Investment Support under Rural Development Policy

Contract 30-CE-0609852/00-41

FINAL REPORT
12 November 2014

Written by Metis/WIFO/aeidl
November – 2014

EUROPEAN COMMISSION

Directorate-General for Agriculture and Rural Development
Directorate E — Economic analysis, perspectives and evaluation; communication
Unit E.4 — Evaluation and studies

E-mail: agri-evaluation@ec.europa.eu

European Commission
B-1049 Brussels

Investment Support under Rural Development Policy

Contract 30-CE-0609852/00-41

FINAL REPORT
12 November 2014

Directorate-General for Agriculture and Rural Development
Investment Support under Rural Development Policy
2014

Europe Direct is a service to help you find answers to your questions about the European Union.

Freephone number (*)**:**

00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Luxembourg: Publications Office of the European Union, 2014

ISBN 978-92-79-35314-7
doi: 10.2762/22502

Printed in Austria
Recommended citation: EUROPEAN COMMISSION - Directorate-General for Agriculture and Rural Development – Unit E.4 (2014): Investment Support under Rural Development Policy. Final Report. Brussels. ISBN 978-92-79-35314-7

Content

Abstract X

Extrait XI

1 Foreword 1

2 Introduction to the study framework 4

2.1 Groups of investment support measures and financial support during 2007-2013 4

2.2 Intervention logic of the four groups of investment support 5

2.2.1 Group A: Productive investment support to private beneficiaries to increase economic performance/business competitiveness 6

2.2.2 Group D: Investment support to private beneficiaries for investments required to meet minimum standards under Axis 1 measures 121 and 123 other than those which improve the economic performance of the holding 8

2.2.3 Group B: Investments in public infrastructures which are complementary to private investments intended to improve economic performance/business competitiveness 9

2.2.4 Group C: Non-productive investments to private beneficiaries for environmental or non-market purposes 10

3 Methodology and tools of the study 12

3.1 Evaluation Questions, judgement criteria and indicators 12

3.1.1 Approach to answering Evaluation Question 1 12

3.1.2 Approach to answering Evaluation Question 2 14

3.1.3 Approach to answering Evaluation Question 3 17

3.2 Overview of evaluation methods for the assessment of investment support 21

3.3 Selected methods tested in the fieldwork phase 22

3.3.1 Input-Output Analysis (IO) 23

3.3.2 Econometric counterfactual design 25

3.3.3 Method for Impact Assessment of Programmes and Projects (MAPP) 32

3.3.4 Programme-theory-based evaluation (TBE) 34

3.3.5 Strategic Environmental Assessment (SEA) 39

3.3.6 Cost-Effectiveness Analysis (CEA) 43

3.4 Choice and justification of the selected RDP territories 46

3.4.1 Criteria for the selection of case study territories 46

3.4.2 Data availability for proposed methods in case study regions 47

3.4.3 Combination of evaluation methods tested in case study regions 52

4 Spotlight on selected case study findings 54
4.1 IO method in case study Austria 54
4.2 Econometric counterfactual method in case study Austria 60
4.3 MAPP Method in case study UK/Scotland 68
4.4 Programme-theory-based method in case study Czech Republic 72
4.5 Strategic Environmental Assessment / Cost-Effectiveness Analysis in case study Greece 80

5 Answers to Evaluation Questions 85
5.1 Answers to Evaluation Question 1 85
5.1.1 Rationale 85
5.1.2 Judgement criteria 86
5.1.3 Answering EQ1 in a “best case“ situation 87
5.1.4 Overview of methods given a "best case” situation 94
5.1.5 The” best case" reference situation compared with the observed situation 97
5.2 Answer to Evaluation Question 2 100
5.2.1 Rationale 100
5.2.2 CEA and SEA results on EQ2 101
5.2.3 IO results on EQ2 102
5.2.4 MAPP results on EQ2 105
5.2.5 Results of PSM to answer EQ2 110
5.2.6 Comparing the efficiency of types of measures 120
5.2.7 TBE results on EQ2 121
5.2.8 Contrasting results and complementarity between methods 123
5.3 Answer to Evaluation Question 3 128
5.3.1 Answers to evaluation sub-question 1-3 on achieving an RDP objective 128
5.3.2 Answers to evaluation sub-question 4-6 on allocating funds to target groups exemplified on the case study M121 in AT 132

6 Conclusions and recommendations 135
6.1 Conclusions on the appropriateness of methods to evaluate investment support measures 135
6.2 Recommendations on identifying appropriate methods to evaluate investment support measures 138
6.2.1 Rationale and overview 138
6.2.2 Capacities, resources, preferences and characteristics 138
6.2.3 Making a decision on an evaluation method 141

6.3 Conclusions on the effectiveness, efficiency and impact of investment support 143
6.4 Conclusions on targeting 145
6.5 Further observations 146

7 Bibliography 149
7.1 International guidebooks 149
7.2 Specific evaluation methodologies for rural development 150
7.3 Specific sources for IO-Method 156
7.4 Specific sources for Counterfactual design 157
7.5 Specific sources for MAPP method 158
7.6 Specific sources for Theory-Based Impact Evaluation 158
7.7 Specific sources for SEA 159
7.8 Specific sources for CEA 160

8 Glossary of key terms 161

Tables and Figures

Table 1. Investment groups and measures covered by the study 4
Table 2. Financial inputs/expenditures of EAFRD for investment measures, 2007-2013 5
Table 3. Sub-Evaluation Questions, judgement criteria and evidence for EQ3 19
Table 4. Overview of evaluation approaches 22
Table 5. Pre-Selection of methods for fieldwork phase 22
Table 6. Factsheet for Input-Output Analysis 25
Table 7. Factsheet for an Econometric counterfactual design 31
Table 8. Factsheet for the MAPP Method 33
Table 9. Factsheet on Programme-theory-based evaluation 38
Table 10. Factsheet on Strategic Environmental Assessment (SEA) 42
Table 11. Factsheet on Cost Effectiveness Analysis (CEA) 45
Table 12. Overview of data availability in case study regions 47
Table 13. Combination of methods per case study and overview of targeting analysis 53
Table 14. Capacity-Adjustment Analysis – national Managing Authority data: Inputs to the Austrian IO Model (in EUR million) 56
Table 15. Capacity-Adjustment Analysis – econometric counterfactual analysis data: Inputs to the Austrian IO Model (in EUR million) 56
Table 16. Impact Analysis, Austria based on national managing authority data (average annual effects compared to 2009) (in EUR million) 57
Table 17. Impact Analysis based on data from the econometric counterfactual analysis, Austria (average annual effects compared to 2009) (in EUR million) 58
Table 18. New Jobs Generated per million EUR of Investment, Austria (FTEs) 59
Table 19. Austria: Effect of M121 on GVA (EUR) 62
Table 20. Austria: Effect of M121 on total farm income (in EUR) 63
Table 21. Austria: Effect of M121 on total farm employment (in AWU) 64
Table 22. Austria: Effect of M121 on farm labour productivity (in EUR/fully employed) 65
Table 23. Influence matrix from North-East Scotland case study 69
Table 24. Conceptual model for the bundle of measures M121, M123, M311 74
Table 25. Implementation of planned activities for the bundle of measures M121, M123, M311 75
Table 26. Direct gross effects for beneficiaries of implemented operations under the bundle of measures M121, M123, M311 77

Table 27. Summary table on the contribution of measures M121, M123, M311 to the intended change 79
Table 28. Completeness of TBE “building blocks” for M121, M123, M311 80
Table 29. M125 projects of the Greek RDP, comparison of CER 84
Table 30. How appropriate are different evaluation methods to analyse efficiency, effectiveness and impact of investment support measures - assuming a “best case” situation 96
Table 31. Ranking of resources used for the fieldwork of geographic experts (most resources: rank 1) 98
Table 32. Subjective assessment of how close case studies in relation to the "best case" 100
Table 33. Shocks of regional or country IO model (million EUR) 102
Table 34. Efficiency, effectiveness and impact results based on the IO analysis 104
Table 35. Aggregated results of MAPP 107
Table 36. Overview of PSM results on efficiency and impact 112
Table 37. Overview of PSM results on effectiveness 113
Table 38. Aggregated TBE results on the *effectiveness* of investment support measures 123
Table 39. Results on labour market related indicators of four methods applied in Czech Republic and measure 121 (M121) 126
Table 40. Number of sub-measures analysed to identify approaches to targeting. 129
Table 41. Distinct statistically significant characteristics of the clusters. 129
Table 42. Correlation between clusters and efficiency criteria from EQ2 130
Table 43. Overview of essential capacities necessary to carry out an evaluation 139
Table 44. Tool for decision makers to state his or her preferences 139
Table 45. Characteristics of generic evaluation methods 141

Figure 1. Intervention logic by type of beneficiary for measures 121, 122, 123 7
Figure 2. Intervention logic by type of beneficiary for measures 311, 312, 313 8
Figure 3. Intervention logic by type of beneficiary for measures 121, 123 8
Figure 4. Intervention logic by type of beneficiary for measure 125 9
Figure 5. Intervention logic by type of beneficiary for measures 216 and 227 11
Figure 6. Workflow of Input-Output Method 24
Figure 7. Quasi-experimental design with treatment and control groups 27
Figure 8. Workflow of PSM 30
Figure 9. Workflow of MAPP Method 33

Figure 10. Workflow of Programme-Theory-Based Impact Evaluation 36
Figure 11. Workflow of Strategic Environmental Assessment 41
Figure 12. Workflow of Cost-Effectiveness Analysis 44
Figure 13. Comparison of IO results on *efficiency* (jobs created per million expenditure) for investment support measures on efficiency across regions 105
Figure 14. Efficiency of investment support measures based on PSM estimates 116
Figure 15. Efficiency indicator on employment effects based on PSM estimates 117
Figure 16. Effectiveness indicator based on PSM estimates 118
Figure 17. Impact indicator based on PSM estimates on a log-scale 119
Figure 18. Efficiency for job creation from IO grouped by measure types 120
Figure 19. Efficiency of GVA estimated with PSM indicator 1a 121
Figure 20. Comparison of IO results on *efficiency* (jobs created per million expenditure) for investment support measures across regions with and without parameters obtained from PSM 127
Figure 21. Dendrogram of the clusters generated from the approaches to targeting. 132
Figure 22. Decision tree 142

List of abbreviations

AEM	Agro-Environmental Measures
AIR	Annual Implementation Report
APR	Annual Progress Report
ATT	Average Treatment on Treated
AWU	Annual work unit
BGS	Biogas
CAP	Common Agricultural Policy
CAPRI	Common Agricultural Policy Regionalised Impact
CBA	Cost-benefit analysis
CEA	Cost-effectiveness analysis
CEQ	Common Evaluation Question
CER	Cost-effectiveness ratio
CGE	Computable general equilibrium
CIE	Counterfactual impact evaluation
CMEF	Common Monitoring and Evaluation Framework
DG AGRI	Directorate-General for Agriculture and Rural Development
DG EMPL	Directorate-General for Employment, Social Affairs and Equal Opportunities
DG REGIO	Directorate-General for Regional and Urban Policy
DiD	Difference-in-differences method
EAA	Economic Accounts of Agriculture
EAFRD	European Agricultural Fund for Rural Development
EIA	Environmental Impact Assessment
ERDF	European Regional Development Fund
ESF	European Social Fund
EDM	Equilibrium displacement model
ENRD	Evaluation Network for Rural Development
EQ	Evaluation Question
EU	European Union
EUROSTAT	Statistical Office of the European Union
FADN	Farm Accountancy Data Network
FBI	Farm birdland index
FFI	Family Farm Income

FFV	Fresh fruit and vegetables
FTE	Full time equivalent
GDP	Gross domestic product
GE	Geographic Expert
GIS	Geographic information system
GPS	Generalised propensity score matching
GRIT	Generation of Regional Input-Output Tables
GVA	Gross value added
Ha	Hectare
HNV	High nature value
IO	Input-Output
IV	Instrument variable
M	Measure
NVZ	Nitrate Vulnerable Zone
LCA	Life-cycle analyses
LFA	Less favoured areas
MAPP	Method for Impact Assessment of Programmes and Projects
MTE	Mid-term evaluation
MS	Member State
PA	Paying Agency
PASMA	Positive Agricultural Sector Model Austria
PRA	Participatory Rural Appraisal
PSM	Propensity score matching
P-TBE	Programme-theory-based impact evaluation
RCT	Randomised controlled trial
RD	Rural Development
RDD	Regression discontinuity design
RDP	Rural Development Programme
SEA	Strategic Environmental Assessment
SME	Small and medium-sized enterprises
SWOT	Strengths, weaknesses, opportunities and threats
TBE	Theory-based evaluation

ToR	Terms of Reference
UAA	Utilised Agricultural Area
WIFO	Austrian Institute of Economic Research
WWC	What Works Clearinghouse

Abstract

This study analyses three questions in relation to the evaluation of investment support in Rural Development Programmes (RDP) of the Common Agricultural Policy (CAP). Different evaluation methods are classified according to their appropriateness and suitability to measure efficiency, effectiveness and impact of investment support measures. In order to evaluate the causality between policy interventions and outcomes a number of specific econometric methods or experiments are necessary. Theory-based assessments and qualitative participatory approaches cannot be used to derive quantitative results. In order to obtain such results, economic modeling approaches like input-output analyses or econometric methods must be used. A further element of the analysis is to estimate efficiency, effectiveness and impact of investment support measures in eleven programme areas of the EU. The quantitative results show a wide range of results that depend on structural aspects of the regions under consideration and programme-specific factors. With the data available, a causal statistical link between efficiency and targeting was not found. However, a case study demonstrated that targeting via eligibility criteria is more transparent than selection through ranking while aid intensity differentiation does not always have statistically significant effects on targeting.

Extrait

Cette étude procède à l'analyse de trois questions relatives à l'évaluation des aides aux investissements dans les programmes de développement rural (PDR) de la politique agricole commune (PAC). Différentes méthodes d'évaluation sont classées selon leur adéquation et leur pertinence afin de mesurer l'efficience, l'efficacité et l'incidence des mesures d'aide aux investissements. Il est nécessaire d'utiliser un certain nombre des méthodes et expériences économétriques spécifiques dans le but d'évaluer la causalité entre les interventions de la politique et les réalisations. Des évaluations basées sur la théorie et des approches qualitatives participatives ne peuvent être utilisées pour obtenir des résultats quantitatifs. Il convient d'utiliser des approches de modélisation économique telles que l'analyse input-output et des méthodes économétriques pour obtenir de tels résultats. Un élément supplémentaire de l'analyse consiste à évaluer l'efficience, l'efficacité et l'incidence des mesures d'aide aux investissements dans onze territoires des programmes de l'UE. Les résultats quantitatifs montrent un large éventail de résultats qui dépendent des aspects structurels des régions à l'examen et des facteurs spécifiques au programme. Avec les données à notre disposition, nous n'avons pas pu trouver de lien de causalité statistique entre l'efficience et le ciblage. Cependant, une étude de cas a démontré que le ciblage au moyen de critères d'éligibilité était plus transparent qu'une sélection via un classement puisque que la différentiation de l'intensité de l'aide n'a pas toujours d'effets statistiques significatifs sur le ciblage.

1 Foreword

Objective of the study

The primary purpose of this evaluation is the methodological advancement with respect to how the impacts of measures to support physical investments are assessed. The task consists in reviewing, analysing and testing a selected sample of methods for evaluating investment measures in RDP (related to impact, effectiveness, efficiency, achievements) across at least ten RDP territories in a minimum of seven Member States. The focus lies on assessing the strengths and weaknesses of the methodologies but also on exploring under what conditions different methods could be applied in the real context of evaluations of RDP. This should take into account the limited data availability or capacity to collect own data for the purpose of an evaluation.

Fieldwork is used to test selected methods and to collect the necessary data. For each method – or combination of methods –assessments are presented, accompanied by an appraisal of the limits and remits of the results. Conclusions related to the different methodologies and the implications for data availability, pre-determining conditions, reliability of the results and advantages and disadvantages of the methods are developed.

The study team

This report has been prepared by Metis GmbH in collaboration with WIFO (Austrian Institute of Economic Research) and AEIDL as partners. The contractors have assembled a team of experts that draw together expertise ranging from micro-economic and macro-economic analysis as well as qualitative research. The research was coordinated by Hannes Wimmer (Metis GmbH), reviewed by Herta Tödtling-Schönhofer (Metis GmbH) and supported by Angelika Kronberger (Metis GmbH), Alexandra Frangenheim (Metis GmbH). The team of methodological Core Team Experts included Jerzy Michalek for econometric counterfactual assessment, Demetrios Psaltopoulos for Input-Output analysis, Andreas Resch (Metis GmbH) for Theory-based impact evaluation, Marili Parissaki for the qualitative methodologies and Angelos Sanopoulos (Metis GmbH) for environmental approaches. Franz Sinabell (WIFO) synthetized their results in order to answer the Evaluation Questions. Ulrich Morawetz (WIFO) conducted the work on targeting approaches. A group of Geographic Experts was established to carry out the fieldwork in the Member States. These experts were Rolf Bergs (Germany), Ricardo Pedraz Gonzalez (Spain), Katalin Kolosy (France), Jerzy Michalek (Poland, Slovakia), Tomas Ratinger (Czech Republic), Franz Sinabell (Austria), Dimitris Skuras (Cyprus, Greece), Morten Kvistgaard (Denmark), Peter Cook/John Grieve (United Kingdom).

Structure and content of the report

The **Final Report** covers the contents as outlined in the Terms of Reference (ToR):

Chapter 2 describes the subject of the evaluation and outlines the intervention logic of the four groups of investment support measures.[1]

Chapter 3 describes the overall methodology of the study in terms of the approach chosen to answer the Evaluation Question; the choice of the methods to be tested in the fieldwork and the selection of the case study territories.[2]

[1] Authors of chapter 2: H. Tödtling-Schönhofer, H. Wimmer
[2] Authors of chapter 3: J. Michalek, U. Morawetz, M. Parissaki, D. Psaltopoulos, A. Resch, A. Sanopoulos, F. Sinabell, H. Wimmer

Chapter 4 includes analytical findings from tested methods and cases.[3]

Chapter 5 develops answers to all three Evaluation Questions based on the analysis of all case studies.[4]

Chapter 6 develops conclusions and recommendations of the evaluation study.[5]

Chapter 7 contains the bibliography of the study and **Chapter 8** develops a critical glossary.

[3] Authors of chapter 4: D. Psaltopolous (chapter 4.1), J. Michalek (chapter 4.2), M. Parissaki (chapter 4.3), A. Resch (chapter 4.4), A. Sanopoulos (chapter 4.5 and 4.6).
[4] Authors of chapter 5: F. Sinabell and U. Morawetz
[5] Authors of chapter 6: J. Michalek, U. Morawetz, M. Parissaki, D. Psaltopoulos, A. Resch, A. Sanopoulos, F. Sinabell, H. Wimmer, H. Tödtling-Schönhofer

Acknowledgements

This study would not have been possible without the co-operation of many people throughout the EU, and the authors would like to thank everyone who has participated in this extensive work which has led to highly stimulating outcomes. We would like to thank in particular all those who have contributed to the accomplishment of the fieldwork in the Member States by assisting our Geographic Experts to establish contacts, facilitating access to data and information and sharing their wealth of knowledge with us.

To mention only a few, we would like to thank:

In **Austria,** Dietmar Weinberger (WIFO - Austrian Institute of Economic Research), Otto Hofer (Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management), Franz Fensl (LBG Austria GmbH); in **Cyprus,** Artemis Antoniades and Constantina Georgiou (Managing Authority of Rural Development Programme); in the **Czech Republic,** Tomáš Medonos, Michaela Lekešová, Zuzana Hloušková (Institute of Agricultural Economics and Information); in **Denmark,** Henrik Skoven (The Danish AgriFish Agency, Ministry of Food, Agriculture and Fisheries of Denmark), Arne Oksen (Danish Agricultural Advisory Services); in **France,** Gilles Allaire and Cédric Gendre (Observatoire du Développement Rural – INRA Toulouse); in **Germany** (Hessen), Joachim Dippel (Hessisches Ministerium für Umwelt, Klimaschutz, Landwirtschaft und Verbraucherschutz), Ulrich Stahl (Landesbetrieb Landwirtschaft Hessen, LLH), Wilhelm Kuckshinrichs (Helmholtz-Forschungszentrum Jülich); in **Greece,** Georgia Kotsakou (Department of Rural Affairs of Western Greece), Kostas Apostolopoulos (Managing Authority of the Rural Development Programme), Vasileios Michalopoulos (Department of Rural Economy of Region of Western Greece); in **Slovakia,** Eva Kolesárová (Slovak Agricultural Paying Agency), Katarina Bradacova and Ruzena Vajcikova (Research Institute of Agricultural and Food Economics), Jela Tvrdonova; in **Spain** (Galicia), Raul Jacome Rodriguez (Xunta de Galicia. Instituto Gallego de Estadística), Martín Alemparte Vidal (Grupo Desarrollo Rural Ribeira Sacra), Juan Manuel Martínez Barral (Xunta de Galicia. Consejería de Medio Rural y del Mar), Javier Blanco Carballal (Xunta de Galicia. Agencia Gallega de Desarrollo Rural), Jose Antonio Nuñez Butragueño (Gobierno de España. Dirección General de Desarrollo Rural y Política Forestal), Paula Maria Rodriguez Andreu (Gobierno de España. Dirección General de Desarrollo Rural y Política Forestal), Julio Álvarez Núñez (Grupo Desarrollo Rural Ribeira Sacra); in **Poland,** Dariusz Osuch (Institute of Agricultural and Food Economics); in the **United Kingdom** (Scotland), Gilly Diggens (Scottish Government).

Finally, we would like to thank the members of the Steering Committee of the European Commission for their constructive comments and support in this work.

2 Introduction to the study framework

2.1 Groups of investment support measures and financial support during 2007-2013

Investment support measures provide funding for investment to increase productivity within the agricultural and forest sector and to diversify production in non-agricultural activities, where tourism is specifically targeted. Furthermore, related infrastructure investment for the agricultural and forest sector and non-remunerative investments supporting agri-environmental schemes are supported. The target groups are economic entities (farm and forestry holdings; micro, small and medium sized enterprises in sectors processing agricultural and forest products) within the agricultural and forest sector as well as individuals trying to diversify and establish businesses or services outside the agricultural sector, including tourism. The measures are part of Axis 1 and 3 (investment support to private beneficiaries and infrastructure investments) and part of Axis 2 (non-productive investments) of the Rural Development Programmes.

The ToR split the investment categories into four groups of investments with specific (but not distinct) measures subsumed under the categories:

Table 1. Investment groups and measures covered by the study

Investment group/type	Code	Description
A: Productive investment support to private beneficiaries to increase economic performance/business competitiveness under measures	121	Modernisation of agricultural holdings
	122	Improving the economic value of forests
	123	Adding value to agriculture and forestry products
	311	Diversification into non-agricultural activities
	312	Support for business creation and development
	313	Encouragement of tourism activities
B: Investments in public infrastructure	125	Improving and developing infrastructure related to the development and adaptation of agriculture and performance/business competitiveness
C: Non-productive investments to private beneficiaries for environmental or non-market purposes under measures	216	Support for non-productive investments in agriculture
	227	Support for non-productive investments in forestry
	313*)	Encouragement of tourism activities
D: Investment support to private beneficiaries for investments required to meet minimum standards under Axis 1 measures 121 and 123 other than those which improve the economic performance of the holding	121*)	Modernisation of agricultural holdings
	123*)	Adding value to agriculture and forestry products

*) overlaps between the investment types

An overview of the public expenditures of the European Agricultural Fund for Rural Development (EAFRD) for investment measures during the period 2007-2013 is presented in the following table.

Table 2. Financial inputs/expenditures of EAFRD for investment measures, 2007-2013

Allocations			Volume of allocated public funds in Mill. EUR		Public funds spent until 31. 12. 2013			
					Mill. EUR		%	
			EU[6]	Case study RDPs[7][8]	EU	Case Study RDPs	EU	Case Study RDPs
Total EAFRD funds allocated for 2007- 2013			96,208.94	35,582.69	71,026.51	26,952.00	73.8	75.7
EAFRD funds allocated for investment measures 2007 - 2013[9]			28,158.53	8,871.05	17,806.09	5,643.94	63.2	63.6
EAFRD funds per measure	Group A[10]	121	11,635.99	4,041.34	8,711.52	3,129.52	74.9	77.4
		122	369.36	101.91	224.82	71.45	60.9	70.1
		123	5,539.96	1,445.17	3,192.68	906.68	57.6	62.7
		311	1,236.75	554.74	763.56	385.78	61.7	69.5
		312	2,046.01	906.38	927.74	303.48	45.3	33.5
		313	1,226.97	329.82	534.57	120.26	43.6	36.5
	Total		**22,055.03**	**7,379.31**	**14,354.89**	**4,917.17**	**65.1**	**66.6**
	Group B	125	4,786.72	1,330.99	2,648.17	631.65	55.3	47.5
	Total		**4,786.72**	**1,330.99**	**2,648.17**	**631.65**	**55.3**	**47.5**
	Group C	216	544.20	100.90	373.73	59.96	68.7	59.4
		227	772.58	59.80	429.30	35.15	55.6	58.8
		313	1,226.97	329.82	534.57	120.26	43.6	36.5
	Total		**2,543.74**	**490.63**	**1,337.60**	**215.37**	**52.6**	**43.9**

Source: http://ec.europa.eu/agriculture/statistics/rural-development/2013/index_en.htm; own calculations

2.2 Intervention logic of the four groups of investment support

The intervention logic of investment support measures on a general level is explained in the Common Monitoring and Evaluation Framework (CMEF) for the 2007-2013 period ("measure fiches"). Investment support measures target a bundle of objectives, aiming to increase productivity and competitiveness of the agricultural and forest sector, diversify and establish businesses or services outside the agricultural sector and improve the environment and the countryside.

[6] http://enrd.ec.europa.eu/enrdstatic/app_templates/enrd_assets/pdf/monitoring_indicators/financial_and_physical_indicators/rdp/b_financial-expenditure-2014_a_eu27.pdf
[7] Case study RDPs: Austria, Cyprus, Czech Republic, DE Hessen, Denmark, ES Galicia, FR Hexagon, Greece, Poland, Slovak Republic, UK Scotland
[8] Calculation based on monitoring data provided by DG AGRI
[9] Investment measures considered in the project: 121, 122, 123, 125, 216, 227, 311, 312, 313
[10] Group A measures 121 and 123 also refer to Group D

2.2.1 Group A: Productive investment support to private beneficiaries to increase economic performance/business competitiveness

Type A represents a cluster of Axis 1 agricultural and Axis 3 non-agricultural activities.

The overall objective of group A investment support is to improve the competitiveness of the agricultural and forestry sector and to encourage diversification of economic activities. In this respect a significant part of the resources of the RDP is devoted towards the creation of employment opportunities in rural areas in non-agricultural activities and services.

Increasing the competitiveness of the agricultural sector requires an improvement of the productivity of physical capital. Modernisation of farms is crucial to improve their economic performance through better use of the production factors including the introduction of new technologies and innovation, farm diversification, etc.

Physical investments are also supported in the forestry sector. As forestry has a significant role to play in the economic activity of rural areas, support is provided for improving the economic value of forests. Possible investments can include all operations at the level of the forestry holding, including investments for harvesting equipment.

Investments in processing and marketing of existing products, as well as in the development of new products, new processes and new technologies can improve the added value to agricultural and forestry products. Such investments could be towards the construction, acquisition or improvement of permanent buildings, the purchase or leasing of new machinery and equipment and towards general costs linked to expenditure such as patent rights and licences.

Moreover under group A members of farm households who diversify into non-agricultural activities are supported. There is a number of different categories of non-agricultural activities that can be supported, for instance: service activities (such as bed and breakfast; education and social activities on farms); craft activities (such as pottery and production of local products), producing alternative energy sources and; trade activities (such as the creation of stores attached to farms, where artisan products are sold directly to the customer).

Support for business creation and development is provided to existing micro-enterprises or to persons who plan to set up new micro-enterprises in non-agricultural businesses. This can help to promote entrepreneurship and develop the economic structure in rural areas, thus contributing to the creation of employment opportunities.

Tourism is a major growth sector in many rural areas, creates new employment opportunities and is aimed to increase the overall attractiveness of the rural area.

Rural development policy actively encourages tourism activities through supporting small-scale infrastructure such as: information centres and sign posting of tourist sites; recreational infrastructure offering access to natural areas; small capacity accommodation and; the development and/or marketing of tourism services relating to rural tourism.

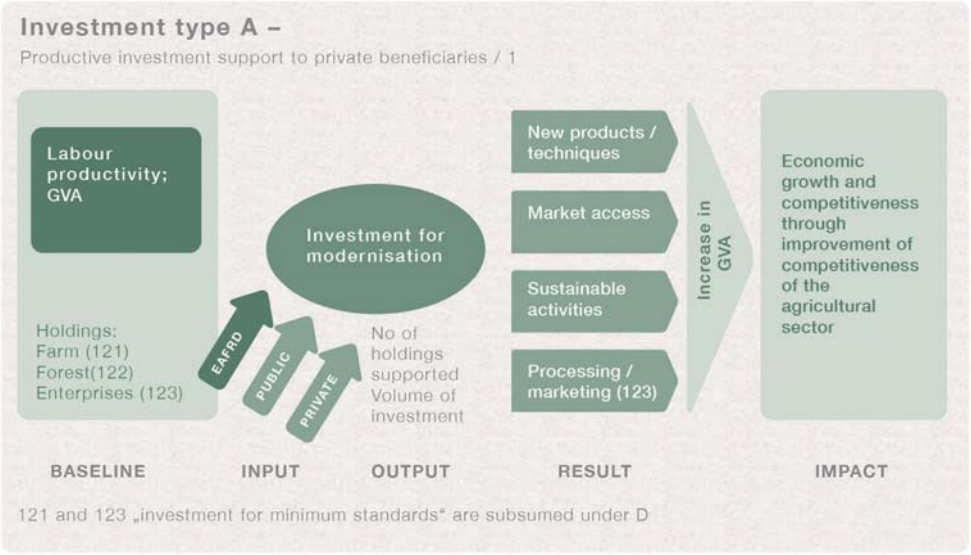
It should be noted that productive investments have also – besides socio-economic impacts – crucial impacts on the environment which in terms of evaluation poses a number of methodological challenges.

The target groups of group A interventions are in general farmers, members of farm households, forest owners, micro, small and medium sized enterprises, organisations / entities gathering primary producers in agriculture and forestry, and the processing industry with specific eligibility criteria under each measure. Under diversification micro-enterprises are defined in the Commission Recommendation 2003/361/EC (i.e. less than 10 workers and less than EUR 2 million of turnover),

The funding rates (specifying the share of private investments) to support investments in physical capital, modernisation and diversification vary according to the needs and priorities of the Member States and have to be further clarified.

The CMEF intervention logic is illustrated in the following figures.

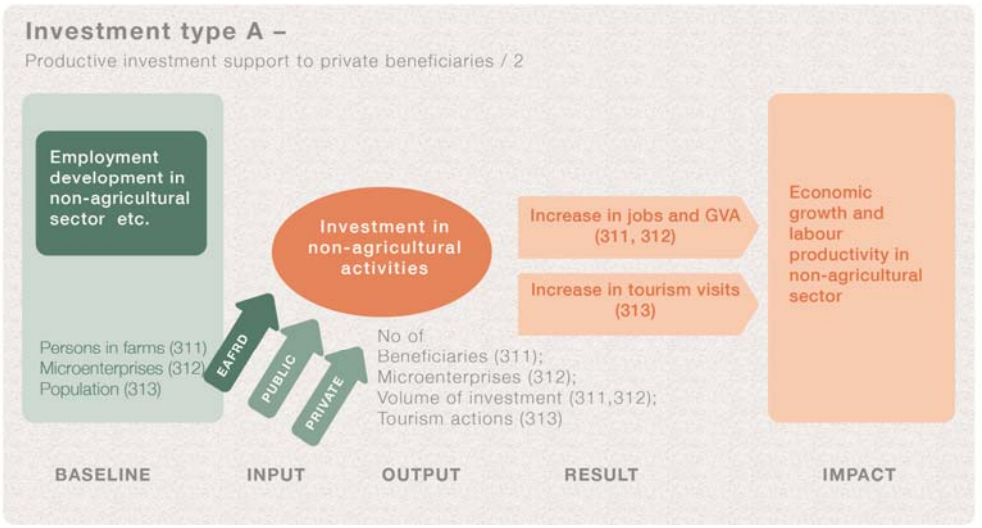
Figure 1. Intervention logic by type of beneficiary for measures 121, 122, 123



Source: Metis (2014)

Measures under Axis 3 have a similar intervention logic and evaluation design, but as the measures support investments in non-agricultural activities, the results and impacts are to be measured outside the agricultural sector only. This is definitely a challenge in terms of availability of data.

Figure 2. Intervention logic by type of beneficiary for measures 311, 312, 313

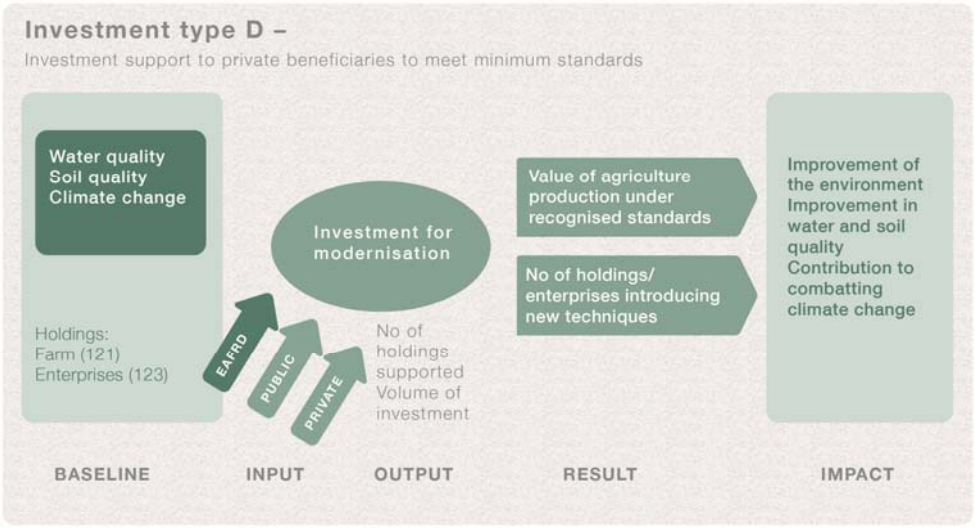


Source: Metis (2014)

2.2.2 Group D: Investment support to private beneficiaries for investments required to meet minimum standards under Axis 1 measures 121 and 123 other than those which improve the economic performance of the holding

This group addresses investments in agricultural holdings to upgrade them to Community and national standards. The intervention logic and target groups of this type are similar to those of group A but have different overall objectives, results and most of all environmental impacts.

Figure 3. Intervention logic by type of beneficiary for measures 121, 123



Source: Metis (2014)

2.2.3 Group B: Investments in public infrastructures which are complementary to private investments intended to improve economic performance/business competitiveness

Investment support type B is composed of one single measure (M125).

The overall objective of group B investment support is to improve the competitiveness of the agricultural and forestry sector.

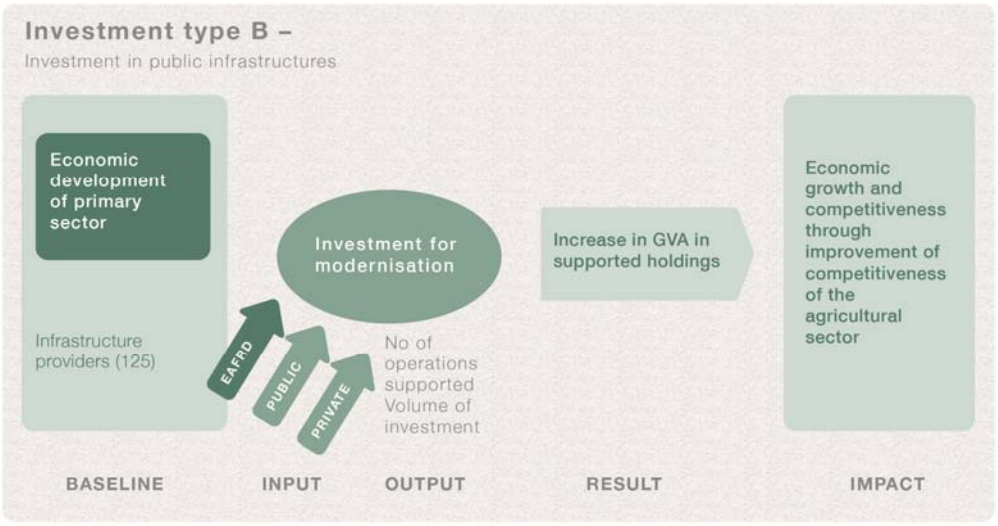
Improvement of the infrastructure related to the development of the agricultural and forestry sector contributes to the competitiveness of farming and forestry. Support under this scheme may cover investments on access to farm and forest land, land consolidation, energy supply, water management such as improvement of irrigation networks, drainage, etc.

The target groups of M125 are "infrastructure providers" such as private and public entities, associations in agriculture and forestry. Infrastructures may have public or "semi-public" character.

The funding rates (specifying the share of private investments) to support investments in infrastructures vary according to the needs and priorities of the Member States.

Through the infrastructural investment an increase in the gross value added (GVA) in supported holdings should be achieved and this, in turn, would contribute to economic growth and competitiveness. Again, the evaluation design is similar to the previous ones (see Figure 4).

Figure 4. Intervention logic by type of beneficiary for measure 125



Source: Metis (2014)

2.2.4 Group C: Non-productive investments to private beneficiaries for environmental or non-market purposes

Type C represents a cluster of Axis 2 agricultural and Axis 3 non-agricultural activities.

The overall objective of group C investment support is to improve the environment and the countryside.

Rural development efforts to preserve natural resources and farm landscapes are – besides agri-environmental measures – pursued through the provision of support to non-productive investments in agriculture (M216) and forestry (M227).

These investments are seen to be necessary for the achievement of agri-environmental and forest-environmental commitments, as well as the on-farm enhancement of the public amenity of Natura 2000 areas, other areas with high natural value and forest areas (e.g. forest holders that make non-remunerative investments or owning forests in Natura 2000 areas).

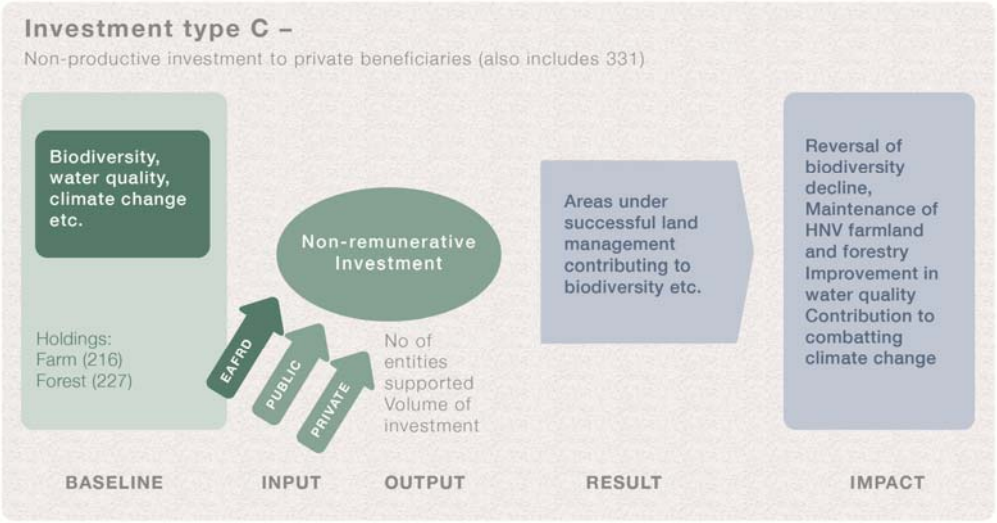
Under M313 tourism activities can include non-productive investments for instance in recreational infrastructure offering access to natural areas. However, it may be a challenge to identify those parts of M313 tourist activities which should be addressed under group C.

The target groups of these interventions are farmers, forest holders and other land managers. In addition other private or public entities may be addressed in the case of tourism activities.

Funding rates, specifying the share of private investments, supporting non-productive investments vary according to the needs and priorities of Member States.

The intervention logic for measures of Axis 2 (M216 and M227) differs from the ones above: Holdings receive non-remunerative investment. The expected result is not an increase in GVA, but an improvement of the areas that are under successful land management. This contributes to an improvement of biodiversity, maintenance of high nature value (HNV) farmland and forestry, an improvement in the quality of water and also assists in combatting climate changes (see Figure 5). Thus the impact is less clearly specified and definitely influenced by many other factors. As the amount of funds spent in this measure is – in comparison to other measures – rather low, a relation between the investment and the potential benefit must be taken into account.

Figure 5. Intervention logic by type of beneficiary for measures 216 and 227



Source: Metis (2014)

3 Methodology and tools of the study

This chapter outlines the methodology for the assessment of each Evaluation Question, including the criteria for the assessment of the different evaluation methodologies, the overview of evaluation methods and the approach for the selection of methods for fieldwork.

3.1 Evaluation Questions, judgement criteria and indicators

The objective of this study is to review, analyse and test different methods for evaluating investment measures in RDP (related to impact, effectiveness, efficiency and achievements). The three Evaluation Questions (EQ) specified in the ToR have an overarching nature and can be complemented with evaluation sub-questions accompanied by judgement criteria which define the success of intervention. EQ1 covers **methodological and practical issues** whereas EQ2 is a **cause-effect** question in the traditional sense. EQ3 links the outcome of EQ2 to the targeting approach applied by the Member States and, as such, is a highly interesting **policy-oriented question** with many practical implications.

3.1.1 Approach to answering Evaluation Question 1

EQ1: To what extent are the different evaluation methods described and/or tested in this exercise appropriate for the assessment of the effectiveness, efficiency and impact of the different types of investment support considered?

Rationale

EQ1 is dealing with *methods* and not with specific outcomes of programmes. By asking for the *appropriateness* of a method it is addressing a meta-level. The task is to analyse findings of various cases and to give guidance on which methods bring good results in a given real world situation rather than coming to an abstract judgement on what may be the best method in a perfect (but unreal) setting. Appropriateness therefore takes also into account the limitations and requirements of methods with respect to data, technical complexity, time necessary to apply, etc.

EQ1 is to be answered in two steps:

- methodological experts involved in this study are asked to conceive a best case situation for a given method and to make a judgement according to criteria which are used to characterise the method;
- after the fieldwork and the analysis, a second look at the methods under consideration provides evidence on the practicability and the ease to apply a given method in real world situations.

This two-step approach makes it possible to assess the appropriateness of a given method for a range of situations that occur in the practice of RDP evaluations.

The topic is investment support. Therefore specific challenges that pertain to this and similar kinds of policy intervention are briefly explored before the way to answer EQ1 is presented in detail. The study of challenges of programme evaluation is followed by an elaboration of the judgement criteria which are considered to be the most important ones[11] to help answer EQ1.

[11] http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html, See the example of ranking evaluation methods, page 92 - 93

Investment support: causes and effects

The standard challenge of the assessment of impacts is explaining causality and determining what would have happened in the absence of investment support. To truly understand the impact of support on a given impact indicator, information would ideally be available on beneficiaries obtaining the support and those same beneficiaries without the support. The indicator could then be compared between these two states to see if the support had an impact. Of course, beneficiary farmers cannot be simultaneously supported and non-supported making it necessary to find a substitute group of farmers to act as the counterfactual; that is, what would happen in the absence of the project. To be a legitimate counterfactual, this counterfactual, or control group would need to be similar to the supported beneficiaries, or treatment group, except they would not have received the support. Thus, any differences in the indicator could be attributed to the investment support. Creating a counterfactual through identifying a reasonable control group and ensuring that an identified impact can be attributed to a given intervention is however a challenge.

Ability to eliminate a possible selection bias

One common issue with evaluating investment support is that it often involves self-selection of beneficiaries (individuals or regions). Self-selection implies that only certain types of farmers, firms or regions may choose to participate in a given programme or are eligible for the programme. If an evaluation attempts to determine the impact of an investment support by comparing those that chose to be in the programme to those that did not, difference in the indicator of interest may not only reflect the impact of the support, but also any innate differences between programme participants and non-participants. For example, suppose the more productive farmers in a region decide to participate. Such farmers are likely to have higher yields or GVA per hectare (Ha) or GVA per labour even without the investment support. A comparison of yields or GVA between these innovative, supported farmers and other non-participants is likely to show higher yields and GVA for the supported farmers not only due to the support but also due to the fact that the supported farmers are innovative. Reliable evaluation methodologies should take all similar confounding factors into consideration.

Ability to isolate the effect of the programme from other factors

Observed change in programme results and impact indicators can be an effect of many factors. Among them are business cycle effects for the whole economy or a sector or price changes that affect the level of output or the costs of inputs. An appropriate methodology used for the assessment of an impact of investment support should be able to identify the different consequences of each factor and to isolate the programme effect from other factors.

Evaluation sub-question	Judgement criteria	Evidence to answer the evaluation sub-question[12]
To which extent the different methods meet the evaluation standards?[13])	• Scale • Rigour • Reliability and robustness • Transparency • Validity • Practicability	The evidence is based on scientific literature and previous studies in which a given method was applied. The method is under consideration; therefore an ideal situation is envisaged.
To which extent the different evaluation methods tested in this exercise allow to explain the causality and determine what would happen in the absence of the investment support?	Method is able to explain the causality and determine what would happen in the absence of the investment support	Evaluation of the type of method and its general approach. The counterfactual can be empirically measured, based on assumptions of stakeholders or the evaluator.
To which extent are different methods able to eliminate a possible selection bias?	Method is able to eliminate selection bias	Exploration of the available data and ways to identify good controls.
To which extent are different methods able to isolate the programme effect from other intervening factors?	Method is able to isolate programme effects from other intervening factors	Are sufficient data available to identify confounding factors and allow the method to make use of this information?

Judgement criteria

The framework for testing the different evaluation methods (micro and macro) regarding their appropriateness for the assessment of the effectiveness, efficiency and impact of different types of investment support (A-D) includes the use of judgement criteria which are further explained in chapter 5.

3.1.2 Approach to answering Evaluation Question 2

EQ2: What is the effectiveness, efficiency and impact of the investment support studied in the selected RDP territories?

Rationale

According to the ToR a range of different methodologies should be reviewed for assessing the impact, effectiveness and efficiency of investment support related to four groups of investment (A, B, C, and D). It is not the objective to undertake a systematic and comprehensive assessment of the overall impact, effectiveness and efficiency of all physical investments but rather to explore how different methods can be applied in various regions. Another objective is to measure the relevant indicators in order to evaluate the performance of investment support measures.

[12] Evidence should be collected for each of the tested method separately via case studies, the same time allowing for comparisons among methods.
[13] These evaluation standards should be in line with the text as mentioned at page 20 at the beginning of 3.2.1, this includes judgement criteria ”verification of reality, validity and applicability” (practicability).

The case studies were not chosen specifically to make it possible to compare the results of different methods in the same situation. Such an approach would be taken if the principal interest had been a methodological assessment.

Rather the regions were chosen to make it possible to explore the methods in a wide range of different contexts. All relevant measures are evaluated with the given methods in each region. As several indicators are used to identify effectiveness, efficiency and impacts the results are numerous.

In some instances, the same measures and regions are evaluated with different methods. Not all results are consistent when different methods are compared. The findings of EQ1 can be used to explain differences and examples show how complementarities between methods can be used to broaden the range of results.

Approach

The core expert team developed a targeted approach for the assessment of the effects of investment support by defining the level of assessment (macro/micro-level), the scope of effects (economic, environmental, others), the related investment support types and proposed different evaluation methods (qualitative, quantitative) which could be potentially used to compare findings at case study level. The proposed approach is shown in the following overview.

Level of assessment	Scope of effect	Investment support types	Methods
Macro-level (effects measured at the programme area level)	• Economic impact (net effects and gross estimates) • Environmental impacts (estimates for specific environmental effects based on the changes of relevant context indicators)	A, B (jointly for all relevant measures) B, C (jointly for all relevant measures)	• Input-Output (national or regional design) • Econometric counterfactual analysis (regional design) • MAPP (regional design) • Theory-based evaluation (programme design) • Strategic Environmental Assessment
Micro-level (effects measured at the level of direct beneficiaries)	• Economic results (net effects and gross estimates) • Environmental results (net effects – if possible – and gross estimates) • Other effects related to Community standards (gross effects and – if possible – net estimates)	A, B, C, D	• Micro-econometric counterfactual analysis (micro-design) • MAPP (micro-design) • Theory-based evaluation (micro-design) • Strategic Environmental Assessment • Cost Effectiveness Analysis

Economic effects at macro and micro-level

Assessment of economic impacts of investment support at the **macro-level**, as defined by the CMEF, refer to RDP benefits in the programme area beyond the immediate effects on its direct beneficiaries. They are linked to the wider objectives of the programme. They shall be expressed in “net” terms, i.e. by subtracting effects that

cannot be attributed to the intervention (e.g. double counting, deadweight), and taking into account indirect effects (displacement and multipliers)[14].

Suitable impact indicators are economic growth, employment creation and labour productivity. In order to calculate the indicators socio-economic data have to be collected. The types of investments concerned are aggregates of A (productive) and B (infrastructure).

For the evaluation of economic impacts at the macro-level (national or regional level) both, quantitative and qualitative approaches can be used. The methods employed are: Input-Output analysis (IO) national or regional design, method for impact assessment of programmes (MAPP) and an econometric approach for regional data analysis. The identification of a counterfactual situation is seen as a very important part of the application of the chosen evaluation methods, which allows expressing the effects of intervention in net terms.

Moreover, based on the assessment of the *impact*, the *effectiveness*[15] to achieve quantified targets at programme level (comparing net impacts results with their target values) and the *efficiency* (comparing net impact with the total public expenditures of RDP) can be calculated.

The assessment of **economic results at the micro-level** refers to effects at the level of direct beneficiaries (e.g. farmers' economic performance). Results are defined in the CMEF as direct effects of the intervention for the beneficiaries linked to competitiveness and the diversification objective.

Suitable indicators are: Increase in GVA in supported holdings; number of holdings introducing new products and/or new techniques; number of farms entering the market; increase in non-agricultural GVA in supported businesses; number of jobs created; and additional number of tourist visits.

The types of investments concerned are A (productive investments) and B (infrastructure), at the detailed measure or sub-measure level with a focus on the main target group (not all target groups have to be covered).

Appropriate methods to evaluate economic results at the micro-level comprise a micro-econometric approach (propensity score matching), MAPP (micro-design) and TBE (micro-design). These methods allow expressing effects in net terms and as gross estimates. An IO analysis cannot be used to evaluate the effects of investment support measures at the level of individual farms or enterprises.

Based on the identified economic results the effectiveness to achieve quantified targets (comparing net results with specific target values set out in the RDP for investment support) and the cost-efficiency[16] (comparing net results with the actual financial costs) will be calculated.

[14] http://enrd.ec.europa.eu/enrd-static/evaluation/library/evaluation-helpdesk-publications/en/evaluation-helpdesk-publications_en.html

[15] According to the CMEF definition, effectiveness assesses the extent to which objectives pursued by an intervention are achieved. An effectiveness indicator is calculated by relating an output, result or impact indicator to a quantified objective.

[16] Efficiency is according to the CMEF definition the best relationship between resources employed and results achieved in pursuing a given objective through an intervention.

Environmental effects

Types of investments concerned for which direct environmental effects have to be demonstrated are B (infrastructure) and C (non-productive).

The study focuses on a very narrow segment of RDP measures with very specific environmental effects such as climate change, water quality (M125, M216, M227), which relate to measures considered by this study. Agri-environmental measures which are sought to contribute prominently to biodiversity are not analysed in this study because its focus is investment support.

Accordingly the isolation of the measure effect at the level of the programme area on the two CMEF impact indicators related to biodiversity and HNV (farm birdland index (FBI) and high nature value farmland (HNVF) is barely possible. Due to the complexity of RDP impacts on the environment a rigorous quantitative assessment of environmental effects for the programme area (macro-level) requires economic-environmental modelling which is cost intensive and out of the scope of the present study (see for instance the modelling of the nitrogen balance at the NUTS 2 level by means of the CAPRI model, www.capri-model.org/).

