empirical research, based on Swedish data, suggests that, if government regional support has a negative effect on regional unemployment, this in turn will have a positive effect on migration to that region, which is of considerable significance for regional policies (see Aronsson *et al*. [2001] and Lundberg [2003]). Other studies show that migration to attractive regions also has a positive effect on migration to neighbouring regions (Lundberg [2006a]). A possible explanation for migration spillover into other regions may be that property prices in attractive regions tend to be higher than in the less attractive regions, with the result that individuals on lower incomes may settle in regions where property prices are lower.

Recently, in an attempt to bring effects of this nature into the equation, more and more studies have begun to apply what is known as spatial econometrics. The term "spatial econometrics" is an overall concept for descriptive statistical tests and explanatory regression models that allow for the fact that events that occur in, say, a company or region, may affect developments in other companies or neighbouring regions. The fact that spatial econometrics enables analysts to consider varying degrees of dependence between companies and regions, has made it possible to model their mutual locations. Previously almost exclusively a method used by regional science, urban and regional economics and economic geography, spatial econometrics has recently become an increasingly useful tool for traditional empirical economics and theoretical econometrics. One explanation for this is that, in principle, spatial econometrics becomes applicable as soon as there is any reason to suppose that one agent's (region's, individual's, company's, authority's, etc.) behaviour is influencing the behaviour or development of another agent. Examples of areas of applica-

* *By Johan Lundberg, Centre for Regional Science (CERUM), Umeå University.*

tion include demand analysis (Case, 1991), international economics (Aten, 1996), labour market economics (Topa, 1996), public economics and local public economics (Case et al., 1993; Murdoch et al., 1993; Holtz-Eakin, 1994; Brueckner, 1998; and Lundberg, 2006b), tax competition between regions (Besley and Case, 1995; Brett and Pinkse, 1997; Bivand and Szymanski, 1997; and Revelli, 2002), and regional growth and migration (Armstrong, 1995; Rey and Montouri, 1999; Fingleton, 2001; and Lundberg, 2006a).

In an increasingly internationalized world in which transport costs tend to represent an increasingly smaller proportion of a company's overall production costs and consumers are increasingly making purchases and obtaining information on different prices and products over the Internet, are we justified in asking whether the location of companies and production facilities is still significant? And, for the companies themselves, do their relative geographical locations matter? Additionally, for many end consumers, where a product is made is not of decisive importance.[27] However, at the same time as it may be argued that the actual location of various industries is no longer of any great significance, there are clear indications that many do tend to gather in certain specific locations. If we take Sweden as our example, we find a strong concentration of glass-making industries in Småland, telecommunications and electronics companies in Kista, the financial market in central Stockholm, medical companies in Lund, etc. There are a number of reasons for this. Disregarding transport costs, a small company lacking the resources to perform a localization analysis may benefit from setting up close to a larger rival, since proximity to a large company often brings access to certain peripheral functions and services, such as special administrative (e.g., legal assistance specific to the industry) and technical expertise, subcontractors and/or retailers. It may also be easier to recruit staff with specific skills in a region in which there are already several companies in the same industry. Large companies can also benefit from setting up close to other companies in the same line of business as themselves. Informal meetings and inter-company information leakages are often more common if the geographical distance between the companies is small. At informal meetings, information on technological achievements, new trends, personnel skills and other information important to the industry may be spread outside the company walls. Other important factors affecting a company's choice of location are the availability of natural resources and raw materials (e.g., in the mining and forest industries), transportation routes and communications (e.g., the forest industry's former preference for locations at the mouths of rivers). A further factor, and one which is perhaps not entirely without significance, is the personal preference of senior management.

From the point of view of redistribution policies, production location is of major importance. This is because production generates jobs and incomes, and if production is unevenly distributed across regions there is a tendency for incomes to become unevenly distributed as well. At the same time as production tends to be located in big-city regions or regions that offer a specific type of expertise or access to important raw materials and natural resources, there are also signs that point to a corresponding increase in regional differences in incomes. Taking Sweden as a concrete example once again, in 1981, the average middle-range income in the municipality with the highest incomes was just over twice that of the municipality with the lowest incomes. By 1999, the figure had risen to

[27] *This naturally depends on whether production is in line with the consumer's own values as regards working conditions, political views and distribution of the incomes derived from production. As it is beyond the scope of this report to discuss these factors in detail, it will be assumed here that production conditions are similar regardless of where production actually takes place.*

over three times the level of the lowest income municipality. Furthermore, while the population of the big city regions is increasing, it is decreasing in Sweden's more sparsely inhabited regions. All in all, trends such as these can result in significant differences in the standard of services, both public and private, available in different regions.

The following section reviews the means by which spatial econometrics may be used to analyse the effects of government support, including both general support for various regions and subsidies more specifically targeted at companies. Spatial econometrics can also be applied to a variety of other fields, although this is beyond the scope of the present report. In other words, the section in no ways claims to present a comprehensive roundup of all tests, estimators or areas of application for spatial econometrics, but should be regarded more as a brief review.

3.2.1 How can spatial econometrics be used to analyse the effects of government support for companies and regions?

Spatial econometrics differs from "normal" econometrics in that it takes into account the fact that what happens in a particular region[28] can also affect events in other nearby regions, without excluding other potentially important factors that may explain a certain process or event. It is also possible to include and model time dynamics in spatial models.[29] Let us take a concrete example.

Suppose that the government decides to provide extra backing for its labour market policies by paying out start-your-own-business subsidies. Although such support may in itself target individual companies, it is not likely that it will be evenly distributed over all municipalities. In this example, it is assumed that the actual purpose of the subsidy is to boost local employment. Here, spatial econometrics enables us to use statistical methods to determine whether the subsidy has any effect on employment levels in the neighbouring regions. For example, the start-up may employ subcontractors in other regions or itself become a subcontractor or supplier to companies located in other regions, or it may recruit its personnel from people living in other regions. Using spatial econometrics enables us to test the physical extent of this dependence. Does it involve only the neighbouring regions, or does it extend also into regions that are some distance away?

Another interesting question that may be resolved by spatial statistical analysis is the assumed impact of the start-up subsidy of our example on the survival prospects of existing companies in the same branch of industry, and, if such an impact can be identified, the extent to which it makes itself felt. The questions that may be answered are whether only companies in the same region are affected or whether companies in other nearby regions are also affected, and, if they are, the extent of the overall impact.

The above questions are of interest from a policy perspective, since they provide insights into how the effects of a given measure will spread. If a measure targeting a specific company or region also tends to affect companies and regions other than envisaged by the subsidy, this should be taken into consideration in establishing the principles and directives applicable to specific forms of support. It is therefore important to gain a greater under-

[28] As noted in the introduction, the phenomenon is not limited to regions alone but may, for example, arise from a dependence between companies, individuals, etc. However, regions will be used throughout in order to provide concrete examples illustrating the overall theme of the report.

[29] A review of statistical methods of capturing and modelling time dynamics is beyond the scope of this report. An excellent survey of dynamic panel data methods may be found in Baltagi (2001).

standing of how the effects of different types of support are spread and to take this into account in analyses of the overall effect of any given measure.

A concrete example of the application of spatial econometrics[30]

To illustrate how spatial econometrics may be put to practical use for this type of analysis, the following presentations and descriptions of various spatial tests and estimators are based on a concrete, albeit hypothetical, example. Our starting point is that, in one particular year, the government has decided to boost its support for its labour market policies through the provision of start-your-own-business subsidies. As the subsidies are payable directly to individual companies, they will not be evenly distributed across all Swedish municipalities. In this example, the actual purpose of each subsidy is to boost employment in the region in which the company is located, although it will also be of interest to study its potential impact on employment in other nearby regions and, if any such impact can be identified, the extent to which it makes itself felt.[31] Implementation comprises four separate stages:

1 Collection of data and use of Geographic Information Systems (GIS).

2 Weighting matrix – definition of spatial dependence.

3 Descriptive tests for spatial correlation.

4 Regression analysis.

Collection of data and use of Geographic Information Systems (GIS)

To perform this type of analysis, information is required on the company's geographical location, the amount of support the company has received, and the regional employment level or change in the employment level from the time the support was paid out until, for example, two or three years afterwards. To check for other potentially important factors that might explain a change in the employment rate, it is essential to have access to data on, say, the availability of other public services within the region, other regional employment schemes, regional demographic data (age distribution), access to human capital within the region, etc.[32] In addition, the data should include all regions and preferably cover a substantial period of time. With access to data covering a long period of time, the analysis will not be dependent on the choice of years and more general conclusions can be drawn from its results.

A GIS is a computer system for capturing and linking location-dependent information to its precise geographical position. Such information might be a company's address, annual turnover, number of employees, etc. By selecting certain pertinent information relevant to

---

[30] *For a more technical review of the test and estimators described below, see Anselin (1988), which is perhaps the most comprehensive textbook on spatial econometrics and serves as an excellent introduction to the field. For later references, see individual articles, above all by Anselin, Cressie, Florax, Haining, Kelejian, LeSage, Pace, Pinkse, Prucha and Rey.*

[31] *Naturally, the ambition could also be to increase the number of companies. In this case, just substitute "region" for "company" in the present text. To simplify the presentation as much as possible, only one example is given here.*

[32] *There are naturally other factors that might explain a change in employment levels in a particular region. However, to simplify our presentation somewhat, these have been ignored. We have also chosen to disregard a more complicated time dynamic.*

its geographical location, it is possible to create maps and, from them, to build up a picture of how various factors and events interrelate in geographical terms. GIS is extremely useful for spatial analyses simply because the information is linked to its geographical coordinates, which makes it relatively easy to visualize with the aid of maps. In the above example, all the information in the data set is bound to a specific geographical location or geographical area. The company is located in a certain place, government support is directly linked to a specific company, the age distribution is bound to a certain municipality, etc.

Weighting matrix – definition of spatial dependence

Once the data has been acquired, the first step of spatial analysis is to define the spatial dependence. It would be desirable to estimate the degree of dependence between the various regions by statistical means along with the other parameters of the model. However, this is almost always impossible since the number of parameters in such an estimate would exceed the number of observations. In most cases, not even if one had access to a succession of observations over time would this be possible in practice, since it would mean that the number of periods exceeded the number of individuals (or regions).[33] The degree of dependence between the various regions must therefore be assumed in advance. This in turn means that the results of all spatial econometrics, both descriptive tests and explanatory regression analysis, are based on pre-assumed notions of spatial dependence, or the degree to which regions are affected by one another. In interpreting the results of these models, it is important to bear this in mind. Usually, dependence is based on geographical distance between the regions or on whether the regions are directly adjacent. In the case of the former, dependence is often defined as decreasing with distance, which in many cases is perfectly reasonable. In the case of the latter, dependence is discrete, i.e., either there is dependence or there is not.

It is often of interest to establish the extent of the impact of a given measure. For example, it might be useful to study whether a particular type of company support spreads to regions within a certain distance of the region in which the company that has received the subsidy is located. Once we have constructed a set of weighting matrices with different cut-off values, spatial econometric testing will enable us to determine the extent to which the effects of the measure have spread. The "cut-off" value here means that dependence is assumed to persist up to a certain distance; after that point, it is assumed that there is no dependence at all. However, there are also other definitions based on, for example, similarity in population size, the size of the regions in question, the proportion of immigrants, etc. In defining the elements that make up the weighting matrix, it is important that they can be assumed to have been exogenously given. If not, there will be methodological problems. Given that the assumption is exogenous, using the tests and regression methods described below, we can test for various types of dependence to see which definition best applies to the question and body of data at hand. Based on the definition of what the dependence between the various regions looks like, we then create a weighting matrix in which each element defines the dependence between two different regions. As noted above, it is important to keep in mind that all results are conditional upon the assumption

[33] *Assume an analysis of the employment level of all Sweden's municipalities, of which there are today about 290. In this case, before we can make an estimate of spatial dependence, the number of periods would have to exceed 290 by a wide margin. In practice, it would be impossible to acquire sufficient data for this purpose. Furthermore, the layout of Sweden's municipalities and communications network has changed during the period in question, thereby creating methodological problems.*

of what the spatial dependence looks like. This places great demands on the plausibility of the definitions of the elements that make up the weighting matrix, which should therefore be accorded much attention. For estimation and interpretation reasons, the lines of the matrix are often standardized such that the sum of each individual line equals 1.

