advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Development and application of advanced quantitative methods to ex-post evaluation of EU RD and structural programmes

IV Evaluation Conference

Warsaw,

17 October 2008

Jerzy Michalek -
University of Kiel

Application of advanced evaluation methodology:

- **Micro- analysis of EU programme results**
- **Regional/macro- analysis of EU programme impacts**

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Answers to EC common evaluation questions

“Quantification is <should be> an essential element in effective monitoring and evaluation”

(EC, Evaluation Guidelines, September 2006, p.14)

Jerzy Michalek -
University of Kiel

Current approach to evaluation

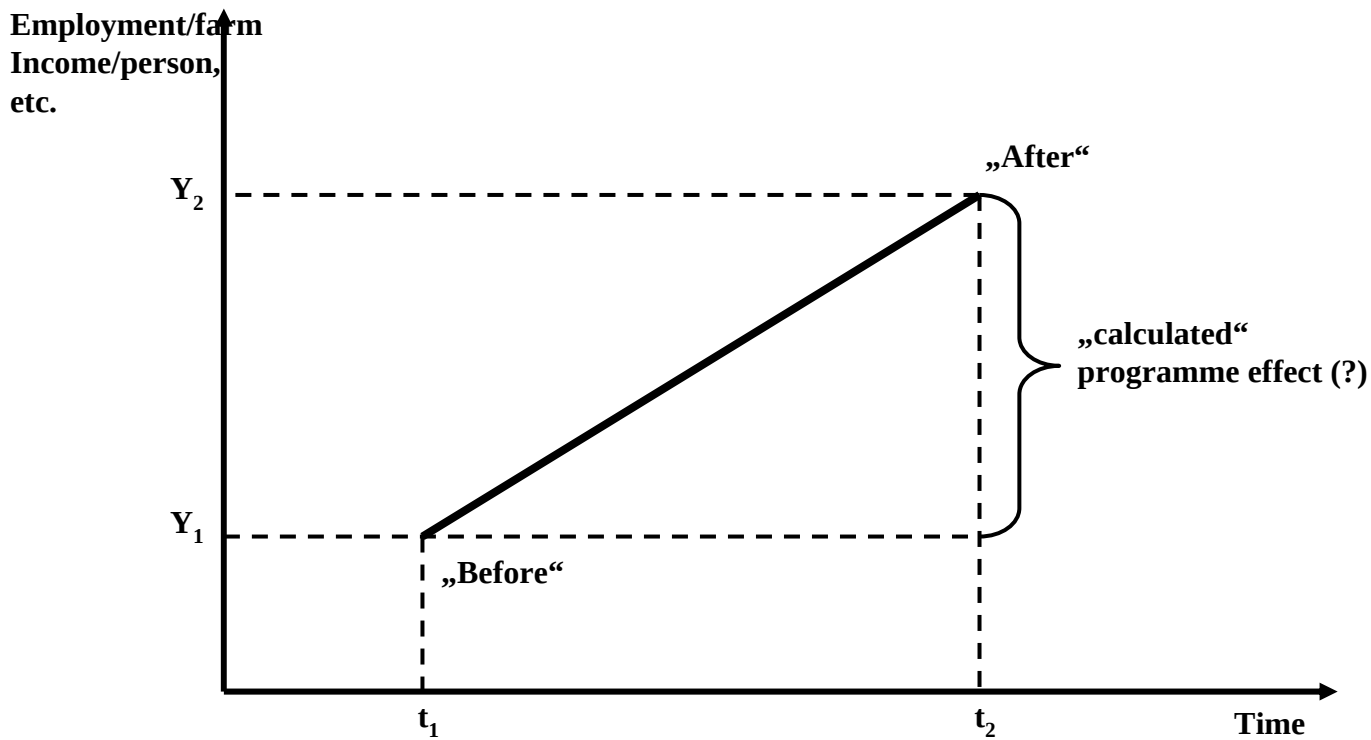
"Normal" evaluation praxis:

a. **Self-assessment of programme participants (interviews with programme beneficiaries)**
   - **It can be erroneous**
   - **Subjectivity**

b. **Assessment of available firm data (before=> after)**
   - **Can be extremely biased**

"Normal" praxis => Problems with distinction of effects

Naive evaluation of programme effects



Jerzy Michalek -
University of Kiel

What do we mean by the programme's effect?

=> The difference between the relevant outcome indicator *with* the programme and *without* it