However the study is not "blind" on environmental impacts at the programme level. By adopting the Strategic Environmental Assessment (SEA) it is possible to establish and discuss the causal link between measure outputs/results and impacts on a series of environmental issues.

The focus of the environmental analysis is on those environmental effects which are directly connected to investment support measures such as the effects on the improvement of water management (which is of major interest of the European Commission (EC) related to this study) and the contribution to climate change mitigation and adaption (e.g. reduction of emissions, production of renewable energy). Environmental effects of investment measures are primarily assessed at the micro-level focusing on result indicators at beneficiary level while the assessment of the environmental impacts at the programme area level by means of common impact indicators is not possible with the tested approach.

Regarding the selection of methods to analyse environmental effects there is a limited number of methods which are feasible within the scope of this study, e.g. Life Cycle Assessment (very data intensive), SEA (report and monitoring reports), Cost Effectiveness Analysis (CEA) (efficiency tool).

3.1.3 Approach to answering Evaluation Question 3

EQ3: To what extent have the different approaches to targeting investment support studied been effective in meeting the general objectives of rural development policy and/or specific objectives included in the relevant RDPs?

Rationale

When planning and implementing farm investment measures in the RDP territories, Member States have various options to target structural and territorial needs and structural disadvantages:

- Eligibility criteria: sector, territory, beneficiaries, investment type, eligible costs,
- Aid intensity differentiation,
- Selection method.

RDPs may, for example, decide to target young farmers, collective actions, certain regions or small and medium-sized enterprises (SME).

Targeting may be achieved by making only certain applications eligible (eligibility criteria with regard to sector, territory, beneficiary characteristics, investment type, eligible costs), giving additional support to certain applications (aid intensity differentiation) or by giving extra scores to certain applications (selection by ranking).

Practically, targeting is a two-step procedure: In the first step, eligibility criteria are used to determine which applications are considered for support, while aid intensity differentiation sets incentives for applications of relevance to the target groups. In the second step, a selection method is applied to decide who is actually supported. With various dimensions of targeting (eligibility criteria, aid intensity differentiation, selection method), each with several sub-methods, a multitude of approaches to targeting are possible, such as eligibility and aid intensity differentiation with respect to sector, territory, beneficiaries and potentially changing criteria during the programming period, e.g. as part of the CAP Health Check.

EQ3 has at least two possible interpretations. The first is the effectiveness of targeting approaches to achieve RDP objectives and the second is the effectiveness to allocate funds towards the targeted groups. The second interpretation is important as it influences the first.

These two interpretations are reflected in the evaluation sub-questions (compare Table 3 for details). Sub-questions 1 to 3 deal with the effectiveness of targeting approaches to achieve RDP objectives. The sub-questions 4 to 6 deal with the effectiveness of allocating funds towards the groups targeted.

The literature on targeting mainly focuses on sub-questions 4 to 6. Bibi and Duclos (Bibi & Duclos, 2007) describe that effectiveness of targeting is mainly described by two types of indicators: leakage rate and non-take-up. Leakage rate is typically defined as the proportion of total transfers going to the pre-transfer non-target group. Definitions of non-take-up vary, but can be defined as the ratio of the number of beneficiaries in the target group not funded to the total number of potential beneficiaries in the target group (Blundell et al., 1988).

Table 3. Sub-Evaluation Questions, judgement criteria and evidence for EQ3

Evaluation sub-question	1 Judgement criteria	2 Evidence to answer the evaluation sub-question[17]
Effectiveness of targeting approaches to achieve RDP objective		
1) To what extent has targeting investment support via ***eligibility criteria*** been effective in meeting general objectives of EU rural development policy and RDP objectives?	EU and RDP rural development objectives have been met most effectively with ***sectors, types of beneficiaries/territories/ types of investment and eligible costs*** selected for the investment support.	Identified clusters of approaches to targeting consist of measures which are significantly more effective than those in other identified clusters and this difference can be attributed to targeting.
2) To what extent has targeting investment support via ***aid intensity*** been effective in meeting general objectives of EU rural development policy and RDP objectives?	EU and RDP rural development objectives have been met most effectively with ***aid intensity*** set up for the investment support with respect to sector, types of beneficiaries, territory, investment type and eligible costs.	Identified clusters of approaches to targeting consist of measures which are significantly more effective than those in other identified clusters and this difference can be attributed to targeting.
3) To what extent has targeting investment support via ***selection*** been effective in meeting general objectives of EU rural development policy and RDP objectives?	EU and RDP rural development objectives have been met most effectively achieved with ***selection through ranking***.	Identified clusters of approaches to targeting consist of measures which are significantly more effective than those in other identified clusters and this difference can be attributed to targeting.
Effectiveness to allocate funds towards the groups targeted		
4) Were **eligibility criteria** effective in allocating funds towards those specified in eligibility criteria?	• Eligibility leakage rate equal 0 with respect to eligibility criteria. • Eligibility non-take-up equal 0 unless funds have been fully used and leakage rate equal zero.	• Eligibility leakage rate: Percentage of subsidies to beneficiaries who do not meet eligibility criteria relative to total subsidies. • Eligibility non-take-up rate: Percentage of eligible beneficiaries not being funded relative to all eligible potential beneficiaries.
5) Was the differentiation **in aid intensity** effective in allocating funds to those with preferential aid intensities?	Average subsidy of preferential aid intensity group higher.	Comparison of participation likelihood after matching.
6) Was **selection by ranking** effective in allocating funds towards those specified in allocation criteria?	Statistical tests on equality of mean and quantiles of the selection criteria of the not-supported and supported farms.	Distribution of selection criteria of not-supported and supported farms.

[17] Evidence should be collected via comparing different types of targeting within similar group of cases studies (among case studies).

Approach

To answer the evaluation sub-questions 1 to 3 it is necessary to identify the "approaches to targeting" of all analysed measures. This can be done by eliciting from RDP documents which eligibility criteria, which aid intensity differentiation and what kind of selection method was used. Care has to be taken as these criteria may have changed during the programming period (e.g. as part of the "Health Check") and that they may differ between sub-measures.

Once the "approaches to targeting" of each measure has been identified they have to be clustered to groups, due to the high dimensionality of the data. One method to do this is to first apply a principal component method (e.g. Multiple Correspondence Analysis for non-continuous data) and then, based on it, a clustering (e.g., a hierarchical agglomerative clustering)[18]. The attractiveness of this method lies in the possibility to a) describe which characteristics of the targeting approach make it distinct from other clusters, and b) to add supplementary variables which are not used for clustering but might nevertheless be typical for those measures in a cluster[19]. A disadvantage of this method is that the identified clusters are not necessarily meaningful and that there might be no association between the clusters and the results of EQ2. Finally, it must be stressed that what is measured are correlations and no causal relationships.

To answer evaluation sub-questions 4 and 6, leakage rate and non-take-up are defined for eligibility. Leakage rate for eligibility criteria is typically the sum of subsidies to those who did not fulfil eligibility criteria relative to the sum of all subsidies. It should generally be expected that the leakage rate for eligibility criteria is zero. Estimating it serves two purposes: checking and establishing a benchmark for comparison with selection criteria. The non-take-up rate for eligibility criteria is the number of eligible beneficiaries who are not funded to the total number of those who fulfil the eligibility criteria. This is not necessarily well defined[20].

For sub-question 5 (aid intensity differentiation) the concept of leakage rate and non-take-up rate the approach is to analyse if the differentiation had an impact on funding, net of the differences in the structure of the groups. This can be done by comparing average subsidies granted to the groups (e.g., those in a less favoured areas (LFA) and those outside an LFA) after securing that only similar farms are compared. One way to identify comparable groups is through matching algorithms.

For sub-question 6 the distribution of the selection criteria among not supported and supported farms is compared by statistically testing for significant differences in the mean, the median and the 0.25 and 0.75 quantile. A possible difference in the distribution is not necessarily due to the selection by the administrative authority but might also be due to self-selection or selection by agricultural extension services. If there is no difference, though, it can be concluded that selection by the administrative authority did not work.

[18] See for example (Husson, Le & Pages, 2011, p. 188) for details.
[19] For example, a cluster A might be distinct from other clusters as only in cluster A the measures have investment specific eligibility criteria. The measures of cluster A might also be different from others because they are assessed as effective according to results of EQ2.
[20] E.g. if the eligibility criteria is "investments below EUR 100,000" the number of those who would fulfil the eligibility criteria is not known.

Judgement criteria

The answer to EQ3 will be based on the results of EQ2 (effectiveness, efficiency and impact of the investment support studied) to answer evaluation sub-questions 1 to 3, and on farm-level data analysis to answer evaluation sub-questions 4 to 6.

3.2 Overview of evaluation methods for the assessment of investment support

The objectives of Rural Development Programmes cover a broad range of economic, environmental and sociological aspects. The attainments of the programmes therefore have to be considered in the context of the heterogeneity of the targeted aspects which may go beyond easy to measure indicators. In some cases, the application of mixed approaches (quantitative, qualitative and mixed methods) may be necessary in order to address all aspects. In other cases, when a programme focuses on few objectives, simpler evaluation approaches may be more adequate. The following challenges need to be addressed in the evaluation of investment support measures:

- The scope of socio-economic and environmental objectives is very broad because of the heterogeneity of rural areas and their specific strengths and weaknesses;
- The outcomes and effects have to be measured at various levels starting from single farms/firms up to the programme area and so cover relatively small areas but also the whole country; and,
- Programmes operate in the real world, not in a laboratory and therefore evaluation has to face limited data availability which calls for the use of less data demanding methods which are not ideal from an analytical point of view.

The screening of relevant studies and literature shows the application of wide range of methods, evaluation designs and approaches as shown in Table 4.

In evaluation studies, frequently **variants of methods** are used, that share some characteristics of the evaluation approaches presented above:

- Some approaches are listed under "computational models" with the highest rank of validity. Such models use routines to estimate parameters in an econometric fashion in order to improve the validity of results (e.g. CAPRI, AGMEMOD);
- Econometric approaches that are used to analyse causal effects, e.g. propensity score matching (PSM), or approaches that are used to analyse causal links between time series can measure past but cannot be used to predict future causalities. When they are used in an *ex ante* context, they are assumed to be "computational" models.

A direct comparison of these groups of methods is complex. When viewed through the lens of policy evaluation, they differ in respect to their purpose and the scope of evidence they can deliver. They may be complementary and each method may unveil aspects that cannot be explored by other methods.

Table 4. Overview of evaluation approaches

Method	Input	Output	Examples of methods
Qualitative methods	Mainly text (spoken or written) and/or theory	Substance of text analysed, effects, impacts (ordinal)	Intervention logic, interviews, MAPP, Delphi method
Theory-based evaluation	Programme theory or any other social/ economic theory	Estimate on effectiveness of the intervention logic	Realist Evaluation[21] Theory-based evaluation[22]
Econometric methods	Economic theory and data at unit level	Estimates of (net) effects (cardinal), hypothesis tests	PSM, regression analysis, DiD
Experimental methods	Designed experiment observations	Estimates of (net) effects (cardinal) hypothesis tests	RCT: Phase in design, pilot project design, encouragement design
Computational economic models	Economic theory and parameters	Estimates of impacts (cardinal)	Regional and national IO, general and partial equilibrium models,farm models, CBA, CEA
Environmental approaches	Scientific theory, figures on unit level, coefficient or parameter	Effects, impacts, text on environment	LCA, integrated modelling approaches, SEA
Combinations of approaches	All of the above	All of the above	GRIT, theory of driving forces, pressures, states, impacts, responses

3.3 Selected methods tested in the fieldwork phase

Based on the review of literature and expert know-how the methods have been selected according to their suitability to address the EQ, the reflection in literature, valid experiences in applying them and their validity.

Table 5. Pre-Selection of methods for fieldwork phase

Method	Econometric	Economic modelling	Qualitative methods	TBE	Env. assessment
Example	PSM	IO	MAPP	P-TBE	SEA, CEA
Method design suitable for evaluation?	yes	yes	yes	yes	Yes
Method part of literature on programme evaluation?	yes[23]	yes[24]	not verified yet	yes	yes[25]
Method applied in RDP evaluation?	yes[26]	yes[27]	yes[28]	yes	yes[29]

[21] Pawson, Tilley, 1997
[22] Weiss, 1997
[23] e.g. Imbens and Wooldridge, 2009
[24] Psaltopoulos et al., 2006
[25] Pearce, 2005
[26] e.g. Michalek, 2012; Pufahl and Weiss, 2009;Chabé-Ferret and Subervie, 2013
[27] e.g. Psaltopoulos, Balamou und Thomson, 2006
[28] Kantor, 2012
[29] Hanley, Whitby, and Simpson, 1999; Pearce, 2005

Method	Econometric	Economic modelling	Qualitative methods	TBE	Env. assessment
Studies on suitability of method for programme evaluation available?	yes[30]	yes[31]	not verified yet	yes	yes
Method has acceptable level of validity in "real world environment"?	yes	yes	yes	yes	yes

3.3.1 Input-Output Analysis (IO)

Input-Output analysis is “an adaption of the neoclassical theory of general equilibrium to the empirical study of the quantitative interdependence between interrelated economic activities”. It is a quantitative technique for studying the interdependence of the producing and consuming units within an economy. An IO table identifies the major industries in an economy and the financial flows between them over a stated time period (usually a year). It indicates the sources of each sector's inputs, which are purchased from the same or other sectors in the economy, imported, or earned by labour (household's wages and salaries). It also provides a breakdown for each sector's output, which can be sales to other industries and to final demand (household consumption, government consumption, capital formation, and exports). The interdependence between the individual sectors of the given economy is normally described by a set of linear equations, representing fixed shares of input in the production of each output.

IO modelling incorporates sectoral analysis into a macro-economic framework, thus creating a basis for an evaluation of sectoral or/and investment policies to national or regional goals such as GDP, employment and the balance of trade. Hence it provides more general information compared to a partial equilibrium model, which concentrates on one sector and more disaggregated information compared to a “pure” macro-economic model.

An IO model can be used to estimate the indirect effects of a change in the level of final demand for the output of a particular sector (impact analysis). These effects may be measured as output, income and employment changes, calculated using sectoral multiplier coefficients, which express the ratio of total effect to the initial change in demand. Impact information is available in a disaggregated as well as total form, and policy makers can thus be provided with information on which industries or sectors are impacted by a specific event and by how much.For any sector, a high level of purchase of domestically (locally) produced inputs leads to strong linkages, and creates significant indirect effects in the output of supplying sectors.

These effects are measured by Type 1 multipliers, for each sector, showing the ratio between direct and indirect effects against the direct effects. Household spending from income generated in the directly and indirectly generated employment creates further economic activity (induced effect), which is included in Type 2 multipliers: here the denominator reads as “direct, indirect and induced effects” and the nominator as “direct effects”.

[30] e.g. Henning and Michalek 2008; Margarian 2008
[31] e.g. Grady und Muller, 1988

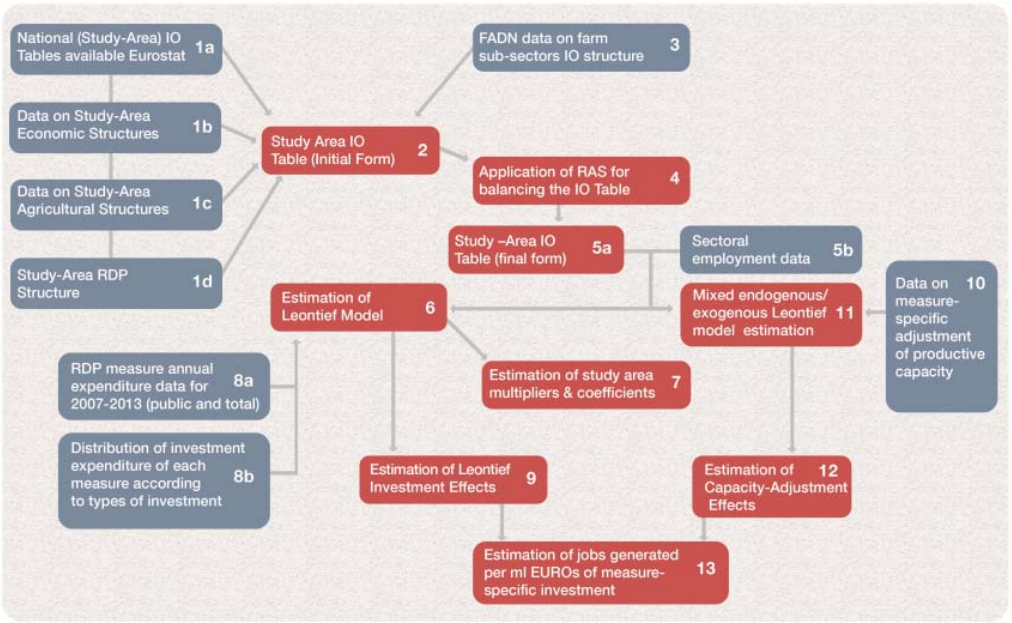
The transformation of an IO table into an economic model facilitates the analysis of the economy-wide impacts of exogenous demand shocks, including policy interventions. Within this context, two types of effect can be estimated:

Investment Effects: Given the structural linkages identified in each case study area (using the IO models and the derived multiplier coefficient values), financial flows associated with specific RDP measures can be inserted to the IO model in the form of sector-specific exogenous demand shocks. Subsequently, following the traditional Leontief procedure, economy-wide growth generating impacts are estimated for each RDP measure, in terms of average annual output, income and employment effects.

Capacity-adjustment effects: Estimating capacity-adjustment effects of Type A and B investments follows the 'mixed exogenous/endogenous variable version of the Leontief model' (indicatively, see Psaltopoulos et al., 2004).

RDP measures' expenditure may relax a constraint on the level of certain economic activities by increasing the capacity of, for example, agri-tourism accommodation facilities. Such expenditure has economy-wide effects not only through the immediate effects (direct, indirect and induced) of the investment activity thus stimulated, but also because other activities that use the constrained capacity can expand to meet demand which was not hitherto satisfied. Of course, expenditure that does not have this effect - either because it does not raise capacity, or does so but the extra capacity is not used - can be ignored in the present context. Again, economic impacts are estimated for each measure, in terms of average annual output, income and employment effects. Thus, it is possible to estimate the economic impact (in the form of changes in the levels of gross output and/or final demand) of an exogenously set change in output, which can amongst others originate from investment associated with RDP intervention.

Figure 6. Workflow of Input-Output Method



Source: Psaltopoulos (2014)

Table 6. Factsheet for Input-Output Analysis

Type of method	Econometric (general equilibrium analysis)
Brief description	IO analysis is a general equilibrium quantitative technique for studying the interdependence of the producing and consuming units within an economy. IO modelling incorporates sectoral analysis into a macro-economic framework, and creates a basis for an evaluation of sectoral or/and investment policies (such as GDP and employment).
Data requirements	• National/regional IO tables for a year close to 2007 • Data on study area economic structures (output, employment), agricultural structures, Study area RDP structure • FADN data on farm sub-sectors IO structure • Sectoral employment data (baseline) • RDP measure annual expenditure data • Distribution of investment expenditure per measure according to types of investment (e.g. machinery, equipment, construction, etc). • Data on measure-specific adjustment of productive capacity (e.g. change in GVA or employment).
Counterfactual (How can the method deal with counterfactuals?)	IO analysis can deal with counterfactuals if provided counterfactual data on measure-specific adjustment of productive capacity. In this study, econometric counterfactual analysis provided estimates on measure-specific changes in GVA. Utilising the mixed exogenous/endogenous variable version of the Leontief model, these estimates were used as inputs to case-study IO models in order to estimate economic impacts of RDP measures.
Scale of indicators	Ordinal and cardinal
How to measure efficiency, effectiveness, impact?	• Efficiency: Jobs created in the economy per annum per million Euro of RDP investment support. • Effectiveness: IO estimates of increase in economy-wide GVA per annum compared to result indicator targets associated with each measure. • Impact: IO estimates on economy-wide employment creation per annum compared to impact indicator targets

3.3.2 Econometric counterfactual design

Brief description

The main challenge of any impact evaluation is to provide evidence of a true cause-and-effect link between the observed indicators and the programme. Solving this problem always has to do with the "attribution" of the change observed to the intervention that has been implemented. Is the observed change in indicators due to the policy or would it have occurred anyway? Clearly, answering these questions is not straightforward. Given that for any unit (e.g. farm, holding, community, region, programme area, etc.) an effect of an intervention is defined as a difference of outcomes for the same unit with and without the programme, the main difficulty in quantifying an effect of a given programme/measure is finding a credible approximation to what would have occurred in the absence of the intervention, and to compare it with what actually happened. The purpose of a counterfactual in evaluations is to address the question "What would have been the situation of the programme/measure beneficiary if the programme/measure had not taken place?"

Main types of counterfactual evaluation designs

The "golden standard" in evaluation is randomised controlled experiments or the so-called "**experimental designs**", where randomly selected groups receive support (or "treatment", as technical term) and a randomly selected control group does not. Random assignment ensures that treatment and controls are comparable in their characteristics before the policy intervention so that then any follow up observable differences can be attributed to the intervention[32]. However, conducting field experiments poses several methodological challenges like external validity, spillover effects, dynamic selection, etc. Moreover, they can also be practically or ethically impossible or socially unacceptable. For example, when it is impossible to randomly assign persons or economic entities to a subsidy or exclude them, other methodological possibilities to design the counterfactual have to be applied, e.g. **quasi-experimental approaches**. The latter are very similar to experimental designs but lack the key ingredient, i.e. random assignment. In quasi-experimental design the standard to base models on individual data is the most effective one and generates results that are much more accurate than those applied at aggregated samples.

An evaluation based on experimental designs needs to be designed before the measure is implemented. Quasi-experimental evaluations, in contrast, may be done even if they have not been planned. Still, quasi-experimental evaluation can be substantially improved if random elements (e.g. discontinuities in eligibility criteria) are included in the design of the programme.

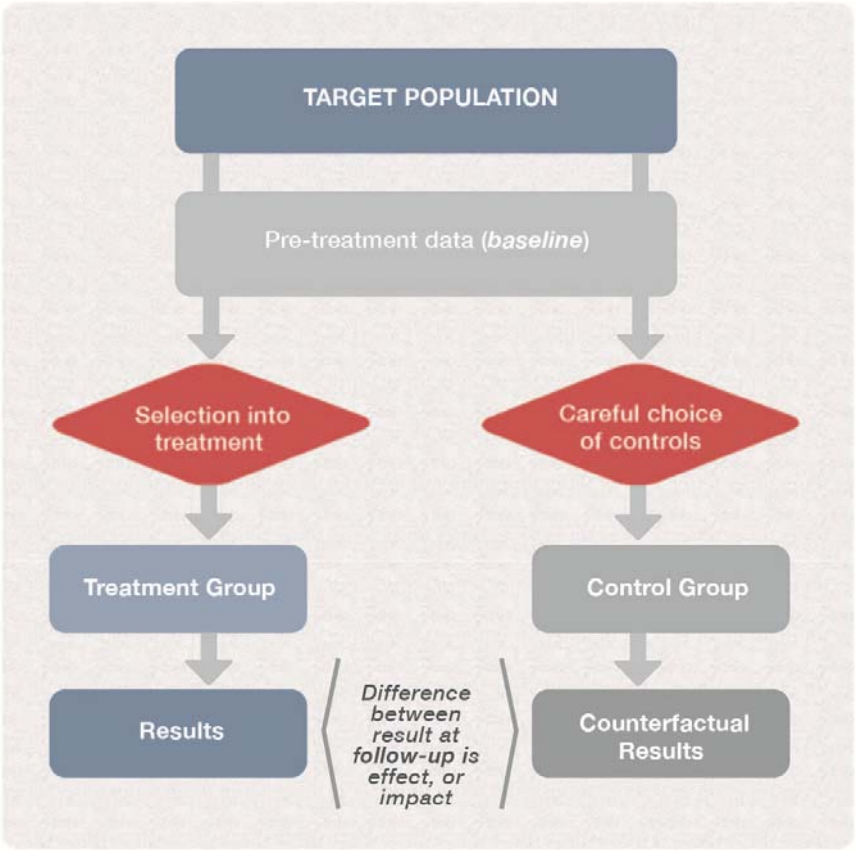
As in rural development random assignments to RDP are not possible, the "**quasi-experimental" design** has to be applied. Under this design a meaningful control group to match those units that received funding has to be constructed in order to compare the outcomes observed for programme beneficiaries with a similar non-supported group. In practice, the evaluator uses an appropriate comparison group to estimate counterfactual in order to find what would have happened to the programme participants without the programme. A crucial issue in an impact evaluation based on quasi-experimental design is to identify a group of programme participants and a group of programme non-participants (the comparison group) that are statistically identical in the absence of the programme. If the two groups are identical (they have the same characteristics), except only that one group participates in the programme and the other does not, then any difference in outcomes must be due to the programme. In principle the similarity between these two groups must be assessed using the following criteria: the supported group and the comparison group must be identical in the absence of the programme; they should react to the programme in the same way; and they cannot be differentially exposed to other interventions during the evaluation period.

How to produce such a control group depends on data availability where several methods may be applied.

Some basic principles of a quasi-experimental design using treatment and control groups are illustrated in Figure 7.

[32] Since in experimental design beneficiaries are randomly selected to receive an intervention, each has an equal chance of receiving the programme support.

Figure 7. Quasi-experimental design with treatment and control groups



Source: EC (2013), own adaptation

Establishing a control group and estimation of programme results/impacts

A credible control group can be developed in a number of ways. First, a matching approach can be applied. Typically data are collected from the supported units and a sample of non-supported units, prior to the programme. A control group is then further constructed from the group of non-supported units. Here, there are several possibilities:

- **Propensity score matching** (PSM) entails estimating a statistical model for the entire sample (treatment and potential controls) that yields an estimated propensity to participate in a programme for each individual or firm - regardless of whether they actually participated in the given programme/measure or not. Treated individuals or firms are then matched to programme non-participants on the basis of the propensity score. A control group identified in such a manner can subsequently be used to derive an estimate of the counterfactual. The critical assumption underlying the matching approach is that the selection process can be characterised by the observable data only.
- The **Regression discontinuity** compares groups around a threshold. This approach may be adopted when access to an intervention is determined by a cut-off point along a continuous rating, scale or measure. For example, by comparing those who received a subsidy and just made the score above the

threshold against those who just missed the threshold. The approach makes use of the fact that those immediately around the cut-off point will be very similar to one another, but also takes into consideration that those just above it are exposed to the intervention whilst those just below are not. Results for those above and below (the cut-off) can be compared to obtain an estimate of the intervention those above the cut-off point.

- For the **instrumental variables** (IV) approach, selection into treatment should be at least partially determined by an exogenous factor (or shock) which is unrelated to results other than through the treatment. Thus the exogenous factor influences participation, but not directly the results. Typically, such exogenous factors can be administrative errors or oversights, or other random variations in treatment receipt.
- The **Difference-in-differences** (DiD) estimator calculates the difference in a result before and after support and compares a control group against a supported group in order to obtain an estimate of an intervention result. Impacts calculated on the basis of DiD are usually derived within a simple computation (or regression framework) that also accounts for other observed differences between treatment and control groups. Moreover, this approach controls for unobserved differences between the two groups which remain fixed over time as well as differences which vary through time but which affect both control and treatment groups equally (for example economy-wide factors). For the assessment of its external validity the DiD estimator needs to be embedded in a wider evaluation framework.
- **Conditional DiD estimator**. Whilst the PSM can be applied as control for the selection bias on observables (e.g. observed farm characteristics) at the beginning of the programme, a combination of PSM with DiD methods allows for a better controlling of selection bias in both observables and unobservables. Combining PSM and the DiD estimator into a (binary) conditional DiD estimator facilitates unbiased estimates of the binary treatments effect.
- **Generalised propensity score matching (GPS)**: In case all units are programme beneficiaries programme effects can be estimated by employing the GPS methodology. Given explicit information on the intensity of investment support (e.g. financial flows into a public investment programme per farm, holding, community or region) programme effects (results/impacts) under this framework can be analysed by means of a **dose-response function** and **derivative dose-response function.** Such a specification allows to estimate not only the *average* impact of public investment support on the selected result/impact indicator (e.g. GVA/farm or GVA/region), but also to assess the *marginal* effects of those programmes/measures, *independent from the support intensity level obtained*. Obviously, such disaggregated programme evaluation results cannot be obtained by employing traditional evaluation techniques, e.g. the binary propensity score matching methodology, regression discontinuity design, or any other techniques utilised in standard evaluation studies. An important empirical property of this evaluation technique is that it allows assessing programme effects/results/impacts (within a counterfactual framework) in situations where all units (e.g. farms, holdings, communities or regions) received a programme support at various intensity levels.

The main steps involved in carrying out a binary propensity score matching (PSM) are as follows:

a) Obtain a dataset which includes information on basic characteristics and performance of programme beneficiaries and non-beneficiaries in two time periods, i.e. prior to implementation of a given RDP and after this programme. One possibility to pre-select individual units of programme non-beneficiaries to a dataset is to apply programme/measure eligibility criteria.

b) Compute differences in all basic characteristics and performance indicators of programme beneficiaries and non-beneficiaries prior to applying matching.

c) Selection of variables (covariates) to be included in the model.

d) Run a participation model (probit/logit regression). Generally, covariates entering the probit/logit function are expected to determine both programme participation and outcomes (the latter are typically measured in terms of relevant result indicators at micro-level).

e) Calculate participation probabilities (propensity score) for each individual unit (programme beneficiaries and non-beneficiaries) included in the dataset.

f) Drop observations outside the region of common support (i.e. individual observations in the group of programme beneficiaries whose probability of receiving support exceeds that of any from the potential comparison group, or those from the control group with probabilities of receiving programme support below those of any members of the group of programme beneficiaries).

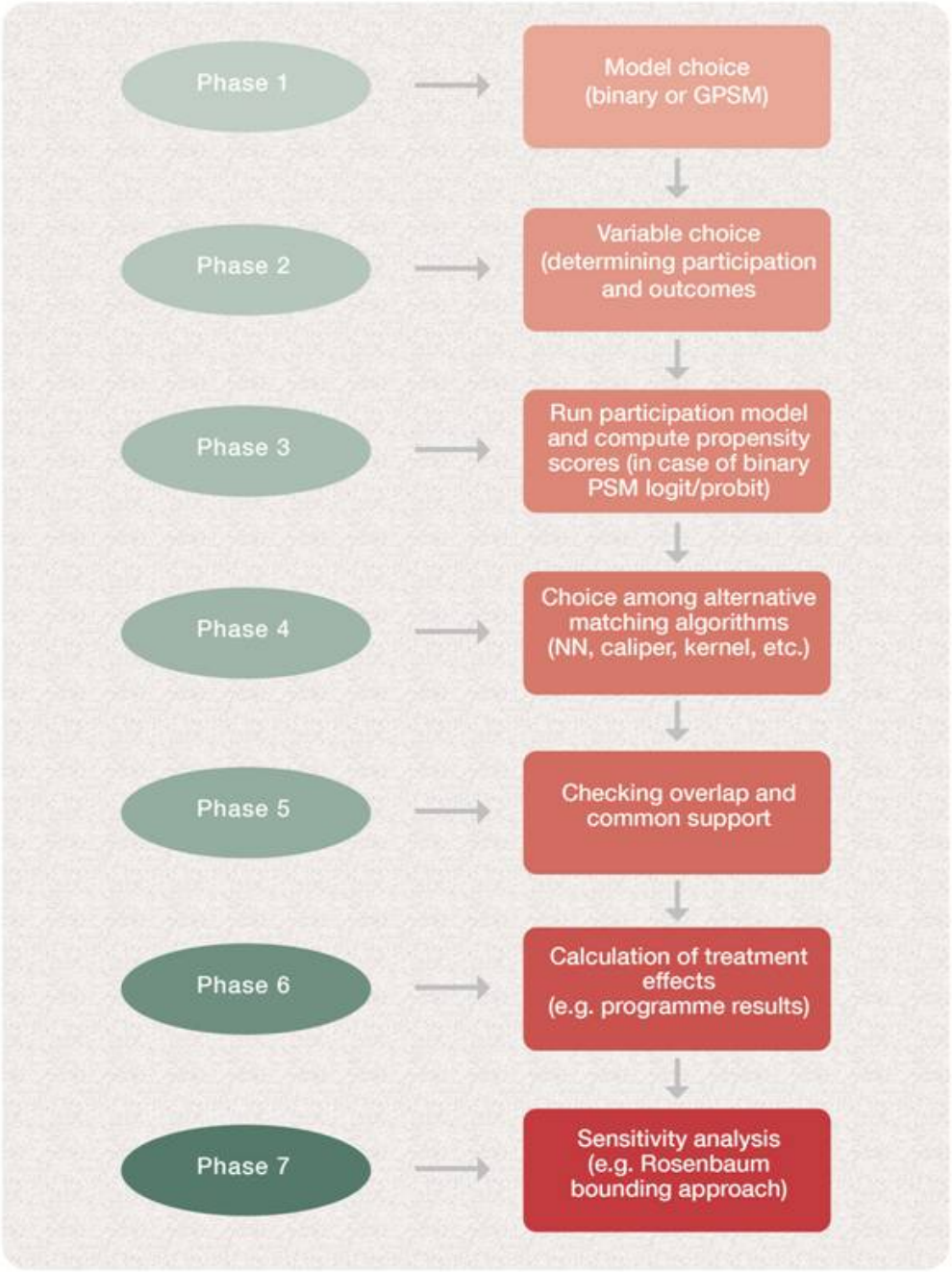
g) Match observations based on participation probabilities (here various matching algorithms, e.g. nearest neighbour, radius caliper, kernel, etc. can be applied). Selection of the appropriate matching algorithm should be subject to statistical analysis of the matching quality, e.g. by i) applying percentage to the standardised bias reduction – after matching; or ii) applying pseudo R2 test after matching - as a matching algorithm selection criterion).

h) Calculate programme results for each pair or set of matched observations.

i) Calculate the average of these differences for a period prior to and after programme implementation using Average Treatment on Treated (ATT) combined with DiD.

j) Performing sensitivity analysis (analysis of potential impact of unobservables on obtained results).

Figure 8. Workflow of PSM



Source: Michalek (2014)

Table 7. Factsheet for an Econometric counterfactual design

Type of method	Econometric method (Propensity Score Matching)
Brief description	The technique is currently one of the most advanced and effective tools applied in evaluation of various programmes, especially if combined with DiD method. One can distinguish two types of PSM approaches: i) standard, conventional or binary PSM; and ii) generalised PSM. PSM is a powerful quasi-experimental approach for finding appropriate controls using counterfactuals and for estimating the programme effects.
Data requirements	Counterfactual design methods are relatively data demanding. Collected economic data should include all relevant information on programme beneficiaries and non-beneficiaries regarding their structure and performance and should cover periods "before" and "after" the implementation of the programme. The biggest part of collected data (approx. 80%) is usually related to the data block "structure" and is used to construct meaningful control groups (e.g. via the application of matching techniques, etc.). The collected results and impact indicators are part of the data block "performance". Data at micro-level should be collected from beneficiaries and non-beneficiaries on the basis of secondary data (e.g. FADN data combined with anonymous data on programme beneficiaries from PA) and/or own surveys.
Counterfactual (How can the method deal with counterfactuals?)	The core is to find, in the case of binary PSM, from a group of non-participants, units that are observationally similar to programme participants in terms of pre-programme characteristics. Participants are then matched on the basis of this probability (propensity score) to non-participants. Each participant is matched with an observationally similar non-participant, and then the average difference in outcomes across the two groups is compared (programme treatment effect).
Scale of indicators	Cardinal
How to measure efficiency, effectiveness, impact?	**Efficiency:** ratio of result/impact indicator for level of support (both at micro, and macro-levels) **Effectiveness:** 1. At micro-level: Outcomes achieved by programme beneficiaries compared to target values⇔ (equivalent to) an increase of a given result indicator due to a given programme measure compared with target values. 2. At micro-level: Outcomes achieved by programme beneficiaries compared to outcomes achieved by programme non-beneficiaries (in %)⇔ (equivalent to) an increase of a given result indicator for programme beneficiaries compared to an increase of the same result indicator for the control group. 3. At micro-level: Structure of a total increase of a given result indicator (% share due to a given measure compared to % share due to other factors). 4. At programme area level: Net outcomes achieved by programme beneficiaries compared to target values⇔ (equivalent to) an increase of a given result indicator due to a given measure compared with target. **Impact** should be computed only at a programme area level. They may be positive or negative, primary and secondary, expected or unexpected, intended or unintended. They are normally expressed in "net" terms (i.e.after subtracting of effects that cannot be attributed to the intervention) and by taking into account indirect effects (e.g. displacement, multipliers, deadweight, etc.). Impacts may be computed by aggregating micro-results to a programme area level.

3.3.3 Method for Impact Assessment of Programmes and Projects (MAPP)

MAPP is a qualitative participative method for assessing impacts based on group discussions. The members of the groups include a range of stakeholders, from several levels (beneficiaries, representatives of beneficiaries, programme managers, etc.). In general it is more interesting to apply MAPP in mixed groups comprising both sexes and different socio-professional groups. This helps to provide a comprehensive picture about an issue, and to assure vertical information flow. Also, in mixed groups, different perceptions can be discussed simultaneously. The participants of the MAPP includes beneficiaries and non-beneficiaries of the investment measures being assessed, namely measures 121, 122, 123, 125, 216, 227, 311, 312 and 313. MAPP is more effective if applied on a limited number of measures to enable also comparisons with impacts from other sources (other programmes, contextual factors, etc.). For this reason, the MAPP method is applied in a limited number of measures in each selected case study territory.

The MAPP method follows a bottom-up approach and therefore is very pertinent for the current evaluation which focuses on specific RDP territories. However, when these RDP territories cover large regions (e.g. Galicia in Spain) or whole countries (e.g. Greece), a sub-territory or sub-territories (e.g. provincial level) should be chosen for conducting the MAPP and preserving its bottom-up character.

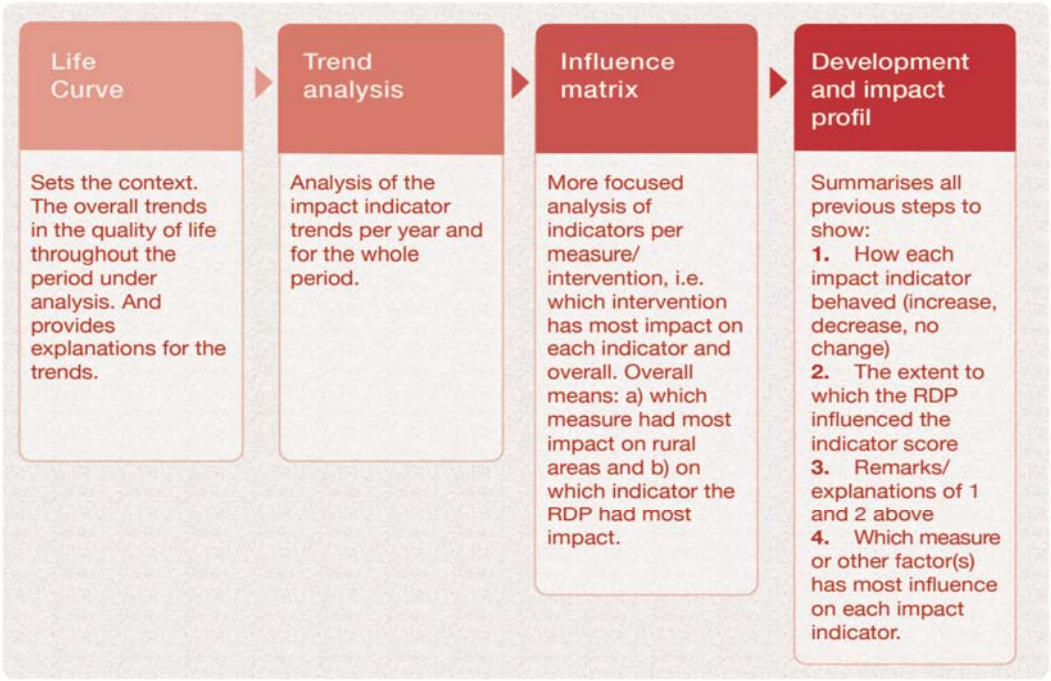
A very important value added of MAPP is the comparison of the respective impacts of different programmes and projects in the same area. For instance it enables the comparison of respective impacts of the selected investment measures with those of other funds e.g. ERDF, ESF and other programmes. In the *ex post* evaluation of RDPs, this proved to be very powerful compared to the interviews, giving a more nuanced, balanced vision of the impacts of the RDP subject of the case study. In particular this method allows us to identify external factors and to discuss the comparative weight of other funds and programmes to better understand the different factors in the evolution of the test area, providing a broader and better view of the probable impacts of the studied measures/RDPs. In this respect, a very important added value of the MAPP in the *ex post* evaluation is to identify critical external drivers of the evolution of rural areas that never came out of interviews.

Basically it allows for better appraisal of the relative weight of the studied measures on the evolution of the test-area in comparison to other factors and policy instruments. For these reasons, it is considered pertinent to be applied in the current context.

The timely and thorough preparation of MAPP focus groups is paramount for its success. A checklist and guidelines are prepared in advance that include: a) the list of steps involved in the implementation of the method; b) the data required, including the indicators to be assessed with the focus group; c) an estimation of data availability (usually finalised during the preparation of the fieldwork); d) the methodological and technical issues related to the preparation of the focus group and the preparation of the experts who will coordinate it; e) the estimated resource implications for preparing and conducting the MAPP focus groups; f) the pertinence of the method for the investment measures and for the evaluation criteria; g) finally a list of open questions that need to be addressed prior to the commencement of the fieldwork.

The following work flow presents the sequence of the MAPP tools used to conduct the focus group.

Figure 9. Workflow of MAPP Method



Source: Parissaki (2014)

Table 8. Factsheet for the MAPP Method

Type of method	Qualitative and participatory method
Brief description of the specific method	MAPP is a participatory method for the assessment of impacts. It uses a list of structured tools to: a) set the overall socio-economic context in the programme area, b) analyse impact indicators trends, c) assess the influence of each measure or programme intervention on each impact indicator as well as the influence of other measures/programmes or other factors on the same indicators, d) assess which measure or other factor had most impact on which indicator and why. By doing so, the method enables a distinction between the measure related impact and the impact due to other factors.
Data requirements	The MAPP method measures impact at the beneficiary level. For instance, the effect on income, job creation, environmental improvements, etc. Simple indicators are therefore used, such as job creation, income increase, water quality improvements, increased tourism, etc.
Counterfactual (How can the method deal with counterfactuals?)	MAPP can measure the counterfactual by including both beneficiaries and non-beneficiaries in the sample of participants.
Scale of indicators	Nominal and ordinal
How to measure efficiency, effectiveness, impact?	MAPP measures impact. Indicators include impact on employment, on incomes, on productivity, on competitiveness, on tourism, on the environment

3.3.4 Programme-theory-based evaluation (TBE)

Brief description

Theory-based evaluation (TBE) came to prominence two decades ago with Chen's book "Theory-Driven Evaluations" (1990). Since then a number of articles, guidelines and textbooks have been published to develop TBE into a detailed methodological framework. The European Commission has also picked up the approach in working papers (see literature list and link) and in the CMEF.

The starting point of a theory-based evaluation design is always a causal chain or theory of change which explains how and why the intervention will work and is expected to lead to the intended outcomes.

The underlying theory of change can refer to a programme theory or to any than other social or economic or political theory that explains how allocating funds will produce outputs through which intended results (specific objectives) and impacts (overall objectives) are to be achieved (the expected change).

In the case of a "programme-theory-based evaluation" a plausible programme theory is established by the intervention logic of RDPs. The intervention logic is an essential cornerstone for the TBE assessment. As outlined in the glossary of the common monitoring and evaluation framework (CMEF) of 2007-2013, intervention logic *"represents a methodological instrument which establishes the logical link between programme objectives and the envisaged operational actions. It shows the conceptual link from an intervention's input to its output and, subsequently, to its results and impacts. Thus, an intervention logic allows an assessment of a measure's contribution to achieving its objectives"*.

Programme-TBE follows each step of the programme´s intervention logic identifying causal links and mechanisms of change. The various links in the intervention logic can be analysed using a variety of methods, building up an argument as to whether the theory of the intervention logic has been realised in practice. It is able to help explain why and how results have been achieved and to appraise the contribution of the RDP activities to the RD objectives in terms of effectiveness.

TBE is a good method to bring light into the **effectiveness** of interventions. The assessment of the effectiveness looks at the extent to which the measures attain the policy objectives. TBE answers the Evaluation Question related to effectiveness: *How and to what extent have the stated policy objectives been achieved?*

In the simple format TBE is based on non-rigorous methods such as monitoring data analysis (quantitative), interviews, surveys, focus groups and case studies (qualitative) which deliver the information to verify the implementation of planned activities in line with the intended change. Hence it relies on quantitative information on financial inputs and outputs and qualitative estimates on results and impacts. Here, a great deal of other information, besides quantifiable, can be used for TBE.

TBE ends up with a **judgement** on the contribution of main outputs and identified results under a certain measure to the intended change by using an ordinal scale. It produces narrative and non-parametric data such as qualitative classifications, e.g. low, medium, high contribution of a measure to achieving its objectives.

The strength of the TBE approach is that it can reflect the effectiveness of a **broad spectrum of objectives** (bundles of sub-objectives) which are pursued under an intervention and the **real implementation mechanism** leading to outputs and results. This is in strong contrast to quantitative methods which are focused on a very limited set of (impact) indicators (such as GVA) and which do not track the micro-steps that lead from programme inputs through to outcomes.

TBE demonstrates success and failure of interventions, however, TBE, cannot examine the "net-impact", i.e. the extent to which the change observed in the programme area in relation to policy objectives can be attributed to interventions. TBE is not able to disentangle the effects of the interventions from the contribution of other factors (it cannot demonstrate on a numerical basis the difference caused by the treatment). Accordingly it is strongly advised to combine theory-based evaluation (in the simple format) with a counterfactual impact evaluation. TBE can deliver the "hypothesis" on investment support effectiveness which can then be verified by other more rigorous methods such as counterfactuals.

Workflow

In the literature different approaches are described how to implement TBE methodically. The approach taken here is tailor-made for the evaluation of EU programmes with a given intervention logic.

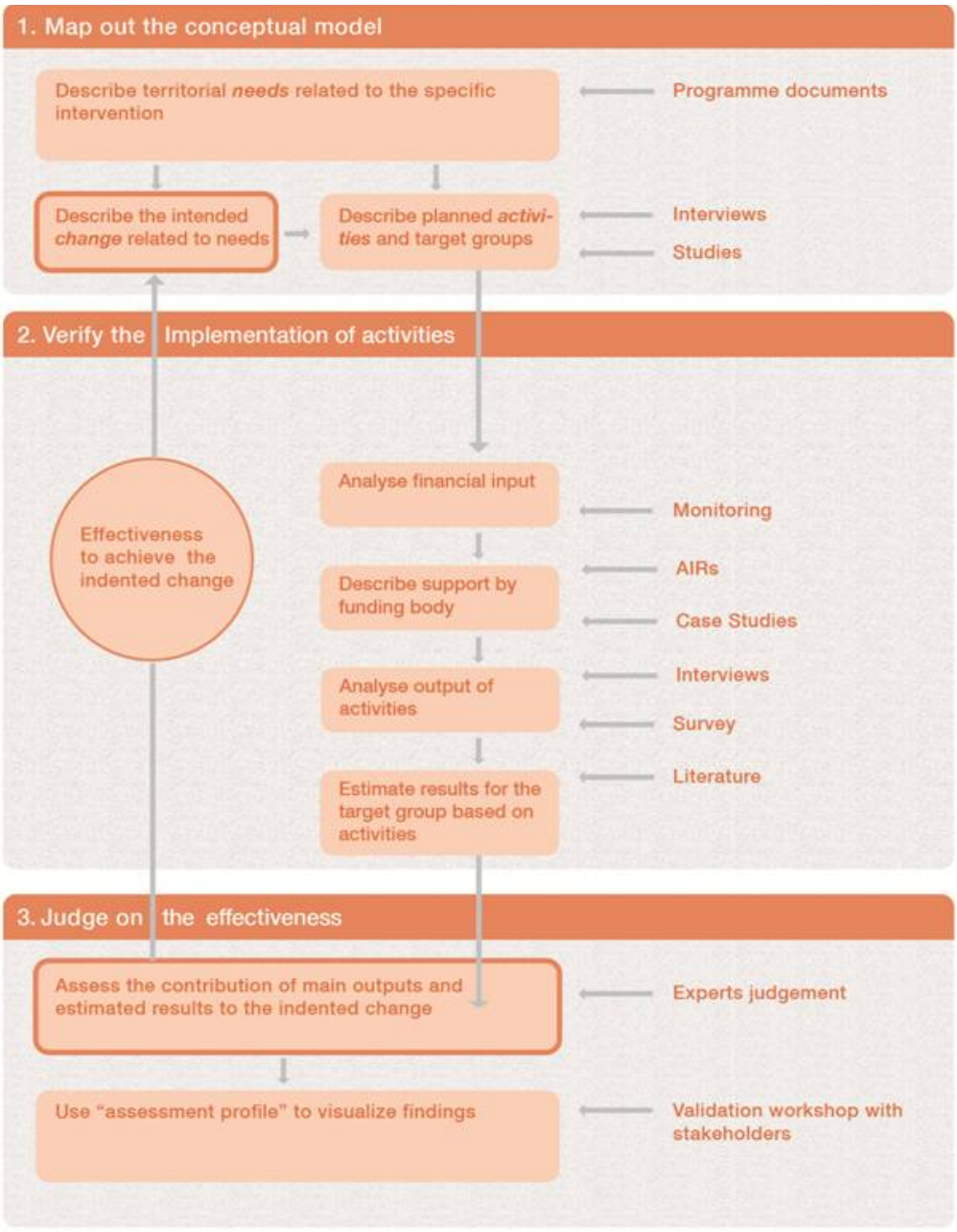
Two types of programme-theory-based evaluation may be applied in the fieldwork: (i) Micro-level TBE verifies the intended change which should be achieved by a specific (sub) measure assessing the effectiveness to achieve specific objectives at the result level; and (ii) Macro-level TBE which deals with a bundle of measures on a more generalised and aggregated level in order to assess the effectiveness to achieve overall objectives on the impact level. A basic description of the TBE workflow is presented in Figure 10.

Following the causal chain of the intervention logic, the TBE verifies each "building block" starting from the intended change.

Map out the conceptual model

When constructing the conceptual model according to the RDP intervention logic, it is necessary to take the following steps. Construct the "theory of change" (= conceptual model for investment support under a specific measure) in a "verifiable way" (based on programme documents and interviews with programme stakeholders if necessary). The conceptual model consists of a brief description of **territorial needs** (taking into account SWOT analysis and needs assessment), related **changes** to be achieved at the end of the programme (specific or overall objectives) related to the identified needs and **planned activities and target groups** per measure (e.g. sub-measures) through which the intended change should be achieved. All elements are interlinked in a causal way. It should be clarified to which level objectives are related (e.g. specific micro-level objectives for the group of direct beneficiaries or overall objectives at the macro-level for whole sector or territory). In the ideal case specific or overall objectives can be more specifically described by judgement criteria (how can success or failure be verified and measured?). This facilitates in later stages the review of objectives. If the intervention logic is not sufficiently described in the programme it has to be re-constructed by the evaluator in close consultation with stakeholders.

Figure 10. Workflow of Programme-Theory-Based Impact Evaluation



Source: Resch (2014)

Verify the implementation of the planned activities (operations level)

The next analytical step is linked to the **planned activities** which were identified in the conceptual model for each of the objectives. Collect empirical data on the projects/operations level to verify the implementation of planned activities (sub-measures) in terms of inputs and outputs based on detailed monitoring data (sub-measures), Annual Progress Reports (APR) or empirical research (survey, case

study). Direct access to monitoring data at project level may be necessary to carry out this step. Financial inputs, output and characteristics of projects/operations supported (e.g. types of investments, sub-sectors and territories affected, size of holdings targeted) may be described on the basis of CMEF provisions (if available). In addition the services given by the funding bodies should be briefly described which may affect the quality and outputs of the supported projects (e.g. business plans are established in most cases or regular meetings with beneficiaries are held or investments are combined with trainings). At the end the main outputs on the project/ operations level per objective are summarised and this leads to the next step.