Descriptive tests for spatial correlation

Having defined the spatial dependence between the regions, we may both run descriptive tests for spatial correlation and incorporate the dependence in regression models. The difference between descriptive statistics and regression analysis is that descriptive statistics describe what data look like, while regression analysis is used in an attempt to explain the underlying causes of certain events. Regression analysis enables us both to consider other factors that may be of importance in explaining a particular course of events and to analyse their relative significance. There are several types of statistical tests that systematically seek to confirm tendencies and patterns in a given set of data. Regional differences in, for example, employment levels or changes in the employment rate can, of course, be visually illustrated *per se* in various types of geographical map. But visually illustrated patterns take on greater authority if they can also be verified statistically.

There are two groups of tests for spatial dependence. One is known as "global", since the tests can establish any general dependence between all the regions included in the data. Examples of global tests are Moran's *I* and Geary's *C*. By applying these two tests, which in terms of interpretation do not differ, we may establish the presence of clusters of high or low values or whether the data are organized such that high and low values lie close to one another. However, these tests do not enable us to determine whether it is the high or the low values that are correlated, and they do not indicate where this correlation is strongest.

The other group of spatial correlation tests is known as "local" and enables us to identify cluster formations within the data. In other words, these tests enable us to identify patterns specific to our observation. Local tests answer the question of whether a certain observation is surrounded by other observations with either similar or different values, either by comparison with- or independent of the value in one's own region. For example, a local test will enable us to ascertain whether a region with a high level of employment is surrounded by regions in which the employment rate is high or low. Local tests are also extensively used to locate extreme observations that might disrupt later regression analysis. Examples of local tests are G_i statistics and LISA (Local Indicators of Spatial Association). Global and local tests are often used in combination with one another in order to identify both general patterns and cluster formations.

Regression analysis

Unlike descriptive tests, regression analysis enables us to consider other factors that may be of importance in explaining a particular course of events, and to identify those that are of the greatest significance. This is of great importance from a policy perspective, since regression analysis can establish whether a given supportive measure is of greater or lesser significance for regional development by comparison with other potentially important explanatory factors.

Although the spatial correlation tests described above enable us to demonstrate spatial dependence, they provide no guidance as to how such dependence should be incorporated in a regression model. Spatial econometrics often distinguishes between two different types of regression model. In one, the so-called "team model", spatial dependence is included as an extra variable in the regression equation. This type of regression analysis

enables us to ascertain whether there is any direct spatial dependence between regions as defined in the weighting matrix. It is also possible to establish whether, for example, a measure taken in a particular region has had a positive or negative effect on developments in neighbouring regions. Using the team model, we may also estimate "reaction functions", i.e., how a certain region reacts to something that happens in another region. Reaction functions are commonly considered in, for example, the literature on tax competition. The other type of spatial regression model is the so-called "error model", in which the spatial dependence occurs instead in the random term of the regression equation. The interpretation of a significant spatial dependence will then be that any disturbance introduced into the system will spread to all regions, not just to the ones immediately adjacent. There are statistical tests designed to discriminate between the team model and the error model, i.e., tests suggesting which of the two specifications is best adapted to the data set in question.

A source of erroneously specified spatial models, and hence erroneous interpretations of the parameter estimates, arises from the occurrence of so-called "structural instability". This means either that the "true" parameter estimates are not the same for all regions or that the actual function form differs between regions. In somewhat simplified terms, this means, for example, that the effects of a particular supportive measure are not the same in all regions, which, in many cases, it is realistic to assume. Spatial models allow structural instability to be taken into consideration.

3.2.2 Concluding discussion

Over the past few years, spatial econometrics has found increasing use as a tool for analysis of economic phenomena owing to its ability to capture the interaction between different players or different regions. Recent empirical research on regional development has to an increasingly large extent begun to apply spatial econometrics both to descriptive tests and to more explanatory regression analyses. Although it is not always statistically possible to prove spatial dependence between regions, research suggests that this type of association should at least be tested for. There are no signs that interregional dependence is likely to fall in the future. One of the challenges facing research in the future will be to develop methods capable of capturing both spatial dependence and time dynamics as the spatial dependence changes over time.

3.3 General equilibrium model

The first part of this section describes, in an intuitive manner, a series of possible models that could be used to estimate the effects of state aid.

General equilibrium tries to provide an understanding of the entire economy using a bottom-up approach, starting with individual markets and agents. The general equilibrium model is traditionally used to analyse the effects of a change in economic policy and helps to predict the consequences of the proposed change on a variety of economically significant variables: price, level of output, government receipts, and the distribution of income among the consuming units. One of the main virtues of the general equilibrium model is its ability to trace the consequences of large changes in a particular sector throughout the entire economy.

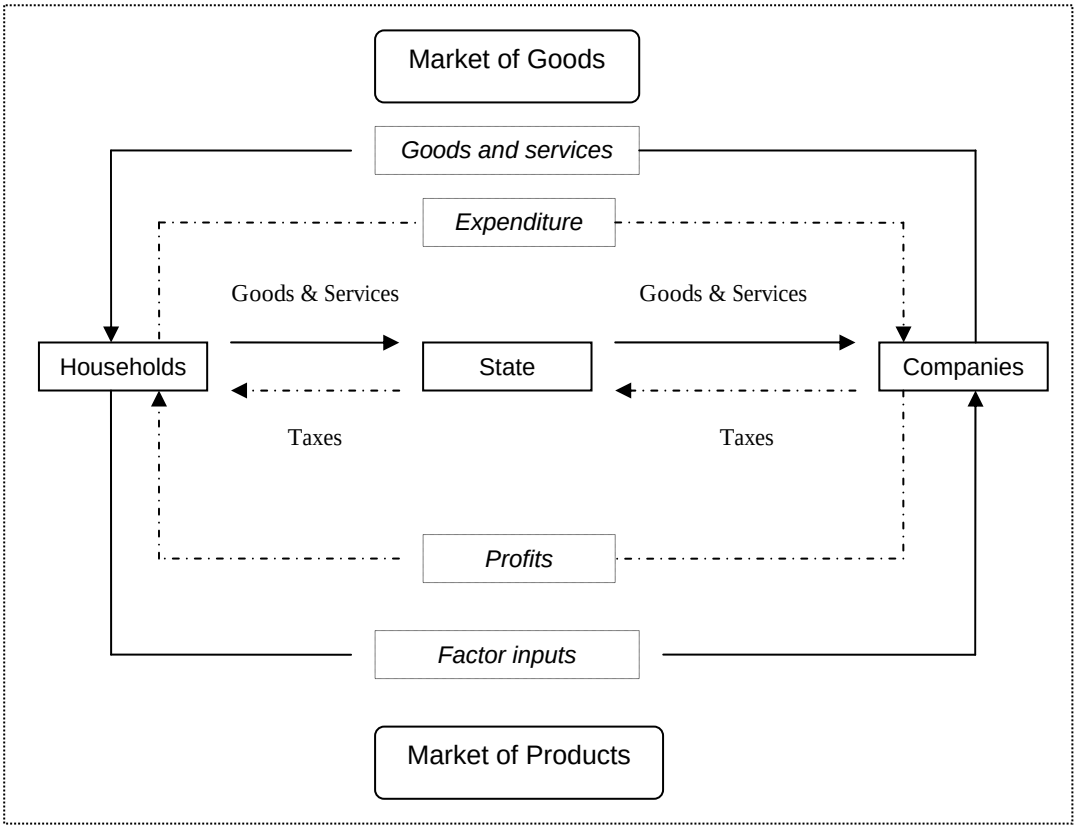
The term *general equilibrium model* presents both a methodological and a theoretical aspect. In fact, general equilibrium models consider the economy as a closed and interrelated system in which changes in one variable lead to changes in all the other variables of

the system. For example, the introduction of tax reduction or aid not only reduces costs for the company but also has effects on employment and production. A higher or lower production in one company can lead to higher or lower production in other companies, such as customers or suppliers. It is therefore important to estimate all the parameters in the system at the same time in order to capture both the direct and indirect effects of a policy. By estimating all the parameters, we can thereby reduce the number of exogenous variables.

The general equilibrium theory considers prices and quantities of equilibrium in a system with perfect competition, which means that it considers all the interactions between markets as functions in the individual market.[34]

In practice, it is almost impossible to carry out general equilibrium models because they usually produce non-linear variables. The presence of non-linear variables in a system makes the system impossible to solve analytically, leading to the use of numerical methods or a "computable general equilibrium" (CGE) model, which uses a symmetric matrix called "social accounting matrix" (SAM) to calibrate the coefficients for the equations in the model.

Figure 3. Cyclical flow in a closed economy.



A CGE model is built according to the issue at hand, the problem to be solved and the data available. The fundamental concept for CGE models is the cyclical flow of goods in a closed economy, represented in Figure 3. Households and companies are the main actors in the economy and the state usually has a passive roll in the cyclical flow, even though the

[34] *For a more in-depth discussion of general equilibrium theory, see Varian (1992) and Mas-Colell, Whinston and Green (1995).*

state is explicitly represented in many CGE models. Households own the factors of production and they consume goods and services produced by the companies. Companies hire factors of production in order to produce goods and services. The state collects taxes and distributes the revenues to households and companies as grants, subsidies, contributions and transfers. The flow in Figure 3 shows that everything produced by the firms is kept or consumed by the households, while firms absorb the entire supply of factors. This means that, for every activity in the economy, the value of the expenses enters the books against the value of the incomes. The double entry bookkeeping system is reflected in the SAM matrix, in which the economy is broken down into different sectors according to the problem in question.[35]

When using partial general equilibrium models such as factor demand models, the focus shifts to a limited number of markets instead of the entire economy. In practice, it is almost impossible to estimate all the parameters in the economy at the same time, and it is necessary to limit the number of variables taken into account and estimated.

CGE models allow us to study the economy at an aggregate level, while the advantage of partial equilibrium models is the ability to use data from the firm level to study the effects of public measures.

3.3.1 Existing models

In recent years in Sweden, a number of CGE models have been developed with different goals. Here, we discuss advantages and disadvantages of the existing models, given that, in the framework of the present work, it is also important to understand which of the existing models might be used and developed to evaluate state aid policies in Sweden.

EMEC - Environmental Medium Term Economic Model

National Institute of Economic Research (*Konjunktursinstitutet*, *KI*) has developed a long run static model called the "environmental medium term economic model" (EMEC), specially designed for the analysis of the implications of the Swedish environmental policy for households and firms. EMEC is a static CGE model with 26 industries and 33 composite commodities, and a public sector producing a single commodity. Products goods and services are exported and used together with imports to create composite commodities for domestic use. Composite commodities are used as inputs by industries and for capital formation. In addition, households consume composite commodities, and there are 26 consumer commodities. Production requires two types of labour and capital as primary factors, as well as inputs of materials, transport and energy. The peculiarity of EMEC is the very detailed specification of emissions from the use of both fuel inputs production and households' fuel consumption: the model includes emissions of carbon dioxide, carbon monoxide, methane, sulphur dioxide, nitrous oxide, nitrogen oxides and particulate matter.

The supply of each type of labour is exogenous for the economy as a whole, while capital is supplied to the economy at a given price. All factors can move freely between domestic sectors. Perfect competition and no economies of scale in production are assumed for all markets. The small country assumption is adopted for tradable goods and the problem of overspecialization is handled by the Armington (1969) assumption for imports, i.e., that products traded internationally are differentiated on the basis of their country of origin and

[35] *For a detailed description of CGE models, see Shoven and Whalley (1984), Kehoe and Kehoe (1995).*

by a relative price-dependent supply function for exported goods. Households are distributed into six subgroups according to disposable income and place of residence. The model runs with exogenous interest rates and is closed with an exogenous ratio of the current account. The foreign price level is chosen as numeraire. The use of energy by firms or households is subject to an energy tax and pollution taxes. Tax exemptions based on the use of CO_2 permits or other reasons are reflected in the estimated tax rates. Consumer goods are also subject to a value added tax, as well as other indirect taxes. The use of labour is subject to social security fees and households pay income tax on labour income. Firms and households react to prices, including taxes, and adjust their mix of inputs or their bundle of consumer goods by substituting away from the most expensive inputs or goods.

The representative firm is assumed to choose, in three stages, an optimal mix of two types of labour and an optimal mix of energy. The firm then decides upon the mix of labour and physical capital in the creation of value added, as well as the mix of energy and materials in the creation of energy-material input. The various inputs and outputs must be transported, and the firm chooses an optimal transport solution in two steps. An optimal mix of value added and energy-material input is chosen at the highest level, to produce the firm's output. Another kind of substitution relates to goods of the same classification. Domestic goods are non-perfect substitutes for foreign goods in domestic- as well as foreign use, i.e., in imports as well as exports.