=> Distinction of programme vs. non-programme effects

Jerzy Michalek -
University of Kiel

Methodological approach to evaluation

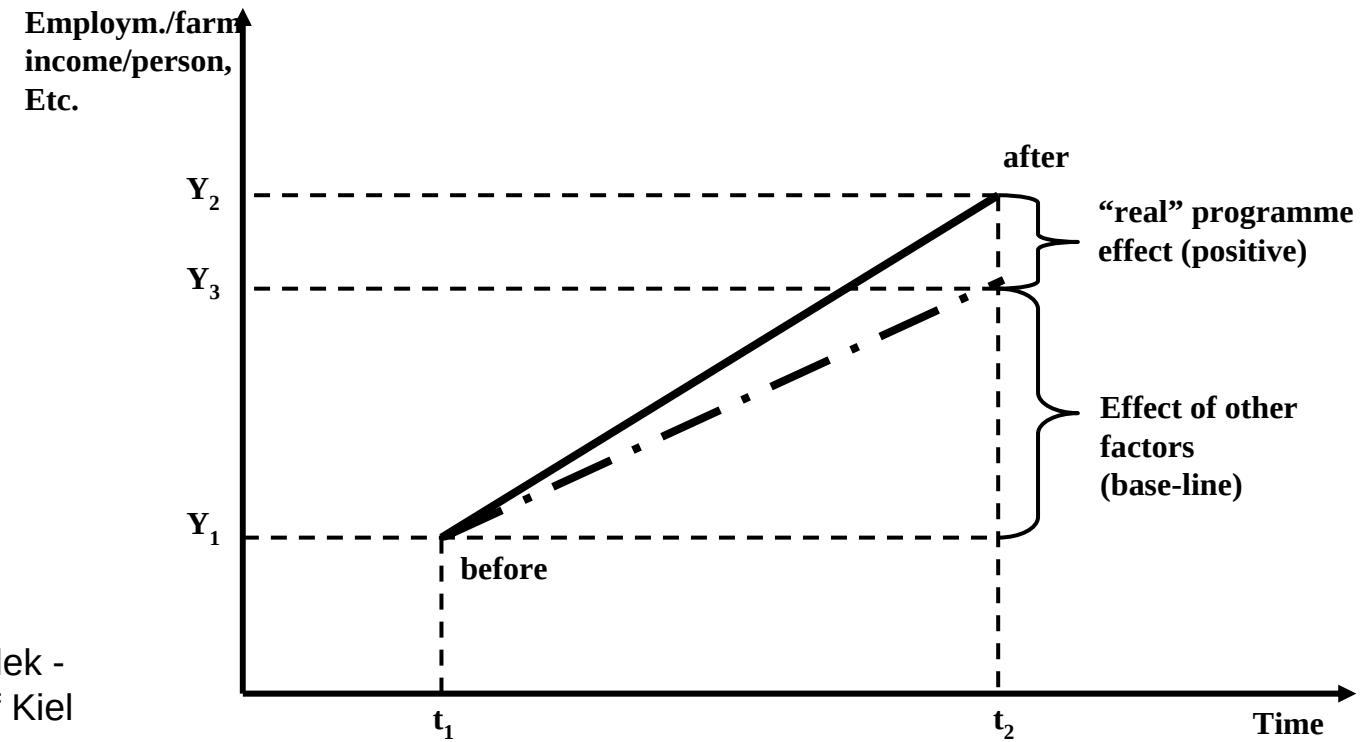
The central issue regarding the evaluation of an effect of a programme:

1. **What would have happened to programme beneficiaries had they not received intervention?**

=> base-line

Problems with distinction of effects

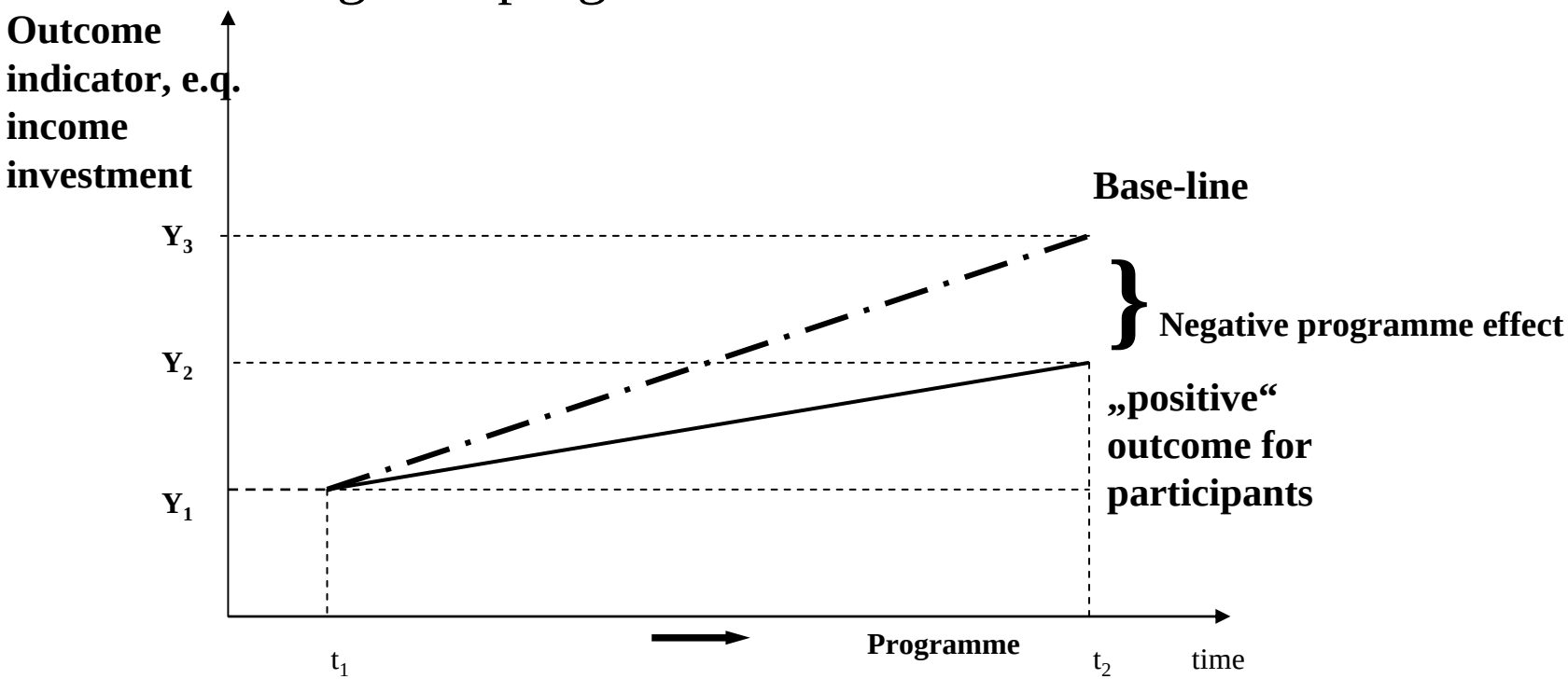
Significance of the relevant base-line ("no programme" scenario)