Transform main project outputs at the operations level into results for the target group (next level)

This step is based on the main outputs on the project/operations level which have been summarised in the previous step. It may be useful to link results to a bundle of activities and main outputs. Collect empirical data on the gross direct effects/results for beneficiaries of implemented project/operations (based on monitoring data, APRs or empirical research). This step is leading from the projects/operations level to gross results for the target group (e.g. holdings, small enterprises). A transformation of main project outputs into results for the beneficiary is necessary given the availability of result indicators in the monitoring system.

CMEF result indicators provide usually information on gross changes in the capacity or performance of direct beneficiaries/holdings (e.g. increase in GVA in supported holdings) which is affected by a bundle of factors and projects implemented (and not only by one single funding operation). Effects of investments are visible only after two or three years; here annual fluctuations have to be considered.

If gross direct effects are not collected by the monitoring systems it is necessary to utilise other quantitative sources (e.g. evaluations) or – in the most limited case – make qualitative estimates. The evaluator should comment on the cause-effect chain between main outputs and direct gross effects. Gross results for the target group are influenced by other factors besides RDP support (e.g. changes in the macro-economic context). Gross results may have been achieved anyway also without RDP support. Therefore it is important to indicate other influencing external factors (social, economic, political or administrative factors).

Judge effectiveness of investment support measures against the expected change

In the final stage of the evaluation, the judging phase, the evaluator draws evidence-based conclusions on the effectiveness of the interventions. According to the CMEF definition, effectiveness assesses the extent to which objectives pursued by an intervention (the expected change) are achieved. The appraisal focuses on whether the programme's intervention (measure) headed towards the expected changes within the programme area and whether the programme objectives (measure objectives) have been achieved.

A judgement on the contribution of main outputs and identified gross direct effects under a specific measure (sub-measure) to the indented change is made by using an ordinal scale,e.g. on a Likert-like scale range from 1 ''very low'' to 5 ''to a great extent''. It is also possible that some achievements cannot be assessed, and this possibility is reflected by allowing evaluators to choose "don´t know" to indicate no evidence whatsoever. A summary table should be established for each measure which provides

a qualitative estimate on the contribution of projects supported to the indented change. The expert´s judgement is based on

- the indented change which should be achieved at the end of the programme;
- main outputs on the project/operations level (quantitative information);
- an indication of other influencing factors;
- direct gross effects for the target group (qualitative estimates).

The rating should be tracked back to main outputs and gross results (if available). The summary table will form the basis to compare the qualitative estimates by TBE with findings of other (more rigorous) methods. If possible the rating should be discussed with stakeholders in a "validation workshop". A key element to make clear statements is the so-called "assessment profile" which is presented in the annexed TBE case study reports. The judgement forms the basis to formulating recommendations for improving theory and practical implementation of the intervention logic.

Table 9. Factsheet on Programme-theory-based evaluation

Type of method	Theory-based approach
Brief description	TBE follows the causal chain of the intervention logic and verifies each "building block" starting from the intended change in order to assess the effectiveness of interventions. In the structuring phase a clear model of the intervention logic has to be established or even re-constructed. In the observing phase qualitative and quantitative data needed for the verification of implementation has to be collected by using different sources. In the analysis phase following the causal chain of the intervention logic, the evaluation verifies each building block (inputs, activities, outputs, and results). The evaluation tries to find evidences for cause-and-effect associations between the different building blocks of the intervention logic. In the judging phase evidences from data collection are cross-analysed to obtain findings on effects which are then used for making a (qualitative) judgement of the contribution to RD objectives (effectiveness) and answering the Evaluation Questions. Based on these statements recommendations for improving the intervention logic are formulated.
Data requirements	In the simple format TBE relies on quantitative information on financial inputs and outputs and qualitative estimates on results and impacts.
Counterfactual (How can the method deal with counterfactuals?)	TBE can provide good insights if and how interventions produce the expected outputs and results in line with the intervention logic, but the focus is not a counterfactual. Theory-based evaluations are more about plausibility and less about causality. Therefore theory-based evaluation should be linked with counterfactual impact evaluation.
Scale of indicators	The judgement on the contribution of main outputs and identified results under a certain measure to the intended change is based on an ordinal scale.
How to measure efficiency, effectiveness, impact?	The contribution of a measure to reaching the specific and overall objectives of RD policy can be assessed in a qualitative way (effectiveness)

3.3.5 Strategic Environmental Assessment (SEA)

Brief description

The **Strategic Environmental Assessment (SEA)** is not an evaluation method *per se* but a structured procedure[33] in which significant effects on the environment are defined *ex ante* and the reasonable alternatives of the proposed plan or programme are identified. The SEA can be used to frame the evaluation of RDP environmental impacts, hence it can play a similar role as the TBE. A significant advantage of the SEA is that it has been a compulsory part of the *ex ante* evaluation of the programme and hence all RDPs are starting from the same point and have, in theory, respected the requirements of the Directive 2001/42/EC on environmental monitoring[34].

The **compulsory monitoring provisions** for the environmental impact of the programme as defined in Article 10 of Directive 2001/42/EC, if done properly, should have provided each RDP with a dedicated database of environmental context indicators. These might be identical to the CMEF environmental baseline indicators or might differ, since some Member States might apply unified SEA Monitoring databases for all programmes subject to SEA.

The reasons for proposing the SEA for this study have been:

- SEA was conducted as an *ex ante* evaluation process for all Member States, so for each RDP results should be available.
- Most environmental reports drafted during the SEA process adopted a measure by measure assessment which fits to the study context, especially considering the bottom-up, micro-approach of RDP monitoring and reporting.
- SEA is similar in its logic to the TBE, so most evaluators can apply the basic frame, even if environmental specialists will be needed at a later stages (e.g. on biodiversity). This feature has been considered as especially relevant for measures of Group B (and in analogy for Group A and D, although they have not actually been subject to the environmental methods in this report but only in exceptional cases), since most RDPs addressed environmental effects only peripherally in the examined measures. In that case SEA can provide a reference framework for the environmental assessment identifying the areas where further research, e.g. via case studies is needed.
- Assessment of environmental impacts in most RDPs is focusing on the evaluation of the Agro-Environmental Measures (AEM) leaving little resources for the other measures. This has been the case for Group C. Similar to Group B the SEA can be used as a reference framework in absence of a better option.
- Last but not least and taking into account the remark that the SEA is not an evaluation method *per se* but a structured procedure, the findings and conclusions generated by this tool can be gradually improved as more advanced methods (e.g. counterfactual impact evaluation with treated/non-treated cases

[33] http://ec.europa.eu/environment/eia/sea-legalcontext.htm

[34] In the SEA procedure the public and the environmental authorities are informed and consulted on the draft Rural Development Programme and the related environmental report. The environmental report and the results of the SEA consultations with RDP stakeholders are taken into account before adoption of the programme. Once the RDP is adopted, the environmental authorities and general public receive relevant information on the potential environmental effects, which programme might cause in course of its implementation. In order to identify unforeseen adverse effects at an early stage, significant environmental effects of the plan or programme are to be monitored.

using the LCA method) can be deployed to "fill in" the reference framework provided by the SEA.

SEA can provide statements only for the **effectiveness and impact** criteria. Its usability on judging **efficiency** is limited. For that reason it was complemented by CEA which focuses mainly on efficiency and partially on effectiveness.

There are two fundamental types of documents:

- The SEA Environmental Report, compiled in 2006/2007 containing the *ex ante* formulation of the causal chain and the "theory of change" among RDP measures or actions and their impact on the so called environmental issues.
- The SEA Monitoring Report tracking down the implementation progress of the RDP and the change on selected environmental context indicators based on the assumptions made in the SEA Report, in some cases the RDP might also make use of reporting sources in relation to other EU reporting curricula such as the Water Framework Directive 2000/60/EC, Natural Habitats Directive 92/43/EEC or Air Quality Directive 96/62/EC.

In various evaluation studies, SEA is applied in slightly different ways; however these approaches are quite similar due to the reference to Directive 2001/42/EC.

A basic description of the SEA workflow is presented in the illustration below; it should be noted however that the approach described here refers to the use of the SEA approach to examine the RDP effects in the context of the present study, hence some characteristic SEA steps used in an *ex ante* application (e.g. scoping, public consultation) are left out.

Workflow

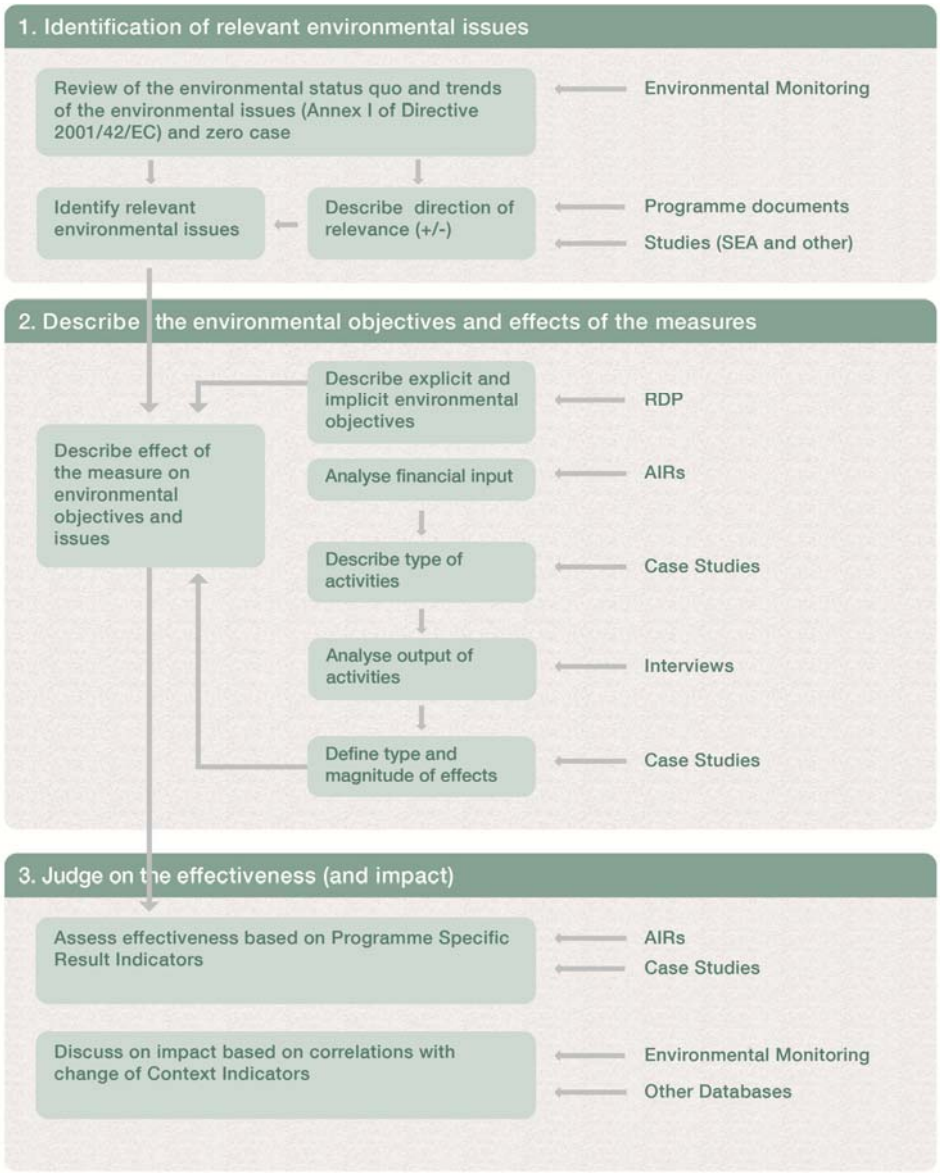
The steps used where:

- Review of the environmental status quo and trends of the environmental issues[35] based on the context indicators of the RDP, the SEA Report, and in an ideal case the SEA Monitoring Report;
- Review of the RDP intervention logic and identification of implicit and explicit environmental objectives defined at measure level;
- Confirmation and if appropriate identification of a relevance between a measure and an environmental issue on the base of the environmental report (yes/no);
- Reporting on direction of the relevance (positive/negative/neutral or negligible);
- Definition of a naive counterfactual (zero case) describing the development without the programme (usually a trend projection) for the illustration of the "gross" effect of the measure;
- Discussion of the effect of the programme using non-parametric data and/or context indicators; conclusions are based on analogy and likelihood of effects based on the description of the RDP measure, experience from the past and other sources;
- Definition of the RDP effects based on the RDP monitoring data (input, output and result indicators) and monitoring of the development of the environmental

[35] Usually those listed in Annex I of Regulation 2001/42/EC i.e. biodiversity, flora and fauna, soil, water, climate, air, landscape, human health and population and cultural and material goods.

issues based on causally, temporally and spatially relevant context indicators in order to identify correlations (this is usually the Achilles' heel of most RDPs and SEAs since the former does not contain or track relevant result indicators while the latter does not implement thoroughly the obligations of Art.10 of the Directive).

Figure 11. Workflow of Strategic Environmental Assessment



Source: Sanopoulos (2014)

Table 10. Factsheet on Strategic Environmental Assessment (SEA)

Type of method	Environmental Approach. N.B.: The Strategic Environmental Assessment (SEA) is not an evaluation method *per se* but a structured procedure, similar to TBE. Hence in a simple form it is a qualitative, theory-based and non-rigorous framework that can be enhanced by more advanced methods (counterfactual, modeling) if possible.
Brief description	SEA is similar to the TBE with the exception that it also examines "change" which might not be part of the intervention logic (unintended effects). Its steps are: • describes the baseline situation for the environmental issues; • reviews the RDP intervention logic, outlines explicit and implicit environmental objectives and identifies intended and unintended effects on the environmental issues; • defines a "theory of change" in forecasting the potential and significant impact the RDP can have on the environmental issues; • monitors the implementation of the RDP and the development of the environmental issues trying to identify programme effects and to take corrective action.
Data requirements	Primary data sources: programme monitoring data (input, output, programme specific result indicators, CMEF baseline indicators), project data if available (as necessary in the environmental permits of projects subject to EIA, especially input-output data, see also ->LCA). Secondary data sources: Context indicators as defined in the SEA Environmental Monitoring Report of the Programme, EUROSTAT and national statistics (e.g. for baseline indicators). Studies.
Counterfactual (How can the method deal with counterfactuals?)	The SEA considers a naive counterfactual in the *ex ante* description of the "zero case", i.e. qualitative description of the situation without the programme. In an *ex post* evaluation, SEA can be enhanced with inputs from a Counterfactual Impact Evaluation in a similar way as the TBE e.g. considering "net results" and "net impacts".
Scale of indicators	Nominal (relation to an environmental issue), ordinal (magnitude of impact), cardinal (change in the context indicator value)
How to measure efficiency, effectiveness, impact?	The SEA does not measure efficiency. Effectiveness and eventually impact are discussed on a qualitative way regarding the contribution to explicitly mentioned or "en passant" addressed environmental objective (nominal and ordinal indicators) or in a quantitative way through the use of explicit (result) indicators or through the comparison and correlation of output indicators and environmental context indicators (cardinal indicators).

3.3.6 Cost-Effectiveness Analysis (CEA)

Brief description

The Cost-Effectiveness Analysis (CEA) is similar in its logic to the Cost-Benefit Analysis but also contains some fundamental differences[36].

CEA is basically designed to compare cost (expressed in monetary values) and effectiveness (expressed with an effectiveness indicator, e.g. a range of 1-100, where the maximum value represents full achievement) for a range of alternatives. This is expressed as the cost/effectiveness ratio (CER), e.g. EUR per Ha of protected area. Hence cost is expressed in EUR, while effectiveness in a "programme objective" unit.

A cost/effectiveness ratio is an important measure for those responsible for allocating resources across programmes. Ratios are obtained from a range of different interventions and enable the allocation of resources to those interventions which provide greater value for money.

However this ratio is based on an "exogenous", expert defined indicator, and does not say if the intervention is worth to undertake. Hence CEA is consistent only within the RDP frame, assuming that the defined effectiveness indicators were relevant in the first place.

If CEA is made in an interdisciplinary context, results can be better communicated to an audience with no background in economics. Researchers with a background in engineering are familiar with discounting and accept positive discount rates while researchers with a background in ecology frequently are not willing to use this concept at all[37].

The CEA can be used as an alternative to CBA in those cases where externalities or benefits are difficult or impossible to express in monetary values. CEA was chosen basically for similar reasons to the SEA: it is simple, can be served by regular monitoring and does not require specific expertise, since it assumes that the environmental objectives of the RDP were sound in the first place. Whenever data and methodology make it possible the CEA should be "graded up" to a CBA.

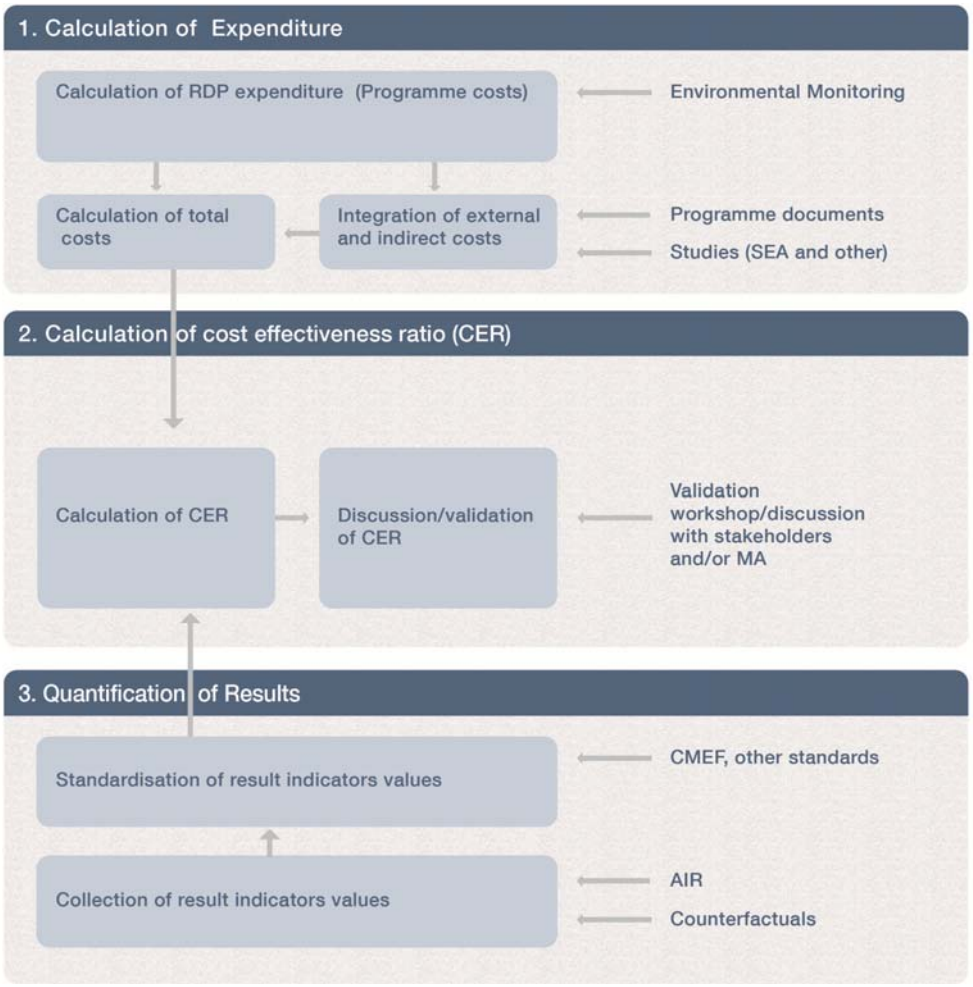
The basic steps are progressing from a basic "worst case scenario" to more complete data sets (see Figure 12).

In the case of the CER numerator (i.e. EUR):

- Programme Costs (i.e. EAFRD + National Contribution) are available. This is common and these numbers are found in the RDP and APR.
- Full costs (i.e. programme expenditure, non-eligible but related expenditure, hidden costs, administrative costs, etc.) are available; this is rarely the case and in the best case reference values are possible through case studies.

[36] Bleichrodt, H., Quiggin, J., (1999). "Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis?", J Health Econ 18 (6): 681–708.
[37] WHO (2003), Making choices in health: WHO guide to cost-effectiveness analysis, Geneva

Figure 12. Workflow of Cost-Effectiveness Analysis



Source: Sanopoulos (2014)

In the case of the CER denominator (i.e. EUR):

a) The RDP lacks result target values, hence no CER is possible;
b) The RDP does not (regularly) report result achievements; in that case a CER is only possible at the end of the programme implementation. For illustration purposes also the *ex ante* target value can be used;
c) The RDP regularly reports result achievements but in a non-standardised indicator form (this is especially the case for environmental results under Group A, Group B and Group D; in theory Group C should use the standardised CMEF indicator "*Ha of land under a given regime*" but even then there is discretionary interpretation);
d) The RDP regularly reports gross result achievements in a standardised indicator form (e.g. m^3 of water saved). This level could be reached if the RDPs monitor thoroughly their projects and establish an environmental indicators set for their investment measures.

e) The RDP regularly reports net result achievements in a standardised indicator form (derived by CIE). This is an ideal case that is seldom achieved.

In practice the combination of the cases 1 and b) or c) is encountered among investment measures.

Table 11. Factsheet on Cost Effectiveness Analysis (CEA)

Type of method	Environmental approach frequently combined with Computational Economic Models.
Brief description	The basic steps are: Collection of the monitoring data on RDP achievements (e.g. result indicators and expenditures); Examination of the available common and RDP-specific indicators on their suitability to express the effectiveness of the action; Calculation of the Cost/Effectiveness Ratio (CER); Putting the CER results into context especially comparing to findings of -> Theory-based evaluation findings or through discussion with stakeholders.
Data requirements	Primary data sources: RDP monitoring data (input, output, result data), interviews/focus groups with beneficiaries, stakeholders, experts and case studies. Secondary data sources: Eurostat and national statistics (e.g. for baseline indicators). Studies and standards for consideration of indirect costs, if available.
Counterfactual (How can the method deal with counterfactuals?)	The CEA does not consider in principle counterfactuals. However the calculation of the Cost/Effectiveness Ratio can be calculated using also “net results” without modification, leading to more reliable conclusions. Therefore CEA should be combined (in an advanced setting) with counterfactual impact evaluation whenever possible.
Scale of indicators (nominal, ordinal, cardinal)	Cardinal (or ordinal in lack of cardinally defined targets)
How to measure efficiency, effectiveness, impact?	The CEA only measures efficiency. Effectiveness and impact are not covered. Efficiency is expressed via the Cost Effectiveness Ratio, as unit of input/unit of result, e.g. in Group C EUR/Ha of land under successful management.

3.4 Choice and justification of the selected RDP territories

3.4.1 Criteria for the selection of case study territories

At least ten different RDP territories had to be selected for testing the methods. The selection criteria in line with the ToR were:

- C1. Geographical balance
- C2. Inclusion of both old and new Member States
- C3. Coverage of businesses and investments across the size range
- C4. Number of projects implemented under the relevant measures
- C5. Programmed and actual expenditure under the relevant measures
- C6. Targeting approaches used to direct investment support towards specified objectives
- C7. Availability of data

Furthermore, the ToR stated a set of requirements for the selection of case studies:

- R1. Conduct case studies in at least 10 different RDP territories
- R2. Each of the measures listed in the TOR should be included in a minimum of two of the territories selected
- R3. The measures accounting for most programmed expenditure overall should be represented more widely
- R4. Within the proposed RDP territories the minimum number in which each group of measures should be evaluated is as follows:
  - The selected RDP territories may be used to assess more than one of the four groups of investments;
  - Group B case studies to be undertaken in RDP territories where measures 121 and 122 are also being evaluated.

Thus the study team had chosen, in agreement with the Steering Group, a list of 11 case study territories to test the evaluation methods. For each of the methods the data requirements had been specified by the respective core team members. Geographic Experts assessed the data-availability in a first screening exercise.

After the screening one case study (Spain/RDP Asturias) was exchanged with Spain/RDP Galicia. The final outcome of the screening of data availability is summarised in the table below and described in more detail on the following pages.

Table 12. Overview of data availability in case study regions

RDP territories	Methodologies for assessing investment support and applicability for investment group							
	Input-Output Analysis	Econo-metric counterfactu-al analysis	MAPP method	Theory-based evaluation design	Environmental Impact Assessment	Cost-Benefit Analysis	Cost Effective-ness Analysis	Life Cycle Assessment
	A,B	A,B,C,D	A,B,C,D	A,B,C,D	A,B,C,D	A,B,D	A,B,C,D	A,D
01_AT	+	+/R	+/R	+	+/R	0	+	0
02_CZ	+	+/R	+/R	+	0	0	+/R	0
03_DE (Hessen)	+/R	+/R	0	+	0	0	+/R	0
04_DK	+	0	-	+	0	0	+/R	0
05_ES (Galicia)	+	0	+/R	+/R	+/R	0	+	0
06_FR	+ (only 4 regions)	0	+ (only 4 regions)	+	+/R	0	+	0
07_PL	+/R	+/R	+/R	+	-	0	+	0
08_GR	+	0	-	+	-	0	+/R	0
09_SK	+/R	+	+/R	+	-	0	+	0
10_CY	+	0	-	+	-	0	+/R	0
11_UK (Scotland)	+/R	-	+/R	+	-	0	+	0

Source: Data checklist filled-in by Geographic Experts

Legend:

+	data largely available
+/R	data partly available, research needed to fill gaps
-	data not available, extensive research would be necessary to establish data sets
0	data are not accessible or only at high cost

3.4.2 Data availability for proposed methods in case study regions

For each of the methods data requirements have been defined by the Core Team Experts. Geographic Experts were asked to verify data availability in the selected case study regions for each of the methods as described below.

Data availability for Input-Output Analysis

According to our investigation, Input-Output Analysis which is suitable for investment groups A and B is broadly applicable. In some cases additional research was necessary to fill data gaps, e.g. data on RDP measures' payments by type of expenditure. Specifically availability by country is as follows:

- Austria: A National IO Table for 2009 (the 2008 IO Table available from Eurostat had some problems) and sectoral employment data at national level have been obtained from Eurostat. Farm structures data utilised to disaggregate agriculture into sub-sectors was provided by the Geographic Expert (GE). The Mid-term Evaluation Report (MTE) 2010 report, the APR for 2012 and RDP financial flows data (2012) were all available from DG AGRI. Data on agricultural sub-sectors

became available from FADN. RDP data on: i) realised expenditure/investment *per annum* for measures of investment groups A and B; ii) distribution of measures expenditure *per annum* according to type of investment action; and iii) direct impacts of RDP measures on production capacity were requested by the GE from the Austrian RDP Managing Authority and were provided. Survey data on measure-specific investment without RDP support has been provided by the counterfactual analysis survey.

- Czech Republic: A National IO Table for 2007 and sectoral employment data at national level have been obtained from Eurostat. Farm structures data utilised to disaggregate agriculture into sub-sectors was provided by the GE. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on: i) realised expenditure/investment *per annum* for measures of investment groups A and B; and ii) distribution of measures expenditure *per annum* according to type of investment were requested by the GE for the Czech Republic from the Czech RDP Managing Authority and were provided. Direct impacts of RDP measures on production capacity and data on measure-specific investment without RDP support has been provided by the counterfactual analysis survey.
- Hessen (DE): Supply and Use Tables for Hessen, 2007 have been provided by the Jülich Research Centre and were utilised to generate a symmetric IO Table for Hessen for year 2007. Sectoral employment data at the regional level has been obtained from Eurostat. Farm structures data has been provided by the GE, and utilised to disaggregate agriculture into sub-sectors. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on: i) realised public expenditure for measures of investment groups A and B; and ii) distribution of measures public expenditure according to type of investment were requested by the GE for Hessen form the Hessen RDP Managing Authority and provided. Direct impacts of RDP measures on production capacity and data on measure-specific investment without RDP support has been provided by the counterfactual analysis survey.
- Galicia (ES): A Regional IO Table for Galicia, 2008 and sectoral employment data at regional level have been obtained from the Galician Regional Government. Farm structures data has been provided by the GE and utilised to disaggregate agriculture into sub-sectors. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on: i) realised expenditure/investment for measures of investment groups A and B; ii) distribution of measures expenditure according to type of investment; and iii) direct impacts of RDP measures on production capacity were provided by the Galicia GE.
- Poland: Supply and Use Tables for Poland, 2007 have been obtained from Eurostat and utilised to construct a symmetric IO Table for 2007. Sectoral employment data at national level have been obtained from Eurostat. Farm structures data has been provided by the GE, and utilised to disaggregate agriculture into sub-sectors.The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on realised public expenditure *per annum* for measures of investment groups A and B has been

obtained. Data on the distribution of measures of public expenditures *per annum* according to type of investment has been requested by the GE for Poland from the Polish RDP Managing Authority. Direct impacts of RDP measures on production capacity and data on measure-specific investment without RDP support were provided by the counterfactual analysis survey.

- Scotland (UK): A Regional IO Table for Scotland, 2007 and sectoral employment data at regional level have been obtained from the Scottish Office. Farm structures data were provided by the GE and utilised to disaggregate agriculture into sub-sectors. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on: i) realised expenditure/investment for measures of investment groups A and B; ii) distribution of measures expenditure according to type of investment; and iii) direct impacts of RDP measures on production capacity was requested by the GE form the Scottish RDP Managing Authority and was provided.
- Greece: A National IO Table for 2008 and sectoral employment data at national level have been obtained from Eurostat. Farm structures data has been provided by the GE, and was utilised to disaggregate agriculture into sub-sectors. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were all available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on: i) realised public expenditure for measures of investment groups A and B; ii) distribution of measures public expenditure according to the type of investment; and iii) direct impacts of RDP measures on production capacity were requested by the GE from the Greek RDP Managing Authority and provided.
- Cyprus: Supply and Use Tables for Cyprus, 2008 have been obtained from Eurostat and were utilised to generate a symmetric IO Table for Cyprus for year 2008. Sectoral employment data at national level has been obtained from Eurostat. Farm structures data was provided by the GE and utilised to disaggregate agriculture into sub-sectors. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DG AGRI. Data on agricultural sub-sectors became available from FADN. RDP data on: i) realised expenditure/investment for measures of investment groups A and B; ii) distribution of measures expenditure according to the type of investment; and iii) direct impacts of RDP measures on production capacity were requested by the GE from the Cypriot RDP Managing Authority and were provided.
- Slovakia: Supply and Use Tables for Slovakia, 2009 (the 2008 IO Table available from Eurostat had some problems) have been obtained from Eurostat and were utilised to generate a symmetric IO Table for Slovakia for year 2009. Sectoral employment data at national level has been obtained from Eurostat. Farm structures data was provided and utilised to disaggregate agriculture into sub-sectors. The MTE 2010 report, the APR for 2012 and RDP financial flows data (2012) were available from DGAGRI. Data on agricultural sub-sectors became available from FADN. RDP data on realised public expenditure *per annum* for measures of investment groups A and B has been obtained. Data on the distribution of measures expenditure *per annum* according to the type of investment has been requested by the GE from the Slovakian RDP Managing Authority. Direct impacts of RDP measures on production capacity and data on measure-specific investment without RDP support has been provided by the counterfactual analysis survey.

Data availability for the econometric counterfactual design

In all EU Member States a counterfactual econometric analysis of effects of RDP support (especially under measure 121 but also other investment measures) was possible and had been **greatly facilitated** by a relatively **good availability** of individual farm time-series data (i.e. records covering number of years) in form of **national farm bookkeeping records or/and FADN data**. The bookkeeping/FADN records could relatively easily be combined with information on RDP beneficiaries available in national Paying Agencies (PA). In this respect, linking FADN or farm bookkeeping data with PA records was **always** (!) carried out by respective national FADN offices. In some cases individual permissions concerning data protection were required.

- In Poland a national FADN office (IERIGZ, Warsaw), responsible for collection and processing of FADN records, carried out a compilation of Polish FADN data with another set of records about programme beneficiaries (available from PA) and provided the requested anonymous individual dataset to the project (5154 farms in years 2006 and 2012; panel data). This dataset was made available at FADN office site only so the whole analysis had to be carried out at this location.
- In Austria: The Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Managements agreed to provide an access to micro-economic farm data necessary for carrying out counterfactual analysis within this study. Available anonymous data combined the farm bookkeeping data (approx 1400 farms over the period 2006 – 2012; panel data) with data from PA on programme beneficiaries (including the level of obtained support from individual RDP measures). The bookkeeping dataset for Austria contains many variables that describe the characteristics and performance of an individual farm household in a detailed manner. As in principle micro-data is heavily protected, analysis of programme effects using counterfactual methodologies were carried out in the premises of the Austrian Institute for Economic Research (WIFO) which regularly undertakes various analytical tasks involving FADN data.
- In Czech Republic: Available anonymous data provided by Czech authorities to the project combined the Czech FADN data (approx 600 farms over the period 2006 – 2012; panel data) with data from the PA on programme beneficiaries (including the level of obtained support from individual RDP measures). As in principle micro-data is heavily protected, necessary calculations of programme effects using counterfactual methodologies were carried out in the premises of the national FADN office, i.e. Institute for Agricultural Economics and Information (UZEI) which regularly undertake various analytical tasks involving FADN data.
- In Slovakia: anonymous farm data was provided by Slovak authorities to the project. Available dataset combined the Slovak FADN data (approx 600 farms over the period 2006 – 2012; panel data) with data from the PA on programme beneficiaries (including the level of obtained support from individual RDP measures). After making all records anonymous all requested data was made available to the project without any further restrictions.
- In Germany/Hessen numerous administrative procedures had to be followed aiming at receiving respective permissions from the Hessian Ministry of Agriculture and other national authorities enabling the use of available FADN and PA data. After all these procedures were successfully completed available anonymous farm bookkeeping data combined with relevant records from the PA was transferred to the project without any further restrictions.

- Spain (Galicia): administrative procedures required to receive requested data appeared as too time consuming.

Data availability for MAPP method

Regarding the MAPP method, a participatory impact analysis method, the application in all chosen case study regions was possible. Data availability for this method is less of an issue vis-a-vis the overall availability and openness of relevant stakeholders to participate in the assessment. However, the presentation of official statistics/monitoring data for the indicators addressed with MAPP helps stimulate the debate amongst MAPP focus group participants and drive the discussion towards an analysis of the effects and causes of changes of the indicators.

Data availability for Programme-theory-based evaluation

TBE design verifies the RDP intervention logic as it is usually done in MTE 2010. A broad applicability is given. Two types of programme-theory-based impact evaluation were applied in the fieldwork:

- Micro-level TBE models and verifies the intended change which should be achieved by a specific (sub)measure (assessment of the investment support types at the detailed level of (sub)measures), and
- Macro-level TBE which deals with a bundle of measures on a more generalised and aggregated level.

The required data are available in all case study regions. However the GE should update findings on investment support from MTE 2010 reports by utilising the Annual Progress Reports 2012. Moreover, in order to verify the implementation of the planned measures at micro-level direct access to RDP monitoring data at project level is necessary. In the case of the macro-level assessment it may be sufficient to relay on generalised data which are presented in the Annual Progress Report. Additional interviews with stakeholder may be necessary to clarify the intervention logic and to fill data gaps.

Data availability for Environmental approaches

The selected approaches related to the assessment of environmental impacts are broadly applicable as data is stemming mainly from mandatory reporting (e.g. data from SEA Monitoring Report according to Art.10 of the Directive 2001/42/EC, data from EIA-environmental permits, reports on Water Framework Directive implementation, reports on Flora Fauna Habitats-Directive, etc.). However data quality was a main issue in the fieldwork phase.

Also the **Cost-Effectiveness Analysis** is broadly applicable; however it is not suitable as a “stand alone method” and needs to be complemented by other approaches.

3.4.3 Combination of evaluation methods tested in case study regions

Based on the verified data availability in the case study regions the following combinations of methods were tested in the fieldwork phase:

Multiple combinations of methods are used for a number of case study areas where data is highly available to test the most advanced combination of evaluation methodologies including counterfactual design. This refers to, for example to, the RDPs Austria (M311, M125), Czech Republic (M121, M123, and M311), Germany/Hessen (M121), Poland (M121), Cyprus (M121) and Slovakia (M121) where a combination of sophisticated input-output analysis, counterfactual design, programme-theory-based evaluation and/or MAPP, in some cases accompanied by a strategic environmental assessment is possible. The adoption of these methods leads to the least biased, the most credible and robust findings (under given data circumstances).

Combination of methods: For a number of case study areas which face some constraints in data availability slightly less advanced method combinations were tested, combining input-output analysis and MAPP or programme-theory-based evaluation, in some cases accompanied by a strategic environmental assessment. This is for example the case in Spain/Galicia, UK /Scotland and Greece where IO and MAPP were combined to assess investment type A measures. The lack of counterfactual design is typical for this group.

Basic set of methods: The third case addresses the specific situation of Denmark and France where only a limited set of qualitative non-rigorous methods such as TBE or SEA could be applied due to data constraints. These cases establish a “baseline” of what can be achieved by using less rigorous methods.

In addition, **targeting** of investment support was analysed across all RDP territories for around 30 measures (see the following table; the measures for which targeting was examined are highlighted in green).

The combination of methods per RDP territory allows for a comparison of quantitative evaluation methods towards less rigorous and less data demanding evaluation techniques. Most methods (with the exception IO analysis) can cater for a broad variety of investment types. The differences in methods are constituted mainly by data needs. Clearly, lower data availability enforces an application of simpler evaluation techniques. However, simpler techniques usually tend to be accompanied by a higher selection bias which is equivalent to a reduction of evaluation rigour.

In total six methods were applied (IO, PSM, MAPP, TBE, SEA, CEA) covering 62 measures under four investment types in 11 RDP territories. Overall 127 method tests were implemented.

The following table shows for each measure which methods were effectively applied in the case study territories. As a minimum three methods had to be tested per case study. In addition, the measures are highlighted in green where also the targeting approach was examined.

Table 13. Combination of methods per case study and overview of targeting analysis

Case studies		Planned application				Realised application of methods (measures where Targeting was analysed are highlighted in green)												
MS	RDP territory	A:	B:	C:	D:	A						B	C			D		no of methods
						M121	M122	M123	M311	M312	M313	M125	M216	M227	M313	M121	M123	
AT	Austria	x	x			IO, PSM	IO, PSM	IO, PSM	IO, PSM, SEA, CEA	IO	IO	IO, PSM, SEA, CEA						4
CZ	Czech Rep.	x	x	x	x	IO, PSM, MAPP, TBE		IO, PSM, MAPP, TBE	IO, PSM, MAPP, TBE	IO	IO	IO		TBE	TBE	n.a.	n.a.	4
DE	DE / Hessen	x		x		IO, PSM, TBE		IO	IO	IO	IO			TBE, SEA, CEA				5
DK	Denmark	x		x		TBE, CEA		TBE						n.a.				2
ES	ES / Galicia	x	x	x		IO, MAPP	IO, MAPP	IO, MAPP	IO, MAPP	IO, MAPP	IO, MAPP	MAPP, SEA, CEA		MAPP, SEA, CEA				4
FR	France			x	x	SEA, CEA							TBE, SEA, CEA	TBE, SEA, CEA	TBE, SEA, CEA	TBE		3
PL	Poland	x			x	IO, PSM, TBE		IO	IO	IO						n.a.	n.a.	3
UK	UK / Scotland	x		x		IO, MAPP	IO	IO, MAPP	IO, MAPP	IO	IO			SEA, CEA				4
GR	Greece	x	x			IO, MAPP		IO, MAPP	IO, MAPP	IO, MAPP	IO, MAPP	IO, SEA, CEA						4
CY	Cyprus	x		x		IO, TBE, SEA, CEA		IO, TBE						TBE, SEA, CEA	TBE			4
SK	Slovakia	x	x		x	IO, PSM, MAPP	IO	IO, MAPP	IO, PSM		IO	IO, MAPP				n.a.	n.a.	3

Source: Metis (2014)

4 Spotlight on selected case study findings

In this chapter the potential results of each method are demonstrated by showing selected case study findings in detail. The selected case studies are not representative but rather show where the conditions for the respective method were particularly favourable. To facilitate the comparison of findings, all cases refer to measure 121. Only for SEA/CEA measure 125 was chosen, since measure 121 is not focused on environmental outcomes.

4.1 IO method in case study Austria

Context and scope

One territory for which the IO method was applied in order to assess the economic impacts of RDP 2007-2013 measures was Austria; hence, this assessment corresponds to the national level. Main characteristics of agriculture and rural areas in Austria include the following:

- 8 million inhabitants of whom 78% live in rural areas and the remaining 22% are located in the predominantly urbanised areas of Vienna and the Rhine valley in Vorarlberg.
- Around 54% of all farms are managed on a part-time basis;
- Almost 80% of the total land area and about 70% of utilised agricultural area (UAA) is classified as less-favoured area.
- High nature value, clean environment and rich cultural and natural heritage prevail.

In the Austrian RDP 2007-2013 a strong emphasis was placed on the environment, nature protection and landscape conservation, taking account of the various environmental effects of agriculture and forestry, responding to the particular geographic and topological situation of an alpine country, and the demands of society. Therefore Axis 2 was the most important axis in financial terms.

Total public expenditure amounted to EUR 7.8 bn. Axis 1 (13.8% of RDP funds) prioritised the modernisation of farm holdings, the setting-up of young farmers, and investment for infrastructure and for adding value to agricultural and forestry products by supporting processing and marketing. The main priorities under Axis 2 (72% of total funds) were the agri-environmental measures with a broad variety of sub-measures. Compensatory allowances in less-favoured areas and payments for agri-environmental measures accounted for 90% of Axis 2. Main priorities under Axes 3 and 4 (in total, 12% of funds) included basic services for the economy and rural population, the conservation and upgrading of the rural heritage and the support to integrated rural development strategies. Nature conservation, national parks, cultural landscape development and awareness raising for potentials of the Alpine region were the most important measures implemented.

Findings

In the case of Austria all Type A and B measures were assessed through IO analysis. This includes measures 121, 122, 123, 311, 312 and 313 (Type A) and 125 (Type B). Also, counterfactual analysis data were provided for the evaluation of measures 121, 122, 123, 125 and 311.

Variants of the methods use in the case of Austria include:

- The use of a national IO table for 2009, since the 2008 IO table had several balancing problems.
- The use of capacity-adjustment data (data on GVA increase) for the period 2007-2012, since data for 2013 was not available.

The basis of the model is an IO table for Austria, where agriculture was disaggregated to field crops, milk, other grazing livestock, granivores and other agriculture. Also, other sectors targeted by the RDP (e.g. food processing, tourism, trade, etc.) were separately recorded in the IO table. In terms of ranking, output multipliers seem high for construction, energy, trade and several service sectors, while values for agricultural sub-sectors seem comparatively low. Income generation capacity seems very high for public services, recreation, trade and construction, but rather low for food processing and agriculture. In contrast to the above findings, the agricultural sub-sectors seem to be the highest performers in terms of employment generation capacity.

Measure 121 is associated with the highest average annual investment flows (EUR 368.4 million in 2009 prices), followed by measure 123 (EUR 60.2 million). About 77% of investment flows were directed to construction, while around 20% were used for purchasing machinery and equipment.

Table 14 and Table 15 present the capacity-adjustment analysis inputs to the IO model originating from national and counterfactual data, respectively. Estimates of GVA increase associated with RDP interventions seem to generate significant annual increases in the output of food processing and agriculture (Table 13)[38]. However as shown in Table 15, counterfactual analysis estimates on additional GVA (and thus, output) were extremely low compared to national data ones. This is because counterfactual analysis indicated far more modest effects on additional GVA associated with measures beneficiaries.

As shown in Table 16, total economic effects of the RDP measures analysed for Austria, indicate increases of 0.43% for output, 0.39% for income and 0.56% for employment. In general, the magnitude of these effects is similar to those estimated in other recent relevant studies.

Main observations associated with these results include:

- The significance of capacity-adjustment effects in employment creation, which can be attributed to the links of measures 121 and 123 with agriculture, i.e. a sector which is able to generate a comparatively very high number of jobs for every additional million EUR of output.
- The fact that investment seems to have generated higher output and income effects than those associated with the operation of investments.
- The fact that measure 121 (important in terms of expenditure) generates very high investment effects and very low capacity-adjustment effects; this pattern is more or less repeated in the case of measures directly targeting agriculture and forestry.

[38] Due to lack of data, capacity-adjustment effects were estimated for agriculture in total.

- The fact that measures targeting non-primary sectors (i.e. M123 and Axis 3) are associated with capacity-adjustment effects which are significantly higher than investment effects.

Table 14. Capacity-Adjustment Analysis – national Managing Authority data: Inputs to the Austrian IO Model (in EUR million)

Measures/Sector	Increase in GVA per annum	Increase in Output per Annum	Corresponding Sector
Measure 121	54,830	202,752	Agriculture
Measure 122	4,900	12,776	Forestry
Measure 123 Food Processing	96,000	474,440	Food Products
Measure 123 Wood Processing	1,144	2,983	Wood Products
Measure 125	13,310	34,700	Forestry
Measure 311 Tourism	3,330	5,767	Accommodation and Food Services
Measure 311 Energy	3,330	15,303	Electricity, gas
Measure 311 Food Processing	3,330	16,457	Food Products
Measure 312 Tourism	3,550	6,148	Accommodation and Food Services
Measure 312 Trade	3,550	5,962	Wholesale-Retail trade
Measure 313 Tourism	4,66	8,071	Accommodation and Food Services
Measure 313 Recreation	4,66	7,826	Recreational services

Source: Austrian RDP Managing Authority; Psaltopoulos (2014)

Table 15. Capacity-Adjustment Analysis – econometric counterfactual analysis data: Inputs to the Austrian IO Model (in EUR million)

Measures/Sector	Increase in GVA per annum	Increase in Output per Annum	Corresponding Sector
Measure 121	34,100	126,096	Agriculture
Measure 122	-6,300	-16,426	Forestry
Measure 123	-0,600	-2,965	Food Products
Measure 125	1,100	2,868	Forestry
Measure 311 Tourism	0,833	1,443	Accommodation and Food Services
Measure 311 Energy	0,833	3,828	Electricity, gas
Measure 311 Food Processing	0,833	4,117	Food Products

Source: Counterfactual Analysis; Psaltopoulos (2014)

As expected capacity-adjustment effects associated with counterfactual analysis are much lower than those estimated through the use of national data (see Table 16). Indicatively, counterfactual analysis output and employment effects estimated are 14% and 34% of those estimated through the use of official national data.

Sensitivity analysis of investment effects involves a 10% increase of the share of construction in total RDP expenditure (for each measure) and an equivalent decrease

in the share of machinery/equipment. Sensitivity results confirm the robustness of the method, as differences in estimated impacts range between 0.7% and 1.2%.

Table 16. Impact Analysis, Austria based on national managing authority data (average annual effects compared to 2009) (in EUR million)

Type of Effect	Change in output	% change	Change in GVA	% change	Change in Employ-ment (FTEs)	% change
a) Investment Effects						
Measure 121	1,194.03	0.186	475.61	0.190	6,822	0.197
Measure 122	28.62	0.004	11.71	0.005	176	0.005
Measure 123	186.30	0.029	81.71	0.033	1,145	0.033
Measure 125	58.94	0.009	22.84	0.009	324	0.009
Measure 311	40.00	0.006	18.32	0.007	256	0.007
Measure 312	8.73	0.001	4.07	0.002	55	0.002
Measure 313	14.46	0.002	5.41	0.002	123	0.004
TOTAL	**1,531.08**	**0.239**	**619.68**	**0.248**	**8,901**	**0.257**
b) Capacity-Adjustment Effects						
Measure 121	295.89	0.046	83.88	0.034	4,950	0.143
Measure 122	14.63	0.002	5.50	0.002	99	0.003
Measure 123	800.10	0.125	205.66	0.082	4,448	0.128
Measure 125	39.73	0.006	14.93	0.006	270	0.008
Measure 311	55.61	0.009	16.10	0.006	261	0.008
Measure 312	19.41	0.003	9.89	0.004	152	0.004
Measure 313	25.49	0.004	12.98	0.005	199	0.006
TOTAL	**1,250.86**	**0.195**	**348.93**	**0.140**	**10,379**	**0.300**
c) Total Effects						
Measure 121	1,489.92	0.23	559.49	0.22	11,772	0.34
Measure 122	43.25	0.01	17.21	0.01	275	0.01
Measure 123	986.40	0.15	287.37	0.12	5,593	0.16
Measure 125	98.67	0.02	37.78	0.02	594	0.02
Measure 311	95.61	0.01	34.42	0.01	517	0.01
Measure 312	28.15	0.00	13.96	0.01	207	0.01
Measure 313	39.94	0.01	18.39	0.01	323	0.01
TOTAL	**2,781.93**	**0.434**	**968.62**	**0.388**	**19,281**	**0.556**

Source: Psaltopoulos (2014)

Finally, as shown in Table 18, (with the exception of measure 313) jobs generated due to investment are generally around 18 -19 per million EUR. However, for measures directly targeting agriculture and forestry and (even) farms, jobs generated (per million EUR invested) in the operation stage of investment projects, are considerably lower compared to those generated by investment action.

The contrary is though observed in the case of measures 123, 312 and 313, which target the non-farm economy; in this case, job generation efficiency is much higher in the stage of operation than in the stage of investment. The above mentioned characteristic is repeated in the case of counterfactual analysis, though estimates are much lower.

Table 17. Impact Analysis based on data from the econometric counterfactual analysis, Austria (average annual effects compared to 2009) (in EUR million)

Type of Effect	Change in output	% change	Change in GVA	% change	Change in Employ-ment (FTEs)	% change
a) Capacity-Adjustment Effects - Counterfactual Analysis						
Measure 121	184.02	0.029	52.16	0.021	3,078	0.089
Measure 122	-18.81	-0.003	-7.07	-0.003	-128	-0.004
Measure 123	-4.97	-0.001	-1.28	-0.001	-28	-0.001
Measure 125	3.28	0.001	1.23	0.000	22	0.001
Measure 311	13.91	0.002	4.03	0.002	65	0.002
TOTAL	**177.44**	**0.028**	**49.08**	**0.020**	**3,011**	**0.087**
b) Difference (Counterfactual Analysis - National Data)						
Measure 121	-111.87	-0.017	-31.71	-0.013	-1,871	-0.054
Measure 122	-33.44	-0.005	-12.57	-0.005	-227	-0.007
Measure 123	-805.07	-0.126	-206.94	-0.083	-4,475	-0.129
Measure 125	-36.45	-0.006	-13.70	-0.005	-248	-0.007
Measure 311	-41.70	-0.007	-12.07	-0.005	-196	-0.006
TOTAL	**-1,028.52**	**-0.160**	**-276.99**	**-0.111**	**-7,017**	**-0.203**

Source: Psaltopoulos (2014)

Table 18. New Jobs Generated per million EUR of Investment, Austria (FTEs)

Type of Effect	National Data	Counterfactual Analysis Data
a) Investment Effects		
Measure 121	18.52	18.52
Measure 122	16.27	16.27
Measure 123	19.02	19.02
Measure 125	18.20	18.20
Measure 311	19.30	19.30
Measure 312	18.36	
Measure 313	22.03	
TOTAL	**18.58**	
b) Capacity-Adjustment Effects		
Measure 121	13.44	8.36
Measure 122	9.18	-11.80
Measure 123	73.87	-0.46
Measure 125	15.16	1.25
Measure 311	19.72	4.93
Measure 312	50.59	
Measure 313	35.64	
TOTAL	**21.66**	
c) Total Effects		
Measure 121	31.95	26.87
Measure 122	25.45	4.47
Measure 123	92.89	18.56
Measure 125	33.35	19.45
Measure 311	39.02	24.23
Measure 312	68.95	
Measure 313	57.67	
TOTAL	**40.24**	

Source: Psaltopoulos (2014)

Assessment

In general, the **quality of results** is considered as very satisfactory, mainly due to the fact that they are based on a satisfactorily comprehensive data set and also because they indicate a rather clear pattern of interpretation.

In terms of **representativeness**, it is rather obvious that a lot depends on economic structures. Sectors which are comparatively “modern” are associated with a higher efficiency of job creation and vice-versa. However, such a finding is not repeated in case studies associated with a lower development context.