As mentioned above, EMEC is a model designed for the analysis of environmental issues and could be useful to evaluate the effects of energy tax reduction. Because of the way it is built, it would require a certain amount of adjustment and would, in any case, be useful in evaluating only certain kind of policies.

At present in Sweden, there is a series of tax reduction schemes in action: a reduction of selective purchase tax for production of carbon dioxide neutral fuels; the manufacturing industry's tax relief on CO_2 emissions; 1.2 per cent rule – 0.8 per cent rule; the manufacturing industry's tax relief on electricity; and full exemption from energy tax for energy-intensive companies.

Through meeting with Göran Östblom and Magnus Sjöström from KI, we have tried to find answers to the following questions:

Is it possible to use and adapt EMEC to study the effects of the noted tax reduction schemes? If so, how long would it take?

Some of the specified issues can be analysed with EMEC. The easiest one to study is the tax relief on CO2 for manufacturing industries. In order to estimate how long it would take to run the regressions, it is necessary to better define what needs to be done and under what assumptions. Given that a certain amount of changes to the model are necessary, we estimate it would take roughly two months for each issue.

How long would it take to develop a model like EMEC?

It depends on the circumstances, in particular on whether there is a similar model to start from and what skills are required. The development of the current version of EMEC has taken 3–4 working years, spread over more than 10 years and 10 researchers.

Is it possible for KI to cooperate with ITPS in the development of a new model or adaptation of a model originating from EMEC?

In addition to the assignments that KI receives from the government, there is room for the Environmental Economics Department to engage in external research. This means that EMEC could be developed and adapted according to the required needs. For external assignments the cost is 900 SEK/hour. It is not possible to develop new models for other agencies, but cooperation between ITPS and KI in order to jointly develop a model is plausible, after the necessary agreements. It is important to underline that external assignments may require more long-term planning.

Can EMEC be used to discuss regional issues? Or is it better to develop a new method to study regional effects?

EMEC reproduces the Swedish economy at an aggregate level and does not include a regional level. The results from the simulation can be decomposed with respect to regions if there is a suitable tool of distribution. This is, however, something completely different from simulating regional effects with a model that uses regional inputs. The main problem is not building a model but accessing the data, such as that contained in the national accounts, which is the basis for simulation in EMEC. If we were to break down EMEC's results at a regional level, we would definitely interpret them with caution.

It is hard for KI to say whether it is better to build a regional model to capture regional effects, given that it also depends on the resources available and on the questions to be answered.

Do you estimate all the parameters used in EMEC? What about data availability?

No statistical estimation is done in EMEC. The basis for the simulation is a so-called "input-output" table (SAM) from the national accounts. Elasticity and coefficients come from different sources and only a limited number are calculated at KI. The basic scenarios are created from the assumptions presented in LU.

FIMO – Financial Model

FIMO is a calculation and simulation model that on, an annual basis, describes financial flows and net lending divided into the institutional sectors of the economy as defined by the national accounts. The sector division in the model comprises central government, pension system, local government, households, corporations, and the rest of the world. The model consists of more than 500 variables at the lowest level, of which the majority describe transactions within the general government and between the general government and households. The model generates a large number of aggregates, on both a sector and transaction basis. Examples of the former are *total income and expenditure*, *net lending*, *disposable income*, *net savings* and *changes in assets and liabilities*, while aggregated payment flows between sectors such as *total tax payments* and *transfers to households* are examples of the latter.

The model also includes an exogenously determined aggregated description of the real side of the economy on which the financial flows are contingent. Possible model applications include calculations of financial flows and net lending based on a given development of the real side of the economy, as well as various types of sensitivity analysis.

The main pitfall can be found in the real side of the model, which is based on a great deal of assumptions and is exogenously determined.

LUMOD – Long-term dynamic simulation

LUMOD is a dynamic model that studies savings and capital formation. It can be seen as a growth model but it also includes the Keynesian consumption function and multiplicator as well as investment accelerator.

LUMOD includes both a demand and supply side, and prices are cost-driven. The model takes into account productivity, profit shares, salaries, income taxes, transfers, interest rates, and import prices.

The exchange rate, as well as the rate of growth of foreign prices, is exogenously determined. Using an import-export function, the foreign trade is related to the internal and external demand and to the price trend. The public sector is treated as an aggregate. On the other hand, public revenues and expenditures are studied in more detail. On the revenue side, we have domestic taxes, income taxes, and indirect taxes. On the expenditure side, we have consumption, investments, foreign transfers and transfers to households, and subsidies to firms.

LUMOD is an open model with a great number of exogenous variables. It is, above all, an instrument to test the consistency of exogenous assumptions and results in an integrated frame.

LUMOD's estimations do not provide forecasting results but illustrate possible patterns of development of the Swedish economy given different assumptions. The purpose is to increase the understanding of the connection between development factors as well as to provide an experimental tool to analyse different policies.

KIMOD

KIMOD is a dynamic general equilibrium model of the Swedish economy to be used for aggregate economic policy analysis and medium-range macroeconomic scenarios. It is a highly aggregated macroeconomic model in which all firms are the same, as are all households. In addition, the general government sector is consolidated and thus not separated into central government, municipalities and county council districts. The model is micro-based in the sense that firms and households make optimal decisions on output and consumption, respectively, given rational expectations about the behaviour of other actors and about the probable future development of the model as a whole.

KIMOD is dynamic, meaning that investment and savings during a period affect future possibilities for output and consumption, respectively, and that all decision-makers take this into account. Time is divided into discrete periods of one year. The projections generated by the model are thus time series with a yearly frequency, and the econometric equations of the model are estimated on the basis of annual data. The national accounts are the preferred source for initial data and for estimating parameters.

KIMOD is an equilibrium model in two respects. First, prices in each period are set so that supply is equal to demand on all markets except the labour market. The latter is modelled instead as a search market, where wages are set in negotiations between employers and employees. Unemployment arises both from search-related friction and from imperfect competition. Second, in the long run, the model approaches a steady state that is independent of the state of the economy at the outset. In this steady state, the economy is on a balanced growth path with a constant relative growth rate.

Thus, KIMOD is intended for use in macroeconomic analysis and in medium-range scenarios, medium range being a time horizon of two to six years. For other time horizons, the lower limit of usability for the model results from the fact that the length of periods is set at one year. This means that there are no seasonal dynamics at all, and that data for parts of the current year cannot be used as the initial state for the model. For long-range purposes, the usability of the model is limited by the fact that demography and other structural developments over time have not been modelled.

ISMOD

ISMOD was originally developed in the early 1980s and has since then been used for long-term economic forecasting for the Department of Finance. During the last ten years, the model has only been used by NUTEK and SIKA. In recent years, at the School of Economics in Jönköping, the model has been adjusted and re-estimated with new data.

The ISMOD model is formulated to generate solutions valid for the time period of 5–15 years ahead. The time period should not be shorter due to investment processes and not longer because the input data (technology alternative) becomes less adequate when the time period is longer.

The model can be described by the following parts:

1. Production structure: The basis is input-output matrices, in which each sector's demand for input delivery is determined and input coefficients are specific for each technical category in the sector. There is a limited capacity in the model. The capacity can be extended during the time period, but requires investments and this investment generates a demand for deliveries from other sectors (according to a vector of investment coefficients). At the same time, there is a reduction of capacity in some of the sectors. The profit level in the sector determines the speed of the reduction (in turn determined by prices and wage level).
2. Supply of goods and services: During the time period, the supply changes due to (i) capacity reduction, (ii) investments in new capacity, and (iii) imports. All three components are dependent on the equilibrium prices during the time period.
3. Demand for goods and services: The demand side has the following five components: (i) current purchase for the production, (ii) private consumption, (iii) purchase for the public sector, (iv) export, and (v) purchase of investment goods. The demand components are directly and indirectly dependent on the equilibrium prices.

One problem with ISMOD is that it operates with industry sectors and not commodity groups. Trade and transport are treated in a very simplistic way. The output of these sectors is assumed to amount to the aggregate trade margins of the other sectors.

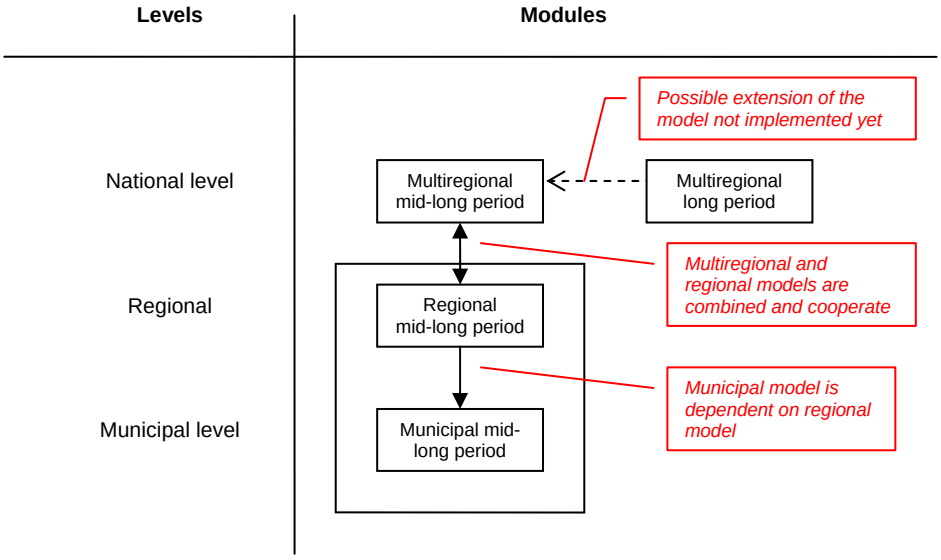
Experience shows that changes in transport costs and infrastructure investment have an impact on the sectoral structure and transported volumes and quantities. In ISMOD, this is not possible. Moreover, there is no geography in ISMOD, and another problem is the time horizon – because analysing infrastructure investments requires long forecast periods.

rAps

rAps is an integrated system for regional analysis built on a central and a local database related to different models. It can be used for both simple analysis and more advanced analysis such as forecasting and scenarios. rAps can be useful both for regional and central

actors and it is built on a system of models. The system consists of a series of modules with specific purposes, which together yield a great degree of flexibility and adaptability to the system.

Figure 4 Modules in the rAps system.



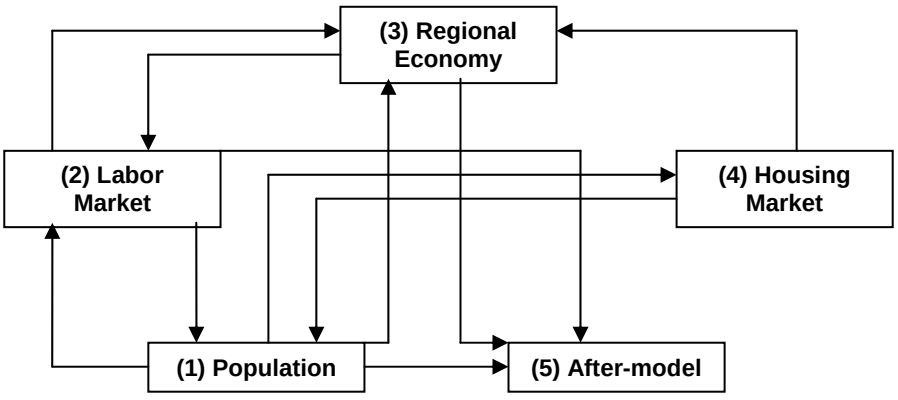
Source: https://www.h.scb.se/raps/Pdf/raps_beskrivning.pdf

As shown in Figure 4, the modules in rAps are grouped by level and by time perspective, and municipalities are the smallest units from which the entire economy is built.

The regional model for the mid-long period (5–10 years) is an economic and demographic model intended for analysis and forecasting at the regional and municipal level. Regions are not geographically static, because they are built as groups of municipalities and can vary according to the issue under investigation.

The regional model is built from the connections at the municipal level and consists of 5 partial models: (1) population; (2) labour market; (3) regional economy; (4) housing market; and (5) after-model by municipality. These five partial models determine employment by municipality and commuting, as well as municipality's revenues and expenditure.

Figure 5 The five partial models of the regional model.