Jerzy Michalek - University of Kiel

Problems with distinction of effects

Negative programme effect



Jerzy Michalek -
University of Kiel

Methodological approach to evaluation

- **Problem => we can never observe a firm in two different states of nature in the same time**

Counterfactual analysis must be performed !

=> Similar control groups have to be used as proxy for the state of beneficiaries in the absence of intervention (absence of a RD programme)

Calculation of a true „base-line“

“Normal” evaluation praxis:

a. **“base-line” => Arbitrary** selected **controls**
   - all *non-participants, or*
   - *national average*

=> Both selections are usually biased (graph)

Why do we need more sophisticated methods?

1. **Base-line** => a problem of a perfect comparability (ideally, the same enterprises should be used for comparisons)

2. As this is not possible, one should ensure that all basic **characteristics** of compared enterprises are **almost identical** with those enterprises supported by RD programme

Why do we need more sophisticated methods?

Important assumption:

1. **Similarity** between supported and non-supported firms **can be statistically verified** on the basis of distributional characteristics of various **observable covariates**

Conventional matching

- Conventional matching is **impractical** due to a large number of variables/covariates to be compared

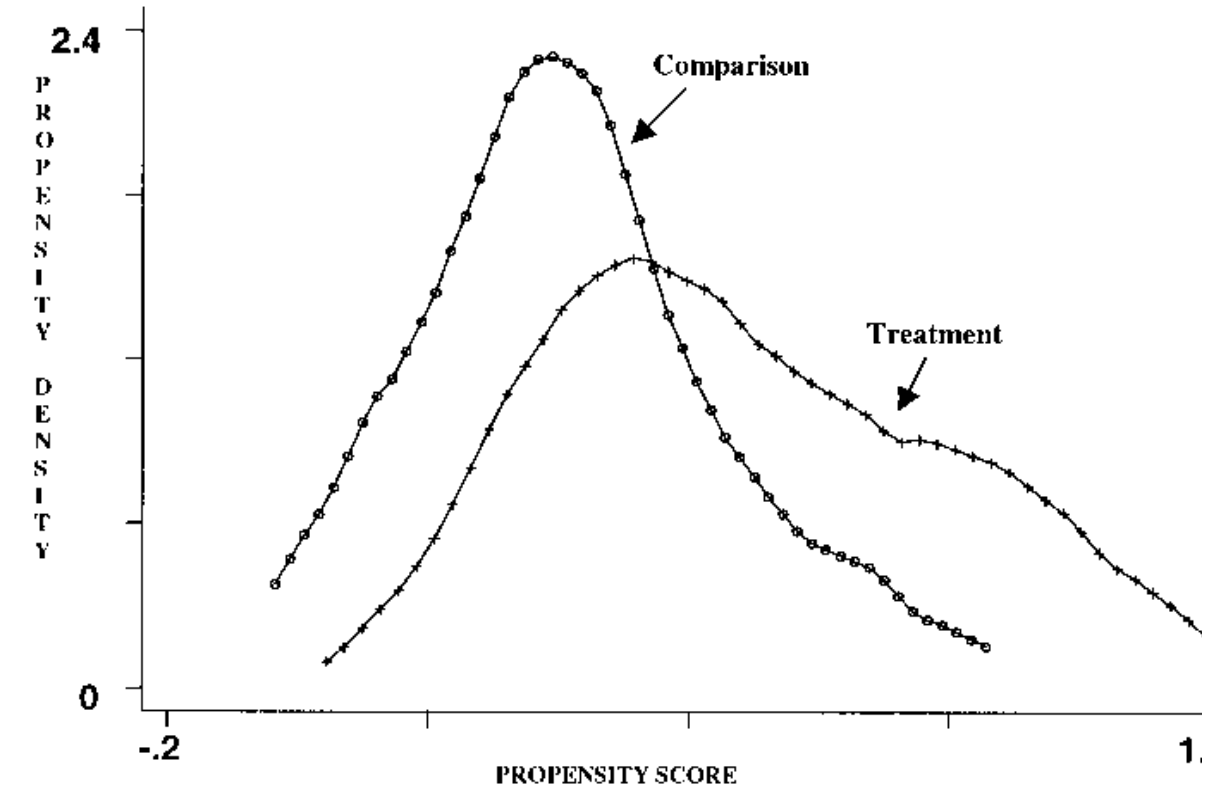
=> solution: **Propensity Score Matching** (Rosenbaum and Rubin, 1983; Heckman, 1998; 2002; 2005) => **one** dimension only!