Problems encountered in the Austrian case study are associated with the base of the IO table (2009, instead of the ideal 2007) and the availability of additional GVA data up to 2012. In general these problems were minimal compared to those encountered in other case studies, and hence, no solutions were pursued (e.g. to estimate annual increments of GVA, total estimates were simply divided by 6 rather than by 7).

4.2 Econometric counterfactual method in case study Austria

Context and scope

This case study covers the entire country (programme area) and is focused on measure 121 which is, from a financial point of view, the most important RDP single investment-type measure in Austria (the public funds allocation in 2007-2013 to measure 121 amounted to EUR 467.5 million)

As in other Member States, a counterfactual analysis of effects of RDP support under measure 121 was greatly facilitated by a good availability of individual farm data in the form of national farm bookkeeping records or/and FADN data (and this is valid also for other investment measures). The bookkeeping/FADN records could be relatively easily combined with information on RDP beneficiaries available from a national Paying Agency.

The Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Managements agreed to provide access to micro-economic farm data necessary for carrying out counterfactual analysis within this study. Available anonymous data combined farm bookkeeping data (approx 1400 farms over the period 2006 – 2012; panel data) with data from PA on programme beneficiaries (including the level of obtained support from individual RDP measures). The bookkeeping data set of Austria contains many variables that describe the characteristics and performance of an individual farm household in a detailed manner. As in principle micro-data is heavily protected, necessary calculations of programme effects using counterfactual methodologies were carried out in the premises of the Austrian Institute for Economic Research (WIFO) which regularly undertakes various analytical tasks involving FADN data.

Important characteristics of the above dataset which can be used for evaluations of RDP are: i) rich information about the structure and economic performance of farms observable over longer period of time (2007-2012) and prior to this period, and ii) clear distinction of farms which received or not received programme support under measure 121 (and other measures).

Assessment of programme effects involving counterfactual analysis was based on application of two different counterfactual methodologies:

a) Traditional DiD method: i.e. comparisons of performance of programme beneficiaries (measure 121) with all or arbitrary selected programme non-beneficiaries, and
b) Combination of DiD with a binary Propensity Score Matching (conditional PSM): i.e. comparison of performance of programme beneficiaries (measure 121) with a control group consisting of matched programme non-beneficiaries.

An important element of this assessment was a comparison of results obtained by applying the above methods with a naïve "before – after" approach which is often used in other empirical evaluation studies[39]

The key in the counterfactual analysis was to construct a control group which is as similar as possible (in observable and unobservable dimensions) to the recipients of the intervention (here it is measure 121). If similarity of both groups is verified (e.g. through carrying out matching analysis supported by respective statistical tests), one can use a group of programme non-beneficiaries as a counterfactual to assess the magnitude of real effects of the programme (measure 121). Generally, comparison based on matching allows for the establishment of causality, i.e. attributing observed changes in outcomes to the programme measure 121, while removing confounding factors.

Estimation of various effects of the given programme measure 121 was carried out using GVA as the main result indicator.

Results based on PSM matching methodologies are characterised by a very high degree of rigour and validity, and a high degree of practicability in comparison with other counterfactual approaches (e.g. with traditional DiD).

Findings

Counterfactual analysis: measure 121

Data used: farm bookkeeping data combined with data from Paying Agency

Place of data compilation: WIFO

Number of observations: Panel of 1393 units/year (2006 and 2012) of which:

- beneficiaries: 546 farms (39%)
- non-beneficiaries: 847 farms (61%)

List of variables used for establishing a control group (2006):

- Gross value added per farm
- Own agricultural area per farm
- Total labour per farm, Family labour per farm
- Cultivated area under grains per farm
- Livestock intensity (per 100 Ha)
- Production of wheat per farm
- Value of assets per farm, assets buildings per farm, assets machinery per farm
- Equity per farm
- Turnover per farm
- Consumption expenditures per farm
- Total level of support received in years 2003-2006 per farm

[39] Sometimes, before-after approach is also referred as to a specific form of "counterfactual". Yet, this assumes that without a programme (counterfactual) beneficiaries' outcomes would be as in a base period (i.e. before the programme) which is in most cases not defendable.

Effectiveness of measure 121 was measured using the following approaches:

- Outcomes achieved by programme beneficiaries were compared to target values ⇔ (equivalent to) an increase of a given result indicator due to a given programme measure compared with target values.
- Outcomes achieved by programme beneficiaries were compared to outcomes achieved by programme non-beneficiaries (in %) ⇔ (equivalent to) an increase of a given result indicator for programme beneficiaries compared to an increase of the same result indicator for the control group.
- Structure of a total increase of a given result indicator was analysed (% share due to a given measure compared to % share due to other factors).

Efficiency was calculated as a ratio of result/impact indicator to obtained level of support in EUR (both at micro and programme area levels).

Table 19. Austria: Effect of M121 on GVA (EUR)

Result indicator	Beneficiaries	Control group	Difference
GVA		**Year 2006**	
Unmatched	41,277	26,604	14,673
Matched (ATT)	40,266	40,193	72
		Year 2012	
Unmatched	56,464	33,450	23,013
Matched (ATT)	55,108	47,581	7,526

- Real effect of measure 121:
  =7,526–72 = **7,454 EUR** in 6 years = **1,242 EUR**/year
- Naïve approach (before-after)
  = 56,464 – 41,277 = 15,187 EUR in 6 years = 2,531 EUR/year (**+104% of a real effect**) = **very strong bias (!!!)**
- **In comparison to naïve approach** (before-after) = 100 => real effect = **49%**
- Estimated factual increase of gross value added due to M121: = (40,266 + 7,454)/40,266 = +18.5% (total over years 2007-2013)
- Targeted value (GVA): + 8,000 EUR/farm (over years 2007-2013), i.e. in comparison with target value = 7,454/8,000 = 93%
- Estimated structure of a total increase of gross value added due to M121:
  = M121 (7,454/14,842 = 50%) + other factors (7,388/14,842 = 50%)
- Efficiency of farm support (measure 121) in Austria (at farm level):
  - Average support per farm (M121) (using bookkeeping data) =20,139 EUR (in 6 years)
  - Efficiency of measure 121: =7,454/20,139 = 0.37 ⇔ 1 EUR brought 37 cents of effects (Gross value added)

Impact of farm support (measure 121) in Austria (at programme area level):

- Estimated gross impact at a country level (on gross value added of farms) =1,242 EUR/year *27,447 farms which received support from M121 (until end of 2013) = 34 mill EUR/year * 6 years = 204 million EUR
- Total public costs (until end 2012): 467.5 million EUR

Efficiency of measure 121 at programme area level: 43.6% (= 204/467.5) <=> 1 EUR brought about 43.6 cents of effects (GVA)

Table 20. Austria: Effect of M121 on total farm income (in EUR)

Result indicator	Beneficiaries	Control group	Difference
Farm Income	**Year 2006**		
Unmatched	54,389	45,577	8,812
Matched (ATT)	53,963	53,952	10
	Year 2012		
Unmatched	67,428	55,097	12,331
Matched (ATT)	66,685	63,350	3,335

- Real effect of measure 121:
  =3,335 – 10 = 3,325 **EUR** in 6 years = **554 EUR**/year
- Naïve approach (before-after) = 6,7428 – 54,389 = 13,039 EUR = 2,173 EUR/year (very strong bias = +292% of a real effect !!!)
- Estimated factual increase of total farm income due to M-121: = (53,963+3,325)/53,963 = +6% (in total over years 2006-2012)
- Estimated structure of a total increase of total farm income due to M121:
  = M121 (3,325/12,722 = 26%) + other factors (9,397/12,722 = 74%)
- Efficiency of farm support (measure 121) in Austria (at farm level):
  Average support per farm (M121) (using bookkeeping data) =20,139 EUR (in 6 years)
  Efficiency of measure 121: =3,325/20,139 = 0.16 ⇔ 1 EUR brought about 16 cents of effects (total farm income)
- Estimated gross impact at a country level (on total farm income) =554 EUR/year *27,447 farms which received support from M121 (until end of 2013) = 15.2 million EUR/year * 6 years = 91.2 million EUR
- Total public costs (until end 2012): 467.5 million EUR

Estimated efficiency of M121 at the programme area level: 19.5% (=91.2/467.5) ⇔ 1 EUR brought about 19.5 cents of effects (increase of total farm income)

Table 21. Austria: Effect of M121 on total farm employment (in AWU)

Result indicator	Beneficiaries	Control group	Difference
AWU	**Year 2006**		
Unmatched	1.861	1.478	0.38
Matched (ATT)	1.841	1.837	0.00359
	Year 2012		
Unmatched	1.864	1.409	0.4542
Matched (ATT)	1.838	1.757	0.0812

- Real effect of measure 121:
  =0.0812 – 0.00359 = 0.078 **AWU** in 6 years
- Naïve approach (before-after) = 1.864 – 1.861 = 0.003 AWU in 6 years (very strong bias = 3.8% of real effect !!!)
- Estimated factual increase of total farm employment due to M-121: = (1.841 + 0.078)/1.841 = +4%
- Estimated structure of a total increase of total farm employment (AWU) due to M121:
  = M121 (+0.078) + other factors (-0.081)
- Efficiency of farm support (measure 121) in Austria (at farm level):
  Average support per farm (M121) (using bookkeeping data) =20,139 EUR (in 6 years)
  Efficiency of measure 121: =20,139/0.078/6 years= 43,032 EUR ⇔ 43,032 EUR/year was necessary to increase level of farm employment by 1 AWU
- Estimated gross impact at country level (on total farm income) =0.078 *27,447 farms which received support from M121 (until end of 2013) = 2,141 AWU
- Total public costs (until end 2012): 467.5 million EUR

Estimated efficiency of M121 at the programme area level: 218,355 EUR/1 AWU (=467,500,000/2,141) ⇔ 218,355 EUR was necessary to increase farm employment by 1 AWU

Table 22. Austria: Effect of M121 on farm labour productivity (in EUR/fully employed)

Result indicator	Beneficiaries	Control group	Difference
GVA/AWU	**Year 2006**		
Unmatched	22,314	17,474	4,839
Matched (ATT)	22,155	21,724	430
	Year 2012		
Unmatched	30,495	25,193	5,302
Matched (ATT)	30,369	28,016	2,352

- Real effect of measure 121:
  =2,352 – 430= 1,922 **EUR** in 6 years= 320 EUR/year
- Naïve approach (before-after) = 30,495 – 22,314 = 8,181 EUR = 1,363 EUR/year (very strong bias = +326% of real effects !!!)
- Estimated factual increase of total farm labour productivity due to M121: = (22,155 + 1,922)/22,155 = +8.7%
- Estimated structure of a total increase of total farm labour productivity due to M121:
  = M121 (1,922/8,214 = 23%) + other factors (6,292/8,214 = 77%)
- Efficiency of farm support (measure 121) in Austria (at farm level):
  Average support per farm (M121) (using bookkeeping data) =20,139 EUR (in 6 years)
  Efficiency of measure 121: =1,922/20,139=0.09 EUR ⇔ 1 EUR of support brought about 9 Cents of an increase of farm labour productivity
- Estimated gross impact at a country level (on total farm labour productivity) =1,922 *27,447 farms which received support from M121 (until end of 2013) = 52.7 million EUR
- Total public costs (until end 2012): 467.5 million EUR

Estimated efficiency of M121 at the programme area level: 11.3% (=52.7/467.5) ⇔ 1 EUR brought about 11.3 cents of effects (increase of labour productivity).

Further analysis of effectiveness regarding investment support under M121: Internal return of investment

Application of counterfactual analysis allows estimating an effectiveness of total investment on farms which benefitted from M121 (542 farms) in comparison with farms which were non-beneficiaries of M121 (846 farms). In all farms above total gross investment during period 2007-2013 was positive, i.e. all farms have invested.

1. Cost of *corrected* total gross investment (for beneficiaries of M121) was calculated as a difference between total gross investment and the value of support received under M121.
2. Investment rate of return for non-beneficiaries of M121 was computed as a ratio of an increase of GVA (dGVA) and total gross investment, whereas investment rate of return for beneficiaries of M121 was computed as a ratio of an increase of GVA (dGVA) and corrected total gross investment.
3. The analysis shows that the calculated rate of investments for:
   a. Non-beneficiaries of M121 equalled 0.0413, whereas
   b. Beneficiaries of M121 equalled 0.0735, i.e. it was by 0.0322 higher

   Interpretation: Investment support under M121 had a significant influence on the reduction of costs of total investments undertaken by an average farm. The analysis shows that a return of total investment on a beneficiary farm could be reduced from 24 years (or 0.0413) to 13 years (0.0735), i.e. by 11 years.

Assessment: The main conclusions for Austria (M121)

EQ1 (appropriateness of methods):

Among the selected three counterfactual approaches (including a naïve “before and after” approach) the conditional PSM method (combination of a binary PSM and DiD) appeared to be the most rigorous, transparent and reliable method with a high degree of validity and practicability.

Conditional PSM method performed also very well in terms of its ability to **reduce** original selection bias occurring while making comparisons between programme beneficiaries and non-beneficiaries (in this case study the selection bias was reduced by 95% !)

The before-after approach generated results which were highly biased (they differed from results obtained by applying a conditional PSM approach by a magnitude of factor 2-25 (!), i.e. results from “before-after” method were found to be overstated by as much as by +326% (see: effect on labour productivity) (!) or understated by a factor of 25 (!) (see: effect on farm employment).

EQ2 (efficiency, effectiveness, impact of investment support):

The study shows that effectiveness of measure 121 in Austria measured by comparing its real effect on GVA with a target value was relatively high, i.e. although due to M121 gross value added at a farm level increased over the period 2007-2012 in comparison to year 2006 by only +18.5% this increase was close to its targeted value set on 19.8% which resulted in effectiveness equalled to 93%.

However, in the case of other result indicators (e.g. employment, labour productivity, etc.) similar comparisons could not be carried out as no target values were provided (programme documents were incomplete in this respect).

Regarding the structure of an increase of result indicators in the period 2007-2012: 50% of a total increase of gross value added at farm level was due to measure 121 (another 50% due to other factors). Concerning other result indicators, it was found that the M121 contribution to a total increase of farm income was only 26%, and to an increase of labour productivity only 23%. On the other hand, it was found that M121 contributed to a slight increase of farm employment (other factors would reduce farm employment).

Concerning the efficiency of M121 measure in Austria it was found that the respective ratios at a farm level were very low. For example, in the case of GVA the efficiency ratio was 0.37. This means that that 1 EUR support from public funds on measure 121 allocated to programme beneficiaries at a farm level resulted in only 37 cents of increase of gross value added. For other indicators the ratios were even lower: 16 cents (total farm income) and 9 cents (labour productivity). Concerning employment, as much as 218,355 EUR was necessary to increase farm employment by 1 AWU.

The gross impact (without indirect effects) on M121 on GVA at the programme area level was also low. At country level total public costs spent on measure 121 (467.5 million EUR) resulted in an increase of gross value added by +204 million EUR (efficiency = 43.6%), total farm income by +91.2 million EUR (efficiency = 19.5%), farm labour productivity by +52.7 million (efficiency = 11.2%). In terms of its effect on farm employment, M121 resulted in an increase of the country's farm employment by +2141 AWU.

An estimated efficiency of total public costs spent on measure 121 (over 6 years) **at country level** on GVA was found to be slightly higher than at farm level, and amounted to 0.43 vs. 0.37, i.e. 1 EUR expenditure on M121 generated 43 cents of GVA at country level and only 37 cents at farm level). One possible interpretation of this result could be a slightly higher amount of M121 support received by farms included in the bookkeeping database (20,139 EUR/farm) compared with a national average (17,033 EUR/farm), while the effectiveness in both groups was assumed to be the same.

Investment support under M121 had a significant influence on a reduction of **costs** of total investments undertaken by an average farm. Our analysis shows that a return of total investment on a beneficiary farm could be reduced from 24 years (or 0.0413) to 13 years (0.0735), i.e. by 11 years.

The above results can be considered as representative for the whole country (on average), yet at a farm level results may be slightly different especially when the analysis is carried out for various specific types and classes of farms (conditional PSM method can also be applied to answer Evaluation Questions regarding differences between farms). However, estimated results have to be considered as "gross", i.e. without taking into consideration programme indirect effects (including substitution, displacement, etc.).

4.3 MAPP Method in case study UK/Scotland

Context and scope

The MAPP focus group was conducted in the North-East of Scotland. This area of the programme territory was selected because it is one of the administrative regions within the quasi-regionalised Scottish RDP, it has a very mixed agriculture and rural economy providing examples of all types of RDP investments and because it has a distinct area focused around one city. In addition, this has been a very active region in the 2007-2013 RDP which absorbed the largest quantity of RDP funds including investment support.

The analysis focused on six of the most important RDP investment schemes. These included: farm investments, diversification, forestry, food processing, agri-environment and Leader. The influence of external factors on each impact indicator was also analysed. Impact indicators chosen comprise: jobs in RDP assisted agriculture/forestry, jobs in non-agricultural sector RDP assisted, jobs in non-RDP assisted agriculture/forestry, jobs in non-agricultural sector non-RDP assisted, farm incomes in RDP assisted farms, farm incomes in non-RDP assisted farms, competitiveness, rural tourism, energy/water efficiency and quality, biodiversity area and quality of the environment.

Findings

Findings are presented for each tool and then summarised.

The **first tool (Life curve)** set the overall context and helped identify three key development phases: a) the 2007-2009 period, which started off with good prices for grain, cattle and sheep, but then the 2008 financial crisis brought difficulties to farmers to obtain credit and those with cash-flow problems suffered the most. Due to high oil prices, Aberdeen boomed and this caused people to move out of rural areas into the city, with negative consequences for instance on the dairy sector that was faced with a shortage of workers to milk cows; b) the 2009-2012 period was characterised by rising sales abroad (e.g. whisky to Asia) and rising demand for forestry products and rural tourism. As a consequence, prices rose and with them the confidence to the Scottish rural economy; c) the 2012-2013 period was characterised initially by a bad winter and falling yields, coupled with high input prices, but followed in 2013 with the uptake of profitable renewable energy options and thus the creation of a major new income source.

The **second tool (trend analysis)** assessed the performance of the chosen impact indicators over the programming period, with the following findings for employment, incomes, competitiveness and environmental indicators.

Agricultural employment (including farmers, working partners and employees) is continuing its long standing slow, but steady decline. At best the RDP investment measures have helped to stabilise numbers in assisted businesses, though often an RDP investment has helped farmers reduce their labour requirement. The only positive labour trend resulting from RDP support is in non agricultural jobs – the input supply trades, construction, primary and secondary food processing businesses, diversified businesses. They have directly benefited from RDP investments (e.g. builders) or indirectly from an increase in output or quality (e.g. meat processors).

The underlying trend in farm incomes has been mildly positive over the period (in consistence with the life curve which displayed the increase in sales and demand and subsequently incomes) and has only been marginally boosted by RDP supported investments.

The trend in competitiveness was positive, partly due to changes farmers have made following a shift away from headage and area payments, but was affected a little by RDP investments toward the end of the period. Investments were on balance investments well planned and improved output per unit of labour.

Trends in energy efficiency, water efficiency and quality, biodiversity area and quality of the wider environment were all assessed as positive over the programming period.Water efficiency is not a big issue in Scotland where there is no lack of supply, but water quality is very important. Most of lowland North-East Scotland is designated a Nitrate Vulnerable Zone (NVZ) and two rivers have had eutrophication problems. Of all environmental indicators water quality was felt to have seen the largest improvement.

The **third tool (influence matrix)** gave a good indication of which RDP schemes (and hence measures) had the most influence on farming and rural areas (see figure 15). The totals for each indicator showed where the RDP had most impact. Ratings for the external drivers were fairly similar across all indicators (except the last two) so this did not skew the conclusion on impact of the RDP schemes.

Table 23. Influence matrix from North-East Scotland case study

	Farm investment	Diversification	Forestry	Agri env	Food processing	Leader	External	
Jobs in Ag/forestry RDP assisted	3	1	1	0.5	3	0	4	**12.5**
Non Ag/ forest jobs RDP assisted	4	2	2	3	3.5	2	4	**20.5**
Jobs in Ag/forestry non RDP assisted	1.5	0.5	0.5	0	0.5	0	4	**7**
Non Ag/forest jobs non RDP assisted	2	1	1	0	1.5	1	4	**10.5**
Farm incomes RDP Assisted	4	2	1	3	2	0	5	**17**
Farm incomes non RDP Assisted	0	0	0	0	2	0	5	**7**
Competitiveness	3	1.5	0	2	3		5	**14.5**
Rural tourism activity	0	2	0	1	0	1	4	**8**
Energy, water efficiency, quality	2	0	1.5	2	0	0	4	**9.5**
Biodiversity area	0	0	3	3	0	0	2	**8**
Quality of environment	2	0	2	2	0	1	2	**9**
	21.5	**10**	**12**	**16.5**	**15.5**	**5**	**43**	
0 = None 5 = A lot								

According to this tool and discussions on it, the interventions with the highest impact comprise:

- The farm investment schemes (mainly measure 121 under what was called the "Rural Priorities" scheme) were felt to have had the most impact i.e. the most success in delivering the aims of the programme as defined by the listed indicators. This was mainly on non-farm jobs, farm incomes of assisted businesses and competitiveness.

- Second most important were the agri-environment schemes closely followed by food processing sector grants. Agri-environment had the expected high scores for water quality and biodiversity, but also impacted on non-farm jobs (fencing, tourism) and farm incomes (ongoing management payments and the margins made on investment work done by the farmers themselves). Food processing support had an important effect on jobs on and off the farm and on incomes and long-term competitiveness of entire sectors. The fragility of the small Scottish primary processing sector was a concern, so anything which strengthens it gets a high rating.
- Leader received the lowest influence rating for these indicators.

It should be noted that the focus group rated external factors (also seen in the life curve) much more highly than any RDP intervention.

The indicators that gained most from RDP support, according to the influence matrix, comprise:

- Non-farm jobs/non primary sector employment benefited most from the RDP i.e. jobs in the building and farm service trades, food sector, tourism, local services through Leader, and environment.
- The next biggest gain from the RDP investments was in competitiveness. This benefit came as expected from support for farm investments, but also from supported investments in the food industry and even agri-environment investment (which created new income streams and new infrastructure – fences, shelter, water supplies).
- Farm and land based sector jobs were also a major beneficiary, especially once again via new farm and forestry level investments and protection/enhancement of marketing chains through food sector support.
- Unsurprisingly, the least benefit from the RDP was in jobs and incomes in unassisted businesses, though of course there was some benefit for these from the wider impacts of rural investment, processing sector, etc.

The **fourth tool (development and impact profile)** summarised and explained the findings of the previous tools.

There are positive trends in non-farm jobs, and the RDP investment support has had an important impact there, especially long-term jobs in the food sector as seen above.

There is also a good positive trend in farm incomes with the RDP having a clear extra benefit for those using the support.

The RDP investment support has helped at best to maintain farm jobs among beneficiaries while they have declined in non-participating businesses. Farm investments often reduce labour demand, but they also allow expansion in output, which maintains labour, and sometimes allow introduction of a new niche farm enterprise. This has been the case in North-East Scotland with some very large investments in welfare friendly egg production units.

There is a strong positive trend in competitiveness of the primary industries. The views of the group suggest that this is definitely being enhanced by investments supported by the RDP, but external drivers – the change in the Pillar 1 subsidy regime and rising labour and energy costs – have a major effect.

Water quality and energy efficiency have shown positive trends, but the RDP has had a small effect. Regulations such as NVZs and other support regimes such as Feed-In-Tariffs are stronger drivers.

Likewise rural tourism has shown positive trends over the 2007 to 2013 period, but the impacts of RDP investments, while positive, are small. The strength of the local economy has helped most.

Assessment

The results of the focus group can be considered of **good quality** since all the tools worked well. The Life Curve proved to be a good participatory tool that got everyone involved and covered topics known by all. The trend analysis tool was used in a simplified way by splitting the programming period in two phases, rather than going through it year by year. The influence matrix was the core of the session and provided useful information on the identification of the strongest and weakest influences. High quality feedback was obtained in particular when totalling the scores in the influence matrix. This enabled to identify the schemes/measures that had most impact and the indicators that had been most successfully delivered by the RDP. The last tool (development and impact profile) proved to be useful mostly for the facilitators to summarise the results of the focus group rather than for the participants themselves. Its major value was helping to draw conclusions from the results of the MAPP focus group.

Problems encountered included organisational and implementation issues. More specifically:

Organisational issues and their solutions:

- Knowing who to invite. It is difficult to find the right participants especially when many relevant ones are busy in their farms. The solution to this is to invite people much in advance (about one month) and then closer to the date (about 10-15 days) identify gaps and invite additional ones.
- Availability of data on programme beneficiaries. Data on “who received what” in terms of RDP investment support is not always available from Managing Authorities, therefore local knowledge is important.
- Difficulties to attend. Clashes with other meetings or weather conditions affect the capacity of invitees to participate. The solution adopted above (early preparation and invitation with a follow-up before the focus group) can help reduce the drop-out rate.

Implementation issues and their solutions:

- Group fatigue. This was the result of having to work through too many permutations with several indicators and years; this was especially the case with the trend analysis table when working through employment trends over seven years for four indicators (agricultural jobs in assisted businesses, non-agricultural jobs due to assisted businesses, agricultural jobs in non-assisted businesses, non-agricultural jobs in non-assisted businesses), implying 28 different decisions on employment effects. To get over this problem the years were grouped into two blocks: 2007 to 2010 and 2010 to 2013.
- Confusing concepts. There was some confusion with the measures, with the trends and with the indicators. Concerning the measures, it is common that

participants do not know which measure they benefited from. An explanation at the beginning is therefore useful while using other terms more familiar to people also helped (e.g. in this case “investment schemes”). In relation to trends, there is a difference between overall trends and the influence of the RDP. Again, a clear explanation at the outset is important for clarifying what each tool is looking at (e.g. the life curve and the trend analysis look at overall trends, while the influence matrix looks at the influence of the RDP measures and other factors on each indicator). In this way, the analysis entails the value of assessing what changes are due to the RDP and which ones are not. Finally, some indicators are not well understood, for example biodiversity, quality of the environment, competitiveness. It proved very important to explain them at the beginning of the session especially by giving real examples.

- Handling the MAPP tools. Physically handling the tables is difficult when they include several years and several indicators. The solution to this can be the use of an electronic whiteboard. In addition, the use of two facilitators is paramount for the smooth implementation of the focus group, since one is devoted to facilitating and the other to documenting.
- Achieving a balance between the scores and the explanations. The influence matrix (see figure above) can be quite laborious to fill in. It proved better to ask participants which schemes they considered had the most and the least impact on which indicators. This gave the starting point for filling in each row of the table. A key point after completion of each row is the discussion to draw out reasons for high or low figures. What can be more important than the actual figures/scores in the MAPP tools is the discussion around the rationale for these scores. Capturing these points can help explain the influence of the RDP measures and other programmes/factors on the impact indicators.
- Data availability. Data is not always available on all indicators. The reason why official data is useful is for assessing the extent to which practical experiences match the official data as well as for explaining the official data through practical evidence. In this case, official data was extracted on Scottish farm incomes and employment impacts of the RDP from monitoring data. The MAPP focus group scores and views closely matched the official data. This can be a useful way to implement the triangulation approach in validating findings from different sources.

4.4 Programme-theory-based method in case study Czech Republic

Context and scope

The case study presented demonstrates the application of the TBE approach in the case study territory Czech Republic. The case study clearly demonstrates how the TBE is used in order to assess the effectiveness of measures related to the investment types A (productive, M121, M123 and M311) and C (non-productive, M227 and M313). The TBE approach taken presents a simplified format (structural model) which cannot in practice replace a comprehensive narrative or “performance story”. The study concludes with a final judgement of the experts on the effectiveness to achieve the intended change with the use of an “assessment profile”. In a real life situation the expert’s judgement would be discussed and – if necessary adjusted – during a “validation workshop” with stakeholders.

The workflow and findings for the assessment of investment support to enhance competitiveness of agricultural holdings (type A) is summarised below. Related to investment type A (productive) a bundle of measures of the RDP of the Czech Republic for the 2007-2013[40] period is considered: M121- Modernisation of Agricultural Holdings, M123 – Adding value to agricultural and food products and M311- Diversification into non-agricultural production.

This bundle of investment support measures responds to a strategic objective to improve the competitiveness of Czech agriculture.

They should support the modernisation and restructuring of agricultural holdings where there is an inadequate level of investments, into both, structures as well as technologies, in crop as well as animal production. Obsolete farm technologies and poor conditions of farm buildings are deemed to hamper productivity growth, to provide insufficient level of animal welfare and to cause environmental damage. Encouraging investment in crop storages, silos, animal housing, manure and slurry storages, biogas stations, farm equipment like milking units, postharvest treatment, biomass pellet processing and food processing should not only address the above particular problems, but it should lead to restructuring of farm enterprises and productions in order to enhance efficiency of farm resources and to add value to farm products and hence improve the overall competitiveness of agricultural production/holdings. In spite of their separate implementation, the bundle of measures is considered as complementary in terms of the objective to improve resource efficiency and competitiveness of the farming sector.

The analysed group of investment measures (M121, M123 and M311) represents about 17% of the RDP budget for Axes 1 to 4. However, only slightly less than ¾ of the budget of these measures is used for investment on farm holdings closely linked to agricultural production; the rest goes to food industry, forestry or tourism; in the case of M123 the uptake of farms accounts only for 10% of the measure budget. Renovation and construction of tourist facilities (even on farms) is not considered (about 30% of the measure budget). M121 has the highest– almost ¾ of the selected group (relevant) budget. The measures were implemented as single measures (no formal link) by the State Agricultural Intervention Fund (SZIF). About half of the submitted projects were approved and financed from the RDP budget.

The following **data sources** were used to carry out the TBE assessment:

- Regulation Fund for Energies (2014): Annual Report on the Czech Electric Power Distribution Network in 2013 (Roční zpráva o provozu ES ČR 2013) http://www.eru.cz/documents/10540/462820/Rocni_zprava_provoz_ES_2013.pdf/20c3f587-a658-49f7-ace9-56be8a66b7b9
- Ministry of Agriculture of the Czech Republic (2014): Report on the State of Czech Agriculture in 2013 (pre-published version)
- MA (2010): Mid-term Evaluation Report of the RDP for the Czech Republic for the period 2007-2013. Ministry of Agriculture of the Czech Republic, DHV, TiMa.
- MA (2013): Annual Report on the Implementation of the Rural Development Programme for the Czech Republic for the period 2007-2013. Ministry of Agriculture of the Czech Republic

[40] http://enrd.ec.europa.eu/enrd-static/country/czechrepublic/en/czechrepublic_en.html

- Medonos, T., Ratinger, T., Hruška, M., Špička, J. (2012): The assessment of the effects of investment support measures of the Rural Development Programmes: the case of the Czech Republic. The journal AGRIS on-line Papers in Economics and Informatics (http://online.agris.cz), Vol. IV, No. 4. 35-47.
- Ratinger, T., Medonos, T., Hruška, M. (2013): An Assessment of the Differentiated Effects of the Investment Support to Agricultural Modernisation: the case of the Czech Republic. The journal AGRIS on-line Papers in Economics and Informatics (http://online.agris.cz), Vol. V, No. 4. 153-164.

Findings

As a first result a **conceptual model** according to the RDP intervention logic was constructed (see the table below). The conceptual model consists of a brief description of territorial needs (taking into account SWOT analysis and needs assessment), related changes to be achieved at end of the programme (specific or overall objectives) related to the identified needs and planned activities (and target groups) per measure. The completeness of the conceptual model very much depends on the given information in the RDP (and accompanying interviews if necessary) which have to be translated into a logical structure by the expert.

Table 24. Conceptual model for the bundle of measures M121, M123, M311

Territorial needs related to the specific investment support measure under the competitiveness objective A need for upgrading technologies and innovations in order to increase competitiveness of agriculture and to improve its sustainability and for restructuring and diversifying activities - on farms as well as in the sector in order to respond to new market opportunities.		
	Intended change related to the identified needs (1a)	**Planned activities through which the intended change should be achieved**
Support to on farm investments (M121, M123, M311)	1. Better use of production factors, both on farms and in the sector (some relevance of LFA /Less Favoured Areas)	Investments in animal housing, feed storages, manure storages and production technologies (better use of labour) - M121, investment in renewable energy (processing wastes, better use of land) - M121 and M311
	2. Enhanced marketing, improved revenue on farms (some relevance of LFA)	Investments in storage capacities (reconstruction or new) for plant products incl. FFV /fresh fruits and vegetables - M121. Investments in food processing (on-farm, but also food industry) – M123, investment in renewable energy (new markets: pellets, electric energy) – M121 and M311
	3. Improved competitiveness, (lower costs) of farms and of the sector	Investment in agricultural buildings (storages, housing), supports for perennial crops, restructuring: i.e. giving up unprofitable productions, replacing them by profitable productions
	4. Enhanced animal welfare – on farms and in the sector	Investments in animal production technologies – the support applies to cattle, pigs, sheep, goats, horses and poultry – M121
	5. Reduction of emissions in water and air – particularly in NVZ/Nitrate Vulnerable Zones	Investment in animal housing and equipment including manure and slurry storages – M121
	6. Contribution to renewable energy production – the sector level	Investments in technologies for the processing and use of intentionally grown, residual and waste biomass for energy and material purposes – M121 and M311. Investment in biogas power stations M311

In the next analytical step the **implementation of the planned activities** was verified. Detailed monitoring data was exploited to show the implementation in the time period from 2008 to end 2013 (in 2007 no project was actually launched) in terms of inputs and outputs.

Table 25. Implementation of planned activities for the bundle of measures M121, M123, M311

Implemented activities	Inputs (total public expenditure, number of projects, funding rate)	Support by funding body when delivering the funding activity	Characteristics of supported activities Output (total investments, no. of holdings, average size)	Summarise main outputs
Investment in animal housing and equipment in cattle production and other ruminants (M121)	#projects =1,435, public exp. = k€ 130,220; fund. rate up to 40% +10% LFA +10% young f. avg. publ. support = €90,723	No particular support was provided. The applicants had to submit investment proposals at their own risk and cost.	Reconstruction and building of cow sheds, milking units, manure and slurry storages/processing, winter housing for beef cattle and sheep #projects =1,435, of it 76% in LFA , total inv. = k€ 338,233 avg. investment = €235,644 act. funding rate = 38.5%	An upgrade of dairy and beef cattle technologies in 1,435 projects, 76% of them in LFA. Total investment k€338,233.
Investment in animal housing and equipment in pig and poultry production (M121)	#projects =564, public exp. = k€ 75,668; fund, rate up to 40% +10% LFA +10% young f. avg. publ. support = €134,248		Reconstruction and building up of new housing for pig and poultry with high welfare standards, manure and slurry storages/processing #projects =564, of it 49% in LFA, total inv. = k€ 189,169 avg. investment = €335,619 act. funding rate = 40.0%	An upgrade of pig and poultry technologies in 564 projects, 49% of them in LFA. Total investment k€189,169.
Investment in farm buildings in plant production (M121)	#projects =508, public exp. = k€ 53,240; fund. rate up to 40% +10% LFA +10% young f. avg. publ. support = €104,803		Reconstruction and building up of new facilities for postharvest treatment and storages for field crops and FFV. #projects =508, of it 34% in LFA, total inv. = k€ 138,286, avg. investment = €272,216 act. funding rate = 38.5%	An upgrade of crop storage and postharvest technologies in 508 projects, 34% of them in LFA. Total investment k€138,286.

Implemented activities	Inputs (total public expenditure, number of projects, funding rate)	Support by funding body when delivering the funding activity	Characteristics of supported activities Output (total investments, no. of holdings, average size)	Summarise main outputs
Investment in support and cover constructions for perennial crops (M121)	#projects =155, public exp. = k€ 10,783; fund. rate up to 40% +10% LFA +10% young f. avg. publ. support = €69,570		New constructions to support and protect crops (hops, FFV) #projects =155, of it 10% in LFA , total inv. = k€ 28,009, avg. investment = €180,701 act. funding rate = 38.5%	Built up support and cover constructions for hops and FFV in 155 projects, 10% of them in LFA. Total investment k€28,009.
Investment in biomass processing for energy purposes (M121, M311)	#projects =79 public exp. = k€ 7,230; fund. rate up to 60%, rules according to M121 or M311 avg. publ. support = €91,513		Mainly constructions of new units and equipment for processing biomass pellets and briquettes. #projects =79, of it 44% in LFA total inv. = k€ 14,536, avg. investment = €183,995 act. funding rate = 49.7%	Emerged new capacities for processing energy biomass in 79 projects, 44% of them in LFA. Total investment k€14,536.
Investment in biogas stations (M311)	#projects =137 public exp. = k€ 100,620; fund. rate up to 60% small, up to 50% medium entr. avg. publ. support = €734,451		New constructions of biogas power stations incl. local distribution of heat. #projects =137, of it 30% in LFA total inv. = k€ 306,090, avg. investment = €2,234,231 act. funding rate = 32.9%	Built BGS /biogas power stations in 137 projects, 30% of them in LFA. Total investment k€306,090.
Investment in food processing (M123)	#projects =130 public exp. = k€ 14,450; fund. rate up to 50% SME, up to 25% other. avg. publ. support = €111,152		Reconstruction and new food processing capacities on farms; relatively high participation of wine producing farmers. #projects =130, of it 26% in LFA total inv. = k€ 33,218, avg. investment = €255,521 act. funding rate = 43.5%	New and reconstructed food processing units on farms in 130 projects, 26% of them in LFA. Total investment k€33,218

In a further working step main project outputs at the operations level which have been summarised in the previous step **were transformed into results for the target group** (next level). This step is leading from the projects/operations level to gross results for the target group (e.g. holdings, small enterprises). Since gross direct effects are only rarely collected by the monitoring systems it is necessary to utilise other quantitative sources (e.g. studies) or – in the most limited case – to make qualitative estimates. Since gross results for the target group are influenced by other factors besides RDP support (e.g. changes in the macro-economic context) it is important to indicate other influencing external factors (social, economic, political or administrative factors).

Table 26. Direct gross effects for beneficiaries of implemented operations under the bundle of measures M121, M123, M311

Main outputs on the project/ operations level; indicate any known context factors that affect the achievement of gross results	Direct gross effects for the beneficiaries	Comment on the cause-effect chain between main outputs and direct gross effects
An upgrade of dairy and beef cattle technologies in 1,435 projects, 76% of them in LFA. Total investment k€338,233.	Higher productivity and efficiency → improved competitiveness.	There is a time lag of effects (perhaps 5 years) which include the transition additional invest. and organisational changes.
An upgrade of pig and poultry technologies in 564 projects, 49% of them in LFA. Total investment k€189,169.	The rate of piglets per swine went up, meeting requirements for animal welfare, higher productivity and efficiency → improved competitiveness.	No specific data available. From a case study: cost dropped by 20% to 80%. Daily meat gains improved by 20%.
An upgrade of crop storage and postharvest technologies in 508 projects, 34% of them in LFA. Total investment k€138,286.	Improved marketing operations, improved quality of products, higher revenue (about 5 to 10%). Improved organisation of labour throughout the year.	Storing products in the best way, supplying crop products according to the market opportunities (best price). Energy costs usually rose, but revenue more.
Built up support and cover constructions for hops and FFV in 155 projects, 10% of them in LFA. Total investment k€28,009.	Reducing hail damages (fruits), renovation of support constructions enabling further production (hops, wine) (no output or GVA effects).	
Emerged new capacities for processing energy biomass in 79 projects, 44% of them in LFA. Total investment k€14,536.	Diversification of income.	Rather marginal (perhaps already when designed); limited market for biomass pellets and briquettes.
Built BGS/biogas power stations in 137 projects, 30% of them in LFA. Total investment k€306,090.	Improved and stable income; income from electricity on average 19% of total sales, better utilisation of land, utilisation of wastes.	<20% of sales - 42% of beneficiaries, 20-40% of sales - 40% of beneficiaries over 80% of sales - 2%.
New and reconstructed food processing units on farms in 130 projects, 26% of them in LFA. Total investment k€33,218.	Better access to markets (new markets for products with added value - quality or traditional food) → improved marketing and higher income.	The investment in food processing on farms is rather small. Relatively high participation of wine making farmers (grape growers), some dairy farmers.

Main outputs on the project/ operations level; indicate any known context factors that affect the achievement of gross results	Direct gross effects for the beneficiaries	Comment on the cause-effect chain between main outputs and direct gross effects
External factors		
Deep market decline in 2009-2010 (a consequence of financial crisis).	Sharp drop of revenue and farm income. Labour released from agriculture, drop of investment activity.	The government relaxed conditions for receiving investment support from RDP in order to encourage the adjustment of the farming sector to tougher economic conditions.
Recovery of prices and their further increase since 2011.	Immediate improvement of farm income. Rather marginal effect on production volume.	No effect on the pig sector, i.e. continuing decline. Likely due to high cereal prices.
Gradual increase of direct payments (SAPs).	Income improvement and stabilisation, better access to bank credits. Rising land prices.	In 2013, Czech farmers received the full area payment (€259)
The government launched policy for energy from renewable sources.	Income from biogas electric power. Press on land use, increase of land prices.	The initial price guarantee for biogas electricity was revised/ reduced.

In the concluding phase of the TBE, the **judging phase**, the evaluator drew evidence-based conclusions on the effectiveness of the interventions. The experts judged the contribution of main outputs and identified gross direct effects under a specific measure to the intended change by using an ordinal scale ranging from 1 ‘‘very low’’ to 5 ‘‘very high’’.

For investment type A (productive) all 6 sub-objectives defined in the conceptual model could be assessed. It turned out that the effectiveness is quite high ranging from “medium”, “high” to even “very high”.

To give an example: related to the need for upgrading technologies and innovations the RDP aims, inter alia, at improved competitiveness (lower costs) of farms and of the sector which should be achieved through investments in agricultural buildings (storages, housing), supports for perennial crops and restructuring, i.e. giving up unprofitable productions and replacing them by profitable productions.

The implementation analysis linked to the objective demonstrated an upgrade of dairy and beef cattle technologies in 1,435 projects, 76% of them in LFA. Total investment was k€338,233. Also an upgrade of pig and poultry technologies in 564 projects could be achieved, 49% of them in LFA. Total investment were k€189,169. Production costs dropped by 20% to 80%, and daily meat gains improved by 20% (case study findings). The rate of piglets per swine went up, meeting also requirements for animal welfare,

These achievements contributed to higher productivity and efficiency leading to improved competitiveness (the prior intention), however, only to a “medium” extent. Although competitiveness of the crop and cattle sector improved and the poultry production managed to deal with new requirements for animal welfare, there was no significant investment support in the pig sector which continued to decline.

Table 27. Summary table on the contribution of measures M121, M123, M311 to the intended change

	Intended change which should be achieved by the bundle of measures at end of the programming period	Experts judgement on the contribution of projects supported under the measure to the intended change (insert X)						Justify by main outputs and identified direct gross effects for the beneficiaries taking into account external drivers (which may reduce the contribution)
		don´t know	very low	low	medi-um	high	very high	
			1	2	3	4	5	
Support to on-farm investments (M121, M123, M311)	1. Better use of production factors, both on farms and in the sector (some relevance of LFA)				X			Productivity improved but it was not equivalently followed by the expansion of the production. An excessive expansion of maize production with a negative effect on environment.
	2. Enhanced marketing, improved revenue on farms (some relevance of LFA)					X		It concerns mainly crop production: improved quality of products due to postharvest treatment and proper/advanced storage technologies; more flexibility. Diversification into the renewable energy sector yielded additional income.
	3. Improved competitiveness, (lower costs) of farms and of the sector				X			Competitiveness of the crop and cattle sector improved. The poultry production managed to deal with new requirements for animal welfare. In spite of significant investment support, the pig sector continued to decline.
	4. Enhanced animal welfare – on farms and in the sector						X	Tremendous improvement of animal welfare with new or renovated housing (all animals).
	5. Reduction of emissions in water and air – particularly in NVZ				X			Ammonia emissions dropped by 25% between 2007 and 2012, however, partly due to the decline of animal herds (cattle 3%, pigs 44%).
	6. Contribution to renewable energy production – the sector level				X			From 470 BGS, about 30% were built with the RDP investment support. In 2013 the production of electricity of BGS was 8% of the national production of renewable electric energy and 1% of the consumption of electricity.

Assessment

Quality of results: The following table demonstrates all the "building blocks" which are necessary to establish TBE along the intervention logic. Depending on the available information, these elements can be described more or less well. In the case of investment type A assessment in Czech Republic it was possible for the complete set of building blocks to be well illustrated. Accordingly, **robust** judgements can be made to the extent to which objectives pursued by an intervention are achieved (gross effectiveness).

This assessment could also be made on the detailed level of the 6 sub-objectives defined in the conceptual model. This would allow even more precise statements about the robustness of the TBE.

Table 28. Completeness of TBE "building blocks" for M121, M123, M311

Map out the conceptual model			**Verify the implementation of activities (input and output of operations)**				**Transform project outputs into results for the beneficiaries/target group (next level)**	**Judge on the (gross) effectiveness to achieve the intended change**
1a	1b	1c	2a	2b	3a	3b	4	5
Describe needs	Describe the intended change	Describe related activities, target group	Analyse financial input per activity and characteris-tics	Describe support by funding body (e.g. additional advisory.)	Analyse output of activity and characteris-tics	Summarise main outputs	Analyse direct gross effects of (multi-) activities for the beneficiary based on a plausible cause effect chain	Judge on the extent to which the intended change (1b) could be achieved based on identified main outputs (3b) and - if available - direct gross effects (4)
sufficient (gaps in the RDP)	good	good	good	not applicable	good	good	good	Robust

Representativeness: There is no research on the external validity of TBE findings, i.e. if the results obtained in Czech Republic are the same if applied to a similar intervention in another region.

4.5 Strategic Environmental Assessment / Cost-Effectiveness Analysis in case study Greece

Context and scope

The RDP has an overall budget of 5.6 billion EUR (out of which 544 million EUR are allocated to measure 125) and covers the entire territory of Greece. Greece is located in Southeast Europe and covers a territory of 13,196,887 Ha, out of which 40.9% are agricultural areas. The primary sector has a special social and environmental role in the overall Greek economy and provides a significant proportion of employment.

Sub-measure M125A foresees two broad categories of operations (which in practice might succeed each other and can be combined with actions of Priority Axis 2, M214, and M226), namely:

A. Operations targeting water storage from surface run-offs in order to reduce underground water use and eliminate the need for water drillings. Within this category of actions there are projects for dams in rivers (or diversions) that form artificial lakes, and water reservoirs that are fed from rain and springs.
B. Operations targeting water savings from the main water distribution network before the water reaches the farms. Such operations include maintenance of channels and ditches to reduce water losses from evaporation and infiltration or the installation of new underground distribution networks with pipes and water under pressure.

Greece has overall sufficient quantity of high quality water resources, but also one can identify significant problems with the use of the water resources, mainly due to uneven distribution of those resources in space and time and the uneven distribution of demand in the field, the mismatch between distribution and supply, the geomorphology of the country, the dependence of Macedonia and Thrace (approximately 40% of the country's irrigated areas) from river run-offs from neighbouring states and the aridness of the smaller islands especially in the Aegean. Investments target regions with estimated water deficits now or in the future (7 out of the 14 water departments of the country).

For the assessment of the environmental impacts of the measure both the SEA and CEA methods have been applied, since they can be seen as complementary, SEA focusing on the review of the intervention logic and the extraction of conclusions on effectiveness (and to impact in a broader sense) and CEA on efficiency based on the result indicator values identified by the SEA.

Findings

As a first step the SEA Report of 2007 was reviewed and the **relevance of the environmental issues "soil" and "water"** was confirmed. Agriculture is still the biggest consumer of water (around 85% of the water abstracted), putting pressure on aquifers through high depth drillings from individuals or group of farmers. Furthermore, the deterioration of the distribution networks remains: failure to manage (store) surface winter run-offs and the losses in the distribution network result to even higher pressure for water in the future specially in relation to the forecasted climate change. In particular the regions of Greece such as Thessaly, Central, East Macedonia-Thrace and the Aegean islands which are characterised either by high demand, low supply or both are at high risk. A specific source of monitoring of the environmental trends was not used (i.e. the SEA Monitoring Report).

Water is addressed extensively and sufficiently, e.g. water protection through water management, storage and distribution, utilisation of winter run-offs, enrichment of aquifers, reduction of the risk of floods and creation of lakes and wetlands. **Soil** is only addressed indirectly and in relation to the reduction of erosion risk. Hence the identification of relevance between the measure and the environmental issues can be seen as sufficient, if somehow one-sided.

In the **intervention logic,** measure 125A is serving the overall aims of Priority Axis 1 on improving competitiveness but also included is the specific objective of **"Water resources protection"**. Further operational objectives are:

- **increased use and storage of run-off water** and the gradual abandonment of water drilling and use of subsurface water;
- **improvement of the water distribution network** (open channels and pipelines) for water saving purpose;
- protection of **soil**;
- **reduction in energy use** through avoidance of deep drills and pump operation and/or avoidance of losses.

As **result indicators**, apart from the CMEF result indicator related to GVA increase, the following were proposed:

- increase in the volume of deposited water by 90 million m3;
- land affected by channel and pipeline improvements at 51,000 Ha.

Hence the **proposed programme-specific result indicators only cover part of the operational objectives** of the measure (e.g. not covering energy and soil). They are rather generic but can offer a base for further refinement. However no values are provided since the APR states that entries of values will be inserted in the monitoring system after the completion of the selected operations. Taking into account that most of them have not been completed yet, no values are delivered. As all operations are subject to environmental permits, *ex ante* detailed indicators and values are available and theoretically can be aggregated. Such a step was however beyond the scope of the present study.

Considering the definition of a **naive counterfactual (zero case)** the measure comprises an **overwhelming share of investment** in the field of irrigation improvements, thus at least indicating that its "net" effect is significant at country level. Since the operations are small in number and geographically distinct, a quasi-experimental counterfactual case could be established at regional level taking into account the large agricultural areas irrigated by groundwater sources and open channels.

When all actions are complete and assuming that the targets of the two result indicators have been achieved (90 million m^3 increase in surface water and 51,000 Ha of land under irrigation, i.e., allowing the installation of modern drop irrigation systems), the surface water for agriculture will increase from 3,577.3 million m^3 to 3,667.3 million m^3 an increase of surface water to total abstraction from 45.2% to 46.3% at national level.

However, if we assume that a significant amount of these investments take place in Thessaly where the corresponding ratio of surface water to total abstracted water is only 7.9%, even if 30 of the 90 million m^3 are added the ratio of surface to total water abstracted will become 10.6% with significant consequences for groundwater and for the cost of irrigation (groundwater is very costly due to the high energy demand). This example also highlights the regional dimension of the investments. If investments are geographically targeted at the regions in need or the regions are expected to present a considerably higher demand of irrigation water in the coming years due to climate change, their impact may be significant not only in terms of resource management and conservation but also as a resilience strategy (the case of islands could be illuminating if data were available).

As a conclusion, the **SEA framework application** has demonstrated that:

- the **relevance and positive effect** of the measure to the environmental issues **water** and **soil** is acknowledged;
- **relevance to other environmental issues** is considered (e.g. losses in biodiversity in flooded areas or increase of utilised agricultural land and use of chemicals) but are considered **negligible**;
- the **intervention logic encompasses the environmental issues** of water and soil but also of energy use (i.e. climate/air) in form of operational objectives;
- the measure defines **specific result indicators**, which are related to specific aspects of the water use but they do not cover the entire objective spectrum;
- the measure addresses in a **qualitative manner the "impact level**[41]**"** e.g. related to the share of surface water on total water abstraction or total water abstraction per Ha. Taken into account the fact that the operations are geographically distinct it could be possible to net effects using advanced **quasi-experimental methods for CIE**; assuming that the relevant services of the Ministry of Rural Development or Environment collect causally, temporally and spatially relevant context indicators. This could be easier for operations targeting water savings from the main water distribution network, but can also be applied with certain conditions also for operations targeting water storage from surface run-offs.

Considering **CEA in the case of the CER numerator** (i.e. EUR) the APRs regularly report on programme costs (i.e. EAFRD + National Contribution).

In the case of the **CER denominator** (i.e. irrigated land or m^3 of water) the RDP provides only the *ex ante* values related to the results; however they can offer the base for interesting comparisons, as shown in the table below. There are large differences among regions; it would be interesting to examine the reasons, e.g. topography, land plots structure, etc. that lead to these deviations.