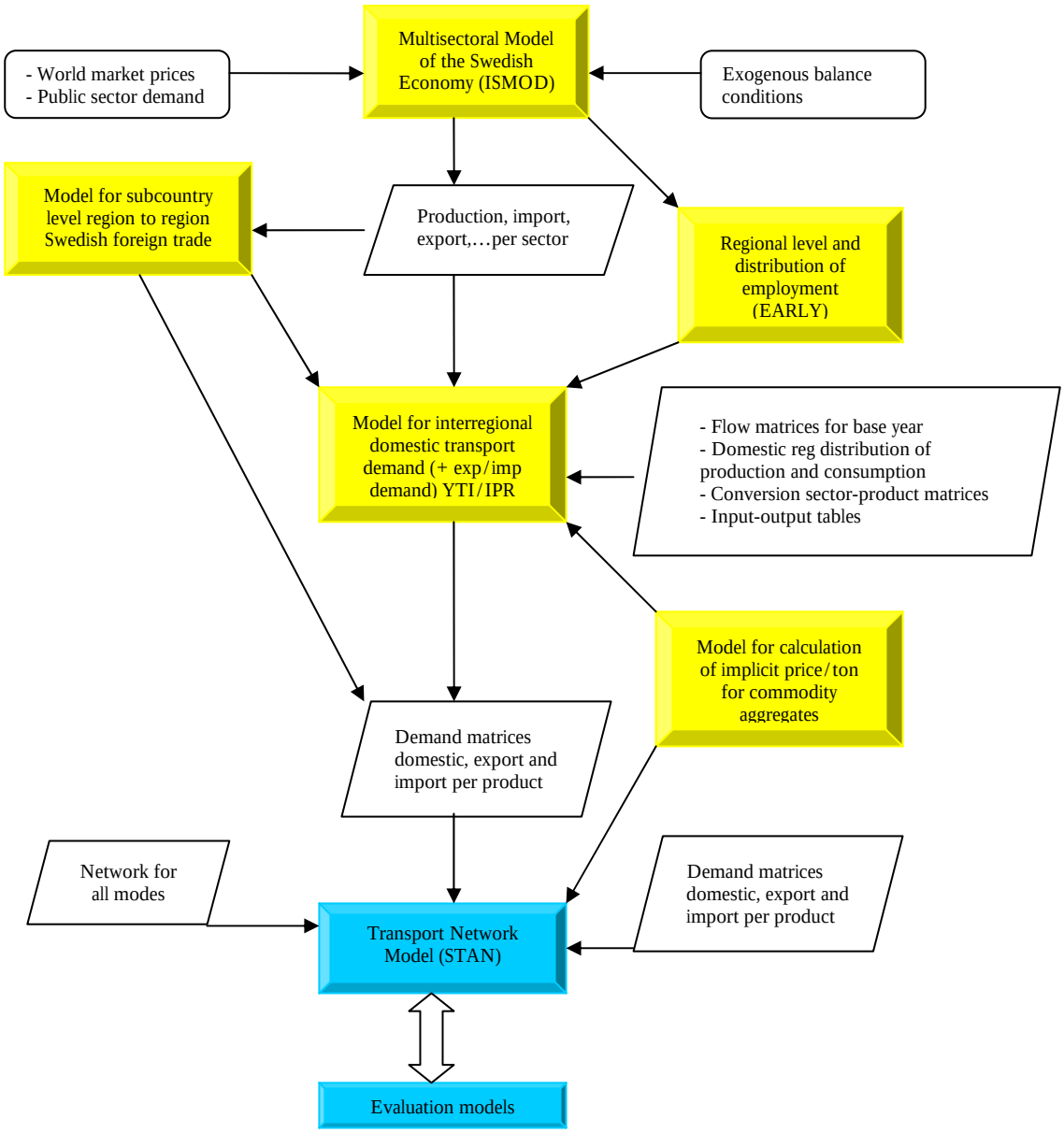


Source: https://www.h.scb.se/raps/Pdf/raps_beskrivning.pdf

SAMGODS

SAMGODS is another model that consists of a number of connected modules, operating at different levels of spatial detail, as shown in Figure 6.

Figure 6 The SAMGODS model.



Source: SAMPLAN 2004:1

The yellow boxes in the figure relate primarily to changes in demand over time, while the blue boxes deal with the impact of particular policies. There are five separate models relating to the forecasting of demand over time:

- A multisectoral system of models of the Swedish economy (ISMOD)
- A model for regional disaggregation of sectoral employment (EARLY) linked to ISMOD

• A model for modelling interregional transport demand within Sweden (VTI/TPR)

• A model for regional forecasting of Swedish foreign trade

• Models for forecasting of implicit commodity value for aggregates of commodities

The SAMGODS model has multiple purposes: to provide demand forecasts, policy and project evaluation information, and analyses of the effects and consequences of alternative strategies. For our purposes, however, we are mainly interested in the use of the model for analysing general transport policy measures as well as policy measures related to specific infrastructures, e.g., road, rail, etc., given that there is a wide and seemingly widening range of decision processes that could make good use of freight demand information.

All the models described above are built as simulation models with the purpose of creating possible future scenarios. Most of the data used are based on a series of assumptions, making them unsuitable for our purposes since they are unable to give a real picture of the world. They are, in fact, basically used to analyse different scenarios, the parameters are not estimated, and the outcomes of the models depend on the user's assumptions.

There are, however, two other models worthy of mention in this study because of their ties and similarities to the "ITPS model" proposed in the next section.

RAMSES – Riksbank Aggregate Macromodel for Studies of the Economy of Sweden

RAMSES is a general model that aims to explain the entire economy, not just a particular component. It consists of numerous different households and firms that interact in markets for goods, capital and labour. As in many other modern New Keynesian general equilibrium models, markets for goods and labour are assumed to be characterized by monopolistic competition. This means that firms and employees, instead of taking prices and wages as given, are aware that they can influence them by their behaviour. However, as price and wage stickiness is assumed to exist, monetary policy is able to affect the real economy (output and labour supply, for example) in the short run because nominal prices and wages do not freely adjust to a change in the nominal interest rate.

The model also includes a central bank that sets the short-term interest rate and a government sector that is assumed to finance its consumption expenditure by taxing labour and consumption. There is also a foreign economy that is assumed to be unaffected by domestic economic developments in Sweden. In the model, a part of consumer and investment goods are imported from the rest of the world and a part of domestic output is exported.

The model can be formulated as a number of mathematical conditions that describe how households and firms act, given the assumption of an optimizing behaviour with rational expectations, and with the short-term nominal interest rate controlled by Riksbank (the central bank of Sweden). RAMSES contains 24 equations for this. Together with the equations for conditions abroad, the conduct of fiscal policy and the course of shocks to the economy, they constitute a consistent mathematical system of non-linear differential equations that produces a fairly acceptable picture of how the economy develops over time.

For our purpose, it is relevant to underline that the theoretical model variables in RAMSES are linked with measurement equations to statistical observations of, for example, output, prices and interest rates. It is not necessary to include observed variables for every one of the model variables. There is in principle no empirical counterpart to some model variables, and there is no satisfactory way of measuring others. The estimation procedure consists of using RAMSES' modelling structure together with the selected observed variables

to form a picture of the variables for which measurements are not available. It is, however, important that the set of observed variables is sufficiently informative to identify the model's parameters and the underlying unobservable model variables.

Estimations of RAMSES are currently performed with the following 15 macroeconomic variables: GDP deflator, consumption, investment, real wages, real exchange rate, policy rate, hours worked, GDP, exports, imports, UND1X, investment deflator, foreign GDP, foreign inflation and foreign interest rate. RAMSES is estimated on data from 1986 Q1 up to the present. The choice of estimation period has to weigh quantity against quality: plenty of data is needed for the accuracy of the parameter estimates but the data should also refer to a period without sizeable structural changes.

As stated in Adolfson, Laséen, Lindé and Villani (2007)

> *As many aspects of the economy are modelled far too simply in RAMSES, this model is not appropriate for a number of purposes. One important example is that RAMSES does not include financial frictions, the importance of which for understanding the monetary policy transmission mechanism is discussed extensively in the literature. Another example is fiscal policy's very limited role in RAMSES. However, the work of developing models is a continuous process at the Riksbank and aims to learn from the shortcomings that are an inevitable feature of this field. In this work it is, of course, also important to document the models' empirical properties.*
>
> *[…] we wish to emphasize that while we do not believe that formal models such as RAMSES can replace the extensive analytical work of sector experts and others, the development of the new generation of general equilibrium models has now proved so successful in various ways that these models have earned a prominent place in a central bank's toolbox.*

Sveriges Konkurrenskraft – Mäta nationell konkurrens (Competitive Edge- Measuring national competitiveness)

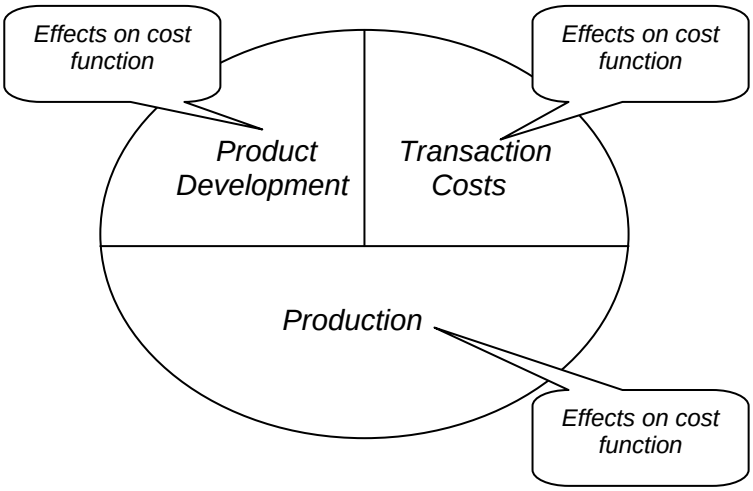
In a recent report by ITPS (ITPS A2007:002), Sandro Scocco highlights that the idea of competitiveness has acquired increasing importance over the past decades. The author has therefore developed an unequivocal definition of competitiveness that makes it easier to measure and understand the concept.

A nation's competitiveness is seen in the ITPS report as the total working population's earnings in relation to the equivalent for a grouping of eleven comparable OECD countries. The measurement of Swedish competitiveness then becomes the change in the earnings value of the working population's production in relation to the corresponding change for countries with similar production preconditions. If the value for Sweden increases faster than the comparison group, then Swedish competitiveness has strengthened.

A country's competitiveness can be analysed starting from the micro level and taking into account the cost function of individual companies. Competitiveness is a relative measure that can not be expressed in absolute terms but, as shown in Figure 7, it can be expressed as a function of companies' costs. A company faces three different kinds of costs: production costs, product development costs and transaction costs, and, given these costs, the company is considered to be competitive when the costs are internationally acceptable and total costs do not exceed those of competitors with the same quality and price. It is important to underline that these are costs that affect the company's value added and are there-

fore not costs related to intermediate goods and services such as electricity or raw materials.

Figure 7 Firms' added value.



Source: ITPS A2007:002

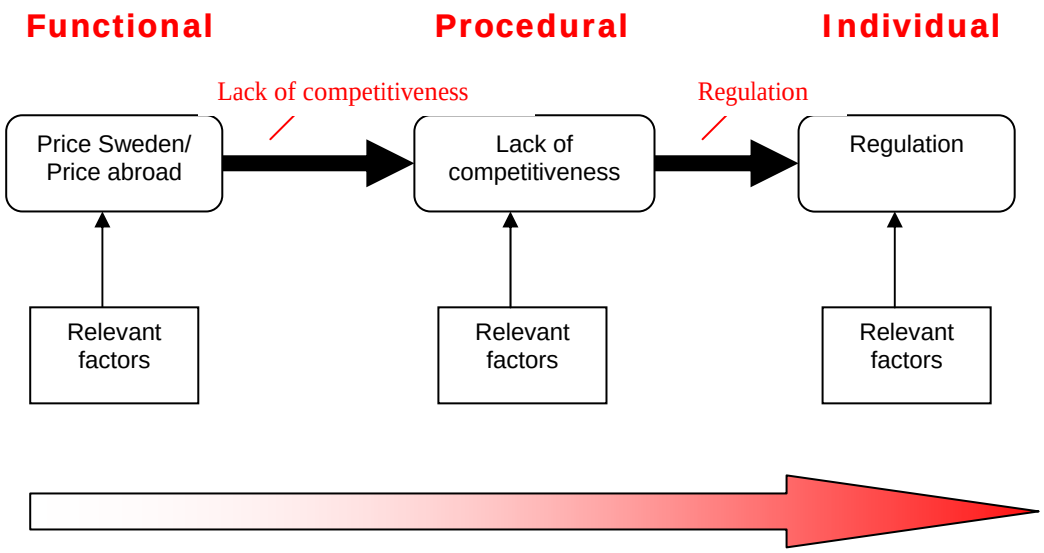
Taking prices and costs as a starting point, it is therefore possible to build an interesting and measurable link between micro level (firm level) and macro level (national accounts).