Matching principle

- Matching => find for each programme **participant** unit a **comparable** programme **non-participant**

Propensity scores before matching

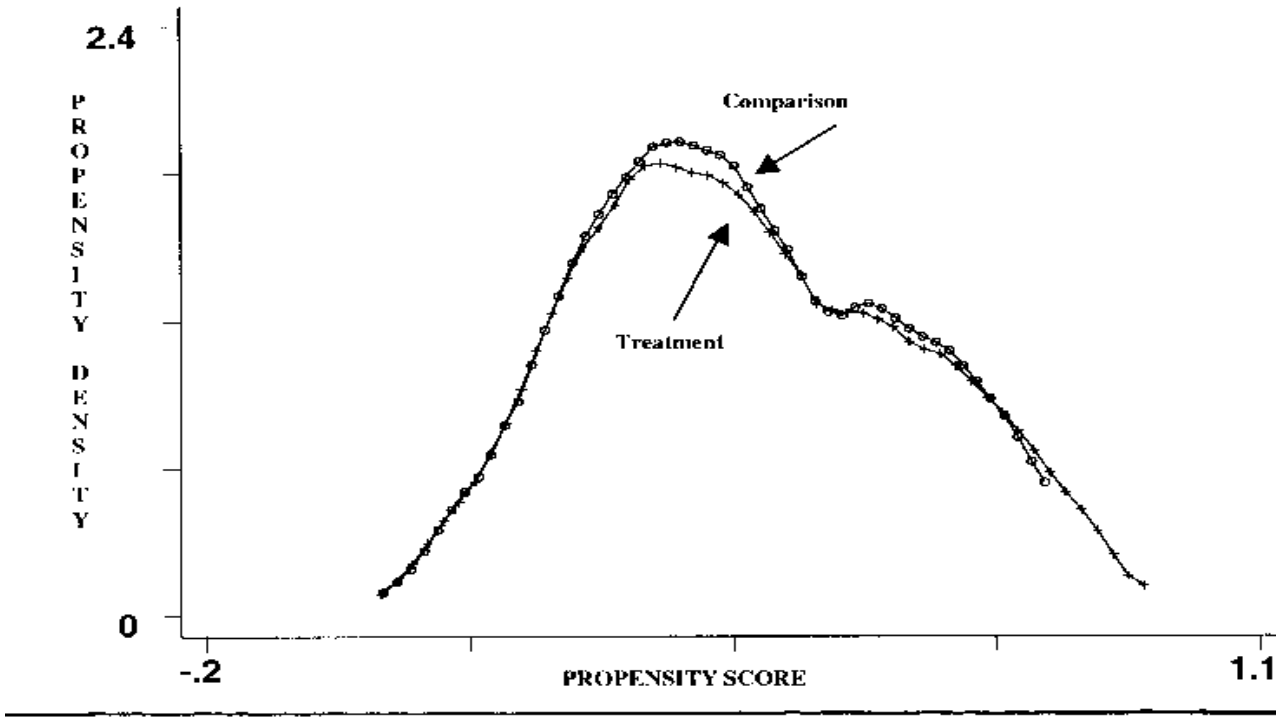
Source: Newman, et.al. 2002



Jerzy Michalek -
University of Kiel

Propensity scores after matching (e.g. nearest neighbour with trimming)

Source: Newman, et.al. 2002



Jerzy Michalek -
University of Kiel

Matching methods

- Nearest neighbour
- Caliper
- 1-to-1 Nearest neighbour within caliper
- 1-to-n Nearest neighbour within caliper
- Kernel Matching
- Local linear weighting function

The key parameters of policy interest:

- **Average Treatment Effect (ATE) (participants & non-participants)**
- **Average Treatment on the Treated (ATT) (beneficiaries)**
- **Average Treatment Effect on non-treated (ATNT) (non-participants)**
- **Marginal Treatment Effect (MTE) (on the margin of indifference between participation and non-participation)**

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Methodological approach to evaluation