The APR states that entries of values will be inserted in the monitoring system after the completion of the selected operations, thus the values below should be updated and debated.

[41] It should be noted here that by "impact" a different context is addressed than the CMEF Impact Indicator on Gross Nutrient Balance and Water Quality. A better description might be "pressure level" based on the DPSIR framework of the EEA.

Table 29. M125 projects of the Greek RDP, comparison of CER

Case	Pre-fecture	Region	Type	Irrigated area (Ha)	Cost (EAFRD+National public contri-bution) in EUR	*Ex ante* CER (in EUR)
Neochorion	Serres	Central Macedonia	Sub-surface irrigation pipeline network (under pressure) for provision of irrigation water	418.3	4,258,434	10,180.33
Damasia	Larissa	Thessaly	Sub-surface irrigation pipeline network (under pressure) for provision of irrigation water	900	1,178,200	1,309.11
Makry-chorion	Larissa	Thessaly	Sub-surface irrigation pipeline network (under pressure) for provision of irrigation water	500	33,000	664.00
Taousanis	Larissa	Thessaly	Sub-surface irrigation pipeline network (under pressure) for provision of irrigation water	700	833,000	1,190.00
Samothraki	Evros	East Macedonia-Thrace	Open irrigation network	100	800,000	8,000.00
Konitsa	Ioannina	Ipiros	Integrated distribution	1,416	15,830,798	11,179.94

Assessment

Overall the **quality of results is satisfactory**, setting aside the lack of reported result indicators due to the lengthy period of implementation of the operations. The **SEA** has highlighted the explicit and implicit intervention logic in relation to the environmental issues and has shown areas for improvement in the programme-specific result indicators and the options for assessment of impacts. The **CEA** is a simple method whose power lies in the credibility of the numbers feeding the numerator and denominator of the **CER**; even when using *ex ante* values the results can be used to trigger the debate for the explanation of deviations among operations and regions.

In the absence of current result indicator values **the discussion of the representativeness of results is not possible**; however the **theory-based assumptions are sound and valid**; the CER is too specific to allow for generalised comparisons.

Both methods of consideration of the environmental dimension have been well executed for a measure under Priority Axis 1 in this case study. The formulation of the "theory of change" on environmental aspects is clear, if incomplete on secondary effects (soil, energy). **Problems** encountered related to the **lack of periodic reporting of result indicator values** and **the lack of tracking of causally, temporally and spatially relevant context indicators**. These indicators, required as a base for the generation of plausible inference of the effect of the measure on the development of the environmental issues, are rarely available in time series or when available they do not perfectly fit in the causal and more importantly in the spatial context.

There is no real solution to this problem apart from time, application forms, environmental permits and other documents have to be screened and discussed, but there is no systematic solution. In the case of the **CEA** the use of *ex ante* values can deliver useful insights.

5 Answers to Evaluation Questions

5.1 Answers to Evaluation Question 1

5.1.1 Rationale

A number of heterogeneous approaches to evaluation are used by experts and researchers of different disciplines such as economics, sociology, regional planning, policy sciences, and environmental sciences which are involved in evaluation. Consequently the range of methods that can be used for programme evaluations is very broad covering theory-based, quantitative and qualitative approaches which focus on aspects that cannot be measured in numbers. This variety offers evaluators a large choice of methods from which to choose the most suitable for a specific evaluation task.

The choice of methods depends on the Evaluation Question under consideration, the data that are available, the resources to bridge data gaps and the means to carry out the analysis. Since every method has specific advantages and draw backs, there is no single method which surpasses all others in all aspects. Therefore it is not possible to expect that every method can be used to answer ***all Evaluation Questions*** in an adequate manner. On the contrary, in most cases methods should be used for clearly specified evaluation tasks.

Evaluation Question 1 (EQ1): *"To what extent are the different evaluation methods described and/or tested in this exercise appropriate for the assessment of the effectiveness, efficiency and impact of the different types of investment support considered?"*

EQ1 addresses this issue by focusing on exploring how appropriate different methods are to answer very specific and well defined Evaluation Questions and to assess effectiveness, efficiency and impact. This task requires the comparison and *if possible* also ranking of different methods while meeting certain criteria. Because the range of involved specialists covers several disciplines it is necessary to develop a common understanding among experts from different fields on how these criteria can be made operational and they should be interpreted in order to allow consensual conclusions. In this respect it is also important to unify the understanding of the key terms used in EQ1, in order to provide sound answer to EQ1.

Judgement criteria for the comparison of methods are: *rigour, reliability, robustness, transparency, validity, practicability.*

This study discusses specific methods that are representative for the following groups of methods:

- econometric approaches for causal analyses (selected method: PSM);
- quantitative (programming) models that represent farms/firms/households/regions (selected method: IO model);
- qualitative and participatory methods (selected method: MAPP);
- theory-based and descriptive approaches (selected method: TBE);
- integrated approaches on environmental and economic indicators (selected methods: SEA and CEA).

One aim is to provide an orientation under which context certain methods and combinations of methods can be used. Applying this to the evaluation of investment support measures is the topic of EQ1.

5.1.2 Judgement criteria

The criteria for comparing different methods to be used in the assessment of *efficiency, effectiveness* and *impact* are: *rigour, reliability, robustness, transparency, validity, practicability.*

Before elaborating the answer to EQ1 in detail it is necessary to consider the following aspects: *Reliability* refers to the property of a method to provide identical results when a measurement is repeated under identical conditions. This definition is pertaining to natural sciences (a reliable thermometer gives the right temperature – a valid result – even if results are different when given in Celsius or Fahrenheit). In social sciences other definitions are used, e.g. "reliability is the extent to which measurements are repeatable - when different persons perform the measurements, on different occasions, under different conditions, with supposedly alternative instruments which measure the same thing"[42]. In inter-disciplinary teams another aspect should be considered: from a user's perspective, lack of reliability may arise if different observers draw different conclusions from the same observations using the same method. In an ideal world this would not happen because a given method would be well specified and using the same information would yield the same result like computer programmes give the same results when identical data are used as input. We make the assumption that qualitative methods would also yield the same results if the same group of persons would be asked to state their opinion on the same issue if all other factors were identical. Different people with different knowledge asked exactly the same question would most likely give different answers. Not getting the same answers would not be due to an "unreliable" method but due to an inappropriate use of a qualitative method. The criterion "reliability" will not be further explored in this study given the complexities described here.

The Evaluation Question refers to *efficiency* of investment support measures. In the context of the evaluation of rural development "efficiency" is defined as the "best relationship between resources employed and results achieved in pursuing a given objective through an intervention" (see glossary). In economics the situation with the "best relationship between resources employed and results achieved" is called to be "optimal". Researchers interested in efficiency would describe an "optimal" situation in the following ways: "In the economic theory of policy evaluation, a comparison between marginal benefits and marginal costs determines the optimal size of social programmes" (Heckman, 2010). Or: the "optimal" scale is where the marginal social benefits of the project/policy are just equal to the marginal costs of the project/policy. "Marginal" here simply means small change in costs and benefits. So the marginal benefit of a policy is the extra benefit that accrues to society from one small change in the 'quantity' of the policy" (Pearce, Atkinson and Mourato, 2006).

In the context of RDP "*efficiency is calculated by dividing the budgetary inputs mobilised by the quantity of effects obtained*". Thus, efficiency as defined by the CMEF may be related to but will neither measure marginal (social) costs nor marginal benefits. The definition is similar to "cost-effectiveness" as used by many economists: "the easiest way to think about CEA [cost-effectiveness analysis] is to assume that there is an effectiveness - E, and this is to be compared to a cost - C" (Pearce, Atkionson and Mourato, 2006).

[42] Drost, E., 2011, Validity and Reliability in Social Science Research. Education Research and Perspectives, Vol. 38, No. 1, 105-123.

5.1.3 Answering EQ1 in a "best case" situation

A "best case" represents the situation under which all conditions needed to apply the given method are ideal, e.g. all data resources (personnel, equipment, time) necessary to use the method properly are available. Although such a situation does not exist in reality, it can be used as a reference point (benchmark) to show what is possible, if all conditions are met. Developing answers to EQ1 in a "best case" world is nevertheless valuable because it shows what can be expected from applying a given method in an ideal situation.

5.1.3.1 The scale of results of different methods

All methods under consideration have been used already in evaluations and are therefore suited to analyse the effects of programmes. But not all of them are quantitative methods which can be used to derive results on a cardinal scale. The quantitative methods have different characteristics, underlying assumptions and limitations.

In this section, a criterion is introduced which can be used to group different methods according to the kind of results: *What is the scale used by a given method to evaluate the effects of investment measures on efficiency, effectiveness, and impact*?

Methods providing results on cardinal scale are preferred to those yielding results on ordinal scales. Results from cardinal and ordinal scales can always be interpreted in nominal terms. Based on the responses of the experts involved in this study, the ranking of methods is as follows:

- cardinal scale: PSM, IO, and depending on implementation: CEA and SEA;
- ordinal scale: all methods;
- nominal scale: all methods.

This result is important because it states that quantitative methods are necessary in order to evaluate the effects of a programme on effectiveness, efficiency and impacts in quantitative terms. Methods that use interviews of experts and stakeholders and data available for desk research are not appropriate to answer the Evaluation Questions of this study in quantitative terms[43]. The best one can expect is a ranking of different measures or a verbal description of observations that are embedded in a coherent structure.

A consequence of this result is that non-quantitative methods cannot be used to evaluate in quantitative terms the programme's effectiveness, efficiency and impacts. However, the non-quantitative methods can be used to describe the logic of intervention, the context of the intervention and many other important aspects that cannot (yet) be quantified.

5.1.3.2 The rigour of the counterfactual

Among the most challenging tasks of an evaluation study is to identify what would have happened if a policy intervention had not taken place. This is the counterfactual situation: "*a situation which would have occurred in the absence of a public intervention, also referred to as "policy-off situation"*. As elaborated previously, in an

[43] This claim is justified by the CMEF definitions of indicators of *efficiency*, *effectiveness* and *impact* which are ratios.

ideal setting, randomised controlled trials (RCT) would allow to statistically estimate the differences of outcomes (among them results and impacts) of treated (= beneficiaries) and non-treated (= non-beneficiaries) units. Even though elements of such approaches are conceivable, to our knowledge, such a study does not exist yet for RDP investment support measures. Following the logic of the CMEF, to specify a counterfactual situation is necessary in order to differentiate a "gross effect" (apparent situation observed relative to baseline situation) and "net effects" (the effect imputable to public intervention and to it alone; see glossary).

RCTs have not been used for the measures under consideration in this report therefore alternative ways have to be defined in order to identify a counterfactual situation.

While it is a relatively easy task to conceive a situation in which a programme was not implemented in thought experiments, it is not possible to make judgements on indicators without further assumptions. Different evaluation methods can be ranked according to how many assumptions are necessary in order to construct a counterfactual situation. The following list gives a ranking starting with the most preferred situation:

- counterfactual situation is derived from unit level data and parameters are estimated by econometric means by securing that treated and non-treated units (individuals, regions, firms) are very similar in observable and unobservable characteristics apart from the treatment;
- counterfactual situation is derived from statistics, monitoring and progress reports, with a quantitative or formal model and key parameters based on econometric estimates, on randomised controlled trials or econometric estimates of counterfactuals;
- counterfactual situation is derived from statistics, monitoring and progress reports in a quantitative or formal model and key parameters are based on expert judgement;
- counterfactual situation is derived by logical reasoning, expert judgement and stakeholders (among them beneficiaries and non-beneficiaries and representatives of Managing Authorities);
- counterfactual situation is derived by logical reasoning and expert judgement based on desk research of progress reports, monitoring reports and other sources without reference to empirically estimated parameters of counterfactual studies.

Classification of methods used in this study according to the ranking assuming a best case situation (=1):

- PSM: the counterfactual situation is derived from observed data;
- IO and CEA: for a second best situation a necessary condition is that PSM, RCT or other adequate methods have been carried and the results are then used as input parameters for an IO or CEA;
- IO, CEA: in a third best situation the counterfactual situation is explicitly stated and based on expert judgement but not derived from an PSM or similar analyses;
- MAPP: counterfactual situation is explicitly stated and based on judgement of stakeholders and experts;

- TBE, SEA: counterfactual situation is derived from documents or based on expert judgement.

The conclusion of this elaboration is: concerning how suitable different methods are in order to identify causal relationships there is a clear and unambiguous ranking: (randomised controlled trials or adequate) econometric methods are necessary to give evidence on counterfactual situations in *ex post* evaluations. In an *ex ante* context empirical evidence on counterfactuals is not possible, it can only be assumed as it is done in computational models[44].

5.1.3.3 How rigorous are the different methods with respect to leverage, deadweight, displacement and multiplier effects

Conclusions in the assessment of effectiveness, efficiency and impact could be biased if the evaluation focuses on gross effects without considering confounding influences which need to be taken in consideration to estimate net effects. In a best case situation the data necessary to estimate deadweight, leverage effects, and displacement effects are available. In a real world situation this is not always given. Additionally, not all quantitative methods can be used to derive all effects mentioned:

a) method(s) suitable to measure deadweight and leverage effects in quantitative terms are: PSM, and in special cases IO and CEA;
b) method(s) suitable to measure displacement effects in quantitative terms are: PSM, and IO;
c) method(s) suitable to measure multiplier effects in quantitative terms: IO.

Further explanations with respect to a):

Deadweight effects and leverage effects are among the most important elements of interest in an econometric counterfactual analysis.

It is possible to take into account deadweight and leverage effects in computational methods like IO even if the parameter has to be taken from other studies. If the parameters are not based on econometric results, they have to be based on assumptions. In several situations it may be legitimate to assume that the deadweight effect is small and to derive the leverage effect from monitoring data. This would be e.g. the case if an investment to improve the environmental quality is not taking place without a policy intervention (e.g. special equipment that is too costly to be adopted under market conditions).

The overview shows that more than one method is needed in order to account for all effects under consideration. IO-multiplier effects are specific to the IO method and therefore are usually not available from econometric methods. One assumption underlying the PSM method is that an intervention is small and does not have an influence on the general equilibrium which restricts its application.[45]

[44] The specification of computational models may be very heterogeneous. Even if the model is specified like an econometric one used to analyse causal effects it cannot predict future causality using parameters derived from observed data.

[45] Rubin (1986) shows that only under the stable unit treatment value assumption the representation of outcomes by the Neyman-Rubin counterfactual model is adequate. Heckman (2005, p.11) argues that this condition rules out general equilibrium effects. [Rubin, D.B., 1986, Which Ifs have causal answers? Journal oft he American Statistical Association, 81, 961-962; Heckman, J.J., 2005, The scientific model of causality. Sociological Methodology, 35, 1-97.]

Conclusion: Quantitative assessments on effectiveness, efficiency and impacts are only possible if quantitative methods are used. Such results are preferred over qualitative results and those based on logical deductions. Nevertheless, the latter methods may reveal important aspects of a given programme and establish a basis for hypotheses that can be evaluated with quantitative tools if sufficient data/resources are available for such assessments.

5.1.3.4 Rigour of different methods with respect to efficiency, effectiveness and impact

More rigorous methods are based on well developed theories, they are more widely applied and accepted in the scientific community or the community of evaluators and it will be described in methodological textbooks. Of course, a new method has to be established at some time in the future and it may take long time until it is well established and a long time until a large group of researchers or experts apply it.

The following methods used in this study are consistent with established and widely used scientific theories:

- PSM: the method is based on statistical theories about causal effects and concepts how to measure them, in particular the “potential outcome model” (see Roy, 1951; Rubin, 1974; Holland, 1986).
- IO: the method was developed by Leontief based on economic theory. Several textbooks and high ranking scientific papers exist on the method (e.g. Leontief, W. (1986), Input-Output Economics, Oxford University Press; Leontief et al. (1953), Studies in the Structure of the American Economy, International Arts and Science Press; Miller, R.E. and Blair, P. D. (2009), Input-Output Analysis: Foundations and Extensions, Cambridge University Press). An aspect that makes this method special is the fact that Input-Output tables are published annually by the statistical offices of all EU Member States (except Croatia).

The following methods are used in many variants and tuned to very specific applications and in many cases are tightly linked to established and widely used theories:

- CEA: The method is well established in economics and environmental sciences and described in most textbooks on environmental economics (see e.g. Pearce, Atkinson, Mourato, 2006). Depending on the context the method uses quantitative results of complex integrated models of various disciplines.

The following methods are - compared to the methods described above - to a lesser extent based on established theories and less widely applied in studies reported in peer reviewed scientific journals:

- SEA: The approach is widely used in the administration of RDP, rather than in research even if some SEA results are based on rigorous methods.
- MAPP: This method is applied to derive results for evaluation reports and not frequently used in studies that are published in peer reviewed scientific journals.
- TBE: This method is applied for evaluation reports and not frequently used in studies that are published in peer reviewed scientific journals.

Transparency about all steps taken during carrying out a given analysis, making evident all assumptions and simplifications and the disclosure of data and computer programme codes should be part of every analysis. High levels of transparency are not limited to specific methods but pertain to all of them.

5.1.3.5 Robustness of different methods with respect to measuring indicators of efficiency, effectiveness and impact

The term "robustness" is defined in different ways in the evaluation literature. In econometrics, robustness is used in various contexts (e.g. biased and unbiased estimators, model and variable selection) and has therefore context-specific meanings. In this study this term is not restricted to the use of econometric methods. Robustness is considered to be high if results are stable and resilient to small but deliberate changes (e.g. an additional year of observations, an additional explanatory variable, another stakeholder, another evaluator). In some methods, the robustness of results can be checked by sensitivity analyses. It is not possible to identify a consistent ranking of methods with respect to a "robustness" scale because such a scale has not been developed (yet).

Typical approaches to increase the robustness of results in the relevant methods (in alphabetical order) are:

- CEA and SEA: the sensitivity of the results on assumptions about specific parameter values is usually checked in a sensitivity analysis;
- IO: the responsiveness of the results on different assumptions about specific parameter values is most frequently checked in a sensitivity analysis; such an analysis can show upper and lower bounds of ranges based on discrete choices of values or it can be based on Monte Carlo simulations - in such a case not only the range but also the distribution of results can be evaluated[46];
- PSM: add more observations; include more variables; carry out specific robustness tests and use strategies to avoid biases; check how sensitive results are to model specification and treatment of outliers; PSM is a method to address the problems associated with *selection bias* (beneficiaries are not chosen at random but participate in a programme voluntarily) in a coherent manner and therefore has advantages over other methods including many other econometric approaches;
- MAPP: the sensitivity of results can be checked if more than one focus groups with a similar composition of stakeholders (and representatives of Managing Authorities) are working on an evaluation of the same measures in the same region; if the outcomes of several different focus groups are yielding similar results one would expect that results are relatively robust even if statistical significance cannot be measured;
- TBE: options to improve the robustness of results based on this method are relatively limited. It is mainly due to the rigour and the skill of the analyst to prevent biased results.

Most of the methods used in this study can employ special techniques in order to increase the robustness of the results or at least give an indication of the range of results. For quantitative methods sensitivity analyses can be used in order to show the robustness. Compared to other judgement criteria used to gauge the appropriateness of methods it is not possible to present a ranking of methods with respect to robustness.

[46] A recently published example that uses this approach in a partial equilibrium model is: Kirchner, M; Schmid, E (2013): Integrated regional impact assessment of agricultural trade and domestic environmental policies.Land Uses Policy. 2013; 35: 359-378.

5.1.3.6 Validity of results of different methods with respect to efficiency, effectiveness and impact

The term validity is used in various contexts and variants:

- "Internal Validity": Results of non-empirical methods are valid if they are logically sound. Results of empirical methods are valid if they are logically sound and factually sound. Logical soundness can be verified and high transparency makes this easier to do so. Factual soundness is verified if the result is identical to the true parameter which often cannot be observed (see counterfactual).
- "External validity" is a quality measure of empirical research. In our context "external validity" is defined as whether the results obtained from a case study will be the same if a similar programme is in place in another context as well. High "external validity" of a method is a feature of great interests because it would give scope to save costs needed for evaluations.
- "Convergent validity" is given if different methods employed to answer the same research question yield similar results.

Characteristics of the methods used in this study with respect to "internal validity":

- Logical soundness: All of the methods used in this study are using logical rigour to derive results. Given that in each case elaborated tools are used to guide the experts who carry out the fieldwork and that core team members cross-check the results there is no reason to assume that methods are different with respect to logical soundness. In the case of IO, logical soundness is given by definition if the data set and computer code are implemented conforming to the method.
- Factual soundness: Given that all the methods are employed in order to analyse what would have happened in case of absence of investment measures the factual soundness cannot be checked directly, therefore other criteria must be used to draw conclusions about the potential factual soundness.

Characteristics of the methods used in this study with respect to "external validity":

- PSM: Special statistics of this method (the area of common support) can be used to make judgements on the validity to transfer results from the analysed sample to the whole sample (for the analysis some observations are disregarded). While this measures how valid the results are based on a sample for the population as a whole, the “external validity” (as defined above) depends on the similarity between the populations where the results are to be transferred to.
- IO and CEA: Statistical tools (e.g. cluster analysis) can be used to identify regions that share common important features (e.g. sectoral structures, income, export-shares, business characteristics; similar environmental conditions). If similar characteristics are prevalent, conclusions may be drawn from results of the analysed regions to potential outcomes in similar regions; the more transparent the underlying results are, the more reliable are conclusions on "external validity";
- MAPP and TBE: According to the views of the researchers working with these methods, it would be problematic to draw conclusions from one case and carry them over to other cases.

When different methods are used to analyse the same investment measures in the same programme regions, it is possible to check for "convergent validity" of the methods under consideration.

5.1.3.7 Practicability of tested evaluation methods

A method to assess the practicability of evaluation approaches does not exist. Given that each method has its own requirements of data, tools, and infrastructure it is hard to assess what is practical or not because the situations may vary from case to case. Practicability also depends on the point of view; what may be very practical from the point of view of an evaluator may be very impractical from the point of view of a Managing Authority (e.g. access to administrative data). The following dimensions of practicability are explored in more depth:

- is the method adequate to be used in *ex ante*, and *ex post* evaluations;
- what qualification is needed by the evaluators and/or reviewers of evaluation reports;
- how much resources are necessary in order to carry out an evaluation of a case study.

Methods and their use in ***ex ante* and *ex post* analyses** (in alphabetical order):

- CEA can be used in *ex ante* situations if the likely environmental outcomes are estimated with ecological models.
- IO is frequently used in *ex ante* (and also *ex post*) evaluations. In *ex ante* evaluations data used to shock the model is based on programme-specific projections and/or expert judgement. The counterfactual situation is constructed by defining the baseline situation and a scenario with a shock that mimics the expected intervention due to the programme.
- SEA is a standard method used during the phase of programme design and therefore well suited for *ex ante* analyses. The method can also be used in *ex post* situations.
- MAPP and TBE: Both can be used in *ex ante* situations but they are more frequently used *ex post*.

PSM cannot be applied for *ex ante* evaluations, nevertheless already during the phase of programme design it is necessary to take care of the data that needs to be collected to provide the necessary empirical basis for this method. More recently computational models specified, like those used to estimate propensity scores, were used in *ex ante* studies. [47]

Another criterion for practicability is the **qualification of the expert** carrying out the analysis or applying a given method, his or her experience in applying it and the complementary toolkit of methods he or she can make use of. There is a differentiation between two categories of methods depending on the experts' background and training bearing also in mind that not only the formal education but also the individual skills and abilities are important:

[47] See: Todd P.E. and Wolpin K. I., "*Ex ante* evaluation of social programmes", in: Annales d'économie et de statistique. – N° 91 – 2010, pp 259-286.

- methods applied by experts with a master degree and several years of experience in programme evaluation: CEA (in case of desk research), MAPP, SEA and TBE;
- methods applied by an expert preferably holding a PhD degree and several years of experience in applying the specified method: CEA (if quantitative work is carried out by the expert), IO and PSM.

Each method has specific **resources requirements** (e.g. data, time, infrastructure, software). It may be very costly to make them available in the case of data on individual firms or in the case of environmental indicators but the costs depend on country-specific procedures and the available infrastructure. Beyond data availability additional resources are necessary as follows:

- The *time* necessary for carrying out the research with a given method mainly depends on the experience of the evaluator as far as quantitative methods are concerned, and it depends mainly on the characteristics of the case study and participants in case of non-quantitative methods. The time of applying a given method ranges from a few days (excluding preparatory work) to a few months (including preparatory work and arranging meetings of several people). Methods that are based on getting people involved (be it stakeholders or representatives of administration) are more time intensive than methods that are using regularly updated data sources.
- *Special infrastructure* is needed in some cases of PSM. This may be necessary when country-specific data protection regulations require the use of safe centres to carry out the quantitative analysis. This infrastructure is provided by the relevant national authorities and statistical offices and needs not to be provided by the evaluator. If safe centres are involved, doing the research is more complicated. In certain instances the collection of environmental data may require specific infrastructure, e.g. a sampling networks, lysimeters, etc. Apart from these specific cases none of the methods used in this study requires specific infrastructure.

Specific *software* is needed for IO and PSM. For the other methods used in this study, standard office software is sufficient to carry out the analyses. Another aspect is the level at which outcomes are measured:

- Programme *results* are derived from micro-data using econometric models, computational models (of farms, households and firms) or non-quantitative methods. IO models, sector models and CGE models are not adequate for such analyses.
- Programme *impacts* are obtained by observations or simulations at programme area level (regional or macro-level). Impacts cannot be measured by computational farm, household and firm models while the other methods are suitable.

5.1.4 Overview of methods given a "best case" situation

In this section the findings of the previous chapters are summarised. An overview that is based on the elaborations of evaluating different methods using a range of criteria is provided in Table 30 below. This overview reflects a "best case" situation which serves as a reference in order to avoid that a given method is ranked lower because of hurdles that are only specific to the case studies under consideration.

Another rationale to refer to the "best case" is that the lessons learned in this study may be used to improve the situations in which the methods are applied. This summary is based on the synthesis of the views of the methods experts of the study team elicited through an extensive survey with feedback loops to make definitions of different methods comparable.

The major findings of the exploration of the appropriateness of the methods analysed in this study are:

- each method under consideration has specific data requirements, specific strengths and specific limitations and there is no method that prevails all other method with respect to all the criteria used in this study;
- the methods that are applied in the fieldwork share many characteristics of a wide range of methods used in scientific literature and by experts in programme evaluation and therefore are representative for the most widely used quantitative and qualitative approaches;
- with respect to evaluating investment measures, methods that can provide quantitative results are preferred; methods that do not provide quantitative results may be very useful when there are restrictions in the use of quantitative methods (e.g. restrictions to data, etc.) or in order to prepare hypotheses that can be tested or further explored with quantitative methods;
- qualitative and quantitative methods are complementary and should not be regarded as substitutes; the analysed quantitative methods are also complementary in many respects because some important results (e.g. deadweight, leverage) can only be estimated by econometric methods while IO yields other results that are method-specific (e.g. multipliers or specific impact indicators of an economy-wide nature); when environmental effects are important, specific, integrated approaches are required;
- in an *ex post* context where sufficient data is available causal effects shall be estimated using adequate econometric methods (e.g. PSM); the results of such estimates can and should be used by other methods in order to improve the validity and their results; results of econometric analyses are a valuable input for other methods, as well;
- in an *ex ante* context, econometric methods cannot be applied because observations are needed that can only be made after a treatment has started;
- in order to fully exploit the capabilities of an econometric causal analysis for mid-term and *ex post* evaluations it is necessary from the outset to prepare the collection of the necessary data; while other methods may not be adequate to fully make use of micro-data, statistics based on them can also be used by various methods therefore it is worth collecting micro-data because they can be used by several methodological approaches.

Table 30. How appropriate are different evaluation methods to analyse efficiency, effectiveness and impact of investment support measures - assuming a "best case" situation

criterion		more preferred		less preferred
reliability		assumed to be given		
		CEA, IO, MAPP, PSM, SEA, TBE		
rigour	causality/counterfactu	estimated	estimates based computations	assumed
		PSM	IO, CEA (in special cases)	CEA, IO, MAPP, SEA, TBE
	scale of indicators	cardinal	ordinal	nominal
		CEA, IO, PSM	IO, PSM, CEA, MAPP, SEA, TBE	CEA, IO, PSM, MAPP, SEA, TBE
	link to theory	closely linked to theory		not closely linked to theory
		CEA, IO	PSM	MAPP, SEA, TBE
	scientific literature	frequently used		not frequently used
		CEA, IO, PSM		MAPP, SEA, TBE
	net/gross effects	can be quantified	estimates based computations	assumed
	deadweight	PSM	IO, CEA	CEA, IO MAPP, SEA, TBE
	leverage	PSM	IO, CEA	CEA, IO MAPP, SEA, TBE
	multiplier	IO		MAPP, SEA, TBE
	displacement	IO, PSM		MAPP, SEA, TBE
robustness		standard procedure	possible with special efforts	not possible
	sensitivity checks	CEA, IO, PSM	MAPP	SEA, TBE
	specific checks	PSM		
validity	internal validity	given		not given
		CEA, IO, MAPP, PSM, SEA, TBE		
	external validity	given		not given
		CEA, IO, PSM		MAPP, SEA, TBE
practicability	qualification	intermediate		high
		CEA, MAPP, SEA, TBE		CEA (special cases), IO, PSM
	time needed	intermediate		high
		CEA, IO, PSM, SEA		MAPP, TBE
	infrastructure	none		specific needs
		CEA, IO, MAPP, PSM, SEA, TBE		CEA (in special cases), PSM
	software	standard tools		specific needs
		CEA, MAPP, SEA, TBE		IO, PSM
frequency	method	more preferred		less preferred
	CEA	11	4	6
	IO	11	4	6
	PSM	13	2	4
	MAPP	5	2	10
	SEA	6	1	10
	TBE	5	1	11

Source: Sinabell/Morawetz (2014) based on expert assessment

Table 30 presents a summary and overview of the verbal descriptions of the methods in the previous chapter. By summarising the findings there is a certain loss of subtle information. The benefit of showing the great picture is that a detailed comparison over a range of criteria can be shown on one page.

The first column of the table lists the criteria used to characterise the different methods applied in this study. In the second column the criteria are specified in more dimensions in order to give enough scope so that methods can be differentiated and distinguished.

The third, fourth and fifth columns are used to make judgements about preferred and less preferred situations. For example, parameters that are estimated using econometric methods are preferred over parameters that are solely based on estimates of experts.

In the first row the criterion "reliability" is listed and, as explained previously, the assumption is made that all the methods in this study are reliable. Further checks could be made in principle but were not carried out in order to concentrate resources on the results on efficiency, effectiveness and impact.

In the second row and the following ones every method is listed under the column "more preferred", "less preferred" or in between according to the criteria and sub-criteria listed in columns 2 and 3. The allocation of methods into different columns is based on the arguments and the reasoning in the previous sections.

At the bottom of the table there is a list of frequencies: how often is a given method mentioned in each of the three columns ("more preferred", "less preferred", and in between). These frequencies are not a verdict on the quality of a given method or the usefulness to analyse and evaluate Rural Development Programmes. Instead they are a reflection on the findings of this study taking into account the user's interests in knowing how appropriate methods are to measure efficiency, effectiveness and impact of investment measures.

5.1.5 The" best case" reference situation compared with the observed situation

In the following sections the "best case" situation described above will be compared to the situation faced by experts in this study. The fieldwork of this study provides the necessary data to assess practicability and transparency. The experts involved in carrying out the methods give a general judgement how close the observed situation is to the "best case".

5.1.5.1 Observations on the practicability of tested evaluation methods

In order to assess the practicability of different methods in the case studies of this analysis, a tool was developed to collect information on resources employed during the fieldwork. Geographic Experts were asked to indicate how many days they had been working on different tasks (methods, measures, other) and how many expenditures were necessary to carry out the analysis. They were also asked to comment and to share their observations on the practicability of different methods and tasks. Experts are classified in two groups:

- The first group (geographic experts) was involved in collecting data and preparing material that was later used by the second group of experts.
- The second group (Core Team Experts) applied the methods based on the input of geographic experts and provided the method-specific answers of EQ1 and EQ2.

The summary of resources spent by geographic experts is given in Table 31. The method with the largest amount of resources (working days or expenditures) has rank 1. The "unweighted" rank is the sum of resources over all case studies. The "weighted" rank is adjusted for the number of case studies *and* measures analysed by each method (there is no differentiation between the measures, e.g. complex and simple measures).

Table 31. Ranking of resources used for the fieldwork of geographic experts (most resources: rank 1)

Method/task		Rank by days		Rank by expenditures	
	case studies	weighted	unweighted	weighted	unweighted
CEA and SEA	27	4	4	4	5
IO	48	5	3	5	4
MAPP	22	3	2	2	1
PSM	13	2	5	1	2
TBE	17	1	1	3	3

Note: The method with the largest amount of resources (working days or expenditures) has rank 1. The "unweighted" rank refers to the sum of resources over all case studies. The "weighted" rank refers to the sum is adjusted to account for the fact that the number of case studies was not the same.

The ranks presented in this overview are observations made under very specific circumstances that have to be considered:

- Experts for each method and Geographic Experts have specific expertise and skills and are very familiar with Rural Development Programmes.
- Some of the experts have been involved in mid-term evaluations and have ready access to material and results that would otherwise be very difficult to obtain.
- In some cases the Managing Authorities carried out specific tasks that had to be performed by experts in other case studies.
- Having established working relationships with Managing Authorities in previous studies made communication easy and work more effective in several situations.
- The methods applied in this study are already established and have been applied in similar situations already several times.

The experts were also asked to estimate their resource input for the analysis of individual measures. According to their feed-back it is however not possible to allocate resources to specific measures when more than one measure is analysed. There are obviously analysis' specific fixed costs that are hard to allocate to specific measures. A more comprehensive analysis of various measures therefore is likely to offer savings compared to a very specific one focusing on fewer measures.

This overview shows the relative resource intensity of the methods for analysing investment measures in a representative number of case studies in the EU. The results show that non-quantitative methods of this study are relatively resource intensive when it comes to manpower to collect the necessary information, to apply the method and to work out the results. It has to be considered, that the quantitative methods can only be applied when adequate data are available. But non-quantitative methods cannot be applied in a vacuum either. Analysts require documents, reports, expertise of persons involved in administration and access to stakeholders. Expenditures for data acquisition have to be considered in this context. The overview shows that for PSM the expenditures were the highest in this study. But the amount is negligible compared to the costs of manpower.

In order to evaluate the practicability of the methods, not only geographic experts were asked to report the resources needed to collect the information, experts who applied the method and summarised the findings were asked in a similar manner. The results of this survey are:

- Applying PSM required the most resources (almost one week per case) in order to derive the results presented in this analysis. Less resource is needed the more standardised the data are (FADN-data are preferred) and the better is the access to the data (an analysis in a safe centre is more time consuming).
- Considerable fewer resources per case study were necessary for handling data and deriving results applying MAPP, IO and TBE. The resource intensity was very similar in this study. The least resources were used for SEA and CEA.

Practicability has more dimensions than just the resource requirements. The other ones were not analysed in a similarly structured manner. There is evidence from the comparison between different programmes that Managing Authorities handle access to data and specific programme information in different ways. The offered data structure can be adjusted with ease for an evaluation in some cases and in others not. This is no wonder because a broad range of evaluation methods is available and a “one fits all” data set does not seem to be feasible or desirable. The best one could expect is standardised data sets for standard methods. Examples are bookkeeping data that are already linked to programme participation information or standardised reports on environmental effects using a small number of environmental indicators.

5.1.5.2 Subjective observations of experts on the quality of the results

The experts applying the different methods were – among others – asked two questions:

- What are the quality attributes of the method under consideration and which of them should a “best case” have?
- How close is a given case study to a “best case” situation (as described in the previous chapter)?

Answers on the first question were the input for the characterisation of the methods and the description of the appropriateness to analyse *effectiveness*, *efficiency* and *impact* of investment measures above.

The second question was answered after the work on the case studies was finished. If a case study is very close to the “best case” situation with respect to the criterions it will be given score 1. If the results are not robust, rigorous or valid it will be given score 5. An overview is presented in Table 32 below.

The answers are inherently subjective. The Likert scale in the table cannot be used to compare the different methods because different people provided the data and each of the persons made judgements only on one method. The rating can be interpreted as how satisfied an expert is with the results given his or her own standards and quality claims.

Table 32. Subjective assessment of how close case studies in relation to the "best case"

Method	Best case			Worst case		Score
	1	2	3	4	5	
	number of case studies in class					weighted average
IO	20	23	5	0	0	1.7
PSM	0	13	0	0	0	2
MAPP	3	6	13	0	0	2.5
SEA and CEA	0	3	7	9	8	3.8
TBE	0	16	27	12	0	2.8

Note: The table contains subjective judgements of one expert per method. A comparison between the different methods can therefore not be made.

5.2 Answer to Evaluation Question 2

5.2.1 Rationale

An overview on the results of the analysis on *efficiency*, *effectiveness* and *impact* of investment support measures of Rural Development Programmes is presented in this section. The focus of the study is to compare methods that are typical for different approaches frequently used in evaluations such as econometric approaches for causal analyses, IO and quantitative/programming models for farm/firm/household/region, qualitative and participatory methods, theory-based and descriptive approaches and integrated approaches. The findings referred to in the previous section indicate that they share some important features but that there are substantial differences between them. In order to acknowledge these findings, key results are listed on a method by method basis.

The focus of the presentation will be solely on *efficiency*, *effectiveness* and *impact* of each measure in each region. All intermediate results that are necessary to come to these final results are not dealt with in this chapter. Such intermediate results are often necessary to understand how the final results are obtained.

Depending on the method, few or many sub-indicators were calculated by the experts applying the methods under consideration. They are reported here in order to show the scope of detail one can expect from the methods applied in this study. As can be seen in the next chapters, the detail of sub-indicators is very heterogeneous. Efforts are made to present the results in a compact manner, therefore graphs and summary tables will be used where possible. This way results can be shown in a condensed manner.

The chapter will be structured in the following manner: In the next sections the findings of the methods will be presented one by one. Following the method-specific presentations, a short overview discusses differences between types of measures (A,B,C,D). In the final section a case study that explores contradictory results and shows how results from two methods can be combined to attain conclusions of higher validity is presented.

5.2.2 CEA and SEA results on EQ2

In total 13 SEA and CEA case studies were prepared covering six measures (121, 125, 216, 227, 311, 313) from three types (A, B, C) and seven regions (Austria, Cyprus, France, Hessen in Germany, Greece, Galicia, Scotland). Each case study focused on environmental aspects of the measures even if the environment was not the main concern of the measure. The most frequently mentioned environmental objective was water (10 times) followed by biodiversity, flora and fauna and landscape (each mentioned seven times).

In those cases where indicators are reported for the individual measures, they are mostly given in absolute values. The following list gives an overview of the indicators in absolute values as they were elicited from official national documents:

- **M125, Austria:** additional access to 8,000 Ha forest to minimize harvesting and erosion damages.
- **M125, Greece:** 90 million m^3 additional use and storage of run-off water and improved water supply for 51,000 Ha (reduced water drilling).
- **M125, Galicia:** water savings equivalent to a value of 390,000€/year (200 m^3/Ha/year) and energy savings equivalent to 108,000 €/year through improvements of infrastructure.
- **M227, Scotland:** increased biodiversity by 3,892 Ha, public access 5,250 Ha, woodland management 833 Ha, woodland access 21 Ha to enhance the environmental and public amenity values of woodlands.
- **M311, Austria:** Reduced heating oil use of 1,646,000 l/a to reduce CO_2 emissions through bio energy production (4,461,000 t/a CO_2 equivalent)

These indicators cannot be interpreted directly to measure *efficiency* as is done by applying other methods in the quantitative case studies. The results make sense in their specific context. Additionally, some of the indicators (e.g. for Austria) are based on case studies only or on assessments as part of the mid-term review (thus not covering the whole scope of a measure, the whole area or the whole period). One way of relating the indicator to context is *effectiveness* which was reported for Hessen:

- **M227, Hessen:** Preservation of countryside including forest soil quality and ameliorating climate change, biodiversity, water protection. Reforestation (52% of target), improved area (6% achievement rate), liming area (63% achievement rate), forest stock area (achievement rate of 22%) , forest protection area (11% achievement).

Measuring *Effectiveness* using SEA or CEA or similar approaches requires that the indicator covers the whole measure, the whole area and the whole period as targets are usually not defined for sub-categories. In many cases this requirement was not met, preventing the calculation of an indicator of effectiveness.

For five case studies (M227 in Cyprus and M121, M216, M227, M313 in France) comparable indicators were not found at all. This was due to insufficient data in official reports that were the only source of data used for SEA and CEA.

Relating the findings from SEA to costs for the specific measures allows the derivation of the cost-effectiveness (CEA). In practice, though, it turned out that a lack of data (with respect to indicators as well as costs not being broken down in sufficient details) made it impossible to apply the CEA for any of the analysed measures.

Given these findings it is not possible to present results on *efficiency*, *effectiveness* and *impact* in a table or a graph as in the case of the other methods in the following sections.

5.2.3 IO results on EQ2

Results of IO analyses are usually presented on cardinal scales and can therefore be presented either in tables or in graphs. An elaborated number of well defined steps has to be taken in order to obtain the final results which are presented here. These just offer a quick glimpse on the large range of intermediate results. IO analyses are data hungry but provide a wide range of insights that cannot be presented adequately in this summary. Readers interested in understanding how the final results are derived are encouraged to go through the analysis step by step using the supplementary material. Intermediate results are multipliers per sector (output, employment, income of type 1 and type 2).

Table 33. Shocks of regional or country IO model (million EUR)

	121	122	123	125	311	312	313	total
AT	368.40	10.83	60.21	17.80	13.25	3.00	5.59	479.08
CY	26.79		4.20					30.99
CZ	92.80	7.23	26.60	18.08		15.07	13.94	173.71
ES/Ga	64.23	5.77	21.43		0.25	7.99	1.28	100.95
GR	55.83		11.43	40.30	0.87	0.96	5.70	115.08
DE/He	53.07	3.75			0.84	0.34	0.10	58.11
PL	618.64		167.87		88.69	87.90		963.11
UK/Sc	35.94	0.09	9.78		6.54	3.04	6.95	62.34
SK	113.88	4.15	49.32	9.47	18.92		2.66	198.40

Source: Psaltopoulos (2014)
Note: Million EUR of average annual investment flows in 2009 prices.

The most important programme related data input for the analysis of the programme effect is shown in Table 33. The model of the national or regional economy (based on Input-Output tables) is "shocked", meaning that additional demand for specific investment products is simulated. The figures are based on the implementation of investment measures in the regions and countries under consideration and are mainly based on monitoring data. The table only shows the overview. For the sake of calculation these figures are broken down to various sectors (e.g. machinery, construction). The additional demand leads to adjustments in the model that simulate the adjustments in the economy. The changes of value added and employment induced by the demand for investment goods are *the investment effects*.

Investments are a bet on future outcomes. After an investment is made, firms may become more productive (unless evidence suggests that wrong investment decisions make them less productive).

More productive farms/firms are one goal of Rural Development. More productive firms usually produce more output because investments are mostly used to expand the capacities of firms[48].

Additional output is available on the market and used as input for other sectors and induces value added and employment there. This is the *capacity adjustment effect*. Together with *the investment effect* the *total effect* of a policy intervention can be quantified.

An important aspect has to be noted: In this section data used for the shocks are derived from programme expenditures that are observed data. How much additional output is induced by these investments is taken from reports of the Managing Authority. No cross-checks were made to corroborate these results. These data are therefore an input for the analyses and taken as given.

In the context of an IO analysis,

- *Efficiency* measures the relation between public expenditures and the intended result of the measure. The indicator of efficiency is calculated by dividing the budgetary inputs mobilised by the quantity of effects obtained.
- *Effectiveness* measures the extent to which objectives pursued by an intervention are achieved. An effectiveness indicator is calculated by relating an output, result or impact indicator to a quantified objective.
- The *impact* of a policy intervention is the average annual effect of the public expenditures of a measure on result indicators for the whole programme region under consideration. Spill-over effects on the whole economy of a country and the rest of the world (mainly EU) can be derived.

The summary results are presented in Table 34. For most of the measures analysed in each region it is possible to provide results on efficiency (see "effi." in the second column), effectiveness (see "effe." in the second column) and impact (see "imp." in the second column). The sub-indicators used for the presentation of the results are:

- *Efficiency*: Jobs created in the economy *per annum* per million Euro of RDP investment support (see "jobs/mil" in column 3).
- *Effectiveness*: IO estimates of increase in economy-wide GVA *per annum* compared to result indicator targets associated with each measure expressed relative to the target(see "Δ GVA" in column 3); targets are not specified for each measure, therefore plausible assumptions were made in some cases (see notes of results table for assumptions).
- *Impacts*: IO estimates on economy-wide employment creation *per annum* compared to impact indicator targets expressed as a share relative to the target (see "Δ jobs" in column 3).

[48] Investments may also be made to improve working conditions for operators, animal welfare or environmental aspects.

Table 34. Efficiency, effectiveness and impact results based on the IO analysis

Measure		indicator	AT	CZ	DE/He	ES/Ga	PO	UK/Sc	GR	CY	SL
121	effi.	jobs/mil	31.95	63.86	16.27	9.08	84.63	33.46	38.00	43.20	53.33
121	effe.	Δ GVA	1.70	12.90		-0.48	0.48	0.45	2.26	10.91	4.83
121	imp.	Δ jobs						0.18	0.57		
122	effi.	jobs/mil	25.45	62.09		30.01		28.45			51.41
122	effe.	Δ GVA	9.56	3.74				0.00			0.23
122	imp.	Δ jobs						0.00			
123	effi.	jobs/mil	92.89	58.49	15.48	-23.33	107.66	26.94	137.58	100.40	52.95
123	effe.	Δ GVA	275.00	7.76		-0.54	0.08	0.32	3.13	3.98	3.45
123	imp.	Δ jobs						0.04	0.42		
125	effi.	jobs/mil	33.35	52.27					36.01		31.61
125	effe.	Δ GVA	3.87	7.45					2.14		
125	imp.	Δ jobs							0.39		
311	effi.	jobs/mil	39.02		15.85	21.51	64.57	26.73	78.12		32.01
311	effe.	Δ GVA	0.43			0.00	0.21	0.02	1.09		5.55
311	imp.	Δ jobs	0.74			0.04	0.40	0.03	0.02		1.85
312	effi.	jobs/mil	68.95	49.24	15.99	52.22	93.18	36.05	118.65		
312	effe.	Δ GVA	0.28	11.33		0.25	0.19	0.01	2.31		
312	imp.	Δ jobs	0.10	0.03		0.37	0.58	0.02	0.03		
313	effi.	jobs/mil	57.67	53.93	13.46	21.49		44.81	76.88		35.41
313	effe.	Δ GVA	1.23	22.03				0.11	1.48		12.75
313	imp.	Δ jobs	0.92	0.03				0.05	0.12		1.21

Source: Psaltopoulos (2014). Results are based on targets defined in programme documents; expenditures and outputs are based on implementation reports of the programmes.

Notes for specific results on investment and capacity adjustment (national/regional data) effects : CZ 121 effectiveness: Target for 121 and 124; CZ 122/CZ 125 effectiveness: Target for 122, 123 and 125; CZ 312/CZ 313 impact: Target for Axes 3 an 4; DE Hessen: Investment Effects only; ES Galicia 122/ES Galicia 313 efficiency: Investment Effects only; PL 311/PL 312 impact: Target for Axis 3 in total; UK Scotland 121/123/311/312/313 impact: Total RDP Target; UK Scotland 122 efficiency: Investment Effects only; UK Scotland 122 effectiveness: Investment Effects only; UK Scotland 122 impact: Total RDP Target, Investment Effects only; GR 121/123/125/311/312/313 impact: Total RDP Target; SK 121/122/123 effectiveness: Target for Axis 1; SK 125 efficiency: Investment Effects only.

Where blank cells for efficiency and effectiveness appear in the table, these are associated with non-estimated impacts due to the fact that the relevant measure was not assessed. In the case of impacts these are due to the fact that the relevant target indicator (jobs created) was not specified at programme level.

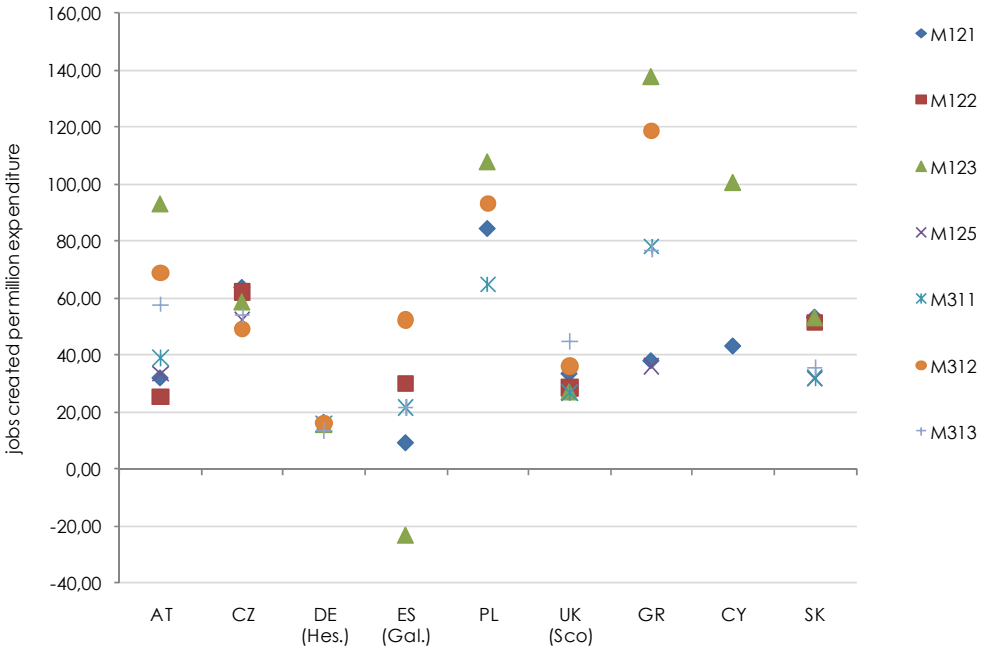
The most frequently used result of IO analysis is the employment effect, measured as number of jobs created per million EUR of public expenditures. This is the indicator for *efficiency* in this analysis. The summary shows that it is significantly different across measures and across countries. All the differences can be explained by looking closely to the intermediate results that are the basis for the indicator reported in this table.

Among the reasons for different results per million expenditures are different structure of the economy, the intensity of the shocks for different sectors, and the assumptions about the *capacity effect*.

The level of the indicators on effectiveness and impact also depends on the levels of targets. Because the targets are set for very specific situations that are hardly comparable across the EU, these results are not well suited for comparisons across countries. The results are of major interest for Managing Authorities and stakeholders.

Efficiency as defined here is an indicator that is well suited for comparisons across different regions (see Figure 13). For the interpretation of the results it is important to consider that inputs that drive the results presented here are based on data about effects on farms or other recipients that may be biased upwards or downwards (see last section in this chapter).

Figure 13. Comparison of IO results on *efficiency* (jobs created per million expenditure) for investment support measures on efficiency across regions



Source: Elaboration of Sinabell/Morawetz based on Psaltopoulos (2014).
Note: The indicator of efficiency is calculated by dividing the budgetary inputs mobilised by the quantity of effects (jobs) obtained.

5.2.4 MAPP results on EQ2

In order to interpret the results presented in this section, it is necessary to recall that the results refer to outcomes each RDP measure had on specific indicators. The overall performance of the indicator is of no interest. For instance, jobs may have increased as a result of Pillar 1, but the impact of the RDP measure may have been small or negative. This has to be borne in mind when the results presented below are interpreted.

Results of MAPP can be presented in two ways, either as narratives (using a nominal scale) or by indicating changes and directions of changes (ordinal scale). The results of a participatory process in meetings where MAPP was applied are documented in tools. Excerpts of them were made in order to summarise the results for a summary table. In this summary table the original definition of indicators and sub-indicators was given as used by the participants of the group. In each of the cases the indicators used were (slightly) different. This was due to different choices made by each MAPP group on what is relevant/important to measure in their area or what indicator they had better knowledge of. For instance, under environment, some groups chose to analyse water quality, while others looked at water quantity. Similarly, some MAPP groups distinguished the impact on jobs between investment support beneficiaries and non-beneficiaries, while some groups did not make this distinction (possibly due to the lack of representatives of non-beneficiaries amongst participants).