For the purpose of this report, the fact that ITPS has developed a measurable concept of competitiveness based on companies' cost function is crucial. It is evident that, by means of state aid and grants, the government affects the costs faced by companies. The state can have an effect on the level of production via resources, rules and regulations and use of authority.

By means of a three-step analysis (functional level, procedural level and individual level) of the three main cost spheres (production, product development and transaction costs), the government therefore has a tool to identify the relevant factors and actors that affect the cost function and, subsequently focus its attention at the firm level.

By implementing the three-step analysis shown in Figure 8, the government could make sure that the questions at hand are economically and politically relevant. Such analysis would therefore lead to more economically suitable policy measures.

Figure 8 Three-step analysis.



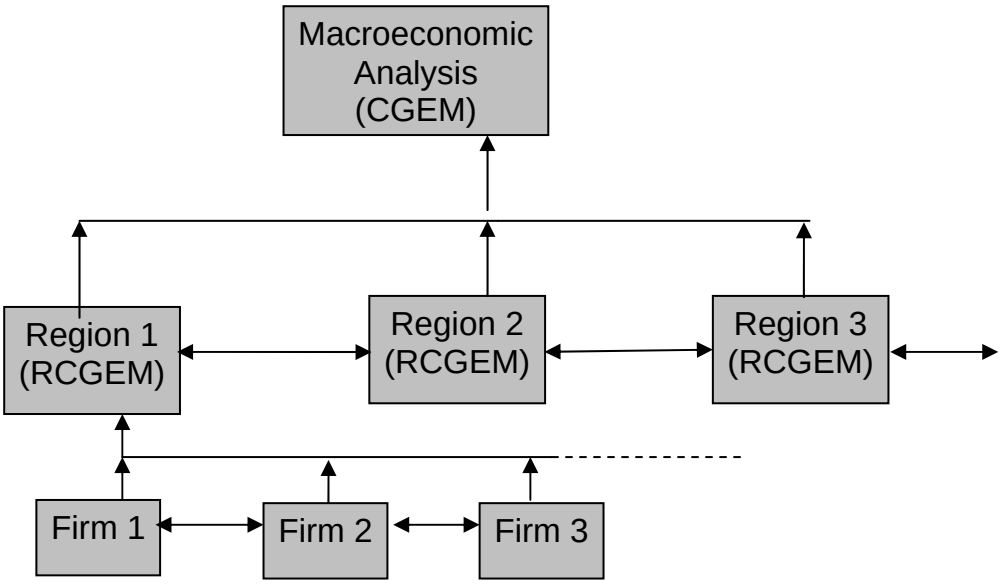
Source: ITPS A2007:002

In the next section, we present in more detail a model that can be considered the technical specification of the concept presented by Scocco. The model proposed is build\t along the lines of the RAMSES model and Scocco's model, and is intended to provide the econo-metrical tool to build an interesting and measurable link between micro level (firm level) and macro level (national accounts).

4 ITPS model

When discussing the effects of public policies, there are different kinds of analysis possible according to the level of aggregation and the recipients of the policy. Effects can in fact be studied at the firm level, the regional level and the macroeconomic level. For example, the introduction of a tax reduction policy addressed to a certain branch not only has effects on the firms belonging to that branch, but on external companies such as suppliers or customers as well. When evaluating the policy's effects, we have to be able to analyse firm-specific spillover effects, which investigate how different firms are interrelated to each other and, consequently, capture the effects on both recipients and non-recipients of policy measures. Moreover, we can not neglect the geographical location of recipients, and in addition to firm-specific spillover effects[36] must also consider regional spillover. Finally, the introduction of a tax reduction affects the general equilibrium of the economy at a more aggregated level.

Figure 9 ITPS model.



As shown in Figure 9, the ideal way to investigate the effects at the three different levels would be to start with a firm-specific analysis of every company in the economy, group them at regional level to investigate regional effects, and, finally, combine the regions in order to study the effects on the entire economy.

However, given the great number of companies in the economy, such an ideal procedure would be both time-consuming and technically impossible.

A possible solution is to build up a flexible system where we can investigate the effects of different policies from time to time on smaller groups of firms. This would allow us, if required, to investigate how changes affect a single firm, although not all the firms at the same time.

[36] *A direct effect, either positive or negative, on someone's profit or welfare arising as a by-product of some other person's or firm's activity.*

Firms could then be grouped to create groups as homogeneous as possible within the groups and as heterogeneous as possible between the groups. For example, if we are interested in studying the sectoral effects of a tax reduction for the manufacturing sector, we could group the companies according to the sector they belong to, while, if we want to investigate the effects of a tax reduction related to emission rights, we may group companies according to emission levels regardless of the sector they belong to. Once we have grouped the companies homogeneously, we can select a subgroup of representative companies for each group.

According to Adolfson, Laséen, Lindé and Villani (2007), until the beginning of the 1980s, the analysis and forecasting of inflation and business cycle was dominated by Keynesian models. Such models assume that players in the economy are governed by various rules of thumb, and they do not believe that expectations of the future are formed rationally; instead, simple projections of earlier patterns in the data are used. The use of Keynesian models was mainly justified by the lack of technical tools such as theories and computers that are essential for solving complex systems of equations with forward-looking expectations. Another reason was that the earliest versions of general equilibrium models for studying macroeconomic developments seemed to be at odds with the data.[37]

However, the Keynesian models, which emphasized the importance of demand for understanding business cycles, ceased to describe data as well as before from the 1970s, when stagflation[38] showed up in the Western world and these models failed to catch a number of structural shifts.

In the past decade, extensive academic research, using technical innovations, has developed a new generation of macroeconomic general equilibrium models where the emphasis in economic description is on the supply side but where demand in the short run also plays a role, through the existence of various market imperfections combined with nominal and real rigidities. Due to market imperfections and rigidities, the responses to various disturbances occur more gradually in the model. These second generation macroeconomic general equilibrium models are commonly referred to as New Keynesian models, and have sound, well-documented empirical properties.

The RAMSES model, as well as the model proposed in this report, are based on BVAR estimation and clearly show that the lack of technical tools such as theories and computers and the lack of data are no longer a problem.

Existing general equilibrium models in Sweden (i.e., rAps and EMEC) have until now not been used to investigate regional issues. This is due to various factors, in particular that: (1) the models were built to study specific questions and consider the economy at an aggregate level, and (2) according to the developers of the models themselves, regional analysis would require data on the flows of goods between regions,[39] such as those available in the national accounts, which are practically impossible to trace.

---

[37] Kydland and Prescott's (1982) fundamental model of the real business cycle, for instance, which emphasized the supply side's importance for understanding macroeconomic developments, was criticized for a lack of empirical realism. The economy seemed to be characterized by much greater rigidities than might be expected if it were governed by market mechanisms and rational households and firms.

[38] A combination of low growth and high inflation.

[39] The only model that takes into account the regional flows of goods is rAps but it does it in a way that is not suitable for ITPS's purpose.

Even if such data should become available, given that ITPS is interested in the effects of public policies on employment and productivity at a regional level, the existing models would still not be suitable and would require a great amount of changes.

In order to serve our purpose, we need:

- A model built on a regional basis.
- A model built on the base of data availability.

Given the existing data and the questions to be answered, we can build a model similar to RAMSES/BVAR developed by Riksbank but on regional levels, in the following way.

Since our interest is to evaluate the effects of policies specifically addressed to firms or groups of firms on a regional level, we need to be able to highlight the different effects on recipients and non-recipients. It may also be of interest to study the spillover and causality effects of a new policy implementation, which may be of use to policymakers interested in an ex-ante evaluation of the range of effects of different interventions.

Since there is data available on annual financial reports, we can use this as a starting point in the study of recipients and non-recipients. Using panel data – a dataset containing data on a certain number of firms (n) over a certain time period – we can estimate demand and supply in different ways. One possibility is to apply a pooling approach based on very strict assumptions, such as that all firms belonging to a given sector adopt the same technology, irrespective of what they produce. This means that the marginal effects of a certain policy are the same for all firms in a given sector, and it would therefore not be possible to investigate how different firms are affected by the policy.

An alternative and less restrictive approach is to allow different technologies in different sectors, meaning that the parameters are sector-specific for a given level of aggregation and, in practice, we estimate different sector-specific partial systems. The advantage of this model as opposed to the pooling approach is that all the parameters can vary among sectors; the disadvantage is that the chosen level of aggregation does not necessarily correspond to the real differences in technology.

Partial CGE models estimate unit and cross-price elasticity of goods and then study how the demand for a given investment good fluctuates due to changes in the price of other goods, everything else unchanged.[40] To develop and maintain a partial equilibrium model is not as time-consuming as for a general equilibrium model, and partial CGE have therefore been used much more often to investigate the effects of different policies.

Instead of panel data, time series data can be used to study the effects of a given policy. In factor demand models, as mentioned above, panel data are used to estimate the effects of policies at an aggregate level. With time series data for every firm, it is possible to investigate the effects of different policies at the firm level.

It can be interesting to study how different policies affect the level of production in different companies and sectors. For example, reduced costs due to state aid are likely to lead to higher levels of production, but companies engaged in the same sector do not necessary have the same cost function, and policy effects may be different for different firms. It can therefore be interesting to study externalities – to investigate how the level of production in

[40] This method has been used in many studies; see for example Dargay (1983), Berndt (1991), Brännlund (1997) and Brännlund and Lundgren (2004, 2005).

one company affects the level in other companies in the same sector – and Granger causality when implementing certain policies. If the current level of production in Firm 1 can be better forecast using data on the production level of Firm 2 for the previous year, we can say that production in Firm 2 Granger-causes production in Firm 1.

Assuming *n* firms in a branch, we define Y_{it} as the total production in firm *i, i=1,…n,* at time (*t*), and assume that Y_{it} depends on the total costs (C_{it}), on the firm's production at time *t-1,* and on the level of production of a concurrent firm (*j*) in the same branch at time *t*-1 (i.e., Y_{it-1} and Y_{jt-1}, where i≠j). We also assume, for simplicity, that there is a log-linear relation between variables.

For example, we can write a time series model, so-called VAR, as follows:

$$\ln(Y_{1t}) = \alpha_{10} + \alpha_{11}\ln(Y_{1t-1}) + \alpha_{21}\ln(Y_{2t-1}) + \ldots + \alpha_{1n}\ln(Y_{nt-1}) + \beta_1 \ln(C_{1t}) + \varepsilon_{1t}$$
$$\ln(Y_{2t}) = \alpha_{20} + \alpha_{21}\ln(Y_{1t-1}) + \alpha_{22}\ln(Y_{2t-1}) + \ldots + \alpha_{2n}\ln(Y_{nt-1}) + \beta_2 \ln(C_{2t}) + \varepsilon_{2t}$$
$$\vdots$$
$$\ln(Y_{nt}) = \alpha_{n0} + \alpha_{n1}\ln(Y_{1t-1}) + \alpha_{n2}\ln(Y_{2t-1}) + \ldots + \alpha_{nn}\ln(Y_{nt-1}) + \beta_n \ln(C_{nt}) + \varepsilon_{nt}$$

Where ε_{it} is the error term for the equation *i* and α_{ji} and β_i, *j*=0,1,…,n and *i*=1,2, …, n are the parameters. α_{ji}, for *i=j*, captures how the production of a given firm in a given period affects the production of the same firm in the following period, while α_{ji}, for *i≠j*, illustrates how the production for firm *j* in a given period is affected by the production for firm *i* in the previous period.

According to economic theory, we would expect the coefficient β_i to be negative, which indicates that if $\ln(C_{it})$ increases by one unit, keeping everything else equal, $\ln(Y_{it})$ decreases by β_i. The described model is a simplified version of many possible models.[41]

If state aid leads to increased production in a certain branch, it could at the same time have an effect on the rate of employment in the same branch, at least in the short run. The effects on employment in a given branch can be evaluated by means of a time series model; moreover, it is possible to develop a time series model to capture effects between different branches of a given policy.