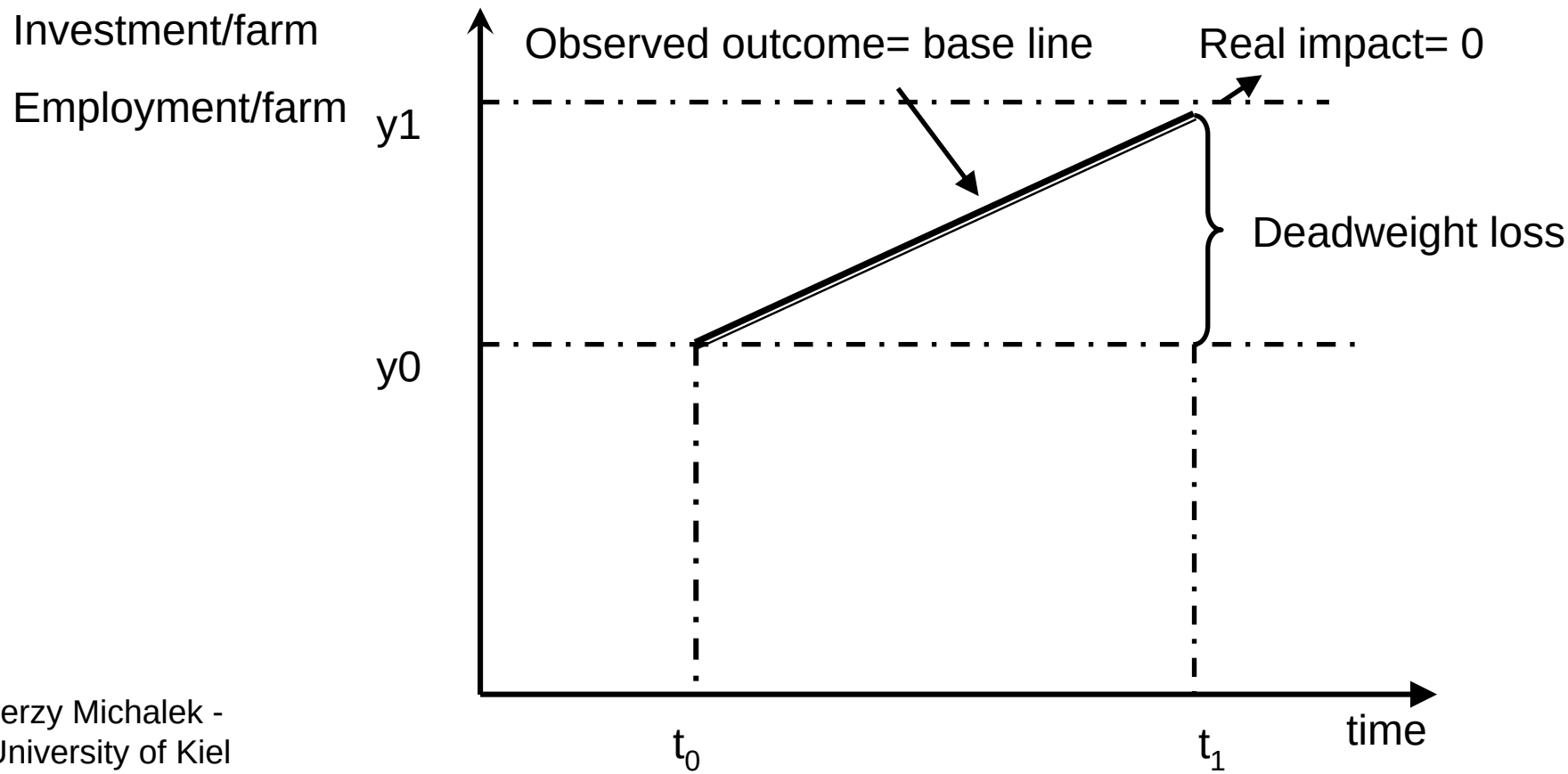
The central issues regarding quantitative evaluation of results and impact of EU programme (1):

1. **Appropriate calculation of a base-line (counterfactual)**

2. **Unbiased calculation of direct programme effects (combination of ATT and DID) including calculation of other programme effects (e.g. deadweight losses)**

Jerzy Michalek -
University of Kiel

Other effects => Deadweight Loss (micro-level)
(if supported activities would have been implemented anyway)



Jerzy Michalek -
University of Kiel

Methodological approach to evaluation

Other central issues regarding quantitative evaluation of EU programme (2):

3. **Unbiased calculation of selected general equilibrium effects (e.g. displacement/substitution effects)**

4. **Calculation of overall (net) programme effects**

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

		Quality of evaluation
Evaluation without a control group	Post-implementation observation	- -
	Observation before and after implementation	-
Evaluation by comparison with a control group	Non-equivalent control group	+
	Equivalent control group	+ + +

Source: adopted from WB (training materials). Based on empirical WB, OECD, etc. studies

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Application within this project

Ex-post evaluation of the SAPARD programme (Poland and Slovakia)

- **Selected measures:**
  - **Investments in agricultural holdings**

Jerzy Michalek -
University of Kiel

Empirical outcomes: comparison of traditional vs. advanced evaluation methodologies

- **Traditional (naive)** approach (Slovakia):

  "effectiveness of the implementation of measure 1 (Investments on farms) has been evaluated as **excellent (income/productivity, etc. of beneficiary farmers improved)**"… (SAPARD ex-post evaluation report, PCM and IBS, Slovakia, 2008) (!!!!!)

Empirical outcomes: comparison of traditional vs. Advanced evaluation methodologies

- **Traditional (naive) approach (Poland)**:

  "Podsumowujac wyniki przeprowadzonych badan ankietowych (…) **projekty SAPARD przyczynily sie** do wzrostu poziomu dochodow u ok. 80% rolnikow (sredni wzrost dochodow przekroczyl 50%)"…SAPARD spelnil swoje zadanie i znaczaco przyczynil sie do poprawy sytuacji dochodowej rolnikow ktorzy z niego skorzystali". (SAPARD ex-post evaluation report, Agrotec, IERIGZ, Warsaw, October, 2007) (!!!!!)