In order to be able to summarise the results so that they can be used to answer EQ2 together with the other results, terms had to be redefined and made more general in some instances. The number of indicators and sub-indicators presented here is very large, in particular larger than in each of the case studies. This decision was made in order to maintain the range of subtleties that were developed during the discussions (see left most columns in Table 35).

The groups involved in working out the results made deliberate choices on important aspects. Some groups differentiated between beneficiaries (coded as b) and non-beneficiaries (coded as n-b). Other groups made statements on impacts without differentiating between beneficiaries and non-beneficiaries (no code used in such a case). Some groups were very specific on changes of the environmental status, others made very general statements. Table 35 acknowledges the views of experts in the regions by maintaining as much detail as possible without sacrificing the primary goal of this chapter, to provide a consistent, comparable overview.

The table is structured in the following way. In the first column starting from the left, the indicator is listed (see column indicator). Most of the indicators for which MAPP provides results are *impacts*. In the second column sub-indicators are listed (see sub-1). They stand for important outcomes like jobs, environmental quality, income and the like (see column sub-2). A further differentiation is given in the fourth column (labelled n-b and b) which indicates whether results hold for non-beneficiaries (n-b), for beneficiaries (b) or for both of them (neither n nor n-b).

The remaining columns list the measures by region using measure and region codes (121 GR stands for measure 121 in Greece).

The table shows at a glance, that investment support measures have effects on a wide range of variables of interest. In most cases economic and environmental aspects are concerned. The MAPP therefore may provide insights that go beyond more narrow methods that focus just on very specific outcomes.

Table 35. Aggregated results of MAPP

Indicator	sub 1	sub 2	n-b b	121 GR	121 UK	121 SK	121 CZ	121 ES	123 GR	123 UK	123 SK	123 CZ	123 ES	125 GR	125 SK	125 ES	Axis 3 GR	Axis 3 ES	311 UK	311 CZ	122 ES
Efficiency							1					1								1	
Effectiveness																					
impact	jobs	agr/prim	b		0	-1				0	-1				-1				-1		
impact	jobs	agr/prim	n-b		-1	0				-1	0				-1				-1		
impact	jobs	agr/prim		-1			-1		-1			0	1	-1		0	-1	0		1	1
impact	jobs	non agr		0			0	0	0			0	1	0		0	0	1		0	0
impact	jobs	non agr	b		1					1									-1		
impact	jobs	non agr	n-b		-1					-1									-1		
impact	incomes	agr/prim	b	0	1	0			0	-1				0			0		-1		
impact	incomes	agr/prim	n-b	0	0	0			0	-1	-1			0	-1		0		0		
impact	incomes	agr/prim					1	1					0			0		0		1	1
impact	incomes	non agr	b			0					0				-1						
impact	incomes	non agr	n-b			0					1				-1						
impact	incomes	non agr						1				1	1			0		1			0
impact	output / tourism	quantity			0		1	1		0		1	1			1		1	-1	1	1
impact	output	quality					1					1								1	
impact	comp./ prod.	agr/prim					1			0											
impact	comp./ prod.				0			1				1	1			1		1	-1	0	0
impact	lower costs		b	0					1					1			1				
impact	lower costs		n-b	0					1					1			1				
impact	enterprises					0					0										

Indicator	sub 1	sub 2	n-b b	121 GR	121 UK	121 SK	121 CZ	121 ES	123 GR	123 UK	123 SK	123 CZ	123 ES	125 GR	125 SK	125 ES	Axis 3 GR	Axis 3 ES	311 UK	311 CZ	122 ES
impact	RD aspects	general																			
impact	environment	quality general		1	-1	a		1	1	0	0			1		1		0	0		1
impact	environment	biodiversity			0		-1	1		0		-1				0		0	0	-1	1
impact	environment	energy prod.										-1								1	
impact	environment	ener. int./eff.				1	a			0	0	1							0	1	
impact	environment	airpollution					a					1								-1	
impact	environment	soilquality					0					-1								-1	
impact	environment	water quality				0	1				1	-1							0		
impact	environment	water quantity			-1	-1					1									0	
impact	environment	other					1					-1								-1	

Source: summary by Sinabell and Morawetz based on Parissaki (2014).
Abbreviations: RD – rural development; b – beneficiary; n-b – non-beneficiary; agr/prim – agriculture and primary sector; non-agr – non-agricultural sectors; comp./prod – competitiveness and productivity; energy prod. – energy production (often renewable); ener.int./eff.- energy intensity and efficiency.

Coding procedure: increase: 1, decrease: -1, response of "no change" or "no impact": 0, a: ambiguous (e.g. "increase" of "reduction of energy intensity"); no answers given or not considered: grey empty field (each discussions in focus groups follows its own dynamic; therefore not all aspects show in the overview were covered in each focus group.),

The indicators measure the effect of investment measures in the regions under consideration. If a given indicator has increased, it is labelled "1" if it decreased it is labeled "-1", no change and no impact is labeled "0", indicators for which no information was available or which were not considered are grey empty fields. The coding is following the rule to be consistent across rows, columns and interpretation. A "1" generally indicates an *improvement* or a change in the direction that was intended by the programme. Hence, "1" indicates an "increase" (more incomes). In order to maintain consistency "1" also indicates lower costs (which is the intended outcome). Some "decrease" (less pollution) were translated into "increase" for another indicator (e.g. increase of environmental quality). Such kinds of recoding were necessary in order to provide consistent results across all cases.

In some cases results were ambiguous (labelled a). The ambiguity is mainly due to aggregation and summarising several answers. For convenience, a colour code is given in addition. The colours are **not** providing additional information but improve visibility of the literal and digit codes.

Effectiveness

There are no MAPP results on the effectiveness of investment support measures, since MAPP is a method for measuring impacts. This is due to the application of the method where participants were asked for assessments with regard to impacts of measures in the region under consideration.

Efficiency

MAPP is a not supposed to measure efficiency, it measures impacts. However, in one case study (Czech Republic), a sub-indicator for efficiency for MAPP was defined to be the **revenue to cost relationship**. An improvement (higher revenues per cost) is observed for the measure 121, 123 and 311 (see overview in Table 35). Because MAPP provides no results on cardinal scales we do not know whether the *efficiency* gains were large or small. Because all other cells are empty and greyed out we do not know how efficient investment support was in the other cases under consideration.

Impact

The results of MAPP are related to *impacts* of the Rural Development Programme. The overview shows that in each case study and with respect to each measure, there are significant differences between the regions. Because the results were not derived in a representative manner one has to interpret the results as the views of the groups working together and reaching a consensus.

Another aspect is that effects of measures are in most cases not indicating improvements for each of the indicators of interest. Participants in the groups working with MAPP obviously expressed their views on trade-offs and countervailing effects. This result is very insightful because trade-offs are frequently ignored when single measures are evaluated.

MAPP also produces results with respect to other factors that influence indicators. Indicators may not show improvement as a result of investment support but as a result of other programmes/measures or external factors. This is valuable information for policy makers.

For instance, if Pillar 1 proves to have more impact on incomes than investment support measures, this information should be used when designing policy. Likewise, if

Axis 3 measures prove to have more impact on jobs than Axis 1 measures, this is also important when policy objectives focus on employment creation.

5.2.5 Results of PSM to answer EQ2

For the counterfactual econometric analysis the propensity score matching method (PSM) was implemented in Poland (121) Austria (121, 122, 123, 125, and 311) Czech Republic (121, 122, 123, and 311) Slovak Republic (121, 311) and Germany (Hessen) (121). The method could not be tested in Galicia (Spain) due to data access problems.

As for all other methods, the CMEF definitions and the nature of PSM imply how *efficiency*, *effectiveness* and *impact* are defined and can be measured. The results discussed here are organised along these terms and therefore it is first explained how these key terms for EQ2 are specified in the PSM case studies.

Efficiency

Efficiency gives the relationship between resources employed and outcomes achieved in pursuing a given objective through an intervention. Efficiency as defined by the CMEF addresses the question whether more effects could have been obtained with the same budget or whether the same effects could have been obtained at a lower cost. This can be done at micro-level and at programme area level. It is measured as the *effect of the measure divided by support* obtained (e.g. additional GVA due to the measure divided by total *public* funds for the measure).

Effectiveness

This is the extent to which objectives pursued by an intervention are achieved. In order to assess programme *effectiveness* one has to ensure causality between a change of relevant outcome indicators and the programme.

Effectiveness should be assessed at the beneficiary level (= micro-level):

- Outcomes achieved by programme beneficiaries compared to target values. This is equivalent to an increase of a given result indicator due to a given measure compared with target values.
- Outcomes achieved by programme beneficiaries compared to outcomes achieved by programme non-beneficiaries (in %). This is equivalent to an increase of a given result indicator for programme beneficiaries compared to an increase of the same result indicator for the control group.
- Structure of a total increase of a given result indicator (% share due to a given measure compared to % share due to other factors).

Impacts

Impacts are - following the terminology of the CMEF - causal effects of an intervention lasting in medium or long-term and refer to the effect of the programme beyond the immediate direct beneficiaries at the level of the intervention. Such effects refer to the benefits of the programme beyond the immediate effects on its direct beneficiaries and should be calculated at the level of programme area only.

Impacts produced by a programme intervention may be positive or negative, primary and secondary, expected or unexpected. They are normally expressed in “net” terms. “Net” means: after subtracting of effects that cannot be attributed to the intervention

(e.g. influence of confounding factors), and by taking into account indirect effects (e.g. displacement, multipliers, deadweight).

5.2.5.1 Overview of PSM indicators

Results on *efficiency, effectiveness* and *impact* are calculated using four indicators:

a) gross value added (GVA) or gross farm income (GFI);
b) family farm income (FFI)
c) labour productivity and
d) farm employment measured in annual work units (AWU) which is equivalent to the work of one person, full time, for one year.

Table 36 presents a selection of PSM results on *efficiency* and *impact.* Table 37 gives an overview of PSM indicators for *effectiveness*. The tables summarise the indicators for all measures and countries that were analysed in this study. In the discussion of the results below the focus is on measure 121 as this is the most important measure in terms of participants and funds in most countries. In addition, it has to be considered that measure 121 is focused on agricultural holdings for which excellent micro-data are available in many countries.

Table 36. Overview of PSM results on efficiency and impact

Country		PL	AT	CZ	SK	DE	AT	AT	CZ	AT	CZ	AT	CZ	SK
Measure		121	121	121	121	121	122	123	123	125	122	311	311	311
	Total public funds (in Mill EUR)	1,487	467.5	265	263.4	99	34.4	124.2	65.1	66	74.6	31.8	99	36
Indicator	Description	**Efficiency**												
1a micro-level:	1 EUR support brought an increase of GVA in EUR per farm	0.20	0.37	0.10	0.03	negative	negative	negative	0.53	0.26	0.78	0.83	0.39	0.06
1b micro-level:	1 EUR support brought an increase of FFI in EUR per farm	0.14	0.16	0.16	nc	nc	negative	0.13	negative	0.7	0.85	1.05	0.44	nc
1c micro-level:	1 EUR support brought an increase of labour productivity per farm by...(EUR)	0.03	0.09	0.01	negative	negative	negative	negative	negative	0.70	0.60	0.41	negative	
1d micro-level:	number of AWU induced by 1 million EUR public expenditures per year	nc	23.24	negative	100.63	3.60	84.70	44.21	45.88	negative	469.04	17.80	20.74	42.03
2a: programme area level	1 EUR support brought an increase of GVA in EUR per farm	0.16	0.43	0.11	0.04	negative	negative	negative	0.09	0.10	0.06	0.48	0.42	0.06
2b programme area level	1 EUR support brought an increase of FFI in EUR per farm	0.11	0.20	0.17	nc	nc	negative	0.01	negative	0.26	0.07	0.20	0.48	nc
2c programme area level	1 EUR support brought increase of labour productivity per farm by...(EUR)	0.03	0.11	0.01	negative	negative	negative	negative	negative	0.25	0.05	0.60	negative	0.03
2d programme area level	number of AWU induced by 1 million EUR public expenditures per year	nc	4.58	negative	24.23	0.40	9.77	44.21	1.49	negative	7.27	1.70	4.47	8.39
Indicator	Description	**Impact**												
1 . programme area level	Mill EUR (over 6 years) GVA	277	204	28	9.8	-1.1	-37.6	-3.7	5.6	6.4	4.5	15.2	41.5	2
2: programme area level	Mill EUR (over 6 years) FFI or Total Farm Income	190	91.2	45.5	nc	nc	-53.8	0.7	-5.4	17	4.9	19.1	47.4	nc
3: programme area level	AWU (over 6 years)	nc	2141	-996	6383	39	336	42	97	242	542	54	443	302

Source: Summary of Sinabell and Morawetz, based on results of Michalek (2014)
Note: GVA=Gross Value Added or Gross Farm Income, FFI = Family Farm Income, AWU = Agricultural Work Unit, negative="negative", nc=" not calculated". Indicators for the years 2007-2012. Cells shaded in grey are extrapolations from the sample results to the regional scale.

Table 37. Overview of PSM results on effectiveness

Country		PL	AT	CZ	SK	DE	AT	AT	CZ	AT	CZ	AT	CZ	SK
Measure		121	121	121	121	121	122	123	123	125	122	311	311	311
Indicator	Description	Effectiveness												
1a micro-level	increase in GVA % (% of target)	19.1 (16%)	18.5 (93%)	4.3 (NA)	4 (NA)	-3 (NA)	-14.4 (NA)	-10 (NA)	11.9 (NA)	6.6 (NA)	0.09 (NA)	57 (NA)	34 (NA)	9 (NA)
1b micro-level	increase of FFI % (% of target)	19 (NA)	6 (NA)	3.9 (NA)	nc	nc	-14 (NA)	2.7 (NA)	-4 (NA)	12.5 (NA)	0.06 (NA)	43 (NA)	21 (NA)	nc
1c micro-level	increase in labour prod. % (% of target)	9.2 (NA)	8.7 (NA)	9 (NA)	-70 (NA)	-4 (NA)	-12 (NA)	-1.6 (NA)	-0.26 (NA)	27.6 (NA)	1.44 (NA)	47 (NA)	-1.2 (NA)	100 (NA)
1d micro-level	increase in employment % (% of target)	nc	4 (NA)	-1.6 (NA)	13 (NA)	3 (NA)	3 (NA)	2.7 (NA)	1.3 (NA)	-5.4 (NA)	15 (NA)	4 (NA)	4 (NA)	7 (NA)
2a micro-level	GVA: % increase beneficiaries & control	64 & 45	37 & 18	8.8 & 4.5	25.3 & 23.2	24.7 & 26.9	6 & 21	20 & 31	80 & 61	9 & 2	0.2 & -11	85 & 31	46 & 12	27 & 21
2b micro-level	FFI: % increase beneficiaries .& control	69 & 50	23 & 17	10.1 & 6.2	nc	nc	6 & 9	23 & 21	24 & 46	8 & -4	12 & 7	64 & 25	31 & 10	nc
2c micro-level	labour prod.: % increase beneficiaries & control	56 & 49	37 & 29	24.5 & 14.8	158 & 295	25.8 & 29.4	9 & 23	25 & 28	22 & 46	24 & -4	112 & -10	83 & 40	23 & 33	100 & 25
2d micro-level	employment: % increase beneficiaries & control	nc	-0.01 & -4.3	-15.1 & -13.7	-45 & -56	15.6 & 13.3	-0.1 & -3.2	0.01 & -2.5	-10 & -13	-1 & 4	-2 & -18	-3 & -6	-11 & -15	-25 & -30
3a micro-level	GVA % or EUR change due to measure & other factors	30 & 70	50 & 50	49 & 51	15 & 85	-1,975 & 18,096	-5,640 & 8,014	-4,318 & 13,193	15 & 85	75 & 25	10,511 & 10,282	67 & 33	73 & 27	35 & 65
3b micro-level	FFI/GFI % or EUR change due to measue & other factors	28 & 72	26 & 74	38 & 62	nc	nc	-8,071 & 4,725	7 & 93	-13,453 & 93,538	6,403 & -2,369	46 & 54	67 & 33	68 & 32	nc
3c micro-level	labour prod. % or EUR change due to measure & other factors	16 & 84	23 & 77	37 & 63	-1,977 & 6,415	-1,122 & 9,012	-2,819 & 4,834	-365 & 5,877	-2,995 & 5,555	6,302 & -860	8,088 & 1,795	56 & 44	-206 & 4051	88 & 12
3d micro-level	AWU or % change due to measure & other factors	nc	0.078 & -0.081	-10 & -90	7.7 & -34.8	19 & 81	0.05 & -0.07	50 & -50	0.24 & -2.24	-0.09 & 0.07	1.26 & -1.45	0.06 & -0.11	1.98 & -7.61	3.37 & -17

Source: Summary of Sinabell and Morawetz, based on results of Michalek (2014)
Note: GVA=Gross Value Added or Gross Farm Income, FFI = Family Farm Income, AWU = Agricultural Work Unit. NA = "target value not available" Effectiveness indicators are for the years 2007-2012.

5.2.5.2 PSM results on EQ2

Efficiency

With regard to efficiency, results suggest that for each EUR of public funding spent on the analysed measures, GVA and family farm income increased by substantially less than one Euro (compare Table 36). This was found for all regions and measures except for measure 311 in Austria (which has an efficiency of family farm income of 1.05). For example for measure 121 in Poland, 1 EUR support from public funds on measure 121 allocated to programme beneficiaries at a farm level resulted in:

- 20 cents of increase of GVA[49](20% efficiency);
- 13.7 cents increase of family farm income (13.7% efficiency);
- 3.4 cents increase of agricultural labour productivity (3.4% efficiency);
- These values are estimates for the period 2007-2012, meaning that one EUR public funding invested during this period lead to 20 cents increase in GVA. For the interpretation of this result it is important to consider that investments might come to effect with a time lag and some investments were made earlier and some later during the programming period. It thus is an **average effect over time and beneficiaries** which are heterogeneous. It is not the sum of added up GVA.

Still focusing on Poland as example, an estimated efficiency of total public expenditures spent on M121 (1,706 million EUR by the end of 2013) at a programme area[50] level was found to be even lower than at farm level, i.e.:

- GVA 16.2% (i.e. 276.6/1,706);
- Family Farm Income 11.1% (189.6/1,706);
- Agricultural labour productivity 2.7% (46.8/1,706);

The reasons for the low efficiency of public support under M121 (considerably less than 1 EUR for 1 EUR of public spending) cannot be explained by PSM. It can be assumed though that the most important one was that in the majority of cases the main beneficiaries of this measure were farms that were on average economically stronger and bigger than the average farm, or farms which were located in areas where access to bank credit was not restricted. In both situations beneficiaries represented those farms which also in a "without support situation" were not credit constrained, i.e. could have an "unproblematic" access to banking loans, if this was economically required. From a perspective of farm development, a high effectiveness and efficiency of *public* investment could be expected in a situation of capital market imperfection, i.e. when access to commercial loans had been restricted, yet the above investments were necessary from an economic point of view. Apparently, in all examined countries this was **not** the case *for programme beneficiaries*. In addition a low efficiency of public support can be better understood by looking at the manner of how these effects are computed, i.e. by comparing effects observed for programme beneficiaries with a similar control group of those who did not received such a support. Clearly, low positive or negative effects of a given measure (e.g. on GVA) mean that a similar control group was able to achieve comparable results even without public support. This means that similar investments would have been carried out also without public support (if it was

[49] More precisely "Gross Farm Income" income was used instead of GVA in case the Polish case-study.
[50] In order to compare the outcomes at programme level it is necessary to extrapolate results from the sample to the whole region.

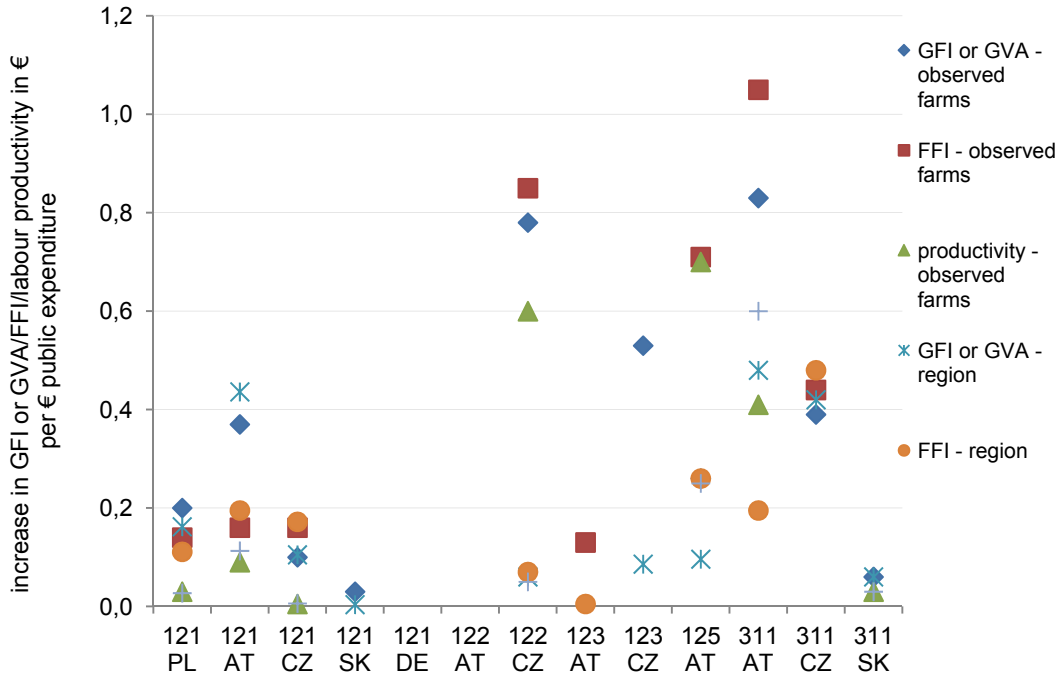
economically necessary) or no additional investments were necessary in order to increase selected result indicators (e.g. GVA/profits). In both cases, public support was redundant, i.e. had similar investments been carried out also without public support means the deadweight loss effects have to be assessed as high or substantial. Alternatively, had a result indicator (e.g. GVA) increased less for programme beneficiaries in comparison with a similar control group who did not invest in an examined period it means that supported investments were not necessary. They were carried out mainly in order to enhance a medium/long-term market position of a beneficiary farm by taking advantage of a higher internal rate of return to investment. One possibility to cross-check whether public support was only a substitution (crowding out) of private investments or if it resulted in additional investments is an analysis of farm leverage effects[51]. In Hessen, for example, there is no evidence that support of farms by M121 lead to an increase of GVA and agricultural labour productivity. On the contrary, estimates show that due to M121 GVA decreased by -3%, and labour productivity by -4% (in comparison to similar control group which did not receive support). At the same time due to M121 farm employment increased slightly (by 3.6 AWU over 6 years for a million EUR public expenditures). However, there is evidence that support of farms from M121 lead to a significant increase of transfers from farm to private consumption (i.e. leverage effect was high). The latter grew by 39% over 6 years (not shown in the table).

Turning to measure 121 in Austria: efficiency of support regarding GVA is estimated to be 37%, FFI 16% and agricultural labour productivity 9%, respectively. Furthermore, public funds of 43,032 EUR per year on M121 resulted in one additional job being created (this is equivalent to 23.24 jobs per million EUR of public expenditure). The number of jobs per million EUR is slightly lower if estimated at the programme area level.

It is important to keep in mind that all efficiency measures are the average effect on those who received funding. It is possible that the measures had very different effects on different types of farms. These differences can be for example with respect to magnitude and/or with respect to time when the effects occur. There is no information with respect to heterogeneity of efficiency in these figures. They are purely the estimated average effect of this particular programme at the year in which the data were analysed.

[51] A detailed analysis of interdependences between deadweight loss (crowding out) and leverage effects in case of RDP in Germany/Schleswig Holstein can be found in: Michalek J. (2012), "Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", Volume 1 – A micro-level approach.", European Commission, JRC Scientific and Policy Reports, pp 1-95 and Michalek J., P.Cialan and Kancs d'A., (2014) "Firm level evidence of crowding out effects of investment support policies", unpublished manuscript.

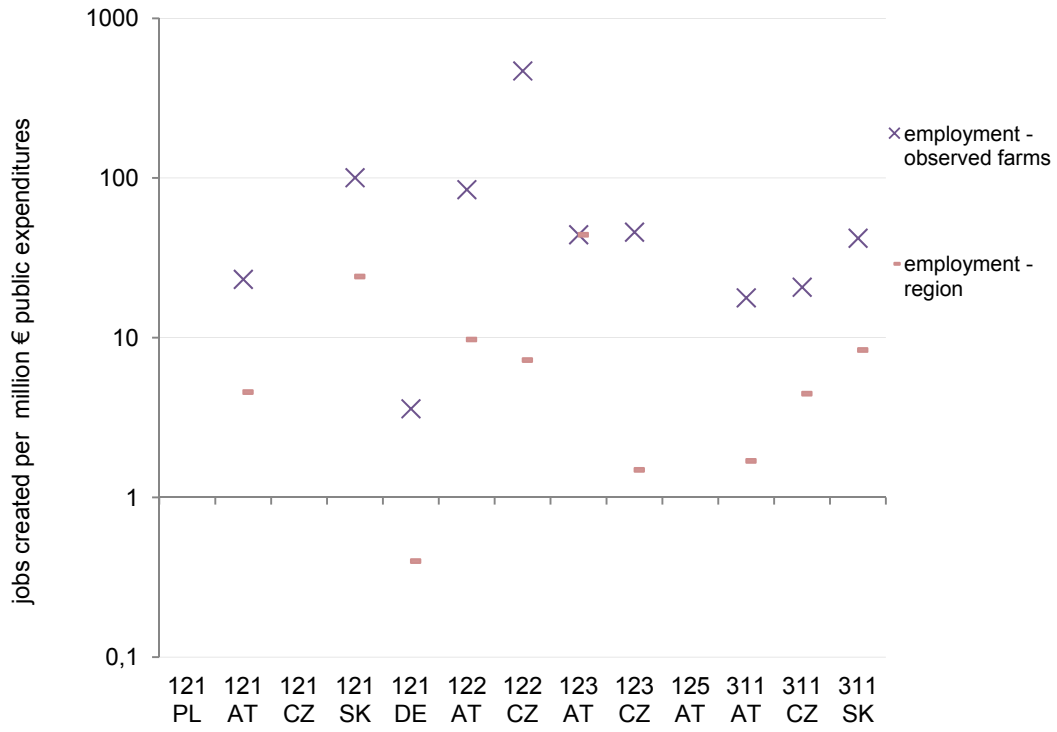
Figure 14. Efficiency of investment support measures based on PSM estimates



Source: Sinabell and Morawetz based on Michalek (2014); Note: detailed description of indicators is given in the tables above.

Figure 14 is a graphical representation of the increase of gross farm income (GFI) or gross value added (GVA) or family farm income (FFI) or labour productivity of beneficiaries in EUR per EUR public expenditures. Observations in the sample of farms are indicated by "observed farms", extrapolations for the average effect on regional scale are indicated by "region".

Figure 15. Efficiency indicator on employment effects based on PSM estimates



Source: Sinabell and Morawetz based on Michalek (2014)
Note: detailed description of indicators is given in the tables above

This figure is a graphical representation on how many agricultural jobs (AWU) have been created per million EUR public expenditures. Observations in the sample of farms are indicated by "observed farms", projections for the average effect on regional scale are indicated by "region".

Effectiveness

The three types of PSM indicators for *effectiveness* are quite different in form (compare Table 37). The first and most intuitive one (% of target value achieved) is available only for measure 121 for the GVA indicator in Poland and Austria.

This is due to the fact that target values from other Member States are unavailable. Poland had an estimated increase in GVA[52] of 19.1% and Austria an increase of 18.5%. While the increase in each of the countries is similar, targets were quite different[53] which results in Poland reaching 16% of the target and Austria 93% of the target. Comparing the percentage increase in GVA, FFI and labour productivity, the increase is mostly higher in Poland and Austria than in the other countries/regions where PSM was applied to quantify the effects of measure 121.

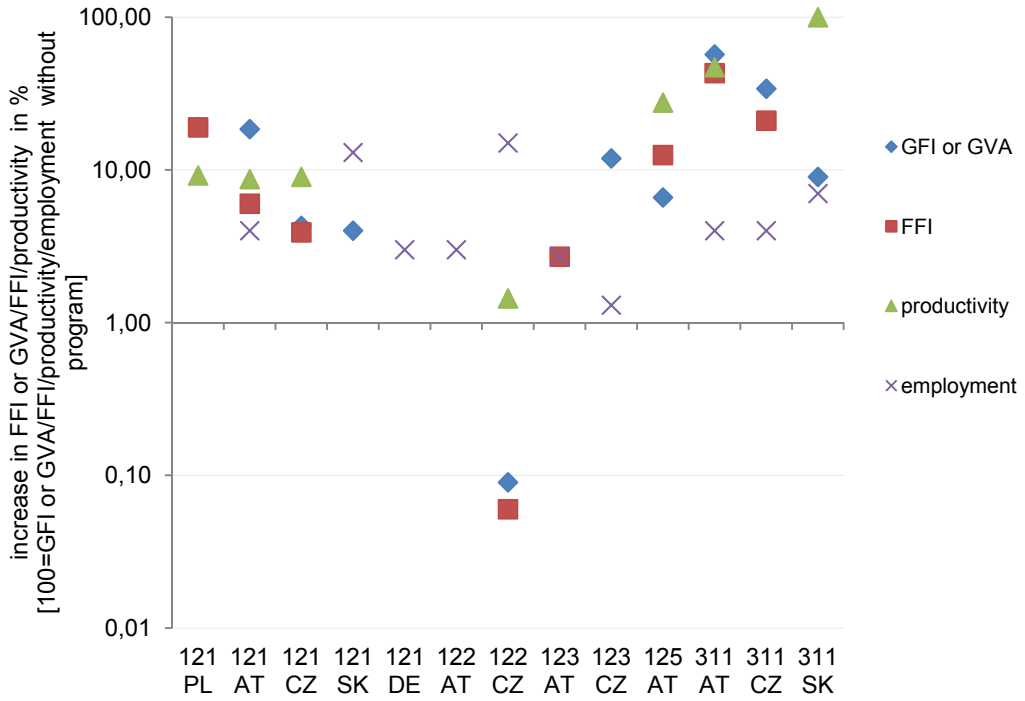
[52] More precisely "Gross Farm Income" income was used instead of GVA in case the Polish case-study.
[53] +119% for Poland and + 19.8 for Austria

The second indicator type compares the change of indicators observed from beneficiaries (participants of a measure) with the change of indicators observed from non-beneficiaries (control group = non-participants). For M121 the increase among beneficiaries was higher (or less negative) than for non-beneficiaries in all analysed regions except for Hessen (for GVA and labour productivity).

The third type of effectiveness indicator is related to the last one and explains how much of the percentage change observed among the beneficiaries is due to the measure and how much is due to other factors. For M121 in Poland, for example, 30% of a total increase of GVA at farm level can be attributed to M121 (70% due to other factors); 28% of a total increase of FFI was due to M121 (72% due to other factors); and 16% of a total increase of agricultural labour productivity could be attributed to M121 (84% was due to other factors). For Austria, 50% of a total increase of GVA at farm level was due to M121 (another 50% due to other factors).

Figure 16 shows *effectiveness* indicators 1a-1d estimated with PSM. The figure allows a comparison across measures and countries. The figure has to be interpreted with care as, due to the lack of target values, no real efficiency is shown but only changes in percentage (compare with first four rows of Table 37).

Figure 16. Effectiveness indicator based on PSM estimates



Source: Sinabell and Morawetz based on Michalek (2014)
Note: description of indicators is given in the tables above

The figure above shows the observed increase of gross farm income (GFI) or gross value added (GVA) or family farm income (FFI) and increase in farm employment (number of AWU) of beneficiaries in percent compared to the counterfactual situation without a programme.

Impact

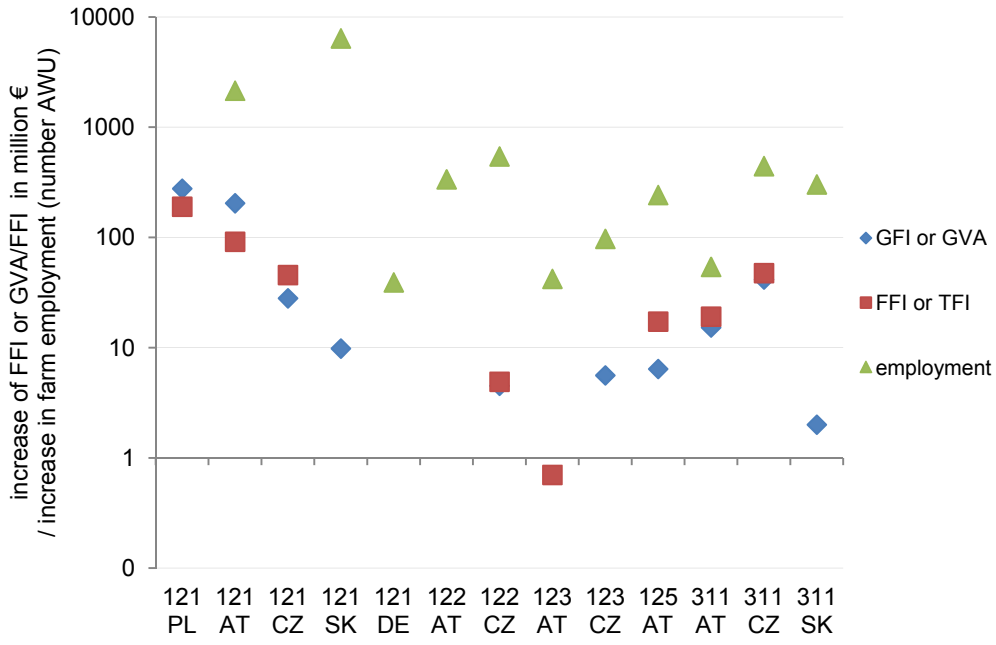
Impacts can be found in Table 36. In order to calculate impacts of the programme it is necessary to extrapolate results from the sample to the region as a whole. The indicators are estimates of the effect of the total public support at country/regional/programme level. The total public funds are reported in the first row of the table. For example for Poland total public funds spent on M121 by the end of 2012 at 1,487 million EUR resulted in an:

- increase of GVA by +276.6 million EUR;
- increase of FFI by +189.6 million EUR;
- increase of agricultural labour productivity by +46.8 million EUR;

For Austria, at the country level total public funds spent on M121 at 467.5 million EUR resulted in an increase of gross value added by 204 million EUR.

Figure 17 gives a summary over the estimated programme *impacts*. The different size in GDP, population and development of the agriculture explains the wide range of different impacts the measure had at an aggregated level. Negative values (e.g. indicator 1 for M123 in Austria) are not reported here.

Figure 17. Impact indicator based on PSM estimates on a log-scale



Source: Sinabell and Morawetz based on Michalek (2014)
Note: description of indicators is given in the tables above

The figure shows the observed increase of gross farm income (GFI) or gross value added (GVA) or family farm income (FFI) or total farm income (TFI) of beneficiaries in million € (over six years) and increase in farm employment (number of AWU) in the programme region compared to the counterfactual situation without a programme.

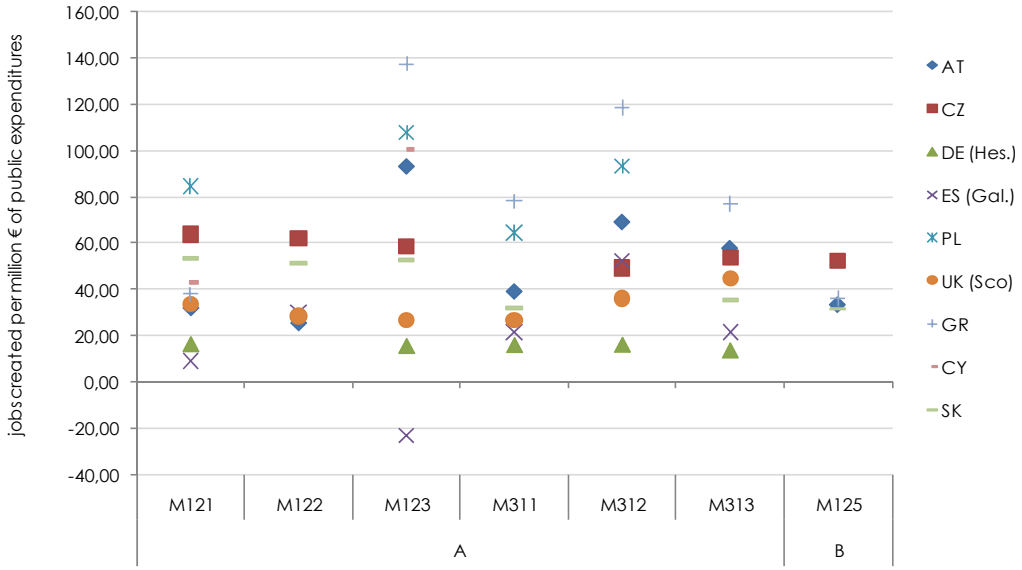
5.2.6 Comparing the efficiency of types of measures

Grouping measures according to whether they are supporting "productive investments" (Type A), "public infrastructure" (Type B), "non-productive investments" (Type C) or "investments to meet minimum standards" (Type D) adds an additional dimension to the results reported in the previous sections. If the measure is very well described by the characteristics of the group it belongs to, one would expect differences in efficiency across types. For example, in the short run Type A measures might have higher efficiency on GVA than Type C measures.

The comparison of efficiency estimates from EQ2 across measure types is one way to find answers to this hypothesis. Using qualitative results which have ordinal or normative scales is likely to be inconclusive and is therefore not attempted here. Using only quantitative results from the IO and PSM method, the results are restricted to measures of Type A and Type B. Two interesting indicators are efficiency of job creation and raising GVA. It allows investigating how "productive investments" relate to "public infrastructure" investments with regard to job creation and higher GVA.

Figure 18 provides an overview about efficiency estimates from IO showing the jobs created in the region *per annum* per million EUR of public funds. There are five measures of Type A and only one measure of Type B (M125). The visual comparison does not support the hypothesis that efficiency in job creation differs between Type A and Type B measures. Due to the small number of observations a statistical appraisal would not be meaningful.

Figure 18. Efficiency for job creation from IO grouped by measure types



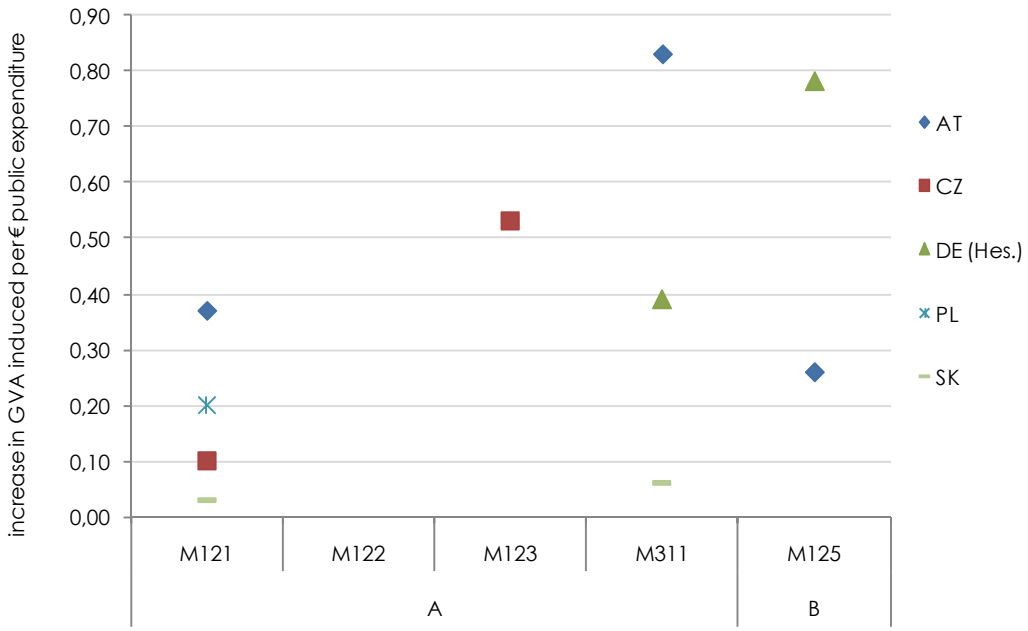
Source: Sinabell and Morawetz based on Psaltopoulos (2014)
Note: The ordinate (vertical axis) measures jobs created in the region *per annum* per million Euro of public funds.

Figure 19 shows the results from the PSM indicator 1a measuring efficiency of the measure in increasing the GVA at farm level. It is interpreted as the increase in GVA

induced by a EUR spent on the respective measure. Again, M125 is the only measure of Type B.

The two efficiency measures for M125 are quite different in magnitude with the value from Austria being less than half of the German value. The variance is thus quite high and no conclusion on the relation between measure type and efficiency is permissible.

Figure 19. Efficiency of GVA estimated with PSM indicator 1a



Source: Sinabell and Morawetz) based on Michalek (2014)
Note: The ordinate measures the increase of GVA induced by a EUR of public funds spent on the respective measure.

5.2.7 TBE results on EQ2

The results of the TBE method are documented in an extensive five-step tool used to derive the results through the combination of theory and empirical evidence. The last step of the procedure is the judgement on the *effectiveness* of projects supported in relation to the main outputs. These judgements are on an ordinal scale with five steps (from “very low” to “very high”). This judgement summarises one aspect of the TBE, disregarding the mostly verbal descriptions of further details. As EQ2 is focused on *efficiency*, *effectiveness* and *impact* the discussion here is focused just on the ordinal judgement summarising the finding with regard to this aspect.

The objectives are measure-specific and differ between measures. To be able to summarise the results so that they can be used to answer EQ2, objectives had to be redefined and made more general in some instances. In addition efforts were made to keep groups similar to indicators used in the MAPP method to increase homogeneity across methods (see previous section). In some cases it was also possible to aggregate objectives (e.g. if *effectiveness* to achieve more competitiveness in all analysed sectors was “low” this was taken to judge that competitiveness is “low”). Consequently summary statistics such as means are not meaningful for interpreting the results.

Table 38 lists the aggregated TBE results on the effectiveness of support measures and is structured in the following way: In the first column, the objectives are listed (see column objective).

They stand for important outcomes like jobs, environmental quality, income and the like. In the second column sub-objectives are listed, if necessary to differentiate between various aspects.

The remaining columns list the results by measure and region using measure and region codes (121 CY stands for measure 121 in Cyprus). All objectives for which TBE provides results are related to *effectiveness*.

The table shows at a glance, that competitiveness is the main concern of M121. It is of interest to note that environmental objectives were part of the intervention logic in three of the five measures including 121. Measures of the type of M227 instead, are exclusively focusing on environmental improvements. Worthwhile noting is that tourism is considered as an output variable which explains why M313 only have output related objectives.

If the effectiveness to reach a given objective in the respective region is considered to be "very low" the indicator of achieving it through the measure is indicated as "very low". Situations where the effectiveness of the measure to reach a goal are “very low” or “low” are coloured red. Those with medium effectiveness are “yellow” and those with “high” or “very high” effectiveness are green. “NA” means that the expert could not judge the effectiveness.

Efficiency

There are no TBE results on the *efficiency* of investment support measures.

Effectiveness

Effectiveness results of the measures analysed by TBE ranges from “very low” to “very high”. Only three environmental objectives are achieved with “very high” effectiveness. Only one measure has scored “very low” for all objectives and effectiveness (DE 121). A closer look at Table 38 reveals that the effectiveness is more related to the measure than to the objective. This is most likely the consequence of the objectives being more heterogeneous than the table suggests (as it was necessary to summarise the objectives as explained above). A country bias may be present as well, but it cannot be verified with statistical tools because the sample is too small.

Impact

There are no results on the *impact* of investment support measures.

Table 38. Aggregated TBE results on the *effectiveness* of investment support measures

		CY 121	CZ 121	DE 121	DK 121	FR 121	PL 121	CY 123	DK 123	FR 216	CY 227	CZ 227	DE 227	FR 227	CZ 311	CY 313	CZ 313	FR 313
Jobs	Quantity								4									
Output / tourism	Quantity								2							3		3
Output/ tourism	Quality	3			3		2	4	3		4						3	3
Income								3										
Competitivenesss	Structure	4	3	1			2	NA			4							
	Technology	3	4	1	4	4	1	3	4									
	Innovation	2		1	4		2		3									
	Lower costs	4	3	1			2											
Diversification							2											
Environment	General						1											
	Environment	2			5				4		5		4	3				
	Energy production						2					3			3			
	Air		3				2											
	Soil																	
	Water quality		3				2			1								
	Other										4							
	Animal welfare		5				2											

Source: Sinabell and Morawetz, based on Resch (2014)

Note: 1 - “very low” effectiveness in red cells, 2: “low” effectiveness in red cells, 3: medium effectiveness in yellow cells, 4: “high” effectiveness in green cells and 5: “very high” effectiveness in green cells; “NA” - expert could not make judgements; no information available - grey cells.

5.2.8 Contrasting results and complementarity between methods

It is not the primary objective of this analysis to identify if results on the same combination of measures and regions obtained by different methods are the same or whether they deviate. The fact that the same measures are analysed in different regions with more than one method calls for an explanation because several results are in fact different.

Contrasting results: Some reasons why methods are not flawed even if their application leads to contradictory results at first glance

Contrasting results are highlighted by looking closer on indicators of job creation due to investment support measures because such indicators are measured with all the methods under consideration except CEA and SEA (which focus on environmental outcomes):

IO: Employment of the regions generally increases; the levels of increase are different among regions (see indicator *efficiency* which measures the jobs created by 1 million of support).

MAPP: In most cases support is reducing farm employment but in some regions farm employment increases. For some regions a differentiation is made between supported and non-supported farms (b for beneficiaries and n-b for non beneficiaries) but in other regions only the farm sector as a whole or the job market as a whole was analysed (indicated by a missing b or n-b).

PSM: This method measures the changes of farm employment (see indicator 1d[54]). The value indicates how much public support is necessary to create one additional job. Results indicate that jobs on farms are in some cases negatively affected. This is an expected outcome after labour saving investments were carried out.

TBE: Employment increases very strongly, but results are only available for Denmark.

Among the reasons that explain results that are conflicting or contradictory are:

The data sources are very heterogeneous:

- For the IO method the input-output tables and administrative data on programme implementation and outcomes were used. In each case adjustments to IO tables needed to be made in order to capture the characteristics of the sector with beneficiaries. The data are representative for aggregates (whole sectors, whole regions, and the whole economy). Data used in IO are taken from various sources but comply with accounting standards and are quality checked.
- For MAPP data were collected by asking focus group members in specific regions. It is evident that methods covering a whole region might deviate from findings of a sub-region. But a more important aspect is the specific way how results were obtained: participants of MAPP focus groups are asked about their perception. Discussions with groups of people lead to the final consensual result. Only one group was consulted in each of the case study regions.
- For the PSM method accountancy data of farms and forestry holdings were used for this analysis. Administrative data on programme participation were linked to the bookkeeping data. The focus is usually on a small sector in the economy and results are specific to it. The data are (more or less - depending on the source) representative for a sector and are available for single farms and firms (micro-data). The focus of the method is on causal effects of measures for specific groups. Therefore results have to be representative for the target groups (programme results are defined as effects on beneficiaries). In many cases the targeted group is only a portion of the whole sector.
- Applying TBE requires using documents and reports as the major source for deriving the results. Expert interviews were used to amend this information base if necessary.

[54] The indicator 1d of PSM can be converted to the efficiency indicator of IO by calculating the inverse and multiplying the result with 1 million. Note that PSM reports data on farms not the whole economy.

The methods are not equally well suited to identify causal effects[55]:

- IO is a standard method to analyse counterfactual scenarios. It can therefore be used to analyse the effect of policy interventions very precisely and at a level of unsurpassed detail. As is the case with any method, the results depend (apart from the skills of the analyst) on the quality of data. IO by itself does not estimate causal effects. To analyse them, it is necessary to look for data from other approaches (e.g. administrative data, literature results and findings of econometric methods).
- For MAPP data are collected by involving focus group members and experts. Such persons frequently have a very good understanding how causal relationships are and which effects one might expect. Unless the persons are explicitly trained in quantitative methods, their judgements may nevertheless be flawed when quantitative results are expected. However, if the persons are well trained in quantitative methods they would not make judgements but look at the results of analyses with micro-data.
- PSM is an econometric method that uses micro-data (a combination of accountancy data and information on participation and non-participation). This method was developed to analyse causal effects.
- TBE uses logical procedures to identify causal relationships. Such a process is often one of the first steps of an analysis that employs quantitative methods.

According to these arguments, there are many reasons why results obtained by different methods may deviate or even be contradictory. However, in order to explain contradictions or deviations it is necessary to analyse a given topic in depth. It would also be necessary to compare every single step to be able to explain why results are method-specific.

A typical example of frequently contrasting results is the yearly statement on farm incomes. There are two major sources, the Economic Accounts of Agriculture (EAA) and the Farm Accountancy Data network (FADN). Results on the change of farm incomes from both sources are different every year and sometimes one source indicates an increase whereas the other indicates a decrease of incomes. There are standard routines to compare results from these sources and contrasting results can be explained very well. Equally established routines to explain differences of methods measuring investment support measures have not yet been developed.

Contrasting results: the case of M121 in CZ

In some of the regions more than one method was used to evaluate the effects of investment support measure. In the Czech Republic even all the methods demonstrated in this study were applied. A short overview of the results on labour market related indicators are presented in the following table.

[55] The methods mentioned here are only those applied in this study. Several other methods are available to identify causal effects such as Randomised Controlled Trials (RCT), Instrument Variable (IV) Regression or Regression Discontinuity Design (RDD).

Table 39. Results on labour market related indicators of four methods applied in Czech Republic and measure 121 (M121)

Method	Indicator	Assessment
IO	number of net jobs created in **all** sectors	… monitoring: 5,926 … PSM results: 5,491
TBE	improved competitiveness (= lower unit costs)	… medium effectiveness
PSM	net jobs created (in agriculture) labour productivity	-996 +9%
MAPP	labour productivity jobs created in agricultural sector jobs created in related industries	… high "impact" … low impact … medium impact

Source: Michalek, Parissaki, Psaltopoulos and Resch (2014)

In the first row of the table are results of the IO method. It measures the effects of M121 in the Czech Republic by summing up two elements: additional demand for investment products and capacity adjustments (= more output of the agricultural sector). Using monitoring data from official sources to shock the IO model results in an additional increase of almost 6,000 jobs. It is important to keep in mind that these figures are for **all** the sectors of the economy. IO uses external data as an input for shocks. If results obtained by PSM are used instead of results from monitoring data to shock the model, the outcomes are a slightly smaller (5,491 instead of 5,926). How these discrepancies can be explained is explored in more depth in the next section.

TBE was not used to measure "jobs created". It focused on (labour) productivity. The result derived from sources on the Czech programme was that the measure achieved a ranking of "medium effectiveness". Comparing this result with those of other methods is not possible because other methods measure efficiency. In the best case M121 can be compared with other measures in the Czech Republic or with similar measures in other countries.

PSM and MAPP are the only two methods that report efficiency on the same indicator, namely labour productivity. The focus group which applied MAPP reached the conclusion that M121 had a positive "high impact" on "labour productivity". PSM found that "labour productivity" increased by +9% which is – with regard to findings in other cases – comparably low. While PSM finds that net jobs in agriculture actually declined due to the measure (by 996 jobs), the MAPP focus group reports that the "impact" of jobs created in the agricultural sector was "low".

A conclusion of this case study on different methods applied for the same investment measure in the same region is that one would expect results that are deviating and sometimes contradictory. Two reasons can be identified:

- the methods do not measure the same thing (e.g. jobs at different levels such as farm level in PSM versus sector level in IO);
- methods use different ways to measure the same thing (e.g. farm accountancy data in PSM versus outcomes of discussion in a single focus group in MAPP).