We should, however, keep in mind that even time series models have their limits, such as the need of a sufficiently long time series and the practical difficulties of estimating a system with many variables. A possible solution is to group homogeneous firms in each branch by means of matching methods (Rosenbaum and Rubin, 1983) and select a representative company or group of companies from each group to be included in the system of equations to be estimated. The grouping of firms can be done by using different kinds of matching estimators such as those described above (i.e., propensity score, difference in difference, Kernel, etc.), according to the purpose.

[41] *For a better description see also Brockwell and Davis (2002) and Hamilton (1994). Fabiani et al. (2000) have applied a structural vector autoregressive (sVAR) model to explain fluctuation in unemployment.*

By means of a time series VAR model, we therefore obtain the parameters α_{ji} and β_i for the representative firms and use them to study the effects of policy implementation at the firm level, by "firm level" meaning the groups of homogeneous firms. Note, however, that VAR models also allow us to study a particular firm, though not all the firms in the economy at the same time because of the large number. As mentioned above, ITPS is also interested in studying regional effects. The easiest way to do this is to apply the VAR model at the regional level in a similar way as described earlier for the firm level. The grouping at regional or firm level is dependent on the objectives of the policy.

VAR models are suitable for short- to medium-term forecasting, and the length of the forecast period depends on the length of the available data.

To study the effects of a certain policy on the entire economy, or to forecast in the long run, we can create a SAM matrix based on the parameters estimated using the VAR model. The advantage of this model is in fact that the parameters are directly estimated from historical data rather than derived from theoretical assumption as in the majority of the existing models.

Such a model is therefore quite complete and allows us to study the economy at different levels. But is it feasible and achievable? If so, what resources would be required? What knowledge and skills are necessary? How many people would be involved? What kind of data is needed?

It is difficult to give clear-cut answers to these questions, but we have tried to estimate the time and costs of developing such models through discussing these matters with people in charge of similar models. From our discussions with Göran Östblom and Magnus Sjöström from KI, we can conclude that developing a model like EMEC would require at least 3–4 full working years and a group of 8–10 researchers. EMEC is not the only model that took time to develop – LUMOD, ISMOD, FIMO and others were also developed over long periods of time. Neither do we expect the proposed model to be an exception to the rule when it comes to the time needed.

Developing a model requires researchers who can dedicate all their time to the project and it is hardly plausible to think of it as a side activity to the analysis work that ITPS carries out on a daily basis. It would therefore be necessary to hire a group of 6–8 people to complete the task, without taking resources away from ITPS's existing activities, for a period of 3–4 years.

As far as data are concerned, most of the data needed are already available from MM Partners, and ITPS has already purchased or collected most of the data that the model requires.

If the project is approved, we expect to carry out the work in three steps:

1. Estimate the variables and test whether the presented model is feasible – 6 months.
2. Carry out a test evaluation of one policy – 18 months.
3. Adapt the model to different policies in order to make it fully functional – 36–48 months.

5 Concluding remarks

This work has illustrated the opportunities for better development of the evaluation methods and culture with respect to the evaluation of the effects of state aid programs on the economy.

The first part of the work has taken into account the existing evaluation culture in different countries – Holland, USA, Ireland and Sweden – and we can summarize the lessons learned as follows:

- Importance of a good policy design that is well-thought-out as an excellent starting point for evaluation.
- Importance of detailed and systematic data collection.
- Programs should be transparent.
- Clear policy goals and description of how to get there.
- Need to increase retrospective analyses.
- Need to incorporate both direct and indirect path analysis in cost-benefit studies.
- Importance of continuous monitoring.
- Relevance of developing new evaluation techniques.
- Identify and address new questions that arise from evaluation.
- Take greater advantage of evaluation results in decision-making processes.

The analysis of the Swedish evaluation culture emphasizes the need to further our understanding of the effects of different policies and to take into account the results of external experts. Monitoring by itself has been shown to be unable to offer relevant insight into policy effectiveness, and a further step towards a more scientific evaluation is required in order to provide the legislators with the tools to design new policies and programs.

Sweden needs to move towards a more complex evaluation culture in order to trigger a learning process in which evaluations provide information and create the necessary knowledge for an improved policy design. The report proposes two alternatives paths that Sweden can take to improve the effectiveness of evaluation: (1) to appoint an external agency to do both ex-ante and ex-post evaluation, or (2) to appoint a single agency responsible for policy implementation, analysis and evaluation.

The second part of the report focuses on evaluation methods and proposes a model that could be developed by ITPS in the future. The main conclusion of this second part is that the counterfactual situation – which answers the dilemma of what would have happened if the policy had not been implemented – can be analysed at both the micro- and the macro level, depending on the objective of the policy, the size of the effects expected, and the data available.

For our purpose, ITPS proposes a so-called “ITPS model” to study the effects of state aid at three different levels: the firm level, the regional level and the national level. The best way to proceed would be to start with a firm-specific analysis of every company in the economy, group them at the regional level to investigate regional effects and, finally, com-

bine the regions in order to study the effects on the entire economy. Unfortunately such a procedure would be both time-consuming and technically impossible. We suggest developing a flexible system, similar to RAMSES/BVAR developed by Riksbank but on a regional level, that we can use to investigate the effects of different policies from time to time on smaller groups of firms.

The development of such a model would require a shift towards incorporating a greater amount of research in ITPS's daily activities and the work of 6–8 people for 3–4 full working years. Developing a model requires researchers who are able to dedicate all their time to the project, and it is hardly plausible to think of it as a side activity to the ordinary analysis work done by ITPS.

References

Adolfson M., Laséen S., Lindé J. and Villani M. (2007). RAMSES-A new general equilibrium model for monetary policy analysis, *Sveriges Riksbank Economic Review 2007:2.*

Ahn, S.C. and Schmidt, P. (1995). Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, Vol. 68.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers.

Armington, P.S. (1969). A Theory of Demand for Products Distinguished by Place of Production, *International Monetary Fund Staff Papers*, Vol. 16.

Armstrong, H.W. (1995). An appraisal of the evidence from cross-sectional analysis of the regional growth process within the European Union. In H.W. Armstrong and R.W. Vickerman (Eds.), *Convergence and Divergence Among European Regions*. London: Pion Ltd.

Aronsson, T., Lundberg, J. and Wikström, M. (2001). The impact of regional public expenditures on the local decision to spend, *Regional Science and Urban Economics*, Vol. 33.

Arrelano, M. and Bond S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, Vol. 58.

Aten, B. (1996). Evidence of spatial autocorrelation in international prices, *Review of Income and Wealth*, Vol. 42.

Baltagi, B.H. (2001). Econometric analysis of panel data. 2d ed. Chichester, UK Wiley.

Bergvall, A., Forsfält, T., Hjelm, G., Nilsson, J. and Vartiainen, J. (2007). KIMOD 1.0 Documentation of NIER's Dynamic Macroeconomic General Equilibrium Model of the Swedish Economy, Working Paper No. 100, National Institute of Economic Research.

Bernt, E.R. (1991). *The Practice of Econometrics: Classic and Contemporary*. Addison-Wesley.

Besley, T. and Case, A. (1995). Incumbent behaviour: Vote seeking, tax setting, and yardstick competition, *American Economic Review*, Vol. 85.

Bivand, R. and Szymanski S. (1997). Spatial dependence through local yardstick competition: Theory and testing, *Economics Letters*, Vol. 55.

Boyle, R. (2005). Evaluation capacity development in the Republic of Ireland. ECD Working Paper Series No. 14, June 2005, Washington: World Bank.

Brett, C. and Pinkse J. (1997). Those taxes are all over the map: A test for spatial dependence of municipal tax rates in British Columbia, *International Regional Science Review*, Vol. 20.

Brockwell, P.J. and Davis, R.A. (2002). *Introduction to Time Series and Forecasting*. Springer.

Brueckner, J.K. (1998). Testing for strategic interaction among local governments: The case of growth controls, *Journal of Urban Economics*, Vol. 44.

Brännlund, R. and Lundgren, T. (2004). Kyoto och Basindustri – modelling och simulering, *Arbetsrapport* 347, Department of Forest Economics, SLU Umeå.

Brännlund, R. and Lundgren T. (2005). Swedish Industry and Kyoto – An Assessment of the Effects of the European CO_2 Emission Permit Trading System, *UES* 668, Department of Economics, Umeå University.

Brännlund, R. (1997). Industrins efterfrågan på energi. Appendix 4, *SOU* 1997:11.

Brännlund, R. (2006). *Grön skatteväxling- Framgångsväg eller återvändnings-gränd? SNS* förlag.

Case, A. (1991). Spatial patterns in household demand, *Econometrica*, Vol. 59.

Case, A., Rosen, H.S. and Hines, J.R. (1993). Budget Spillovers and fiscal policy interdependence: Evidence from the States, *Journal of Public Economics*, Vol. 52.

Cornet M., Vroomen, B. and van der Steeg, M. (2006). Do innovation vouchers help SMEs to cross the bridge towards science? *CPB* Discussion Paper No. 58.

Cox, C.J., Ingersoll, J.E.Jr. and Ross, S.A. Ross (1985). An Intertemporal General Equilibrium Model of Asset Prices, *Econometrica*, Vol. 53.

Cozzens S. (2003). Frameworks for evaluating science and technology policies in the United States. In Philip Shapira and Stefan Kuhlmann (Eds.), *Learning from Science and Technology Policy Evaluation – Experiences from the United States and Europe*. Edward Elgar, Cheltenham, UK, and Northampton, USA.

Dargay, J. (1983). The Demand for Energy in Swedish Manufacturing, *Energy in Swedish Manufacturing*. IUI Stockholm.

Department of Enterprise, Trade and Employment (2004). Expenditure review of IDA Ireland's property programme.

Duignan, P. (2007). Seven possible outcome designs, http://www.parkerduignan.com/oiiwa/model/

ESV (2001). Verksamhetslogik. ESV 2001:16.

ESV (2006). Måluppfyllelseanalys. ESV 2006:7.

European Community Treaty, http://www.europa.eu.int/comm/competition/state_aid/legislation/aid3.html

Fabiani, S., Locarno, A., Oneto, G. and Sestito, A.P. (2000). The Sources of Unemployment Fluctuations: An Empirical Application to the Italian Case. European Central Bank: Working Paper No. 29.

Fingleton, B. (2001). Equilibrium and economic growth: Spatial econometric models and simulations, *Journal of Regional Science* 41, p. 117–147.

Forfás (1995). Evaluation of the Applied Research Programme 1988–1994.

Forfás (1996). Evaluation of AMT Ireland.

Forfás (2003). Review of the IDA's research and development capability grants scheme.

Forfás (2004). Evaluation of agency supports for R&D in the business sector.

Hamilton, J.D. (1994). *Time Series Analysis*. Princeton University Press.

Heckman, J., Ichimura, H. and Todd, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *The Review of Economic Studies*, Vol. 64 (4).

Holtz-Eakin, D. (1994). Public-sector capital and the productivity puzzle, *Review of Economics and Statistics*, Vol. 76.

Holtz-Eakin, D., Newey, D. and Rosen, H.S. (1988). Estimating vector autoregressions with panel data in *Econometrica*, Vol. 56(6).

ITPS (2006). Statligt stöd till näringslivet 2005. S2006:007.

ITPS (2007). Sveriges Konkurrenskraft – att förstå och mäta nationell konkurrenskraft, A2007:002.

Jansson, P., Hansson, J. and Löf, M. (2005). Business survey data: Do they help in forecasting GDP growth? *International Journal of Forecasting*, Vol. 21.

Kehoe, P.J. and Kehoe, T.J. (1995). A Primer on Static Applied General Equilibrium Models. In P.J. Kehoe and T.J. Kehoe (Eds.), *Modeling North American Economic Integration*, Boston: Kluwer Academic Publishers, 1–31.

Kydland, F. and Prescott, E. (1982). Time to build and aggregate fluctuations, *Econometrica*, Vol. 50.

Lundberg, J. (2003). On the determinants of average income growth and net migration at the municipal level in Sweden, *The Review of Regional Studies*, Vol. 33(2).

Lundberg, J. (2006a). Using spatial econometrics to analyze local growth in Sweden, *Regional Studies*, Vol. 40(3).

Lundberg, J. (2006b). A spatial interaction model of benefit spillovers from locally provided public services, *Regional Studies*, Vol. 40(6).

Mas-Colell, A, Whiston, M. D. and Green, J.R. (1995). *Microeconomic Theory*. Oxford University Press.