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Empirical outcomes: comparison of traditional vs. Advanced evaluation methodologies

- Advanced approach: (advanced-eval)

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Profit/company	2003	2005	DID (2005-003)
Participants (1)	-880	1589	+2496
Non-participants (0)	-3338	-839	+2499
Country Average Ø	-2798	-305	+2493
Difference (1-0)	2458	2428	-30
Difference (1- Ø)	1918	1894	-24
Matched participants (1)	-1705	131	+ 1836
Matched control group (0)	-1264	815	+ 2079
ATT	-440	-683	↘ -243

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Table: 2 Agriculture companies: Profit/ha.
- Difference in Differences (DID) of Average Treatment Effects on Treated (ATT)

Matching algorithms	ATT		DID (2003-2005)	Bias	% bias reduction
	2003	2005			
Nearest neighbours N (1)	0.119	-0.001	↘ -0.12	9.6	51%
Nearest neighbours N (5)	-0.314	-0.486	↘ -0.172	10.7	45%
Nearest neighbours N (10)	-0.361	-0.063	↗ 0.298	9.56	51%
Radius caliper (max distance 0.01)	0.558	-0.463	↘ -1.021	19.75	0%
Local linear regression matching (bandwidth 0.2)	-0.303	-0.392	↘ -0.089	9.67	50%
Kernel gaussian	-0.139	-0.333	↘ -0.194	9.01	54%
Kernel biweight	-0.009	-0.338	↘ -0.329	7.51	61%
Kernel epanechnikov bandwidth 0.06	0.0165	-0.345	↘ -0.3615	7.45	62%
Kernel epanechnikov bandwidth 0.1	-0.185	-0.316	↘ -0.131	9.06	54%
Kernel epanechnikov bandwidth 0.2	-0.125	-0.376	↘ -0.251	9.14	53%
Kernel epanechnikov bandwidth 0.01	0.579	-0.446	↘ -1.025	19.61	0%
Hodges-Lehmann	0.527	-0.379	↘ -0.906	.	.

Jerzy Michalek -
University of Kiel

Application of inadequate evaluation methodology

1. Traditional evaluation results (published in ex-post evaluation studies) appeared as **heavily biased**

2. Improvement in performance of **similar non-supported** (by SAPARD) companies was often stronger than those supported (=> programme effects were questionable..)

3. Inefficient allocation of **public and private** resources (better alternatives existed for both)

Evaluation ⇔ Monitoring system

- **The role of an appropriate monitoring system for programme evaluation (!!!)**

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

WP3 – Derivation of RDI / QLI

The main objective of WP3 is to:

- **Derive a Rural Development Index/Rural Quality of Live index to be used as an *impact indicator* in ex-post and on-going evaluations of EU RD policies and programmes**

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

WP3 – Derivation of RDI / QLI

Evaluation questions for all measures under Axis III, (EC Guidelines, 2006), the overall programme objectives, etc.:

"To what extent have (..the measure…programme...) contributed to improving the quality of life in rural areas?"

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

WP3 – Derivation of RDI / QLI

Multifold effects of RD/structural Programmes

1. **Affected areas:**
   a. Employment
   b. Production
   c. Income
   d. Investments
   e. Competitiveness
   f. Environment
   g. Technical and social infrastructure
   h. Skills and education
   i. Other living conditions

2. **Often multiple and adverse effects (positive⇔ negative)**

Jerzy Michalek -
University of Kiel

WP3 – Derivation of RDI / QLI

3. **Net effect** of EU RD/structural programmes is usually unknown because:

- Unknown the **extent** to which EU RD programme affects **individual indicators** (economic, environmental, social etc.)

- Unknown **individual weights** of above indicators to an **overall** impact (inclusive positive and negative impacts)

4. Current studies on rural development index:

- Equal weights for specific RD domains (economic, infrastructure, social, etc.)

- Weights derived from “expert” surveys

Jerzy Michalek -
University of Kiel

WP3 – Derivation of RDI / QLI

Current evaluation practice:

- **Up to 37 RD measures (42 measures 2007-2013)**
- Several quantitative indicators for each measure => hundreds of indicators to be looked at

=> **Overall** impact of RD programme is unclear (aggregation and weighting system)

Jerzy Michalek -
University of Kiel

WP3 – Derivation of RDI / QLI

Two research challenges:

1. **Derivation of a suitable synthetic RDI Index with objective and statistically verifiable weights**

2. **Measurement of the impact of EU programme in terms of the change of this index**

=> Problem: EU programme in almost all regions (but various intensity!)

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

WP3 – Derivation of RDI / QLI

Project approach:

- Deriving a **synthetic RDI** Index as an **outcome indicator** for RD programme impact

- Application of a generalized propensity score (**GPSM-ATT-DID)** with RDI as outcome indicator

- Complementary analysis using **GPSM**-**ATT-DID** on selected **other outcome indicators** at regional level, e.g. employment, income distribution, GDP, etc.

=> Analysis feasible at **single** and/or **aggregated** measures and **regional** levels (NUTS2-NUTS5)

Jerzy Michalek -
University of Kiel

Construction of RDI Index

Empirical results (Poland – NUTS 4)

⇒ **314 non-urban regions**

⇒ **992 regional variables/indicators (econ, social, environ, etc.)**

⇒ **17 factors + 2 transaction costs estimates**

⇒ **Derived RDI weights (panel estimate, logistic regression model with nested error structure, 393128 observations (years 2002 - 2005)**