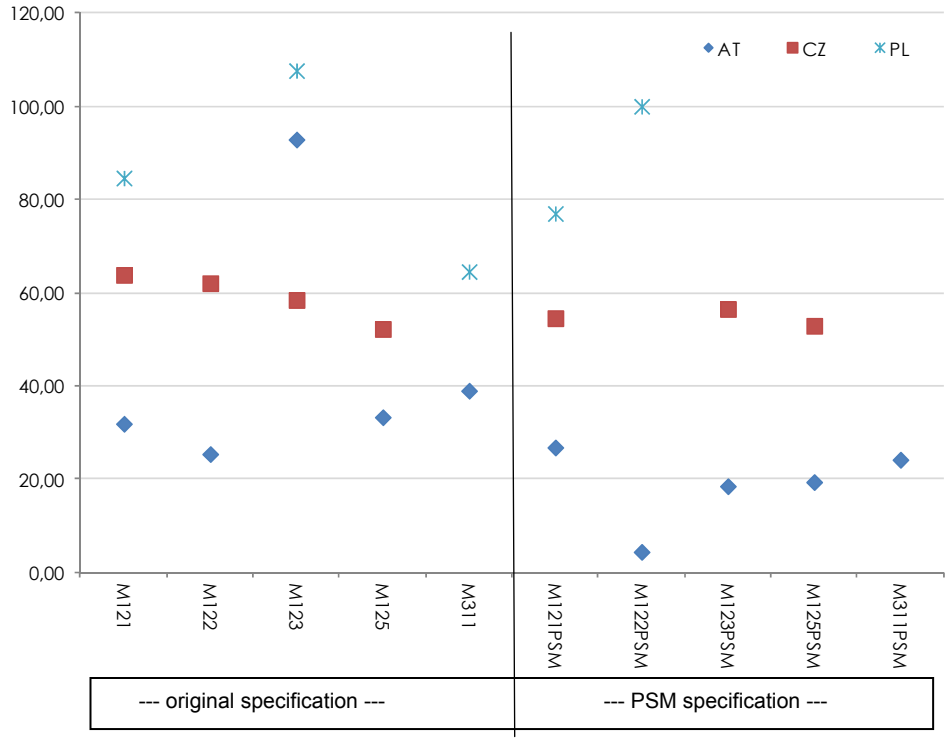
This conclusion is not unexpected. It shows that even in the context of a single study with standardised procedures and a group of well instructed and experienced experts the same phenomenon is perceived in slightly different ways. This does not necessarily mean that different methods yield arbitrary results.

It means that subtle details of definitions (e.g. farm employment *versus* employment of all sectors) and assumptions (monitoring data *versus* data derived from other sources) need to be considered with great care.

Contrasting results: take the better of two methods in order to derive more robust evidence

As outlined above, IO uses exogenous parameters to "shock" a model that is in equilibrium. The shock induces a new equilibrium which represents the economy in another state after the shock. Shocks are given exogenously and come in many instances from administrative data sources. Results of this approach are shown in the section on IO (see above).

Figure 20. Comparison of IO results on *efficiency* (jobs created per million expenditure) for investment support measures across regions with and without parameters obtained from PSM



Source: Sinabell and Morawetz (based on Psaltopoulos and Michalek(2014)
Note: A comparison for M131 is only possible for Austria; comparison for M123 and M125 is only possible for Austria and Czech Republic

Results on the same combinations of measures and regions are different when compared across methods. As explained above, one reason is that parameters may be from different sources. In order to address this possible source of biases, a set of case studies was analysed with IO for a second time. This time parameters obtained from PSM based on micro-data of farms were used in order to shock the model.

Results on the indicator *efficiency* are shown in Figure 20. In the left panel are the results which are also reported in Table 34 (*efficiency* based on parameters from reports

by Managing Authorities). In the right panel of the figure are the results that are based on parameters derived from PSM on the same combination of measures and regions.

The expected but nevertheless interesting observation is that results based on econometric findings identifying causal effects are lower than those based on administrative sources (compare for instance M121 with M121 PSM). When it comes to making judgement on the robustness and validity of results it should be acknowledged that those obtained with more rigorous methods are ranking higher than others.

This example shows an interesting aspect: Depending on the topic of interest it may be worthwhile to use method combinations. As shown here - it may also be adequate and reasonable if results of high validity are looked for. Whether the additional efforts (and costs) are worth it, depends on the level of validity deemed necessary to justify public expenditures.

5.3 Answer to Evaluation Question 3

The work on answering EQ3 was focused on specific aspects. A total of six sub-questions were analysed. The first three aimed at finding out whether there is a measurable relation between the targeting approach taken and the efficiency levels estimated as part of EQ2. The second three sub-questions deal with the approaches available for targeting (eligibility criteria, aid intensity differentiation, selection criteria). An attempt was made to find out whether they are suitable to allocate public funds to the targeted recipients. The rest of this section is structured according to sub-questions as outlined in the methodological chapter:

Sub-questions on *effectiveness* of targeting approaches to achieve an *RDP objective*:

1) To what extent has targeting investment support via *eligibility criteria* been effective in meeting general objectives of EU rural development policy and RDP objectives?
2) To what extent has targeting investment support via *aid intensity* been effective in meeting general objectives of EU rural development policy and RDP objectives?
3) To what extent has targeting investment support via *selection criteria* been effective in meeting general objectives of EU rural development policy and RDP objectives?

Sub-questions on *effectiveness* to allocate funds towards the *groups targeted*:

4) Were *eligibility criteria* effective in allocating funds towards those specified in eligibility criteria?
5) Was the differentiation *in aid intensity* effective in allocating funds towards those with preferential aid intensities?
6) Was the *selection* effective in allocating funds towards those specified in the criteria for ranking?

5.3.1 Answers to evaluation sub-question 1-3 on achieving an RDP objective

The approaches to targeting in the (sub-) measures were elicited from programme documents by experts on these programmes. A summary of the approaches to targeting is shown in Table 40 which lists the frequency of criteria that are used in the investment measures.

The total numbers of analysed sub-measures is compared to the three targeting approaches. The frequency reveals that a wide range of approaches to targeting is applied.

In particular the eligibility criteria *sector* is applied in most of the sub-measures with the exception of M123. Frequently applied are also the *eligibility criteria* restricting support to certain types of investment and limits in eligible amounts, *aid intensity differentiation* (different support for different groups of beneficiaries) and the *selection criteria* through ranking by the administrative authority. Territory and beneficiary-specific *eligibility criteria* are less frequently applied in the sample.

Table 40. Number of sub-measures analysed to identify approaches to targeting

Approach to targeting	M121	M122	M123	M125	M216	M227	M311	M313	Sum
Eligibility criteria									
Sector	16	3	10	4	1	1	7	2	44
Territory	2	0	3	0	1	0	3	1	10
Beneficiary type	9	3	9	2	0	1	2	1	27
Investment type	14	3	13	4	1	1	6	1	43
Limits to support	12	3	9	2	0	0	5	1	32
***Aid intensity* differentiation**	12	2	10	1	1	1	4	0	31
Selection method									
Ranking	14	3	11	4	1	0	6	2	41
Number of sub-measures	16	3	13	4	1	1	7	2	

Source: Morawetz (2014)

The number of approaches to targeting was reduced to a smaller number of dimensions through clustering. The procedure grouped sub-measures together which share many and/or rare targeting approaches. For the 47 analysed sub-measures this resulted in five clusters as shown in Table 41. Each cluster can be characterised by statistically significant distinct characteristics. Some clusters are associated with *regions* or *measures*. For instance, all Polish sub-measures are in cluster 2 as no ranking was applied at the *selection method* (or at least not during the whole programming period). None of the sub-measures from Hessen are found in cluster 3. All of the sub-measures of type 123 (increase of efficiency) are in cluster 3 as these measures are not restricted to the agricultural or forestry sector only.

Table 41. Distinct statistically significant characteristics of the clusters

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Sub-measures	4	5	26	9	3
Characteristics	1. No Investment type eligibility criteria	1. No Ranking as selection method	1. No territory as eligibility criteria 2. Ranking as selection method 3. No aid intensity differentiation 4. Investment type as eligibility criteria	1.Territory as eligibility criteria	1. No sector eligibility criteria

Source: Morawetz (2014).

Figure 21 below shows the dendrogram with the structure of the clusters. The choice to analyse five clusters is based on the drop in inertia captured (variance) possible at a lower nod level.

The judgement criteria for evaluation sub-questions are statistically significant associations of these five clusters with *efficiency* related criteria from EQ2. From PSM and IO cardinal measures of *efficiency* are available from EQ2 (see previous chapter). From MAPP and TBE, the indicators differ between individual field studies and are thus not useful for comparisons between field studies. Indicators from SEA are either not available or increasing. This is not meaningful for cross -ield study comparison as one would explain how targeting is related to data availability. Thus, only the *efficiency* indicators from PSM (indicator 1a, 1b, 1c, 1d) and the IO results can be used. In total, *efficiency* indicators for 13 sub-measures from PSM and *efficiency* indicators for 35 sub-measures from the IO model estimation are available (not for all sub-measures targeting approaches were elicited).

Table 42. Correlation between clusters and efficiency criteria from EQ2

Cluster	Measures	Mean in cluster	Overall mean	p-value
Cluster 1	Efficiency PSM1d	1.01e+05	4.27e+04	1.07e-03
Cluster 2	Efficiency IO	67.52	40.37	0.031
Cluster 3	Efficiency PSM1b	0.46	0.40	0.012
	Efficiency PSM1c	0.22	0.18	0.017
	Efficiency PSM1a	0.40	0.36	0.039
Cluster 4	No sign. efficiency indicators			
Cluster 5	No sign. efficiency indicators			

Source: Morawetz (2014).

The characteristics of the clusters can be described in the following way:

Cluster 1: It has a statistically higher overall mean in the efficiency indicator 1d from PSM. This indicator estimates the public expenses necessary for an additional job. Since only Austria and Hessen are in this category the higher value might be attributed to a higher wage level in these two countries as compared to the Eastern European countries which are part of the other clusters. It is therefore likely to be unrelated to the characteristic "No investment specific eligibility criteria" of this cluster.

Cluster 2: It only consists of a single observation from PSM and a comparison is therefore not meaningful for PSM criteria. For IO, instead, there are three *efficiency* criteria for the sub-measures in this cluster. All three are from Poland. The average jobs created per public expenses in million EUR is significantly higher than in the average over all clusters. Again, this seems to be a Poland specific issue.

Cluster 3: It is characterised by a significantly higher indicator 1a, 1b and 1c from PSM than the other clusters. These indicators measure *efficiency* in GVA, family income and labour productivity change through investment support. The characteristics of this cluster are *territorial* and *investment* type *eligibility criteria*, no aid intensity differentiation, and ranking as *selection*

method. The worse performance of this cluster is unlikely to be causal to the combination of criteria. It is more likely to be a correlation which is driven by confounding variables (e.g. country-specific results). The average IO *efficiency* indicator for cluster 3 is not significantly different for this cluster than for the overall average.

Cluster 4: It does not differ significantly from the other factors with respect to any of the PSM or IO *efficiency* indicators. Thus, we do not see a correlation between territorial restriction and efficiency. This does not mean it does not exist as it might be hidden by the influence of other factors. For cluster 5 the situation is alike.

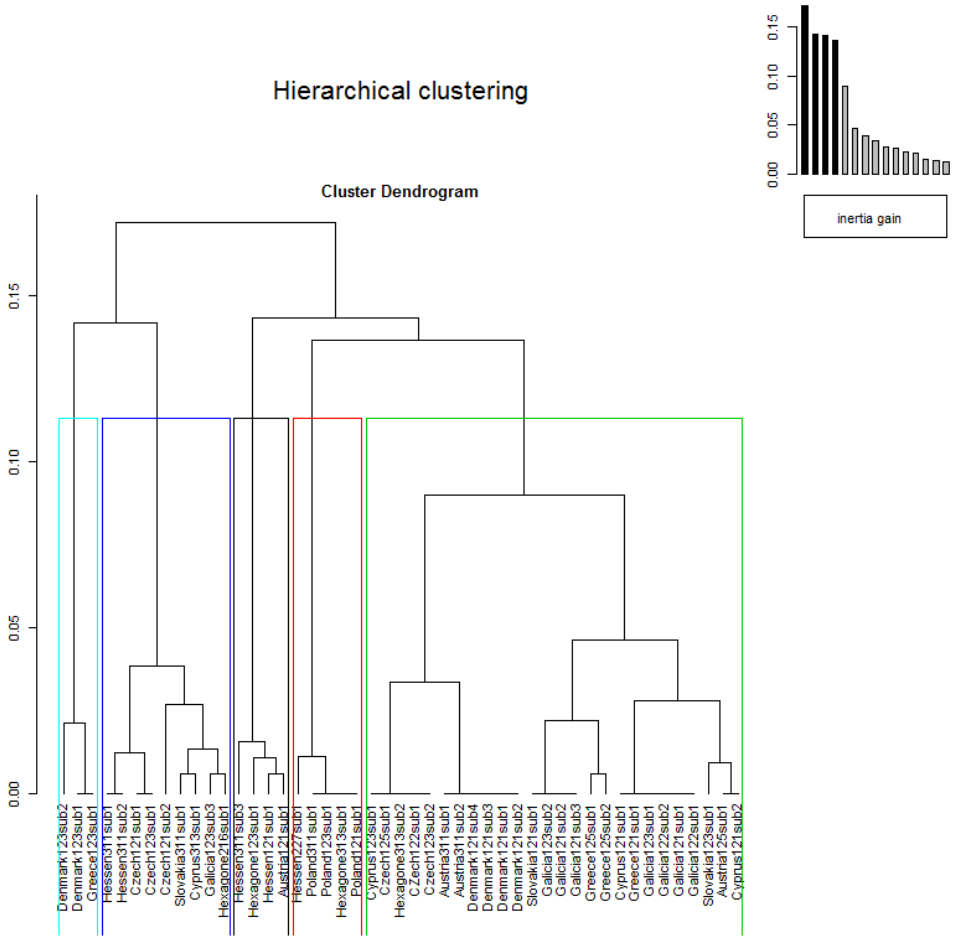
The answer to sub-question 1-3 of EQ3: based on the data available, no causal relationship of the approach to targeting and efficiency could be established.

This finding might be due to an actually weak link, or because the eligibility criteria, aid intensity differentiation and selection criteria have not been applied correctly (which is discussed in sub-question 4-6) or due to shortcomings in the analysis. More specifically, the following shortcomings should be considered to explain this finding:

- estimates of *efficiency* are based on results of EQ2 which introduces additional uncertainty;
- in some cases EQ2 results are not available at a sub-measure level (only at measure level), whilst identification of approaches to targeting has to be undertaken at a sub-measure level;
- the levels of incomes are very different across the EU therefore many quantitative EUR values are hardly comparable across EU regions;
- the relatively small number of observations requires an analysis of all measures together, even though an analysis of one type of measure (e.g. M121) would allow deeper insights.

The dendrogram in Figure 21 illustrates how similar the approaches to the targeting of different sub-measures are. Similar sub-measures are clustered in groups and for each number of clusters a level of inertia (sum of within-cluster and between cluster variance) is shown on the scale on the left. As the gain in inertia is highest for the first five clusters (compare small figure on the right), the total number of clusters is limited to five. The coloured boxes around the sub-measures frame the clusters is described next.

Figure 21. Dendrogram of the clusters generated from the approaches to targeting



Source: Morawetz (2014)

5.3.2 Answers to evaluation sub-question 4-6 on allocating funds to target groups exemplified on the case study M121 in AT

The objective of sub-question 4-6 is to understand if the approaches to targeting used in the RDP actually divert support towards the targeted group. The indicators, leakage rate, underinvestment rate and differences between support up-take by different groups have been developed to answer sub-question 4-6. The indicators where calculated, where applicable, for M121 of the Austrian RDP.

Measure 121 of the Austrian RDP

The Austrian version of M121 has a complex set of targeting and therefore serves as an example. *Eligibility criteria* secure that beneficiaries are active farmers. The *eligibility criteria* are designed not to exclude any actually active farmer (>0.3 workers per farm, >3 Ha cultivated land or > 2 livestock units with exceptions for fruits-, wine-, honey- and hop farmers, qualified craftsman or 5-year experience). *Aid intensity differentiation* is of two kinds: territorial and investment-specific.

The territorial one allows a maximum aid intensity of 50% in less favoured areas (LFA) compared to 40% elsewhere. The investment-specific one consists of two parts. First, maximum aid intensity of investment and credit support is higher for investments in alpine meadows, for animal welfare, horticulture, biomass heating, and marketing projects. Second, there are top-ups of 5% for investments in barns if the applying farm is organic. *Selection by ranking* is based, amongst others , on the economic viability of a beneficiary farm. It is assessed by the administrative authority based on information provided by the applicant.

Evaluation sub-question 4 (eligibility criteria)

The *leakage rate* (percentage of funds diverted to non-eligible applicants) and *non-take-up* need to be calculated. Based on the FADN data from the year 2006 (the year before the programme started) weights and strata are used to estimate the population values of the leakage rate. The *non-take-up rate* (the ratio of the number of beneficiaries in the target group not funded relative to the total number of potential beneficiaries in the target group) is not meaningful for this kind of eligibility criteria as they are basically only introduced to secure that beneficiaries are active farmers.

Leakage rate for the eligibility criterion “minimum of 0.3 workers per farm" is zero. The *leakage rate* for the eligibility criterion “minimum cultivated land” or “livestock units" is 2.25. This means that 2.5% of the funds were provided to farms not meeting this criterion. This is likely to be related to the exceptions for fruit, wine, honey and hop farmers. The *leakage rate* for qualification of the applicant is not calculated as it is not directly supported by the available data. Eligibility criteria for M121 thus were effective, as one would expect.

Evaluation sub-question 5 (differentiation *aid intensity*)

The average support in LFA (less favoured areas) is compared with average support outside of LFA. To compare comparable samples, genetic matching based on the pre-period covariates used in EQ2 was undertaken for the Austrian case study. The matched sample was used to calculate the difference in the support between those in the LFA and those outside. Matching was done with the flexible genetic matching algorithm, still some of the covariates were unbalanced, e.g. area of agricultural land owned. Thus, not all of the bias could be avoided. The average support to those in the LFA is 600 EUR higher than outside, after adjusting for differences due to the differences in the structure between the regions. The Abadie-Imbens standard error of the mean difference is 610 (p-value 0.32) making this difference not statistically significant. Aid intensity differentiation, thus, made no significant difference in average funding after matching. Though, some uncertainty remains, because some of the covariates are unbalanced.

Answer to evaluation sub-question 6 (*selection criteria*)

Selection by ranking is based, among other indicators, on the economic viability of the farm. The assessment of economic viability is to a substantial part exposed to subjective judgements. For simplicity viability is measured by the average profit over the years 2003-2006. We define profit as income from agriculture and forestry (including CAP payments) plus gross investments minus labour and capital costs as defined in the FADN data. Comparing the distribution of not-supported and supported farmers the estimated mean and the estimated quantiles (0.25, 0.5 and 0.75) are significantly higher for the supported farms.

This can be the consequence of 1) the selection process of the administrative authorities, 2) the pre-selection by some agricultural extension services, or 3) the self-selection of the farmers. Data on who applied and who was rejected would be needed to find out more.

Several afterthoughts to the method to answer sub-questions 3-6 are important. First, the procedure is not always applicable (e.g. *non-take-up rate* for investment type eligibility). Second, the indicators presented here are based on the dimension of *farms* and one could do a similar evaluation on the dimension of *projects* instead: some approaches to targeting are related to the funded project (e.g. funding only of certain investment types). The analysis here is based on a sample with weights. Thus the population statistics are estimates from the weighted sample. A non-estimated statistic would require linking administrative data from the payment agency with a census which is detailed enough to derive *eligibility* criteria and *selection criteria performance* for each unit. While preferably, this seems to be hardly ever feasible.

Conclusions

Testing for a casual relationship between efficiency to achieve RDP objectives and the targeting approaches is difficult, if not impossible, if the sample size is small and measurement errors are substantial. Understanding how well eligibility criteria, aid intensity differentiation and selection by the Managing Authority work to divert support to the targeted beneficiaries is more promising. The case study for Austria revealed that this kind of evaluation is feasible with the data available. Eligibility criteria are relatively easy to administer and evaluate. Aid intensity differentiation not necessarily makes a significant difference in the average support. Most likely, this depends on the magnitude of the aid intensity differentiation. Finally, if there is no difference in the distribution between those supported and those not supported with respect to the selection criteria (or if it has the wrong sign), it can be concluded that selection by the administrative authority or take-up by the target group did not work. On the other hand, if there is a difference, this is not necessarily due to the selection done by the Managing Authority.

6 Conclusions and recommendations

6.1 Conclusions on the appropriateness of methods to evaluate investment support measures

The topic of the first Evaluation Question (EQ1) is on methods, not on results of evaluations:

> **To what extent are the different evaluation methods described and/or tested in this exercise appropriate for the assessment of the effectiveness, efficiency and impact of the different types of investment support considered?**

In order to answer this question a first step was made by selecting methods which are representative for groups of methods which are frequently used in evaluations:

- **E:** econometric approaches for causal analyses exemplified by the "propensity score matching method (PSM)"
- **IO, M**: IO and quantitative (programming) models (farm/firm/household/region) exemplified by the "input output model (IO)"
- **QP**: qualitative and participatory methods exemplified by "MAPP"
- **TB**: theory-based and descriptive approaches exemplified by "theory-based evaluation (TBE)"
- **SEA/CEA:** integrated environmental and economic assessments exemplified by strategic environmental assessment (SEA) and cost-effectiveness analysis (CEA)

To make judgements about the *appropriateness* of methods to evaluate investment support measures in the context of Rural Development Programmes, two major aspects have to be considered:

- the generic characteristics of a given method in general;
- the adequacy of a method in the context of the specific case.

In this study both aspects were analysed and evaluated. Based on the judgement of experts having applied a given method for years, the characteristics were classified and an overview was made that shows how well a given method is suited to address aspects that are important in evaluations. The general points of interest for evaluation are:

1. Users of evaluation have an interest that results are reliable, scientifically sound, robust and valid. Using these criteria, it is possible to classify methods and approaches with respect to their appropriateness to evaluate investment support measures.
2. The results of different methods are available at different scales: econometric and quantitative IO or programming methods provide results on cardinal scales, the other on ordinal scales or on nominal scales.
3. In order to make an evaluation it is necessary to develop a counterfactual. The counterfactual situation is a conceived or observed scenario that is used to compare firms/farms with and without an investment programme. It has to be very well specified and the challenge is to make judgements about an effect of the programme which by definition cannot be *directly* observed in most cases.
4. Among the biggest challenges in evaluation studies is to identify and test causal relations between the policy intervention and the outcomes. Only a small set of methods (typically econometric models) are suited to provide such results. If such

results are not available, assumptions about the causal effects need to be made to analyse a counterfactual situation.

5. The combination of methods contributes to the validity of results. This study reveals that the qualitative MAPP method can set the context and contribute to the development of hypotheses which can subsequently be tested with quantitative methods, such as counterfactual and input-output methods. Alternatively, the results of quantitative methods can be validated and causal relationships explained by MAPP focus groups.
6. Several effects are specifically important to consider when investment support measures are evaluated, such as leverage, substitution, displacement effect and deadweight (windfall profit).
7. For the evaluation of similar measures in many different regions, the feature of "external validity" is important. It means that results of one region or a particular sample can be used to make judgements for another region or the population which was sampled. External validity is higher the higher structural similarities between the dimensions (e.g. regions, farms) under consideration are.

The conclusions with respect to the appropriateness of the methods to address the above mentioned points of interest are:

1. When results have to be expressed in numbers (cardinal scale), the scope of methods is limited because qualitative and participatory methods (QP) and theory-based and descriptive approaches (TA) allow ordinal statements at best. Quantitative methods are therefore strongly preferred if the *efficiency*, *effectiveness* and *impact* of investment support measures is of interest.
2. Any method can be used to analyse counterfactual scenarios. But the usefulness depends on the approach. Most IO and other quantitative models (IO, M) are developed for this purpose.
3. In case it is necessary to identify the causal effects of measures on the investment behaviour of firms (e.g. for *ex post* evaluations) there are two alternatives: results can be based on statistical evidence of randomised controlled trials or results are based on adequate econometric assessments (E). In case causal effects cannot be identified by observations (e.g. for *ex ante* evaluations), assumptions need to be made possibly through the application of qualitative methods.
4. Every method can be used to take account of leverage and deadweight effect, but only econometric methods (E) can be used to quantify their size.
5. All approaches under consideration meet the criteria of soundness, robustness, and validity, but at different levels. In scientific literature not all types of approaches are equally well accepted and the robustness of results may be questioned if results are based on a small sample size or a model specification that has not undergone peer reviews or if results were obtained in a non transparent manner. The validity of results can be checked when the research process that leads to results is made as transparent as possible. Restriction on access of individual data may impair a high level of transparency but model specification, estimation techniques and test statistics should be transparent even in such cases.

6. All methods can be used to support assumptions about regions that were not analysed in detail. How well this can be done and how valid the results are depends not on the method but on the specific circumstances of the regions under consideration and their similarities.

As stated above, not only the generic characteristics of a method are important. The context in which they are applied has to be considered as well. In this study the *practicability* of methods in the context of Rural Development Programmes in the EU was a topic of interest and several conclusions can be drawn upon the findings of six methods in eleven regions

1. All methods need data such as programme documents, reports, evaluation studies, etc. Such data are partly readily available (like input-output tables at country level), available but scattered on different places (programme details and data on implementation and outputs), available but accessible only with restrictions (e.g. many micro-data), or have to be collected (e.g. interviews, participatory activities). In the context of this study, expenditures for data acquisition were surprisingly low. Much more resources were necessary to locate, collect and prepare data.
2. The application of quantitative methods, such as econometric analyses (E) or IO needs data that have to be maintained and regularly updated in a structured manner. Some of the data (e.g. IO tables, FADN data) are available. But nevertheless it is necessary to put additional efforts into data preparation (e.g. linking micro-data from various sources). Only in regions where specialised institutes are involved in data handling, data for specific micro-econometric analyses are accessible. This is the case for 27 Member States but not for all programme regions.
3. Non-quantitative methods need data as well. Most frequently they are obtained not from large databases but acquired *ad hoc*. This process is very time consuming and costly.
4. The applications of qualitative methods that are based on group discussions are more effective if they use structured tools (including indicators) that focus and drive the discussion towards the identification of specific trends (as is the case when MAPP is applied). Its reliability however, depends on the sample size and the number of focus groups per programme territory.
5. Standardised data sets (e.g. IO tables, FADN-data) are very well suited for large scale applications. Costs for an additional case study are low compared to non-quantitative methods.
6. The nature of some methods makes their applicability dependent on the design of the measure and most importantly on the number of participants and non-participants. This number determines the sample size for statistical analyses. Not every econometric method (E) for causal analysis can be used to estimate the efficiency of measures if data of non-beneficiaries are not available or if beneficiaries are very different from non-beneficiaries.

6.2 Recommendations on identifying appropriate methods to evaluate investment support measures

6.2.1 Rationale and overview

Decision makers bearing responsibility to allocate public resources in an efficient manner may need guidance on selecting an appropriate method for evaluating investment support measures. One insight of this study is that there is not **one** method that is best suited to address every aspect of such an evaluation. Rather there may be several appropriate methods.

In order to acknowledge this, the tool presented in the next section will not make any prior assumptions on the preferences of the user of the results. Rather, the user will be encouraged to specify her or his *preferences* in a structured way. The methods under consideration are specified according to their *characteristics* in a similar way. The comparison between *preferences* of the user and *characteristics* of the methods will show if there is a *match* between the two. If there is no *match*, new *capacities* or *resources* need to be made available in order to take another step towards matching *preferences* of users and *characteristics* of methods. The outcome of such a decision making process that may run in several loops and is visualised in a decision tree in order to make the rationale clearer.

6.2.2 Capacities, resources, preferences and characteristics

In order to evaluate investment support measures it is necessary to employ resources. They may be available in-house or need to be employed. In any case such resources, such as expertise and data, are costly. Therefore it is evident that two types of costs have to be considered, variable and fixed costs:

- The establishment of a data infrastructure is very costly, it needs staff with special skills and it has very high sunk costs. But once, a good infrastructure for data management is there and once structured data are updated regularly it is relatively low cost to produce results. One drawback of a data warehouse is that the data structure predetermines the set of methods that can be used.
- Data can also be bought or collected *ad hoc* in order to make just one analysis. In such a case the variable costs are high but there are practically no fixed costs. The drawback is that data will be used once and lost because there is no infrastructure (physical or organisational) to store them and make them accessible for future analyses.

There is evidently a trade-off between fixed and variable costs and the right balance is different for every organisation.

Table 43. Overview of essential capacities necessary to carry out an evaluation

	Type	Capacities and resources
1	Expertise	Expertise on quantitative methods
2		Expertise on participatory methods
3		Specific expertise on RDP details
4	Data	Detailed data of participants (e.g. IACS, FADN)
5		Detailed data on participants and non participants
6		Detailed data of programme expenditures and outputs
7		Detailed data on firm structure/sector output
8	Funds	Investment in data infrastructure
9		Continuous operation of data warehouse
10		Funds for *ad hoc* expertise

The next overview (Table 44) is a tool for the decision maker who has to select an appropriate method to analyse the efficiency of investment measures. The rationale of the tool is to help the decision maker first to make explicit her or his *preferences*. Only the decision maker knows if the purposes of an evaluation should go beyond figuring out one efficiency indicator or not. Therefore the *preferences* may be very heterogeneous. To start the decision making process the user of the evaluation should just indicate by an "x" on the Likert scale which aspects correspond only low (column labelled "low") or very high (column labelled "high") with hers or his *preferences*. After reaching line 25 there will be a pattern of "x" which will be the starting point for the first choice.

Table 44. Tool for decision makers to state his or her preferences

Preferences			Less preferred				most preferred
			1	2	3	4	5
1	Costs	Willingness to bear high fixed costs					
2		Willingness to bear high variable costs					
3	Results	Requirement to identify causal effect					
4		Requirement to get quantitative results					
5		- for specific types firms/farms					
6		- for sectors/regions/countries as a whole					
7		Ease to identify leverage and deadweight					
8		- leverage and deadweight					
9		- impacts beyond sample					
10		- displacement effects					
11		Requirement to get results on ordinal scale					
12		Requirement to get narratives instead of indicators					

Preferences			Less preferred				most preferred
13		Requirement to have access to intermediate results					
14		Standardised results					
15	Transparency	Quantitative methods					
16		- primary data access					
17		- access to computer code					
18		Non-quantitative method					
19		- primary data access					
20		- access to supporting material for results					
21	Rigour	Link to a specific theory (e.g. Economics)					
22		Link to science literature					
23		Ease of sensitivity analyses					
24	Other	Requirement for stakeholder involvement					
25		Ease to translate result to other regions					

The methods used in this study fall into the classes introduced above (environmental methods are not included as they typically are a combination of several methods):

- E econometric approaches for causal analyses;
- IO,M IO and quantitative (programming) models (farm/firm/household/region);
- QP qualitative and participatory methods;
- TB theory-based and descriptive approaches.

The methods within these groups share similar characteristics while there is significant difference between them. The following overview (Table 45) states the characteristics that were identified in this study and considered to be important. The columns are used to order the methods into various classes. The order is based on judgement criteria that were developed and shaped during the work carried for this report.

There is a one-to-one relationship between the *characteristics* of the methods (see below) and the structure of *preferences* of the decision maker identified above. While in the case of *characteristics* the columns represent classes, in the case of *preferences* columns represent a Likert scale.

Table 45. Characteristics of generic evaluation methods

	Characteristics		Low				High
			1	2	3	4	5
1	Costs	Fixedcosts	QP,TB			E	IO,M
2		Variable costs		IO,M	E		QP,TB
3	Results	Measure causal effect	QP,IO, M, TB				E
4		Quantitative results	QP,TB				IO,M,E
5		- for specific types firms/farms	QP,TB			IO,M	E
6		- for sectors/region as a whole	QP,TB			E	IO,M
7		Identify leverage and deadweight					
8		- leverage and deadweight	IO,M				E
9		- impacts beyond sample			IO,M[1)],		
10		- displacement effects					E,IO,M
11		Results on ordinalscale					All
12		Narratives instead of indicators					All
13	Transparency	Access to intermediate results		E[2)],QP			IO,M,TB
14		Standardised results	QP,TB		E		IO,M
15		Quantitative methods					
16		- primary data access			E		IO,M
17		- access to computer code			E		IO,M
18		Non quantitative method					
19		- primary data access	QP				TB
20		- access to supporting material					All
21	Rigour	Link to a specific theory	TB,QP				IO,M,E
22		Link to science literature	TB,QP				IO,M,E
23		Sensitivity analyses	TB,QP				IO,M,E
24	Other	Stakeholder involvement	Other				QP
25		Translate results to other regions	QP		E	TB	IO,M[1)]

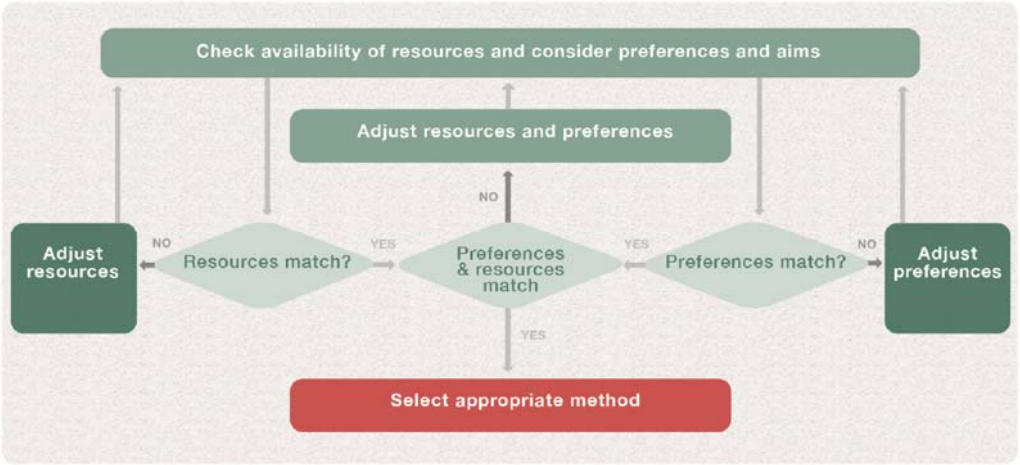
Source: own elaboration

Notes: [1)] – one may expect similar results in regions with similar structures and policy interventions; [2)] - in most cases micro-data used for the analysis are accessible only to the analyst not to reviewers or users of the results.

6.2.3 Making a decision on an evaluation method

The first decision that needs to be made is on the purpose of the evaluation. Typically evaluations are *ex ante*, *ex post* and mid-term (or concurrent). A decision maker may wish to use different methods for the different purposes or may prefer to use one method or a combination of methods for all purposes.

Figure 22. Decision tree



Source: Sinabell (2014)

After this pre-selection, the next step can be taken. By comparing how many matches between "x" preferences and cells in the characteristics overview there are, the decision maker can see at a glance if there is a small correspondence between *preferences* and *characteristics of* methods.

A decision maker may face several situations:

- the "x" are equally distributed over many types of methods;
- the "x" are concentrated on one type of method;
- the "x" are concentrated on two types of methods.

Applying more than one method is very costly. Therefore, there is a trade-off between making specific *capacities* or *resources* available (e.g. data) that may help to reduce the need for further methods. It has to be considered, that even making substantial resources available will not help to overcome method-specific limitations (quantitative results on cardinal scales can only be obtained by certain methods).

Typically, a decision maker has the choice either to adjust the preferences or to make more resources available. By running through a loop of matching processes of *preferences* and *characteristics* and adjusting *resources* and *capacities,* the right class of methods is identified in a few rounds (see decision tree). Depending on the *preferences* and the available *resources* and *capacities* the best choice can be a combination of two or more methods[56]

Once the class of preferred methods is identified it is necessary to make a decision on the specific method (e.g. IO model, CGE model, farm model, and agent based model). Such a decision can be structured in a similar way but the overview of *characteristics* and the tool for making transparent the *preference* structure need to be adjusted.

[56] A more detailed elaboration on this subject is provided in chapter 4.3.4.2 of the Guidelines for the *ex post* evaluation of 2007-2013 RDPs (European Evaluation network for Rural Development, 2014)

6.3 Conclusions on the effectiveness, efficiency and impact of investment support

Six methods were used to assess the effectiveness, efficiency and the impact of investment support measures in order to answer the second Evaluation Question:

> **EQ2: What is the effectiveness, efficiency and impact of the investment support studied in the selected RDP territories?**

Quantitative methods including PSM, a method which can be used to identify causal effects, show a wide spectrum of outcomes.

A major finding - derived from the analysis with the IO method - is that support of investments creates additional demand in the regions where the beneficiaries are located. As econometric evidence confirms, in most cases the capacities expand and output increases. Both effects together have positive yet rather small effects on value added and employment for the whole region or country, not just for the targeted enterprises or sectors. This result is the *impact* of support measures and it is different in scale across the regions. It mainly depends on the volume of expenditures and the structure of the economy.

All regions and their firms are different and all programmes are different as well. In most cases efficiency was at levels expected and observed already in previous studies. However, the results are very heterogeneous. This becomes very evident when considering the *efficiency* of investment support measures. *Efficiency* of investment support (results relative to expenditures) depends on three aspects: the combination of region and measure, the specific result of interest (value added or labour) and the method:

- In several cases the same investment measure was analysed with different methods in the same region. As expected, the results deviated. Even contradictory results were obtained. Looking at such results with scrutiny showed that contradictions are either due to subtle differences of definitions between the methods or due to the fact that one method was qualitative while the other one was quantitative.
- For PSM results for *efficiency* show that for each EUR of public funding spent, the GVA and family farm income increased by substantially less than one EUR.

In order to explain such methodological differences, specific case studies have to be analysed in depth. There are several plausible reasons for a low level of *efficiency* as observed in this study:

- Investment support schemes have some objectives that are not adequately represented in the number of jobs or the value added (e.g. health of workers, animal welfare considerations, etc.). In order to measure the efficiency, such targets need to be evaluated and trade-offs to GVA and jobs have to be identified.
- The time horizon of investments is very important. In the analyses here, this aspect was neglected. The results indicate that this simplicity may imply that the effects of investments with long time horizons may not be adequately measured.
- Firms are heterogeneous and one advantage of micro-data is that the heterogeneity is visible in the data. The heterogeneity should be explicitly accounted for in the analysis. Thus it would be possible to go beyond *average effects* as done in this study. It has to be considered, that such a differentiated

approach requires considerably more resources. In this analysis the deliberate decision was made to prefer a larger number of cases over a smaller number of cases with deeper insights.

How targeting affects *efficiency* of investment support is now better understood. By using statistical approaches it is possible to identify clusters of programmes that share similar characteristics. But the analysis also shows that given the information available now, it is not possible to establish causal links between targeting approaches and programme outcomes.

Another important finding relates to the *effectiveness* of the investment support measures. This indicator relates an outcome of the programme to a target. Frequently targets are not reported and if they are reported there is not a very good correspondence between achievements and target. This is not surprising. Targets are set when the programme is designed and under specific assumptions about the development of the economy and the sector as a whole. Events like the financial crisis in 2008 and 2009 and the price slump of farm commodities are a major shock for the economy and were not anticipated when the programme was designed. Such events may change the structure of a regional economy and thus change the underlying path of development.

Apart from the answers to EQ2 several observations were made that show important aspects concerning the methods involved in this study:

- Several contradictory results on the same case by different methods can be explained by taking into account the different sources of data used by the methods under consideration (e.g. regional statistics, judgements from stakeholders, micro-data) and the methodological characteristics (e.g. causal effect is assumed or measured).
- Some other contradictory results need clarification but this requires further explorations scrutinising the cases in detail.
- In several instances, two methods can be complementary. Because only randomised controlled trials or econometric methods (E) can be used to estimate causal effects empirically, such results are very valuable. They can be used to specify parameters that are used in quantitative models (IO, M). Results of combinations of methods are more robust and reliable and cover a broader scope of indicators than would be possible by using just one method.

Not every method can be used to measure all three indicators of interest, *efficiency*, *effectiveness* and *impact* of investment support measures. Only the studies with quantitative methods (IO and PSM) covered the whole range. Studies applying the other methods reported either mainly on *impacts* (MAPP) or on *effectiveness* (TBE). The evidence provided by a survey of various sources on results on environmental outcomes (SEA and CEA) was scant.

Even if results for all indicators from IO and PSM were presented in this study it has to be considered that neither of these methods is suitable to analyse everything:

- Analyses at aggregate levels are fundamentally different from analyses based on micro samples.
- Methods similar to IO cannot be used to make conclusions at micro-level. Such methods therefore are only useful for analyses at regional or national scale.

- Likewise, to derive aggregated results from econometric analysis of small micro samples is only valid under very specific conditions. Such conditions are hardly met when non-representative samples are used for the analysis or when the sampling design is not explicitly modelled for extrapolation. When explorations of samples to the whole population are made, the approach chosen for the extrapolation must be justified and at least the range of uncertainty has to be stated.
- In order to explain some contradictory results further clarification is needed but this requires explorations scrutinising the cases in detail.

6.4 Conclusions on targeting

The third Evaluation Question deals with a procedural aspect of programme implementation:

> **EQ 3: To what extent have the different approaches to targeting investment support studied been effective in meeting the general objectives of rural development policy and/or specific objectives included in the relevant RDPs?**

There are three elements to divert funds towards the targeted groups. All elements are currently combined in various ways in the RDP leading to a high number of approaches to targeting. The criteria are:

- *eligibility criteria* (selector, territory, investment type, beneficiary characteristics, funding range),
- *aid intensity differentiation* (more funding for targeted groups through top-ups or higher maximal funding) and
- *selection* among all eligible applicants through ranking by the administrative authority.

Different approaches to targeting might differ on how efficient they are in achieving the objectives of the RDP. It might therefore be possible to directly link the results for EQ2 where the efficiency is estimated.

Though, even after reducing the number by clustering according to similarities of the approaches, we did not find a significant relationship which is likely to be explained by the approach to targeting. This might be due to at least three reasons:

- First, and most likely, not enough observations were available. The measurement combined with insufficient variation prohibited uncovering possible relationships between *efficiency* and approach to targeting.
- Second, the approach to targeting does not have a substantial effect on *efficiency*.
- Third, the approaches to targeting described in the programme documents did not divert the funds to the targeted groups.

In order to further investigate EQ3, an in depth analysis was made. The case study chosen for this exploration is investment M121 in Austria. The major findings are:

- Measure 121 in Austria includes *eligibility criteria*, *aid intensity differentiation* and *selection by ranking*. To test if these criteria divert funds to the groups specified, FADN survey data were used. Using the survey design information (weights and

strata) allows to estimate population statistics like *leakage rate* (percentage of funds diverted to non-eligible applicants) or *non-take-up rates* (the ratio of the number of beneficiaries in the target group not funded relative to the total number of potential beneficiaries in the target group).

- *Eligibility criteria* for beneficiaries of the M121 are used as minimum standards securing, that only active farmers can benefit from the investment funding. We observe *leakage rates* of zero or very close to zero. The maximum aid intensity is 40% except for applicants located in a less favoured area (LFA) where it is 50%. Comparing a matched sample and thereby taking account of structural differences, we do not find significant differences in the average funding between applicants from a LFA and those outside which are due to the difference in the maximum aid intensity. The *selection method* applied is ranking by the administrative authority of the economic viability of the applicant (next to other criteria). Using a simple criterion for economic viability (three-year average profit) we find that the supported farms actually are those which were economically more viable before the programme period started. This is not necessarily a consequence of the selection by the administrative authority but might also be due to self-selection or the influence of agricultural extension services.
- We thus find that not all elements of the approach to targeting necessarily work. Eligibility criteria, as they were easy to check worked, whilst for aid intensity differentiation criteria the case study did not find a statistically significant effect on average investment support. Interestingly, in the targeting of the analysed measures eligibility criteria were not used to exclude farms. Supported farms were found to be more profitable before the start of the programming period, which confirms that the selection process worked. Though, it is not known if it was self-selection, selection by agricultural extension services or selection by the Managing Authority. Clearly, these results are based on one sample only and depend on the definitions (e.g. economic viability) and have some uncertainties (comparability of matched groups).

The major conclusion from this analysis is that targeting approaches are as complex as programmes are. In order to meet specific goals, a certain level of complexity seems to be unavoidable but as the case study has shown, our understanding on how effective different approaches are, is scant. In order to better understand the ways targeting approaches contribute to the effectiveness of programmes it seems to be necessary to work further on deepening the knowledge by analyses as presented here for other case studies.

6.5 Further observations

A list of observations and recommendations that are not related to a specific Evaluation Question but may be of relevance for evaluations of Rural Development Programmes in general follows:

- The number of users of evaluation results is large and each user has her or his expectations. While some users need quantitative results, others are more interested in subtleties that cannot be expressed in figures. Serving both prototypical groups calls for a mix of methods. On the positive side they are available but some contradictory results are likely to turn up. Further methodological improvements may help to overcome this situation.

- One important observation is that non-quantitative methods impose high variable costs while quantitative methods exhibit scale economies. These economic implications are important to consider when an evaluation plan is developed.
- In order to analyse measures of the Rural Development Programme, specific knowledge is necessary. The programmes are very complex and the situations are heterogeneous, even within countries. This is not unexpected because the programmes are finetuned to the situations in a given region. One has to be aware that high specificity may increase the administrative burden and diminish the scope of adequate evaluation methods.
- A situation that makes it difficult for experts from other fields (e.g. labour market economics) to analyse and evaluate measures of the Rural Development Programme is the fact that a very specific evaluation terminology is used. This increases the costs of expertise and the need for communication and may lead to a slower uptake of state of the art methods.
- The availability of standardised data sets (e.g. input-output models, FADN data) is a big advantage for quantitative methods. There are significant economies of scale. Given data availability for all regions it might even be possible to analyse a large number of measures for all regions with a small team of experts.
- Such an approach even seems to be necessary in cases where sample sizes of beneficiaries or non-beneficiaries are too small in one region. Having more regions where the same measure is applied might increase the sample size and thus the prerequisite to apply methods that can measure causal effects
- FADN data are very useful; however their usefulness is contingent upon the link to other administrative data. For the purpose of evaluation such a link should be the standard procedure in all regions and programmes. This would be a way to identify causal effects for a large number of cases and improve the validity of the evaluation studies considerably.
- The findings of the IO case studies show that this method and methods that use the same data may deliver a wide range of interesting results. The IO method is a standard approach in economics and because it is so standardised there is large expertise available. Similar, but more advanced approaches like regional CGE models would allow even better insights into the effects of Rural Development Programmes. What is not available are sufficient data for all regions. One recommendation is to establish the necessary data that are used for IO or similar models already at the beginning of the programme (in particular regional input output tables). Such an approach would allow to derive targets consistently and to make mid-term and *ex post* analyses within the same methodological framework for a number of important indicators.
- How targeting affects the outcomes of a programme is now better understood. Building on the findings presented in this study should be part of *ex post* evaluations.
- When programmes are designed it must be already clear which methods will be used for the evaluation. Identifying the data needs at the beginning and designing the reporting in the right way saves considerable costs. Including some subtle administrative kinds of randomness in the measures would help to identify important coefficients for evaluation (e.g. different starting points for random samples of beneficiaries; recording information on responsible officials; allocating applications of beneficiaries randomly to different offices checking applications).

- If it is not possible to identify a method that can be used to evaluate its intended effects it is strongly recommended to reconsider the implementation of a measure. This recommendation is based on the insight that "you cannot control what you cannot measure".
- Many investments have a period of operation that spans over many years (e.g. the planting of an orchard). The methods used to analyse the effects of such investment are different from approaches that focus on investments with short term payoffs (e.g. an automatic milking system). It has to be considered that some effects of the investment support measures will be measurable only after a considerable delay. This may be well beyond the programme phase. In order to address this argument, the time horizon of the investment should be taken account of in an analysis.
- In order to improve the validity of the results, authorities commissioning evaluation studies should put emphasis on as much transparency as possible. This is an effective way to improve the quality of results in the longer run and does not impose further costs.

7 Bibliography

7.1 International guidebooks

EVALSED: *The resource for the evaluation of Socio-Economic Development,* Updated version[57], EC, 2012.

Gertler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch., *Impact Evaluation in Practice Washington*: *The World Bank*. 2011, (available in English, French, and Spanish).

Handbook on Common Monitoring and Evaluation Framework, Guidance Document. DG Agriculture and Rural Development, Brussels, EC, 2006.

IEG (Independent Evaluation Group)., *Impact Evaluations in Agriculture: An Assessment of the Evidence*. Washington, DC: World Bank, 2011.

Khandker S. Et al., *Handbook on Impact Evaluation, Quantitative Methods and Practices*, World Bank, 2010.

Morra-Imas, Linda G., Ray C. Rist., *The road to results: designing and conducting effective development evaluations*, The World Bank, 2009.

Morris St., Tödtling-Schönhofer H., Wiseman M., *Design and Commissioning of Counterfactual Impact Evaluations,* Commissioned by DG Employment, Social Affairs and Inclusion, Unit A3, EC, 2013.

NONIE *Impact Evaluation Guidance*, January 2008.

Rossi, P. H., Lipsey, M. W. and Freeman, H. E., *Evaluation: A systematic approach*, (7th edition), Thousand Oaks: Sage Publications, 2004.

Shadish, W. R., Cook, T. D. and Campbell, D. T., *Experimental and quasi-experimental designs for generalised causal inference*, Houghton Mifflin Company, Boston, US, 2002.

The Green Book: Appraisal and Evaluation in Central Government. HM Treasury (United Kingdom). London: The Agency. 2006.

The Magenta Book: Guidance for evaluation[58]. HM Treasury (United Kingdom). *London: The Agency*. 2011.

WK Kellog Foundation, *Logic Model Development Guide,* 2004.

[57] Comprehensive overview of a wide spectrum of evaluation methods.
[58] The "Green" book discusses the place of evaluation in what the Treasury calls the "policy cycle". The "Magenta" book provides detail on evaluation methodology. These documents are interesting as examples of within-government evaluation perspective.

7.2 Specific evaluation methodologies for rural development

Allen B. Et al., *Common Agricultural Policy Regionalised Impact – The Rural Development Dimension*, 2013, 7th Framework Programme, Project N°: 226195.

Approaches for assessing the impacts of the Rural Development Programmes in the context of multiple intervening factors, European Evaluation Network in Rural Development, Commissioned by the EC, 2010.

Bateman, I., Carson, R. T., Day, B., Hanemann, W. M., Hanley, N., Hett, T., Swanson, J., *Economic valuation with stated preference techniques: a manual,* MA: Edward Elgar, Cheltenham, UK; Northampton, 2002.

Bergschmitt et al., *Proceedings of the European Workshop on the Evaluation of: Farm Investment Support and Investment Support for Improvement of Processing and Marketing of Agricultural Products,* 2006.

Berriet-Solliec, M., Labarthe, P., Laurent, C., Baudry, J., *Empirical validity of the evaluation of public policies: models of evaluation and quality of evidence,* 2011.

Bibi, S., & Duclos, J.-Y., *Equity and policy effectiveness with imperfect targeting*, Journal of Development Economics, 2007, 83(1), 109–140. doi:10.1016/j.jdeveco.2005.12.001.

Boardman, A., *Cost-Benefit Analysis – Concepts and Practice* 4th Edition, Prentice Hall, New Jersey, 2010.

Bonfiglio, A. and Chelli, F., *Assessing the Behaviour of Non-Survey Methods for Constructing Regional Input-Output Tables through a Monte Carlo Simulation, Economic Systems Research*, 2008, 20: 243 – 258.

Carter, R., *Theory-based evaluation approach* (GSDRC Helpdesk Research Report), Governance and Social Development Resource Centre, University of Birmingham, Birmingham, UK, 2012.

Chabé-Ferret, S., & Subervie, J., *How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DiD-matching,* Journal of Environmental Economics and Management, 2013, 65(1), 12–27.

Chen, H.T., *Theory-driven evaluations*, Thousand Oaks, CA, Sage, 1990.

Ciobanu, C., Mattas, K. and Psaltopoulos, D., *Structural Changes in Less-Developed Areas: an Input-Output Framework*, Regional Studies, 2004, 38: 603-614.

Coryn, C. et al., *A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009*, 2011.