Miles, I. and Cunningham, P. (2006). SMART Innovation: A Practical Guide to Evaluating Innovation Programmes, DG Enterprise and Industry, European Commission.

Ministry of Economic Affairs (2005). Monitoring Innovatievouchers 2004.

Murdoch, J.C., Rahmatian, M. and Thayer, M.A. (1993). A spatially autoregressive median voter model of recreation expenditures, *Public Finance Quarterly*, Vol. 21.

Nutek (2006). Effektutvärdering av produktutvecklingsprogrammet, Dnr 70-2006-435.

Nutek (2006). Nytta – Nuteks system för uppföljning och utvärdering.

Poot T., den Hertog P., Grosfeld T. and Brouwer E. (2005). Evaluation of a major Dutch Tax Credit Scheme (WBSO)
www.visioneranet.org/files/55/evaluationofDutchR_Dtaxcreditscheme.pdf

Restad, T. (2006). LUMOD – En Långsiktig UtvecklingsModell för Svensk Ekonomi, Stockholm University.

Revelli, F. (2002). Local taxes, national politics and spatial interactions in English district election results, *European Journal of Political Economy*, Vol. 18.

Rey, S.J. and Montouri, B.D. (1999). US regional income convergence: A spatial econometric perspective, *Regional Studies*, Vol. 33(2).

Rosenbaum, P and Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, Vol. 70.

Rosenbaum, P. and Rubin, D. (1985). Construction a control group using multivariate matched sampling methods that incorporate the propensity score, *American Statistician*, Vol. 39.

Rubin, D. (1978). Bayesian inference for causal effects: The role of randomization, *Annals of Statistics*, Vol. 7.

Rubin, D. and Thomas, N. (1992). Characterizing the effects of matching using linear propensity score methods with normal distribution, *Biometrika*, Vol. 79.

Shoven, J.B. and Whalley, J.L. (1984). Applied General Equilibrium Models of Taxation and International Trade: An introduction and survey, *Journal of Economic Literature*, Vol. 22.

SIKA (2004). SAMPERS och SAMGODS – Nationella Modeller för Prognoser och Analyser inom Transportsektorn, SIKA 2004.

SIKA (2004). The Swedish National Freight Model, Samplan 2004:1.

Smets, F. and Wouters, R. (2003). An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area, *Journal of the European Economic Association*, Vol. 1.

Storey, D.J. (2000). Six Steps to Heaven: evaluating the impact of public policies to support small businesses in developed economies, In H. Landstrom and D.L. Sexton (Eds.), *Handbook of Entrepreneurship.* Blackwells, Oxford, pp. 176–194.

Topa, G. (1996). *Social interactions, local spillovers and unemployment*. Ph.D. Dissertation, Department of Economics, University of Chicago.

Varian, H.R. (1992). *Microeconomic Analysis*, 3rd Ed. W.W. Norton & Company, Inc.

Vedung, E. (1997). *Utvärdering i politik och förvaltning.* Studentlitteratur.

Vinnova (2007). Vinnovas samlade strategi för uppföljning, utvärdering och effektanalys" Dnr 2006-04079.

Östblom, G. (1991). *An Environmental Medium Term Economic Model-EMEC*. Working Paper No. 69. National Institute of Economic Research.

Appendix 1 Discussion topics for interviews

A. Context of Evaluation (General context in which policies are made and reflected upon)

1. How is the evaluation of state aid regulated in the country? Norms, regulation, laws…
2. What kinds of evaluations are mainly performed? Ex Ante – Mid term – Ex post.

B. Relation between policymakers-policy implementers (The way in which new policies are formed and how evaluative information is considered in this process)

1. Describe the process of elaborating, forming and specifying a new state aid program. Actors involved. Power relations and hierarchy. Level of communication and coordination between the actors.
2. Degree of specification of programs and policies. Do you use the idea of program logic or event chain modelling?
3. Degree of freedom at the implementation level.
4. Can you give a concrete example of the decisional process in a state aid program?

C. Program's objectives (Relationship between objectives and goals and the process of setting goals)

1. Describe the process of setting objectives and suitable indicators for these objectives, the resources available to achieve them, and the stated expectations to attain the objectives. This process often couples with a process of setting milestones, that is, setting indicators to verify that you are on the right track.
2. Do you follow guidelines such as SMART – specific, measurable, adequate, relevant and time-specific?
3. Do you find that the resources allocated for the program are in general suitable for the objectives stated/chosen?

D. Monitoring and Evaluation (Evaluation practices: which ones are the most effective according to your experience?)

1. What kind of evaluations do you carry out? Examples (goal attainment, side effects, impact assessment, productivity, efficiency,…).
2. Monitoring? Do you find there is a problem of lacking or unsuitable monitoring data? Conflict between gathering information and administrating (distributing support for) the program? How do you deal with this?
3. Evaluation? At which level? Comparison of performance of the assisted firms with typical firms? Comparison with "matched" firms? Selection bias taken into account?

4. Who carries out the evaluation? Private consulting firms, universities, governmental agencies?
5. Common approaches to evaluation. Standard method for all programs?
6. Who defines the individual steps of the evaluation path? The principal or the agent? How much can the people in charge of implementation steer the program? How good are they at it? Do they know what they are trying to do at each step?

E. Learning process (What decisions have been taken due to information generated in evaluations?)

1. How does previous experience in evaluation enter the picture? Discriminate between ex-ante, mid-term and ex-post?
2. Is the evaluation useful in making decisions? Is it used effectively and taken into account?
3. Does the evaluation lead to changes in conduct?
4. What are the most common problems you face in evaluation? Which ones arise in ex-ante evaluations? How do you solve them?

Please respond to the previous questions by giving examples concerning real programs.

Appendix 2 Some programs and methods from other countries

A2.1 The Netherlands

Innovation vouchers

The 2004 Dutch Innovation Vouchers program aims to increase the interaction between SMEs and public knowledge institutes, such as universities and technology transfer institutes. The voucher is a credit note worth €7500 to be spent at such research institutes and is intended to address a perceived deficit in interaction between commercial firms and research institutes.

The program is particularly interesting because of its peculiar set up, which offers a good starting point to assess the causal impact of the policy instrument. In September 2004, 100 vouchers were allocated randomly among 1044 applicant SMEs. Because of the random allocation, any difference in innovative behaviour between firms with a voucher and firms without is purely the causal effect of the voucher and not a correlation that can be explained by other factors.

The central question is therefore whether SMEs with a voucher commission more assignments from research institutes than SMEs without vouchers. To answer this question, Marten Cornet, Björn Vroomen and Marc van der Steeg[42] from CPB measure this "interaction" as the number of assignments SMEs commission to public knowledge institutes, and define "effectiveness" as the difference in the number of assignments commissioned by SMEs with- vs. without vouchers. The overall effect is distinguished in number, size and timing of the assignments.

In order to determine the effectiveness of innovation vouchers, CPB gathered information on the research assignment of those firms that participated in the voucher round in 2004. Some information was obtained from the application form itself and some from a specially prepared survey conducted by telephone among a sample of firms that were allocated a voucher (winners) and firms that were not allocated a voucher (losers).

With this information-gathering arrangement, it was possible to distinguish between the total group of applicants and the subgroup of survey participants. The application form provides information on a firm's turnover, size (in terms of staff numbers), industry and region. By contrast, information on research assignments is only known for those firms that participated in the survey. Information obtained from the latter group was used to examine the effectiveness of the innovation voucher.

The questionnaire was prepared in cooperation with SenterNovem in order to obtain information on the research assignments SMEs placed with research institutions. The firms were asked to provide information on the timing of the assignment, the value of the assignment, and which institution answered their research question; they were also asked to answer several questions aimed at gauging their satisfaction with various aspects of the assignment. In addition, the firms were asked to respond to several statements about the

[42] *Cornet M., Vroomen B. and van der Steeg M. 2006 .Do innovation vouchers help SMEs to cross the bridge towards science? CPB Discussion Paper No. 58.*

innovation voucher, as well as what they would have done if, as applicable, they had or had not been allocated the voucher. The interviews were qualitative and semi-structured, which offers an opportunity to discuss more complex subjects and to make more intensive use of "open" questions.

Research by the Ministry of Economic Affairs[43] shows that the group of 1,044 innovation voucher applicants reflects a cross-section of SMEs in the Netherlands. 313 firms of the 1044 that applied for the vouchers were contacted. Of the 313 survey participants, 142 had commissioned one or more assignments from research institutions, yielding a total of 158 reported assignments. Of the 71 voucher winners who took part in the survey, 66 had commissioned a total of 76 assignments. Most of the firms that had not commissioned any research assignments said that cost had been a serious obstacle. Among the firms that were allocated a voucher, one-third were not satisfied with the relationship between price and quality.

From the questionnaire, however, CPB was able to gather two types of information about the behaviour of firms concerning research assignments: actual assignment commissioning (actual behaviour) and reported opinions (hypothetical behaviour). The two information sources are used complementarily to provide answers to the three research questions, namely, the effect on the number of assignments, the effect on the value of assignments, and the effect on the timing of assignments.

Both sources yield the same results with regard to the additionality of the voucher, which is estimated at eight out of ten. One out of ten vouchers is not used and one out of ten is used for assignments which would have been commissioned anyway. On the basis of the responses to propositions, there is no indication that the voucher has had an effect on the value of assignments. These responses do give some indication of a small timing effect, however, in the sense that a limited number of assignments were brought forward.

WBSO – Promotion of Research and Development Act

The Promotion of Research and Development Act (WBSO) in the Netherlands took effect in 1994. The WBSO provides a fiscal facility that reduces wage costs for R&D employees by reducing wage tax and social insurance contributions, and represents the most important measure for the promotion of corporate R&D activities in terms of scope and budget. Given the aim of the WBSO, a central question in the evaluation[44] is whether- and to what extent the WBSO leads companies to conduct more R&D activities (1st order effect) and to become more innovative (2nd order effect). The evaluation also looked into: the effects of WBSO on firm innovation (2nd order effect); the effects of WBSO on firm performance (3rd order effects); the degree to which the circumstances that led to the introduction of WBSO still pertain; WBSO's target group penetration; the perceptions of the implementation of WBSO; and, finally, the perceptions of WBSO users regarding potential (budgetary neutral) changes to the design of the WBSO scheme.

In order to investigate the questions mentioned above, a combination of methods was used:

- Econometric analysis. In 1998, the Bartels Bureau conducted an analysis of the WBSO together with Statistics Netherlands (CBS). One of the methodological findings was that data restrictions made it difficult to quantify the effects of the WBSO. In the

[43] *Ministry of Economic Affairs (2005). Monitoring Innovatievouchers 2004.*

[44] *Poot T., den Hertog P., Grosfeld T. and Brouwer E. Evaluation of a Major Dutch Tax Credit Scheme (WBSO), www.visioneranet.org/files/55/evaluationofDutchR_Dtaxcreditscheme.pdf*

evaluation by Poot *et al.* (2002), an econometric analysis was performed, building on the experience of the aforementioned study and on the basis of an improved dataset on WBSO user profiles, built up over a number of years. This made it possible to evaluate the primary and secondary effects of the WBSO.

- Telephone survey. In a detailed field study, 500 companies (net response) that use or had used the WBSO facilities were asked about decision-making on R&D, the effects of using the WBSO scheme, experiences with the implementation of the WBSO scheme and potential improvements in its design. In the processing of the results, the authors distinguished between various dimensions, such as size category, sector, R&D intensity, WBSO intensity, type of WBSO user (e.g., structural, occasional, newcomer), type of project, use of an intermediary/subsidy advisor, and whether or not the company was a high-tech start-up.
- Desk research. In addition to a detailed analysis of the WBSO evaluation conducted in 1998, the most recent scientific insights and policy studies in the field of quantitative evaluation research and the use of tax credit schemes was listed and included in the design of the evaluation and analysis of the results.
- Interviews. Semi-structured interviews were conducted with representatives of a limited number of companies and research institutes. These interviews were useful mainly to shed light on the initial insights and to gain further understanding of the use of the WBSO, particularly by companies that were not included in the field study or the econometric analysis.

Notwithstanding the completeness of methods and the objectivity of the evaluation, the authors have had to make several assumptions on substitution and additionality[45] and face certain practical methodological problems:

- Despite using a better dataset than the one used in the 1998 evaluation, the data still has limitations that affect the econometric analysis.
- One of the most important issues discussed is the causality between the WBSO and an increase in the scale of corporate R&D activities. This point received an exceptional amount of attention in the evaluation, but a more definitive finding on the causality would require panel data enabling companies to be monitored over a longer time period.
- The data available do not allow studying the effect of the WBSO on company performance by econometric means.
- Econometric evaluations of the WBSO and similar schemes would benefit from the availability of panel data over a sufficiently long period.