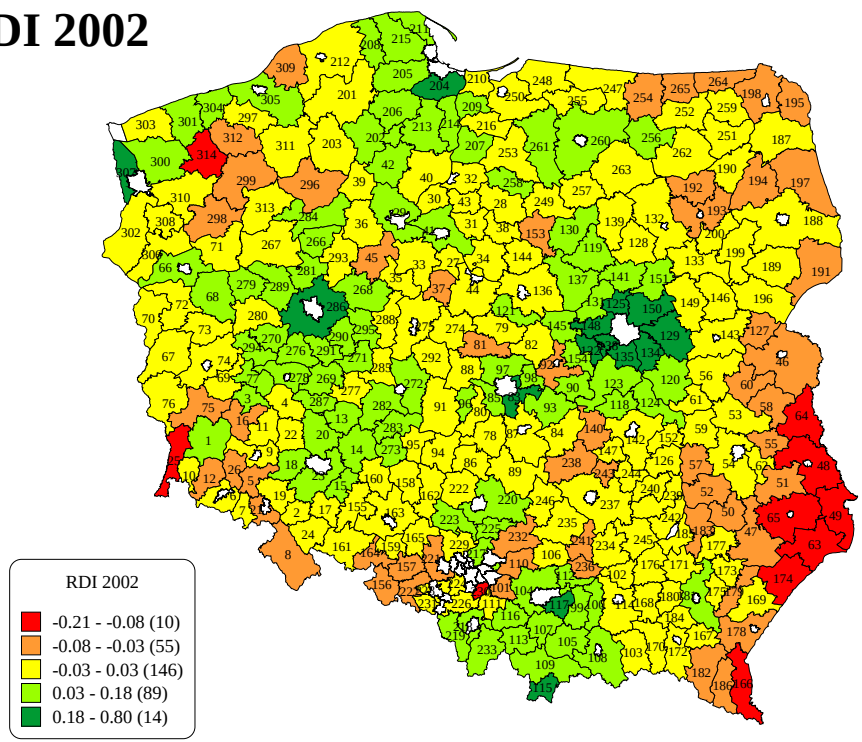
⇒ **RDI calculated for years 2002-2005 (NUTS-4 level)**

advanced-eval

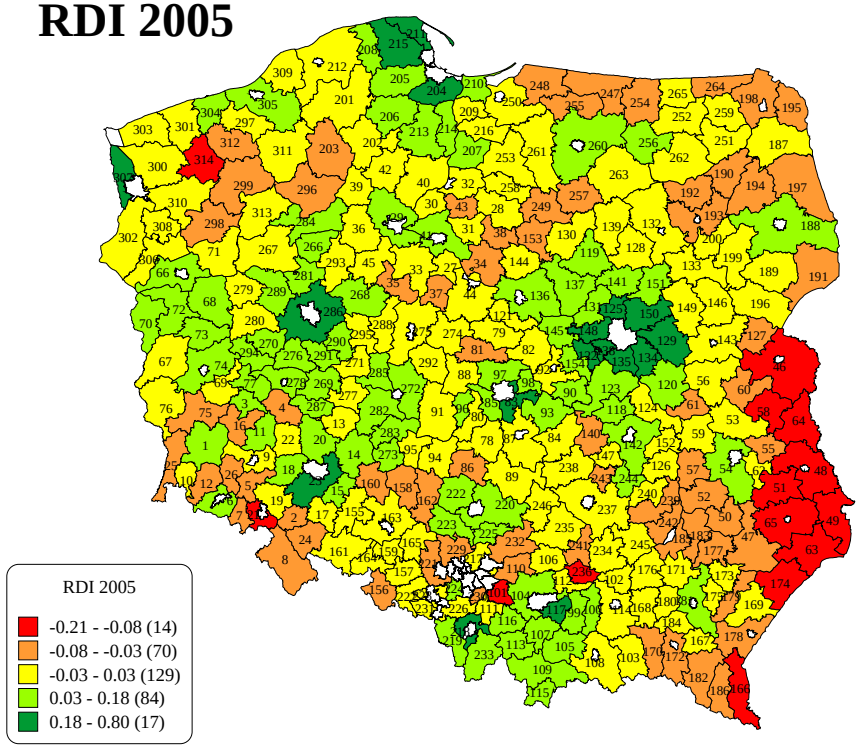
EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Poland: Rural Development Index (2002-2005) – NUTS 4

RDI 2002



RDI 2005



Jerzy Michalek -
University of Kiel

Main advantages of RDI (1/2)

- **Comprehensive consideration of various development domains (economic, social, environmental, etc.)**
- **Easily adjustable (easy inclusion of a new indicator/variable)**
- **Weights empirically derived and statistically verified**
- **Suitable to reflect programme general equilibria effects, e.g. multiplier effects, replacement effects, substitution effects**

Main advantages of RDI (2/2)

- **Powerful both at aggregated level (e.g. NUTS 2) and commune levels (NUTS 5) and even village level (if data exists)**
- **Applicable both for analysis of RD programmes as well as structural programmes**
- **Applicable for an analysis of effects of large projects at all regional levels**

Application of Propensity Score Methodology to evaluation of economic/social/structural policies

- **Standard PSM (=> implementation described under WP2; Rosenbaum and Rubin, 1983; Heckman et. al. 1997;)**

- **Generalized PSM (Imbens, 2000; Lechner, 2001; Hirano and Imbens, 2004; Kluve et. al. 2007)**

Application of the Generalized Propensity Score (GPS) Methodology to evaluation of EU programmes

Implementation:

=> Calculation of the marginal dose-response function as

$$E[Y(t+1)- Y(t)]$$ **in bootstrapping procedure**

Where:

Y = **potential outcome, e.g. RDI Index, unemployment, environmental indicator, etc.**

t = **level of treatment**

Jerzy Michalek -
University of Kiel

Empirical Application RDI & GPSM Methodology to evaluation of EU programmes (SAPARD) in Slovakia and Poland

Poland (314 non-urban regions NUTS 4):

a. **SAPARD (total financial allocation = 1.068 bill EUR in years 2000-2004 of which 680 Mill EUR from EU)**

b. **SAPARD measures (7)**

c. **Financial allocation by *measures* and regions (1-0) => sometimes only 2 or 3 regions with 0 allocations (!)**
   => lack of common support area

d. **Total financial allocation by regions => continuous variable**

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Poland (PSM): Positive impact of measure M3 (development of rural infrastructure) on the RDI Index
(Yet, "0-1" if less than 66% of the mean ! => arbitrary)

RDI	2002	2005	DID
Unmatched (1)	0.0065	0.0025	neg
Unmatched (0)	0.0498	0.0512	pos
Unmatched (1- 0)	**-0.0433**	**-0.0487**	**-0.0054 (negative)**
Matched (1)	0.0123	0.0088	neg
Matched (0)	0.0126	0.0078	neg
ATT (1-0) (PSM)	**-0.0003**	**0.0010**	**0.0013 (positive)**

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

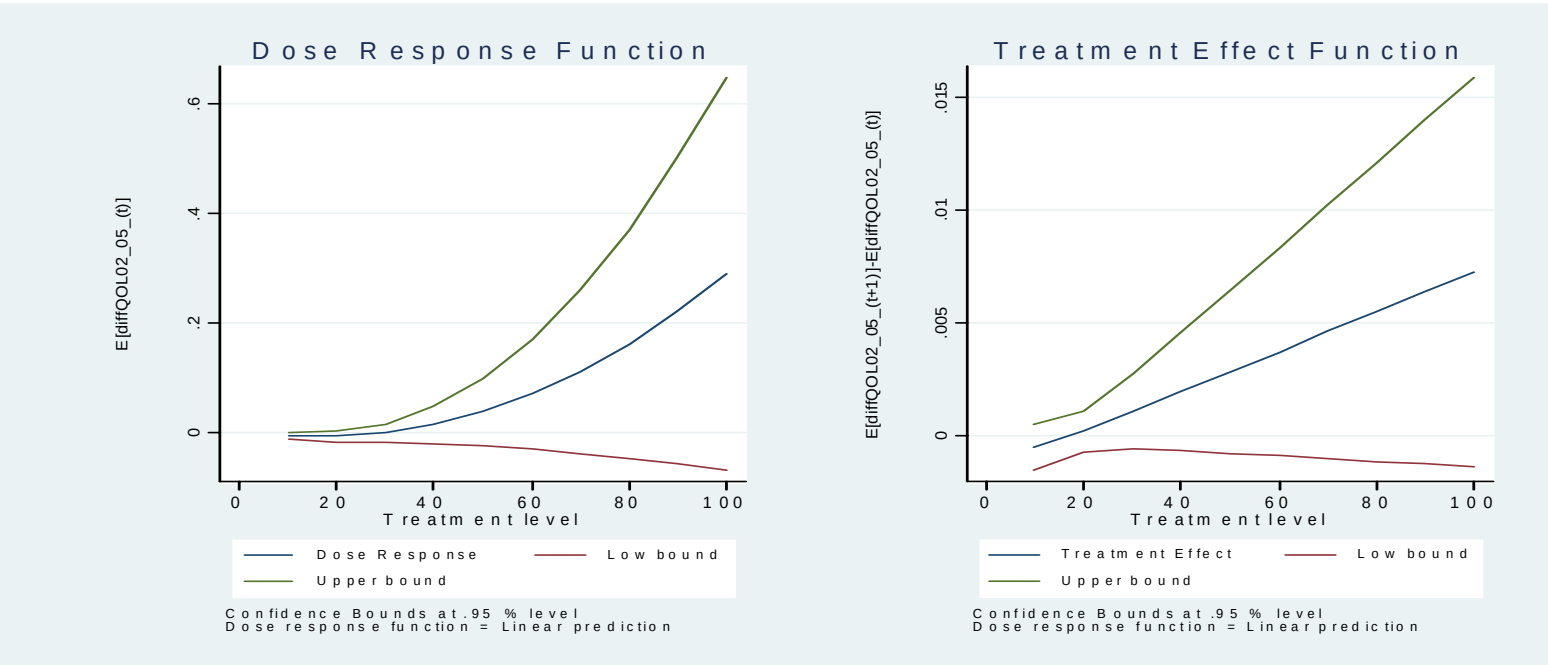
Poland (GPSM): Positive impact of measure M3 on RDI Index => no arbitrary thresholds!

Jerzy Michalek -
University of Kiel

advanced-eval

EVALUATION OF EU-RURAL DEVELOPMENT PROGRAMS

Application of GPSM Method - Poland



Jerzy Michalek -
University of Kiel

Conclusions

- **Naive (traditional) quantitative evaluation methods are inadequate for assessment of real effects of EU structural/RD programmes**
- **EU policies based on naive (traditional) programme evaluation methodology may be economically inefficient**
- **Advanced evaluation methods perform much better but require supplementary technical skills and abundant data**
- **Setting up of an appropriate monitoring system (including non-beneficiaries !) is crucial**