Crawford, I. M., *Marketing Research and Information Systems,* Food and Agriculture Organisation of the United Nations, Hrsg., 1997.

Crawford, I. M., *Marketing Research and Information Systems,* Food and Agriculture Organisation of the United Nations, Hrsg., 1997.

Deaton, Angus., *Instruments, Randomization, and Learning about Development.* Journal of Economic Literature, 2010, 48(2): 424-55.

Diamond, A., & Sekhon, J. S., *Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies,*

Review of Economics and Statistics, 2013, 95(3), 932–945, doi:10.1162/REST_a_00318.

Doyle, C., Mitchell, M., and Topp, K., *Effectiveness of Farm Policies on Social and Economic Development in Rural Areas,* European Review of Agricultural Economics, 1997, **24:** 530-546.

Eiser, D. and Roberts, D., *The Employment and Output Effects of Changing Patterns of Afforestation in Scotland*, Journal of Agricultural Economics, 2002, **53:**65- 81.

Environmental management – Life cycle assessment – Principles and framework, International Organisation for Standardisation, ISO 14040, Geneva, 2006.

Environmental management – Life cycle assessment – Requirements and guidelines, International Organisation for Standardisation (ISO), Geneva, 2006.

Esposti, R., Sotte, F., *Evaluating the effectiveness of agricultural and rural policies: an introduction.*Eur Rev Agric Econ, 2013, 40, 535–539.

Esposti, R.; Bussoletti, St., *Impact of Objective 1 Funds on Regional Growth Convergence in the European Union: A Panel-Data Approach*, Regional Studies, March 2008, v. 42, iss. 2, pp. 159-73.

Ex post *evaluations of Rural Development Programmes 2000 – 2006*, Kantor, Commissioned by European Commission, DG AGRI, EC, 2012.

Farm Investment Behaviour Under the CAP Reform Process, Commissioned by JRC Sevilla, IPTS, EC, 2011.

Feil, J.-H., O. Musshoff, A. Balmann., *Policy Impact Analysis in the Dairy Sector - An Agent-Based Real Options Approach*, Paper presented at the European Association of Agricultural Economists, 122nd Seminar, Ancona, Italy, February 17-18 2011.

Flegg, A. T. and Webber, C.D., *On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables: Reply*, Regional Studies, 1997, 31: 795-805.

Flegg, A. T. and Webber, C.D., *Regional Size, Regional Specialization and the FLQ Formula*, Regional Studies, 2000, 34: 563-569.

Flegg, A. T. and Webber, C.D., *Using Location Quotients to Estimate Regional Input-Output Coefficients and Multipliers*, Local Economy Quarterly, 1996, 4: 58-86.

Flegg, A. T., Webber, C.D. and Elliott, M.V., *On the Appropriate Use of Location Quotients in Generating Regional Input-Output Tables*, Regional Studies, 1995, 29: 547-561.

Franzel, M., St. Kirchweger, T. Moser, M. Kapfer, M. Sandbichler und J. Kantelhardt., *Bedeutung der Investition sförderung für auf biologische Landwirtschaft umstellende Betriebe in Österreich*, Jahrbuch der Österreichischen Gesellschaft für Agrarökonomie, 2013, Vol. 22(2): 107-116.

Frolich, M., *Programme evaluation with multiple treatments*, Journal of Economic Surveys, 2004, 18(2), pages 181-224.

Frolich, M., *Programme evaluation with multiple treatments*, Journal of Economic, 2004.

Froschauer, U., & Lueger, M., *Das qualitative Interview: zur Praxis interpretativer Analyse sozialer Systeme*, WUV, Wien, 2003.

Greening Regional Development Programmes Network, Handbook on SEA for Cohesion Policy 2007-2013, Brussels, 2006.

Hanley, N., Spash, C., *Cost-Benefit Analysis and the Environment*, Edward Elgar, Cheltenham, (993.

Hanley, N., Whitby, M., and Simpson, I., *Assessing the success of agri-environmental policy in the UK,* Land Use Policy, 1999, 16(2), 67–80.

Heumesser, Ch. and U. Morawetz., *Analyzing the effect of agri-environmental policies on nitrate concentration in groundwater*. Presented at the 52ndGEWISOLA Annual Conference, Universität Hohenheim, September 2012, 26-28, 2012.

Husson, F., Le, S., & Pages, J., *Exploratory Multivariate Analysis by Example Using R*. CRC Press, (2011).

Ifs et al., *Assessing the Impact of Rural Development Policies*, FP7 project (no. 213034), 2010, various reports.

Jensen, R. C., Mandeville, T. D. and Karunarante, N. D., *Regional Economic Planning: Generation of Regional Input-Output Analysis*, Croom Helm, London, 1979.

Johns, P. M. and Leat, P. M. K., *The Application of Modified GRIT Input-Output Procedures to Rural Development Analysis in Grampian Region*, Journal of Agricultural Economics, 1987, 32: 243-256.

Karanikolas P, et al., *The modernization process in Greek Agriculture: The case of investment aid*, Agricultural economics review, 2007, Vol 8, nr 1.

Karunarante, N. D., *Australian Development Issues: An Input-Output Analysis*, Newcastle upon Tyne, Avebury, 1989.

Keenleyside, C., Allen, B., Hart, K., Menadue, H., Stefanova, V., Prazan, J., Boatman, N., *Delivering environmental benefits through entry level agri-environment schemes in the EU*, Institute for European Environmental Policy, London, Report Prepared for DG Environment, EC, 2011.

Kirchweger, S. andKantelhardt, J., *Improving Farm Competitiveness through Farm-Investment Support: a Propensity Score Matching Approach*, Paper presented at the 131st EAAE Seminar, Prague, Czech Republic, September 18-19, 2012.

Kirchweger, St., M. Eder, M. Kapfer and J. Kantelhardt., *Using a matching analysis to evaluate the structural effects of farm-investment support in Austria*, Paper presented at the 85th Annual Conference of the Agricultural Economics Society at Warwick University, 18 – 20 April 2011.

Leontief, W., *Studies in the Structure of the American Economy*, Oxford University Press, London, 1953.

Lumley, T., *Analysis of Complex Survey Samples,* Journal of Statistical Software, 2004, 9(8), 1–19.

Mattas, K., *An Estimation of the Impacts of Integrated Rural Development Policies in Greece: An I-O Approach,* Ministry of Agriculture, Athens, 2001.

Medonos, T., Ratinger, T., Hruška, M., Špička, J., *The Assessment of the effects of investment support measures of the Rural Development Programmes: the case of the*

Czech Republic, The journal AGRIS on-line Papers in Economics and Informatics (http://online.agris.cz), 2012, 4: 35-47.

Michalek J. and N. Zarnekow., *Construction and Application of the Rural Development Index to analysis of rural regions*, European Commission, JRC Scientific and Policy Reports, 2012, pp 1-96.

Michalek J., *Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", Volume 1 – A micro-level approach*, European Commission, JRC Scientific and Policy Reports, 2012, pp 1-95.

Michalek J., *Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States", Volume 2 – A regional approach*, European Commission, JRC Scientific and Policy Reports, 2012, pp 1-83.

Michalek J., *Counterfactual impact evaluation of EU Rural Development Programmes propensity score matching methodology applied to selected EU Member States*, Publications Office, Luxembourg, 2012, Volume 1.

Miernyk, W. H., *Elements in Input-Output Analysis*, Random House, New York, 1969.

Miller, R.E and Blair, P.D., *Input-Output Analysis: Foundations and Extensions,* Cambridge University Press, Cambridge, (2009).

Morawetz, U. B., *A concept for a randomized evaluation of agri-environment measures,* In E. Schmid & S. Vogel (Editor), The Common Agricultural Policy in the 21st Century (S. 113–130). Vienna, Austria, 2014.

Morgan, S. L. and Winship, C., *Counterfactual and causal inference: Methods and principles for social research*, Cambridge University Press, New York, 2007.

Morris St., Tödtling-Schönhofer H., Wiseman M., *Design and Commissioning of Counterfactual Impact Evaluations,* Commissioned by DG Employment, Social Affairs and Inclusion, Unit A3, EC, 2013.

ÖIR., *Synthesis of the Mid-term evaluations of Rural Development Programmes 2007 – 2013*; Commissioned by European Commission; DG AGRI, 2012.

Partidário, M., *Strategic Environmental Assessment, Good Practices Guide*, Methodological Guidance, Portuguese Environment Agency, Amadora, 2007.

Pearce, D., et al., *Cost-Benefit Analysis and the Environment, recent developments*, OECD, Paris, 2006.

Pearce, D., *What Constitutes a Good Agri-environmental Policy Evaluation?*, In Evaluating Agri-Environmental Policies (S. 71–97). OECD Publishing, 2005.

Petrick M, Zier P., *Regional employment impacts of Common Agricultural Policy measures in Eastern Germany: a difference-in-differences approach,* Agricultural Economics, 2011, 42(2):183–93.

Psaltopoulos, D. and Efstratoglou, S., *An Empirical Evaluation of EU and National Structural Policies in Remote Rural Areas: The Case of Evrytania*, Agricultural Economics Review, 2000, 1: 7-18.

Psaltopoulos, D. and Thomson, K. J., *Input-Output Evaluation of Rural Development: a Forestry-Centred Application,* Journal of Rural Studies, 1993, 9: 351-358.

Psaltopoulos, D., Balamou, E. and Thomson, K. J., *Rural/Urban Impacts of CAP measures in Greece: an Interregional SAM Approach*, Journal of Agricultural Economics, 2006, 57(3), 441-458.

Psaltopoulos, D., Phimister, E., Ratinger, T., Roberts, D., Skuras, D, Balamou, E., Espinosa, M., Gomez y Paloma, S., Mary, S., Nohel, F., Santini, F., Ex ante *Spatial Policy Impact Assessment of the Rural Development Policy in European Rural areas,* (Sieber S. editor), European Commission, JRC, 2012, IPTS. EUR 25238 EN, Pp 180.

Psaltopoulos, D.,Skuras, D., Thomson, K.J. (2011). Employment Effects of Private Investment Initiatives in Rural Areas of Southern Europe: A Regional SAM Approach, Agricultural Economics Review, 12: 50-61.

R Core Team., *R: A language and environment for statistical computing,* R Foundation for Statistical Computing.Vienna, Austria, 2012.

Richardson, H. W., *Input-Output and Regional Economics*, Halsted Press, John Willey and Sons, New York, 1972.

Salhofer K, Hofreither MF, Sinabell F., *Utilising Equilibrium-Displacement Models to Evaluate the Market Effects of Countryside Stewardship Policies: Method and Application,* Die Bodenkultur, 1999, 50:141–51.

Schaffer, W. (ed.)., *On the Use of Input-Output Models for Regional Planning*, Leiden: Martinous Nijhoff, 1976.

Schönhart, M; Schauppenlehner, T; Schmid, E; Muhar, A., *Analysing the maintenance and establishment of orchard meadows at farm and landscape levels applying a spatially explicit integrated modelling approach,* Journal of Environmental Planning and Management, 2011, 54(1): 115-143.

Schröder L. Et al., *Impact of second Pillar funding – Validation from a modelling- and an evaluation perspective,* FP6 project.WP6 Policy Impact Assessment, Deliverable: D6.7 – Comparison of selected impact assessment results for Germany with the findings from evaluation reports, 2013.

Sekhon, J. S., *Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R,* Journal of Statistical Software, 2011, 42(7), 1–52.

Shaxson, L., *Is your Evidence Robust Enough? Questions for policy makers and practitioners,* Brokering Digest, 16, June 2006.

Sinabell F, Bock-Schappelwein J, Mayer C, Kniepert M, Schmid E, Schönhart M, Streicher, G., *Indikatoren für die Auswirkungen des Programmems der Ländlichen Entwicklung 2007/2013 in Österreich,* Wien, 2011.

Sinabell F., Streicher G., *Programme evaluation with micro-data: the use of FADN data to evaluate effects on the market situation of programmeme participants,* In: Ortner K. Assessing rural development policies of the Common Agricultural Policy Selection of papers from the 87th Seminar of the European Association of Agricultural Economists (EAAE) April 21-23 2004, Wien, Austria. Kiel, 2005, 2004, S. 265–78.

Sommer, A., *Strategic environmental assessment: From scoping to monitoring, Content requirements and proposals for practical work*, Hallein, (2005).

Spaziante, A., et al., *Strategic Environmental Assessment (SEA) of Rural Development Programmes in the European Union-Towards a More Efficient Monitoring of the Environmental Effects of Agricultural Policies*, Inter-University Department of Regional and Urban Studies and Planning, Torino, 2010.

Stone, R. and Brown, A., *A Programme for Growth, Part I: A Computable Model of Economic Growth*, London, 1962.

The Programming Period 2014-2020 - Draft guidance document on monitoring and evaluation for ERDF - Concepts and Recommendations, EC, April 2013.

Thomson, K.J. and Psaltopoulos, D., *Incorporating Output Projections into a Regional Input-Output Model: The Case of Forestry in Rural Scotland*, Journal of Applied Input-Output Analysis, 2000, 6: 1-16.

Weingarten, P., Neumeier, S., Copus, A., Psaltopoulos, D., Skuras, D., and Balamou, E., *Building a Typology of European Rural Areas for the Spatial Impact Assessment of Policies (TERA-SIAP)*, JRC Scientific and Technical Reports, Office for Official Publications of the European Communities, Luxembourg, 2010.

West, G. R., Wilkinson, J. T. and Jensen, R. C., *Generation of Regional Input-Output Tables for the Northern Territory,* Report to the Northern Territory Department of the Chief Minister, University of Queensland: Department of Economics, 1980.

What Works Clearinghouse, Procedures and Standards Handbook, IES, 2014, Version 3.0.

White, H., *Theory-Based Impact Evaluation: Principles and Practice*, 3ie Working Paper [New Delhi: International Initiative for Impact Evaluation], June 2009, http://www.3ieimpact.org/

WHO, *Making choices in health: WHO guide to cost-effectiveness analysis*, Geneva 2003.

7.3 Specific sources for IO-Method

Deaton, Angus., *Instruments, Randomization, and Learning about Development*, Journal of Economic Literature, 2010, 48(2): 424-55.

Doyle, C., Mitchell, M., and Topp, K., *Effectiveness of farm policies on social and economic development in rural areas,* European Review of Agricultural Economics, 1997, 24: 530-546.

Eiser, D. and Roberts, D., *The employment and output effects of changing patterns of afforestation in Scotland,* Journal of Agricultural Economics, 2002, 53:65- 81.

Karunarante, N. D., *Australian Development Issues: An Input-Output Analysis*. Newcastle upon Tyne, Avebury, 1989.

Leontief, W., *Studies in the Structure of the American Economy,* Oxford University Press, London, 1953.

Mattas, K., *An Estimation of the Impacts of Integrated Rural Development Policies in Greece: An I-O Approach*, Ministry of Agriculture, Athens, 2001.

Miernyk, W. H., *Elements in Input-Output Analysis,* Random House, New York, 1969.

Miller, R.E and Blair, P.D., *Input-Output Analysis: Foundations and Extensions,* Cambridge University Press, Cambridge, 2009.

Psaltopoulos, D. and Efstratoglou, S., *An Empirical Evaluation of EU and National Structural Policies in Remote Rural Areas: The Case of Evrytania*, Agricultural Economics Review, 2000, 1: 7-18.

Psaltopoulos, D. and Thomson, K. J., *Input-Output Evaluation of Rural Development: a Forestry-Centred Application,* Journal of Rural Studies, 1993, 9: 351-358.

Psaltopoulos, D. Thomson, K.J., Efstratoglou, S., Kola, J. and Daouli, A., *Regional Social Accounting Matrices for Policy Analysis in Lagging EU Rural Regions*, European Review of Agricultural Economics, 2004, 31: 149-178

Richardson, H. W., *Input-Output and Regional Economics*, Halsted Press, John Willey and Sons, New York, 1972.

Schaffer, W. (ed.)., *On the Use of Input-Output Models for Regional Planning.* Leiden: Martinous Nijhoff, 1976.

Thomson, K.J. and Psaltopoulos, D., *Incorporating Output Projections into a Regional Input-Output Model: The Case of Forestry in Rural Scotland*, Journal of Applied Input-Output Analysis, 2000, 6: 1-16.

Weingarten, P., Neumeier, S., Copus, A., Psaltopoulos, D., Skuras, D., and Balamou, E., *Building a Typology of European Rural Areas for the Spatial Impact Assessment of Policies (TERA-SIAP)*, JRC Scientific and Technical Reports, Office for Official Publications of the European Communities, Luxembourg, 2010.

7.4 Specific sources for Counterfactual design

ADB, *A review of recent developments in impact evaluation*, Asian Development Bank, 2011.

ADB. *Impact Evaluation: Methodological and Operational Issues,* Asian Development Bank, Manila, 2010.

Becker S. A. et. al., *Too much of a good thing? On the growth effects of the EU regional policy*, European Economic Review 56, 2012, 648–668.

Bryson, A. et.al, *The use of propensity score matching in the evaluation of active labour market policies*, Policy Studies Institute, U.K. Department for Work and Pensions Working Paper No. 4., 2002.

Caliendo, M. et. al., *The Employment Effects of Job Creation Schemes in Germany: A Micro econometric Evaluation*, IZA Discussion Papers 1512, Institute for the Study of Labour (IZA), 2005.

Chabé-Ferret, S., & Subervie, J., *How much green for the buck? Estimating additional and windfall effects of French agro-environmental schemes by DID-matching,* Journal of Environmental Economics and Management, 2013, 65(1), 12–27.

Dantler M. et al., *Analyse der Investitionsförderung für landwirtschaftliche Betriebe in Österreich*, Universität für Bodenkultur, Wien, 2010.

Dehejia, R. H. and Wahba, S., *Propensity Score-Matching Methods For nonexperimental Causal Studies*, The Review of Economics and Statistics, 2002, 84, pp. 151-161.

Gertler P.J. et al., *Impact Evaluation in Practice*, International Bank for Construction and Development and the World Bank, 2011.

Guo, S.Y. and M.W. Fraser, *Propensity Score Analysis*, Statisctical Methods and Applications, Sage Publications Inc., Thousand Oaks, 2010.

Heckman, James J., Ichimura, Hidehiko and Todd, Petra E., *Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme*, Review of Economic Studies, Blackwell Publishing, 1997, 64(4), pp. 605-54

IEG (Independent Evaluation Group)., *Impact Evaluations in Agriculture: An Assessment of the Evidence,* World Bank, Washington DC., 2011.

Imbens, G. M., and J. M. Wooldridge., *Recent Developments in the Econometrics of Programme Evaluation,* NBER Working Paper No. 14251, National Bureau of Economic Research, Massachusetts, 2008.

Khandker, S. R. et al., *Handbook on Impact Evaluation. Quantitative Methods and Practices*, World Bank, Washington DC., 2010.

Lokshin, M. and Yemtsov, R., *Has Rural Infrastructure Rehabilitation In Georgia Helped the Poor?,* World Bank Economic Review, 2005, 19(2), pp. 311-333.

Michalek J. et. al., *Capitalization of CAP Single Payment Scheme into Land Value: Generalized Propensity Score Evidence from the EU*, Land Economics, May 2014, (forthcoming).

Michalek J., *Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States, Volume 1 – A micro-level approach*, European Commission, JRC Scientific and Policy Reports, 2012, pp 1-95.

Michalek J., *Counterfactual impact evaluation of EU Rural Development Programmes - Propensity Score Matching methodology applied to selected EU Member States,*

Volume 2 – A regional approach, European Commission, JRC Scientific and Policy Reports, 2012, pp 1-83.

Morawetz, U. B., *A concept for a randomized evaluation of agri-environment measures,* In E. Schmid & S. Vogel (Ed.), The Common Agricultural Policy in the 21st Century (S. 113–130). Vienna, Austria, 2014.

Morgan, L.S. and Ch. Winship, *Counterfactuals and Causal Inference: Methods and Principles for Social Research* (Analytical Methods for Social Research), Cambridge University Press, Cambridge, 2007.

NONIE, *Impact Evaluation Guidance*, January 2008.

Pufahl, A., Ch. R. Weiss, *Effekte von Agrarumweltmaßnahmen und der Ausgleichszulage auf den betrieblichen Faktoreinsatz: Ergebnisse einer Propensity-Score-Matching-Analyse*, In Agrarwirtschaft, 2010, 59(1): 13-29.

Ravallion, M., *Evaluating Anti-Poverty Programmes*, in Handbook of Development Economics, Volume 4, edited by T. Paul Schultz and John Strauss, Amsterdam, North-Holland, 2006.

Rosenbaum, Paul, and Donald Rubin, *The Central Role of the Propensity Score in Observational Studies of Causal Effects,* Biometrika, 1983, 70 (1): 41–55.;

Sianesi, B., *An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s*,The Review of Economics and Statistics, MIT Press, 2004, 86(1), pp.133-155.

Todd, P., *Evaluating Social Programmes with Endogenous Programme Placement and Selection of the Treated,Handbook of Agricultural Economics*, Volume 4, North Holland, (eds.) R. E. Evenson and T. P. Schultz, 2006.

Van de Walle, D., and Cratty, D., *Impact Evaluation of a Rural Road Rehabilitation Project*, World Bank, Washington DC., 2002.

7.5 Specific sources for MAPP method

Description and examples of MAPP: *A participatory method for impact assessment of programmes and projects*.

Susanne Neubert, *MAPP – a participatory method for impact assessment of programmes and projects*.

7.6 Specific sources for Theory-Based Impact Evaluation

“The Center for Theory of Change” / Helen Clark: http://www.theoryofchange.org/ (description of general approach but not directly linked to the intervention logic of EU programmes)

Baker, J., *Evaluating the Impact of Development Projects on Poverty*, A Handbook for Practitioners, World Bank, (2000).

Carter, R., *Theory-based evaluation approach* (GSDRC Helpdesk Research Report), Governance and Social Development Resource Centre, University of Birmingham, Birmingham, UK, 2012.

Chen, H.T., *Theory-driven evaluations*, Thousand Oaks, CA, Sage, 1990.

Coryn, C. et al., *A Systematic Review of Theory-Driven Evaluation Practice From 1990 to 2009*, 2011.

Giel, Susanne, *Theoriebasierte Evaluation*, Waxmann Verlag, 2013, (in German language only).

Morris St., Tödtling-Schönhofer H., Wiseman M., *Design and Commissioning of Counterfactual Impact Evaluations,* Commissioned by DG Employment, Social Affairs and Inclusion, Unit A3, EC, 2013.

TBE section of the Evalsed Sourcebook (DG Regional Policy, 2013/2014).

The Programming Period 2014-2020 - Draft guidance document on monitoring and evaluation for ERDF - Concepts and Recommendations, EC, April 2013.

White, H., *Theory-Based Impact Evaluation: Principles and Practice,* 3ie Working Paper [New Delhi: International Initiative for Impact Evaluation], June 2009. http://www.3ieimpact.org/

7.7 Specific sources for SEA

A Practical Guide to the Strategic Environmental Assessment Directive, Office of the Deputy Prime Minster, 2005.

Austrian RDP „Anlage IV a, *Umweltbericht im Rahmen der Strategischen Umweltprüfung* gem. RL 2001/42/EG“ (in German).

Greening Regional Development Programmes Network, Handbook on SEA for Cohesion Policy 2007-2013, 2006.

Guidelines for the ex ante *evaluation of 2014-2020 RDP*, European Evaluation Network for Rural Development, Brussels, 2012.

Hungarian RDP, *Part II of the* ex ante *evaluation of the “New Hungary” Rural Development Strategic Plan and Programme*, Strategic Environmental Assessment”.

Irish Environmental Protection Agency, *Review of Effectiveness of SEA in Ireland: Key Findings & Recommendations*, Wexford, 2012.

Lower Saxony-Bremen RDP, *Umweltbericht der Strategischen Umweltprüfung zum Niedersächsischen und Bremischen Programm für die Förderung der Entwicklung des ländlichen Raumes 2007-2013,* (in German).

Sommer, A., *Strategic environmental assessment: From scoping to monitoring, Content requirements and proposals for practical work*, Hallein, 2005.

Spaziante, A., et al., *Strategic Environmental Assessment (SEA) of Rural Development Programmes in the European Union-Towards a More Efficient Monitoring of the Environmental Effects of Agricultural Policies*, Inter-University Department of Regional and Urban Studies and Planning, Torino, 2010.

Study concerning the report on the application and effectiveness of the SEA Directive, (2001/42/EC), Final report, Brussels; EC, 2009.

Umweltbundesamt, *Leitfaden zur Strategischen Umweltprüfung* (SUP), Dessau-Roßlau, 2008, (in German)

Wales RDP, *Rural Development Plan for Wales 2007-2013: Annex 3 Strategic Environmental Assessment.*

7.8 Specific sources for CEA

Balana , B., Vinten, A., Slee, B., *A review on cost-effectiveness analysis of agri-environmental measures related to the EU WFD Key issues, methods, and applications*, Ecological Economics, 2010, 70 (2011) 1021–1031.

Bleichrodt, H., Quiggin, J., *Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis?*, J Health Econ, 1999, 18 (6): 681–708;

Böhm, E., et al., *Kosten-Wirksamkeitsanalyse von nachhaltigen Maßnahmen im Gewässerschutz,* Umweltbundesamt Texte 12/02, Berlin, 2002.

Fährmann, B., Grajewski, R., *How expensive is the implementation of Rural Development Programmes? Empirical results of implementation costs and their consideration in cost-effectiveness analyses*, Ancona, 2011.

Majewski, E., *Cost-effectiveness assessment of improving animal welfare standards in the European Agriculture*, Selected Paper prepared for presentation at the International Association of Agricultural Economists (IAAE) Triennial Conference, Foz do Iguaçu, Brazil, 2012.

The implementation Costs of Agricultural Policies, OECD, Paris, 2007.

8 Glossary of key terms

This glossary is based on "Glossary of key-terms concerning Evaluation of the Evaluation Expert Network" from May 2012. Many of these definitions are based on the existing glossary annexed to the Handbook on Common Monitoring and Evaluation Framework (CMEF) (see Guidance Note N: Glossary of Terms).

Aid intensity differentiation

Definition:	Making maximum funding or top-up dependent on applicant or project specific characteristics (e.g. higher funding in less favoured regions or top ups for investments in organic barns).
Comments:	"A first level of targeting may be achieved by setting restrictive eligibility criteria and **differentiated aid rates** for different types of investment project, which should be detailed in the RDP…."[59]

Appropriateness

Definition:	This term is used in the context of this study to describe the degree to which a given method can be used to provide valid results to answer the Evaluation Questions under consideration. A set of criteria (among them reliability, robustness) is used to allow a judgement on how appropriate a method is for the analysis of investment support measures. A verbal description is used for the judgement.

Baseline

Definition:	"State of the economic, social or environmental situation relevant in the context of a programme, at a given time (generally at the beginning of the intervention), and against which changes will be measured."
Comments:	Establishing of a reasonable baseline against which changes (e.g. RDP programmes) will be measured is one of the most important tasks in evaluation. It is therefore useful to call it a "policy-off" scenario. When results of a given programme are evaluated at a micro-level a baseline is a performance of adequate control groups during the entire period of analysis, e.g. 2007-2013. At a macro-level a relevant baseline is a state of economy without a given programme. Because changes of policies should NOT be evaluated by comparing observed results "after" the programme with "before" the programme our suggestion is to delete the following part of the CMEF definition ", at a given time (generally at the beginning of the intervention), and".
Source:	CMEF

Baseline indicators

Definition:	"Baseline indicators reflect the state of the economic, social or environmental situation, at a given time (generally at the beginning of the intervention)." ... "They fall into two categories: 1) Objective related baseline indicators. These are directly linked to the wider objectives of the programme." ... " 2) Context related baseline indicators. These provide information on relevant aspects of the general contextual trends

[59] European Court of Editors special report on targeting 121, 2012, p. 13.

	that are likely to have an influence on the performance of the programme."
Comments:	see comments to “baseline”
Source:	CMEF, shortened

Beneficiary

Definition:	A person or organisation directly benefitting from the intervention whether intended or unintended. Some people may be beneficiaries without necessarily belonging to the group targeted by the intervention. Similarly, the entire eligible group does not necessarily consist of beneficiaries.
Source:	CMEF

Control-group design (Experimental design)

Definition:	The “Golden standard” in evaluation is randomised controlled experiments or the so- called “experimental designs”, where randomly selected groups receive support (or “treatment”, as technical term) and randomly selected control group does not. Under experimental design control group is selected randomly from the same population as the programme participants. The randomised assignment process itself creates comparable treatment and control groups that are statistically equivalent to one another, given appropriate sample sizes. As treated and controls are comparable in their characteristics before the policy intervention so then any observable differences afterwards can be attributed to the intervention. However, conducting field experiments (randomised selection) poses several methodical challenges like external validity, spill-over effects, dynamic selection, etc. Moreover, they can also be practically or ethically impossible or socially unacceptable. For example, as it is mostly impossible to randomly assign persons or economic entities to a subsidy or exclude them, other methodological possibilities to design the counterfactual have to be applied, e.g. quasi-experimental approaches.

Cost-Benefit Analysis (CBA)

Definition:	Cost-Benefit Analysis is often used to estimate benefits and costs of the project, decision or policy and compare and justify various types of interventions. Cost-benefit analysis (CBA) starts with the fundamental question: “To what extent does the expected (social) benefits for a given action surpass (social) costs?” This is the fundamental decision rule; other considerations, e.g. political, ethical etc., are “ignored”. The comparison of benefits and costs is commonly expressed as Benefit/Cost Ratio: Sum of all monetised benefits/sum of all net costs

Cost-Effectiveness Analysis (CEA)

Definition:	CEA is basically designed to compare cost (expressed in monetary values) and effectiveness (expressed with an effectiveness indicator, e.g. a range of 1-100, where the maximum value represents full achievement) for a range of alternatives. This is expressed as the Cost/Effectiveness Ratio (CER), e.g. EUR per Ha protected area. Hence cost is expressed in EUR, while effectiveness in a "programme objective" unit.
Comments:	The Cost-Effectiveness Analysis (CEA) is similar in its logic to the Cost-Benefit Analysis but cannot answer the question if benefits surpass costs.

Counterfactual situation

Definition:	A situation which would have occurred in the absence of a public intervention, also referred to as "policy-off" situation. By comparing the counterfactual and real situations, it is possible to determine the net effects of the public intervention. Various tools can be used for the construction of the counterfactual situation: shift-share analysis, comparison groups, simulation using econometric models, etc. At the baseline, the real situation and the counterfactual situation are identical. If the intervention is effective, they diverge.
Comments:	The assessment of programme effects should involve counterfactual analysis. The key in the counterfactual analysis is to construct a group which is as similar as possible (in observable and unobservable dimensions) to those receiving the intervention. This comparison allows for the establishment of causality – attributing observed changes in outcomes to the programme, while removing confounding factors.
Source:	CMEF

Counterfactual design (micro-level approach)

Definition:	The main challenge of any impact evaluation - is to provide evidence of a true cause-and-effect link between the observed indicators and the programme. Solving this problem has always to do with the "attribution" of the change observed to the intervention that has been implemented. Is the observed change in indicators due to the policy or would it have occurred anyway? The crucial consideration in impact analysis is a fact that programme's impact/result can never be directly observed as it always requires estimation and comparison with outcomes in situation when a given programme is not implemented. Yet, results and impacts can be inferred, as long as the available data allows a credible way to approximate the counterfactual. The purpose of a counterfactual in evaluations is to address the question "What would have been the situation of the programme/measure beneficiary if the programme/measure had not taken place?" Generally, there are various ways on how to conduct evaluation of policy interventions. They relate to evaluation design, which is determined by the choice of methods used to identify a comparison or control group (counterfactual). Evaluation design can be broadly classified into three categories:

	a) Randomised design (sometimes called experimental approach); b) Quasi-experimental design; and c) Non-experimental design.

Deadweight

Definition:	"*Changes observed in the situation of beneficiaries following the public intervention, or reported by direct addressees as a consequence of the public intervention, that would have occurred, even without the intervention….*" The deadweight can be estimated by econometric modeling of counterfactual situations.
Source:	CMEF, shortened

Efficiency

Definition:	*"Best relationship between resources employed and results achieved in pursuing a given objective through an intervention."*
Comments:	When assessing programme efficiency, evaluation looks at the relationship between the allocated resources and achieved programme outputs and consequently results. An analysis of programme efficiency reveals whether more outputs and results could have been obtained with the same budget, or whether the same outputs and results could have been reached with a lower cost. Efficiency is studied by looking at the size of the budget and its division between axes and individual measures, as well as the examination of budget sufficiency to achieve policy objectives and contribute to the Community priorities. The efficiency of the RDP is also affected by the delivery mechanisms and implementation procedures, so these factors should also be taken into account in the *ex post* evaluation. In evaluating the efficiency of a programme it is useful to consider the following questions: To what extent were the allocated resources able to produce expected programme results and impacts? Could the obtained results/impacts be produced at lower costs? What are the factors influencing cost efficiency of programme implementation? In economics the situation with the *"best relationship between resources employed and results achieved"* is called to be "optimal". Heckman (2010)[60]: *"In the economic theory of policy evaluation, a comparison between marginal benefits and marginal costs determines the optimal size of social programmes."* Pearce, Atkinson and Mourato (2006)[61]: the *"optimal scale is where the marginal social benefits of the project/policy are just equal to the marginal costs of the project policy. "Marginal" here simply means 'small change in'. So the marginal benefit of a policy is the extra benefit that accrues to society from one small change in the "quantity" of the policy".*
Source:	CMEF, shortened; *Ex post* evaluation guidelines, 2014.

[60] Heckman, J. J., "Building Bridges between Structural and Programme Evaluation Approaches to Evaluating Policy", Journal of Economic Literature, 2010, 48(2), pp. 356–398.
[61] Pearce, D., Atkinson, G., Mourato, S., Cost-Benefit Analysis and the Environment, Organisation for Economic Co-operation and Development, Paris, 2006.

Effectiveness

Definition:	This is the extent to which objectives pursued by an intervention are achieved. An effectiveness indicator is calculated by relating an output, result or impact indicator to a quantified objective.
Comments:	The main indicator of the effectiveness of a given programme measure should be a ratio of outcomes achieved by programme beneficiaries due to the programme compared to target values (equivalent to) an increase of a given result indicator due to a given programme measure compared with target values. However, in case of absence of programme target values above indicator cannot be used. In such a situation some more information about programme effectiveness can be obtained by analysing the following supplementary indicators: a) Outcomes achieved by programme beneficiaries compared to outcomes achieved by programme non-beneficiaries (in %) (equivalent to) an increase of a given result indicator for programme beneficiaries compared to an increase of the same result indicator for the control group; and b) structure of a total increase of a given result indicator (% share due to a given measure compared to % share due to other factors).
Source:	CMEF

Eligibility criteria

Definition:	Criteria defining who is eligible for funding. Fulfilling eligibility criteria does not imply funding, as the Managing Authority may apply selection criteria among the eligible applicants.
Comments:	*"A first level of targeting may be achieved by setting restrictive **eligibility criteria** and differentiated aid rates for different types of investment project, which should be detailed in the RDP…."*[62]

GRIT procedure

Definition:	GRIT (Generation of Regional Input-Output Tables) is a "variable-interference" hybrid technique for generating regional IO tables based on the concept of "holistic accuracy". A mechanical procedure (application of location quotients) is initially applied to adjust national IO tables to the regional level. Then, the analyst can determine the extent to which he/she "interferes" by the insertion of "superior" data from surveys or other sources. As a result, GRIT includes the advantages of both "survey" and "non-survey" techniques.
Comments:	The GRIT technique was developed at the University of Queensland, Australia, and was originally applied to the estimation of IO tables for the regions of Queensland.
Source:	Jensen et al. (1979)

[62] European Court of Auditors special report on targeting 121, 2012, p. 13

Gross effect

Definition:	Change observed following a public intervention, or an effect reported by the direct beneficiaries. A gross effect appears to be the consequence of an intervention but usually it cannot be entirely imputed to it.
Source:	CMEF, shortened

Gross Value Added (GVA)

Definition:	The concept is used in the European System of Accounts. Gross Value Added (ESA 1995, 8.11) is the net result of output valued at basic prices less intermediate consumption valued at purchasers' prices. Gross value added is calculated before consumption of fixed capital. It is equal to the difference between output (ESA 1995, 3.14) and intermediate consumption (ESA 1995, 3.69).
Source:	CMEF

Impact

Definition:	Effects of an intervention lasting in medium or long-term. Some impacts appear indirectly, (e.g. turnover generated for the suppliers of assisted firms). Others can be observed at the macro-economic or macro-social level (e.g. improvement of the image of the assisted area); these are global impacts. Impacts may be positive or negative, expected or unexpected.
Comments:	Impacts refer to the benefits of the programme beyond the immediate effects on its direct beneficiaries both at the level of the intervention but also more generally in the programme area. They are linked to the wider objectives of the programme. They are normally expressed in "net" terms, which means after subtracting effects that cannot be attributed to the intervention (e.g. double counting, deadweight), and taking into account indirect effects (displacement and multipliers).
Source:	CMEF

Impact indicators

Definition:	These refer to the benefits of the programme both at the level of the intervention but also more generally in the programme area. They are linked to the wider objectives of the RDP. Impact indicators are perhaps the most important of all in assessing the success of the RDP. However, they cannot be understood in isolation, and in order to explain observed outcomes, reference may need to be made to output and result indicators. Example: increase in employment in rural areas, increased productivity of agricultural sector, increased production of renewable energy
Source:	CMEF, shortened

Input

Definition:	Financial, human, material, organisational and regulatory means mobilised for the implementation of an intervention.
Source:	CMEF, shortened

Input indicators

Definition:	These refer to the budget or other resources allocated at each level of the assistance. Financial input indicators are used to monitor progress in terms of the (annual) commitment and payment of the funds available for any operation, measure or programme in relation to its eligible costs. Example: expenditure per measure declared to the Commission.
Source:	CMEF

Input-Output Analysis (macro-level approach)

Definition:	Input-Output (IO) analysis is a quantitative technique for studying the interdependence of the producing and consuming units within an economy. An I/O table identifies the major industries in an economy and the financial flows between them over a stated time period (usually a year). It indicates the sources of each sector's inputs, which are purchased from the same or other sectors in the economy, imported, or earned by labour (household's wages and salaries). It also provides a breakdown for each sector's output, which can be sales to other industries and to final demand (household consumption, government consumption, capital formation, and exports).
Comments:	In an IO context, interdependence between the individual sectors of the given economy is normally described by a set of linear equations, representing fixed shares of input in the production of each output. Despite this rather simplistic assumption, IO models have been extensively used for studying economic interdependence, structural change and especially, for assessing the economic impacts of polcy intervention.
Source:	Miller and Blair (2009)

Intervention logic

Definition:	An intervention logic represents a methodological instrument which establishes the logical link between programme objectives and the envisaged operational actions. It shows the conceptual link from an intervention's input to its output and, subsequently, to its results and impacts. Thus an intervention logic allows an assessment of a measure's contribution to achieving its objectives.
Source:	CMEF

Judgement criteria

Definition:	The judgement criteria specify the aspects against which the merits of the intervention are judged. They are used in conjunction with evidence collected (indicators and other relevant information), to answer the Evaluation Questions.
Source:	Guidelines for the *Ex Post* Evaluation of 2007-2013 RDPs, June 2014, p. 5

Leverage effect

Definition:	In the context of investment support: The leverage effect *"…occurs if public funding (e.g. in form of RD programme) induces private spending among the programme beneficiaries"*.[63] Leverage effects can be estimated by econometric modeling of counterfactual situations.

Life Cycle Assessment (LCA)

Definition:	Life cycle assessment (LCA) is a method of analysing the environmental impacts of a process, product or activity along its life cycle, for example from "the cradle to the grave" and identifying where better environmental performance can be (or was) achieved. LCA does not just look at the impacts directly arising from a project or policy, but at the whole "life cycle" of impacts. The process is defined in detail in the ISO 14040ff of the International Standards Organisation[64] and has already found application in many agricultural case studies[65]. The LCA approach also forms the basis for a range of well-known "footprint" assessments.
Comments:	In the context of the present study LCA could theoretically provide the environmental equivalent to the FADN and bookkeeping data allowing also for counterfactual analysis. In praxis these data are not recorded or their collection would be time-consuming and costly; in best cases one would have to work with case studies and best estimates.
Source:	International Organisation for Standardisation, (2006) ISO 14040ff, "Environmental management -- Life cycle assessment", Geneva

[63] Michalek, J (2012) "Counterfactual impact evaluation of EU Rural Development Programmes – Propensity Score Matching Methodology applied to selected EU Member States. Volume 1: A micro-level approach. JRC Scientific and Policy Reports.
[64] International Organisation for Standardisation, (2006) ISO 14040ff, "Environmental management -- Life cycle assessment", Geneva.
[65] Pergola, M, et al., (2013), Sustainability evaluation of Sicily's lemon and orange production: energy, economic and environmental analysis, J Environmental Management. 2013 Oct 15;128:674-82.

MAPP method (participatory impact analysis method at the micro-level)

Definition:	MAPP is a participatory method for the assessment of impacts of programmes and projects. It is a methodological framework combining a qualitative approach with participatory assessment instruments, but it also includes a quantification step. With MAPP, beneficiaries are evaluating the impacts of rural development interventions following a logical structure. By doing so, they also express their ideas on how programme interventions can be improved for further development. MAPP is useful for conducting case studies, which are particularly valuable to capture individual differences or unique variations from one intervention experience to another.
Comments:	MAPP orients itself towards principles and procedures of Participatory Rural Appraisal (PRA) methodology, such as: • *Triangulation*: the collection of distinct data with different tools in order to prove or raise the validity of the data; • *Optimal ignorance:* the capability to select relevant data and to avoid an information overkill; • *Communal learning: the findings of an assessment are the result of a communication process among relevant groups.*
Source:	Susanne Neubert, Description and Examples of MAPP, 2010, German Development Institute. http://www.seachangecop.org/node/1558

Micro data

Definition:	Micro data are to be collected/obtained for individual units such as farms, food processors, forestry enterprises and agro-tourist farms/enterprises. For example, for farms a part of relevant micro data could come from the anonimised farm accountancy system (bookkeeping farms) or own surveys.

Multiplier (Input-Output Multiplier)

Definition:	Multipliers are the result of Input-Output model estimations. They are differentiated in Type 1 and 2 multipliers: For any sector, a high level of purchase of domestically (locally) produced inputs lead to strong linkages, and create significant indirect effects in the output of supplying sectors. These effects are measured by Type 1 multipliers, for each sector by the ratio between direct and indirect effects against the direct effects. Household spending occurring from labour income generated in the sectors of the economy under study creates further economic activity (induced effect), which is included in Type 2 multipliers: Here the denominator reads as “direct, indirect and induced effects” and the nominator as “direct effects”.
Comments:	Multipliers are obtained on output, income and employment. Multiplier analysis provides valuable information on economic structures as well

	as on the capacity of each IO sector to generate economy-wide effects.
Source:	Miller and Blair (2009)

Multiplier effect

Definition:	Secondary effect resulting from increased income and consumption generated by the public intervention. Multiplier effects are cumulative and take into account the fact that part of the income generated is spent again and generates other income, and so on in several successive cycles. In each cycle, the multiplier effect diminishes due to purchases outside the territory. The effect decreases much faster when the territory is small and when its economy is open.
Source:	CMEF

Net effect

Definition:	Effect imputable to the public intervention and to it alone, as opposed to apparent changes or gross effects. To evaluate net effects, based on gross effects, it is necessary to subtract the changes which would have occurred in the absence of the public intervention, and which are therefore not imputable to it since they are produced by confounding factors (counterfactual situation).

Non-take-up rate

Definition:	Number of eligible beneficiaries not being funded relative to all eligible potential beneficiaries[66].

Output

Definition:	Action which is financed and accomplished (or concretised) with the money allocated to an intervention. A project promoter undertakes to produce an output in immediate exchange for the support granted. Outputs may take the form of facilities or works (e.g. building of a road, farm investment; tourist accommodation). They may also take the form of immaterial services (e.g. training, consultancy, information).
Source:	CMEF

Output indicators

Definition:	These measure activities are directly realised within programmes. These activities are the first step towards realising the operational objectives of the intervention and are measured in physical or monetary units. Example: number of training sessions organised, number of farms receiving investment support, total volume of investment.
Source:	CMEF

[66] Blundell, R., V. Fry, I. Walker. (1988). Modelling the Take-up of Means-Tested Benefits: The Case of Housing Benefits in the United Kingdom. Economic. Journal. 98(390) 58.

Practicability

Definition:	Practicability is a term used in this study to judge the ease with which a method can be applied. Practicability relates to data availability, and any other resources necessary to use a given method for evaluation in a given context (e.g. availability of software packages, willingness to beneficiaries to participate in workshops). Measures of practicability are the amount of money and/or time spent to do an analysis.

Reliability

Definition:	Under the reliability of a given method applied for assessment of effectiveness, efficiency and impact of a given programme/measure one has to understand the degree of stability exhibited when a measurement is repeated under identical conditions. Lack of reliability may arise, for example from divergences between observers and is mostly linked to application of qualitative methodologies which heavily rely on subjective opinions of supported beneficiaries (common in surveys).
Comments:	In social sciences definition of reliability is as follows: *"reliability is the extent to which measurements are repeatable - when different persons perform the measurements, on different occasions, under different conditions, with supposedly alternative instruments which measure the same thing"*[67]. The CMEF defines reliability as: Quality of the collection of evaluation data when the protocol used makes it possible to produce similar information during repeated observations in identical conditions. Reliability depends on compliance with the rules of sampling and tools used for the collection and recording of quantitative and qualitative information.
Source:	Drost, 2011.

Result

Definition:	Advantage (or disadvantage) which direct beneficiaries obtain at the end of their participation in a public intervention or as soon as a public facility has been completed. Results can be observed when an operator completes an action and accounts for the way in which allocated funds were spent and managed.
Source:	CMEF, shortened

[67] Drost, E., 2011, Validity and Reliability in Social Science Research. Education Research and Perspectives, Vol. 38, No. 1, 105-123.

Result indicators

Definition:	These measure the direct and immediate effects of the intervention. They provide information on changes in, for example, the behaviour, capacity or performance of direct beneficiaries and are measured in physical or monetary terms. Example: gross number of jobs created, successful training outcomes.
Source:	CMEF

Rigour

Definition:	In the context of this study a "rigorous" method is understood as a method introduced, applied and further developed in referred scientific journals. A more rigorous method is more widely applied and accepted in the scientific community or community of evaluators and it will be described in methodological textbooks. Methods that can be used to quantify causal effects are considered to be more robust than other methods.
Comments:	The degree to which an evaluation design produces reliable and valid evidence is determined by whether the evaluation is 'rigorous'. Evaluation of the different types of investment support requires first of all a causal analysis, but it also comprises two other important tasks, i.e.: • Identification and prioritising of outcomes and effects that are valued, e.g. including intended and unintended, positive and negative, short-term and longer-term, economic, social, environmental ones for farms, communities and regions; • Gathering evidence of these outcomes and effects, e.g. retrieving existing data, collecting and creating new data, and addressing challenges in the adequacy and feasibility of measures and indicators, particularly for multi-dimensional and longer-term impacts.

Robustness

Definition:	The term "robustness" is defined in different ways in the evaluation literature. In econometrics robustness is used in various contexts (e.g. biased and unbiased estimators, model and variable selection) and has therefore context-specific meanings. In the context of this study, robustness is considered to be high if results are stable and resilient to small but deliberate changes (e.g. an additional year of observations, an additional explanatory variable, another stakeholder, another evaluator). In some methods, the robustness can be checked by sensitivity analyses.

Scale

Definition:	A scale is used to measure or gauge an indicator of interest. There are three types of scales: (1) cardinal scales: given in numbers (e.g. GDP = 200 bn EUR), (2) ordinal scales: ranking (e.g. GDP in 2008 was lower than in 2007); (3) nominal scales: verbal description (e.g. GDP is an aggregate measure of output of an economy in a given year).

Strategic Environmental Assessment (SEA)

Definition:	The Strategic Environmental Assessment (SEA) is a structured procedure[68] in which significant effects on the environment are defined *ex ante* and the reasonable alternatives of the proposed plan or programme are identified. The SEA can be used to frame the evaluation of RDP environmental impacts, hence it can play similar role as the Theory-Based Evaluation. A significant advantage of the SEA is that it has been a compulsory part of the *ex ante* evaluation of the programme and hence all RDPs are starting from the same point and have, in theory, respected the requirements of the Directive 2001/42/EC on environmental monitoring[69].

Targeting

Definition:	Procedure of diverting funds towards specific targeted groups, territories or objectives.
Comments:	*"For investment measures, Member States shall ensure that support is targeted on clearly defined objectives reflecting identified structural and territorial needs and structural disadvantages."*[70]

Targeting Approach

Definition:	Combination of eligibility criteria, aid intensity differentiation and selection criteria applied in a rural development measure to divert funds towards the target group.

[68] http://ec.europa.eu/environment/eia/sea-legalcontext.htm

[69] In the SEA procedure the public and the environmental authorities are informed and consulted on the draft Rural Development Programme and the related environmental report. The environmental report and the results of the SEA consultations with RDP stakeholders are taken into account before adoption of the programme. Once the RDP is adopted, the environmental authorities and general public receive relevant information on the potential environmental effects, which programme might cause in course of its implementation. In order to identify unforeseen adverse effects at an early stage, significant environmental effects of the plan or programme are to be monitored.

[70] Article 43 Reg. 1974/2006

Programme-theory-based evaluation

Definition:	TBE verifies the stepping stones of the intervention logic that lead from inputs to outcomes based on the "theory of change" which underpins the policy intervention. In a simplified format TBE includes three key working steps (i) Map out the conceptual model of investment support in order to capture the goals at different levels and the planned activities and target groups to achieve the desired change. The explicit statement of the "programme theory" is important as it provides the underlying logic for evaluation (ii) Verify the implementation of investment support and tell the "performance story" at a detailed activity level through empirical research which explores how the conceptual model has worked in practice (iii) Draw evidence-based conclusions if implementation and practice actually fit with expected goals and theory of change. Based on the collected evidence a judgement is made on the effectiveness to achieve strategic and operational goals mapped out in the beginning. TBE demands a multi-method evidence base. In the best case situation qualitative methods are required to generate process data, quantitative approaches to measure outputs and outcomes, and comparative observations are required for contextual information. The TBE poses particular challenges to the evaluator since it involves evaluating interventions according to logics that may not be present at the programme design stage. This implies the need to identify and verify the theory of change implicit in the underpinning of policy interventions even if such objectives are not made fully explicit.
Comments:	At present there is no uniform definition of the method. In the literature different approaches are described on how to implement TBE methodically (e.g. Realist evaluation, Theory of Change, Contribution analysis, Performance story). The simplified approach taken here is tailor-made for the evaluation of RDPs with a given intervention logic and takes into account a proportionate effort to carry out the case studies.
Source:	An overview of TBE approaches is given in the TBE section of the Evalsed Sourcebook (based on material produced by Frans L. Leeuw): http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf

Transparency

Definition:	Transparency of an evaluation methodology requires that users know exactly its main elements, structure, parameters, rules and functional responses. A user can therefore monitor that they are followed. A valid estimate of the counterfactual should be based on clear and transparent assignment rules.
Comments:	The criterion transparency is used in this study to gauge the quality of results. The more transparent an analysis, the easier it is for peers to reproduce results. In this study it is measured by identifying which elements of a case study are accessible for the readers.

Under coverage rate

Definition:	Number of eligible beneficiaries not being funded relative to all eligible potential beneficiaries[71].

Validity

Definition:	"Internal Validity": Results of non-empirical methods are valid if they are logically sound. Results of empirical methods are valid if they are logically sound and factually sound. Logical soundness can be verified and high transparency makes this easier. Factual soundness is verified if the result is identical to the true parameter which mostly cannot be observed (see counterfactual). "External validity" is a quality measure of empirical research. In our context "external validity" means whether the results obtained from a case study will be more or less the same if a similar programme is in place in another context as well. "Convergent validity" is given if different methods employed to answer the same research question yield similar results.
Comments:	This term is defined in various variants.

[71] Bibi, S., & Duclos, J.-Y. (2007). Equity and policy effectiveness with imperfect targeting. Journal of Development Economics, 83(1), 109–140.

KF-01-14-028-EN-N

Publications Office

doi: 10.2762/22502