[45] *(1) First and foremost, that there is a direct link between R&D labour costs and the total R&D expenditure. When considering feasible effects, they assume that this ratio is constant and regard R&D labour costs as a proxy for the total R&D expenditure, and do not take the methods used to appropriate WBSO funding into consideration. (2) Secondly, they assume that firms with their own R&D expenses (including R&D labour costs) will also apply for WBSO. In other words, firms that failed to apply for WBSO funding in a given year had no, or an extremely low level of R&D expenditure. (3) The third assumption is that the planned R&D labour costs in a given year are the decisive factor for the total amount of WBSO R&D labour costs applied for. (4) Fourthly, that the reduction received is the same as the reduction applied for.*

- Effects on corporate performance (tertiary effects) can only be determined if longer time series are available for more variables on a disaggregated level, for example, with regard to market expectations.

A2.2 Ireland

At the agency level, established policies are put into operation through various types of programs. Enterprise Ireland runs a variety of different support programs, including:

- A program for the development of e-trade in companies and businesses, consisting of consultancy cheques for companies up to a maximum value equivalent to 10,000 SEK.
- A program aimed at improving the competitive ability of small and medium-sized companies by improving their export potential, whereby a single company can receive support to a maximum equivalent to 4,650,000 SEK, no more than 1,860,000 SEK (equivalent) of which may be an infusion of new capital.
- A program to support product development at manufacturing companies. The total support payable per company is the equivalent of about 6,000,000 SEK, of which about 2,000,000 SEK must be repaid.
- Support for company research projects whose estimated costs will exceed 27,000,000 SEK (equivalent). The maximum support payable amounts to 45 per cent of the costs qualifying for support.
- Consultancy cheques for small companies with the aim of stimulating innovative solutions to business problems. Total support equivalent to approx. 45,000 SEK.
- Enterprise Ireland has interests in a number of venture capital companies in which the amount of support varies. If support is given for the first time, the capital infusion varies between the equivalents of 700,000 SEK and 20,000,000 SEK.
- A program aimed at supporting the development of new business ideas or working up new markets. The maximum support available is approx. 600,000 SEK (equivalent) for a period of two years.

IDA Ireland, the equivalent of the Invest in Sweden Agency, offers four types of support to companies wanting to make a direct investment in Ireland – employment support, skills enhancement support, R&D support, and capital support. Employment support varies from €1,250–€12,500 per employee. Skills enhancement support covers wage costs for an employee for an agreed period of training. R&D support covers up to 35 per cent of the costs qualifying for support in an R&D project, to a maximum of about 4,000,000 SEK (equivalent). Capital support is intended to assist in the purchase of land or construction of new production plants and varies from 7 per cent to 40 per cent of the costs qualifying for support.

Apart from this, Ireland offers one of Europe's lowest corporate income tax rates at just 12,5 per cent. In the matter of government aid, over the past few years, the Department of Enterprise, Trade and Employment has been working on modification of existing directives in order to ensure that companies in need of aid actually get it.

The following brief review shows that qualitative methods dominate the evaluations.

- *Evaluation of the Applied Research Program 1988–1994.*[46] The evaluation comprises an examination of performance levels by using data generated by the program, a questionnaire sent out to participating companies, interviews with key informants and *in situ* visits to the technical colleges and universities involved.
- *Evaluation of AMT Ireland*.[47] Ireland's oldest program for the support and promotion of advanced technology. Methodologically, the emphasis is on peer reviews of projects, interviews with program users, and review of overall control of the project's progress. The evaluators comment on the initiatives selected, an approach that has also been used previously.

  > *In 1993, we returned to monitor the implementation of our recommendations and the subsequent progress of the program. We believe that this process of evaluation and follow-up has been among the major factors leading to the significant improvements in program performance. It reinforces our faith in so-called 'formative' evaluation, where the job of the evaluator is not merely to award marks for performance, like an Olympic judge, but to intervene to help improve performance, like a sports coach.*

- Evaluation of the IDA's support program for R&D potential.[48] This program aims at providing support for companies already established in Ireland, with a view to developing their R&D operations. The evaluation uses interviews with various groups and reviews all the information available about the program.
- Evaluation of R&D support provided to the business community by agencies and authorities.[49] Evaluation is based on both qualitative and quantitative methods. Interviews, focus groups, *in situ* visits and interviews, and studies of the literature are among the methods used for qualitative evaluation. Project data have been used for quantitative analyses. Deadweight discussions are held on the basis of interview data.

[46] *Evaluation of the Applied Research Programme 1988-1994. 1 Dec. 1995, Forfás.*
[47] *Evaluation of AMT Ireland, 1996, Forfás.*
[48] *Review of the IDA's research and development capability grants scheme, 9 May 2003, Forfás.*
[49] *Evaluation of agency supports for R&D in the business sector, 25 October 2004, Forfás.*

Appendix 3 Brief description and review of different econometric programs

There are a variety of econometric program packages, all of which are more or less specialized for various purposes. Generally speaking, the more user-friendly a program is the less specialized it tends to be – and also less flexible in terms of its scope for programming estimators not already built in. Application of the very latest estimation methods within a given field often requires a flexible program allowing a high degree of personal program input. This, however, demands greater skills of the user. Not only must users be capable of familiarizing themselves with the latest research findings, but they must also be able to put them to practical use by constructing their own programs.

Generally speaking, programs that allow user programming, matrix algebra and non-linear optimization can be used for estimating spatial econometric models and tests. The great advantage of specialized programs is that they offer a simple means of creating various types of weighting matrices. For example, it is by no means always simple to program procedures for standardizing lines or creating cut-off values in a weighting matrix. Here, too, user programming of both descriptive tests for spatial dependence and spatial dependence regression models requires considerable skills in econometrics.

General econometrics programs

A general econometrics program may be used to estimate other types of models than spatial models. It may naturally be that there is no spatial dependence, in which case a general econometrics program is to be preferred, since specialized programs often offer only a limited number of other types of estimators. The disadvantage is that general programs often have few built-in applications to estimate spatial models and hence place greater demands on the user's programming ability.

Stata

Stata is a highly capable general econometrics program that offers the user considerable scope for programming, a large number of pre-prepared estimators, and various types of econometric tests. In its basic version, however, there are no special applications for spatial econometrics, although Moran's *I*, Geary's *C*, the various *G*-statistics and maximum likelihood estimators for both the team and error models can be downloaded from Stata's website. In theory, the program's considerable flexibility, enabling both user programming and matrix algebra, makes it perfectly possible for users to program and modify weighting matrices. However, this places great demands on the user's own expertise.

GAUSS

GAUSS is essentially a matrix algebra program that to start with served as the framework for SpaceStat (see more detailed information on SpaceStat below). GAUSS is extremely flexible at the same time as it is fairly non-user-friendly in that users mostly have to program it for themselves. The advantage of this is that, in principle, the program sets no restrictions on what can be done or the estimators that can be used, meaning that experienced users can easily program in new tests and estimators. The disadvantage of this program is that it expects users to be highly skilled in both econometrics and programming.

SAS

SAS, too, is an extremely powerful econometrics and statistics program giving users considerable scope to design their own programs. The program's strength lies in its time series analysis and forecasts. It also enables matrix algebra and optimization, although there are no ready-to-use applications for spatial econometrics.

Limdep

Limdep is a fairly flexible program and offers a large number of pre-programmed estimators and procedures. Limdep was originally developed to calculate so-called Tobit estimators, hence the name *Lim*ited *Dep*endent variable models. The program also offers a number of analytical tools and estimators for cross-section and panel data. Although Limdep can be used to estimate spatial models, its lack of pre-programmed applications means that everything must be programmed by the user.

SPSS

SPSS is perhaps the most user-friendly of the programs discussed here and is hence also the most limited. The program's strength lies in other types of models, not spatial econometric models.

Matlab

Matlab is a highly flexible mathematics program comprising a number of built-in optimization routines. In theory, Matlab is capable of handling all types of mathematical calculations in applied, symbolic and theoretical mathematics, and also offers a set of well-developed simulation routines. Although there are no specific functions for spatial econometrics in the basic version, everything can be programmed. In addition, there is a wide range of applications that can be downloaded from the Internet, so that more and more users involved in spatial econometrics at an advanced level are now switching from other programs to Matlab.

TSP and E-views

E-views was originally a variant of TSP adapted to what in the early 1980s were known as "microcomputers". Since at that time the original version of TSP was too large to be installed on microcomputers, a version known as micro TSP was developed instead. It was this that later became E-views. The strength of both of these programs lies in dynamic models and panel data. They also offer a certain degree of flexibility in that they enable certain matrix algebraic operations. They are, however, not suitable for converting weighting matrices or more sophisticated algebraic functions. Neither of these programs has built-in modules for spatial econometrics and must be considered fairly unsuitable for models of this type.

S-PLUS

S-PLUS is an excellent econometrics program as it enables users to add a specific module for spatial models. S-PLUS can be linked to the ArcView and Grassland mapping programs so as to produce a visual illustration of the results in the form of a map. Among other things, S-PLUS incorporates global and local tests for spatial correlation (Moran's *I*, Geary's *C*, LISA) and a variety of different regression models. The program also offers pre-programmed applications for both defining a weighting matrix and modifying it later on the basis of various criteria. Used together, S-PLUS and, say, ArcView must be considered an excellent combination for spatial analyses.

Special econometrics programs for spatial models

SpaceStat

SpaceStat is a specialized program for spatial econometrics and has a variety of built-in estimators and tests. One of its strongest points is that once a weighting matrix has been loaded, the matrix can be very simply transformed to meet various new criteria. For example, lines can easily be standardized, cut-off values can be introduced, and the matrix can be manipulated in several other ways. In addition, Moran's *I*, Geary's *C*, the various G_i statistics and several types of ML and IV estimators (including GMM) are all pre-programmed in SpaceStat. It is also possible to perform simple matrix algebra and introduce structural instability into the regression models. Another of SpaceStat's advantages is that it can be linked to ArcView, which is a GIS-based mapping program. This enables, for example, local tests for spatial dependence to be simply visualized in the form of maps.

GeoDa

GeoDa is a fairly user-friendly program intended primarily for descriptive analysis and graphic renderings of data in the form of maps. Tests for spatial correlation and indicators for spatial outliers (both local and global tests for spatial correlation) are built into the program. There are also well-developed routines for transforming the weighting matrix to meet selected criteria. GeoDa is not capable of regression analysis, however, which must be regarded as a severe limitation for more advanced analysis. In spite of this, it may still be useful as a complement to other programs, especially in view of the fact that it can be downloaded free from the Internet (https://www.geoda.uiuc.edu/).

STARS

STARS, Space-Time Analysis of Regional Systems, is a program for descriptive analysis of data sets in which there is a dependence both between regions and over time. The program is relatively user-friendly and incorporates applications for graphic representations of various processes. The advantages of STARS lie in its ability to produce graphic representations of regional development trends over time and to run descriptive tests, both local and global, for spatial correlation. Its weakness is its lack of tools for regression analysis. Therefore, like GeoDa, STARS should be regarded more as a complement to other programs. STARS can also be downloaded free from the Internet (http://regionalanalysislab.org/?n=STARS).

The Swedish Institute for Growth Policy Studies (ITPS) is a Government Agency responsible for providing policy intelligence to strengthen growth policy in Sweden. ITPS primarily provides the Government Offices, Members of the Swedish Parliament, other state authorities and agencies with briefings based on statistical material, policy papers and key analyses. Business policy and regional development policy are areas given high priority.
Changes in policy should be based on:

- Statistic data and analyses of the structure and dynamics of industry – to obtain an up-to-date view of future challenges and opportunities.

- Evaluation of results and effects of policy measures and programmes – to provide benchmarks and learn from measures implemented earlier.

- Policy intelligence in order to look outwards and ahead – what issues are likely to come on the growth policy agenda in the future?

These represent the principal missions of ITPS.

ITPS, Swedish Institute for Growth Policy Studies
Studentplan 3, SE-831 40 Östersund, Sweden
Telephone: +46 (0)63 16 66 00
Fax: +46 (0)63 16 66 01
info@itps.se
www.itps.se
ISSN 1652-0483 (web), 1652-8684 (print)

itps SWEDISH INSTITUTE FOR GROWTH POLICY STUDIES