ESSAYS ON LABOUR MARKET POLICY EVALUATION

Ville Vehkasalo

ESSAYS ON LABOUR MARKET POLICY EVALUATION

Ville Vehkasalo

University of Turku

Turku School of Economics
Department of Economics
Economics
Doctoral Programme of Turku School of Economics

Supervised by

Professor Heikki Kauppi
University of Turku

Professor Janne Tukiainen
University of Turku

Reviewed by

Ph.D. Kari Hämäläinen
VATT Institute for Economic Research

D.Sc. Jani-Petri Laamanen
Tampere University

Opponent

D.Sc. Jani-Petri Laamanen
Tampere University

Custos

Professor Heikki Kauppi
University of Turku

UNIVERSITY OF TURKU
Turku School of Economics
Department of Economics
Economics
VILLE VEHKASALO: Essays on Labour Market Policy Evaluation
Doctoral Dissertation, 108 pp.
Doctoral Programme of Turku School of Economics
January 2022

ABSTRACT

This thesis presents three previously published case studies and a new summary article, with the overall aim of improving our understanding of the use of natural experiments in studying causal relationships in social sciences. The overarching theme of our case studies is distinct: we have evaluated public policies aiming to improve citizens' employment, productivity, and re-employment possibilities. We have scrutinised the effects of EU regional policy measures in 2007–2013, assessed the 2011–2014 vocational schools' dropout prevention programme, and evaluated the 2013 public employment service reform. Overall, our findings are not that surprising from the point of view of policy implementation. When we increase public spending allocated to certain eligible areas, desired results can be achieved, if all goes well (article I). On the other hand, when we decrease government resources, this might lead to unexpected disadvantages in the affected areas (article III). Poorly implemented government programmes may not benefit the recipients at all (article II) and represent wasted resources. In the summary article, we take a closer look at the internal and external validity of the case studies. We also suggest some topics for further research on these particular policy issues.

KEYWORDS: Treatment effect, regional policy, unemployment, vocational education, microeconometrics

TURUN YLIOPISTO
Turun kauppakorkeakoulu
Taloustieteen laitos
Taloustiede
VILLE VEHKASALO: Esseitä työmarkkinapolitiikan arvioinnista
Väitöskirja, 108 s.
Turun kauppakorkeakoulun tohtoriohjelma
Tammikuu 2022

TIIVISTELMÄ

Väitöskirja koostuu kolmesta aiemmin julkaistusta alkuperäisartikkelista ja niiden pohjalta kirjoitetusta yhteenvetoartikkelista. Tutkimuksen tavoitteena on ollut edistää ja kehittää luonnollisten koeasetelmien käyttöä kausaalisuhteiden arviointiin yhteiskuntatieteissä. Tarkasteltujen politiikkainterventioiden tavoitteena on ollut parantaa alueellista työllisyyttä (artikkeli I), opiskelijoiden valmistumista ja tätä kautta tuottavuutta (artikkeli II) sekä työttömien uudelleentyöllistymistä (artikkeli III). Yleisesti ottaen kyseisillä interventioilla on siis pyritty edistämään kansalaisten hyvinvointia. Arviointien tulokset eivät sinänsä ole yllättäviä. Julkisten resurssien voimakas lisäys interventioalueilla voi johtaa tavoiteltuihin tuloksiin työllisyydessä, ja ainakin osa aluepolitiikan tavoitteista näyttäisi toteutuneen. Toisaalta taas julkisten palveluresurssien vähentäminen voi johtaa haitallisiin vaikutuksiin, kuten työttömyyden pidentymiseen. Huonosti toteutetut ohjelmat eivät välttämättä hyödytä ketään, ja ohjelmiin uhratut resurssit ovat lähinnä yhteisten varojen tuhlausta. Yhteenvetoartikkelissa arvioidaan tarkemmin tapaustutkimusten sisäistä ja ulkoista validiteettia sekä ehdotetaan mahdollisia jatkotutkimusaiheita kyseisten politiikkalohkojen osalta.

ASIASANAT: Kausaliteetti, aluepolitiikka, työttömyys, ammatillinen koulutus, mikroekonometria.

Preface

After a nearly 20-year hiatus from academic research, I got a second wind almost accidentally. For the inspiration, I must thank our former Head of Performance Audit Unit, Mr Marko Männikkö, for assigning me a pressing task in early 2015. The task was to produce, in a week's time, a preliminary idea for a performance audit topic concerning the Europe 2020 strategy; the topic was urgently needed for an international audit collaboration network.

Grasping at straws, I searched the Internet for various documents on EU regional policy and found these two maps of the European Regional Development Fund's support areas in Finland: one before the year 2007, and another after the year 2007. As I was comparing the maps, you could have seen a light bulb turn on above my head (figuratively speaking of course – our office had fluorescent lights). Mr Männikkö had no clue then, or later, but his immediate assignment in 2015 was really the start of my first published article and finally this dissertation.

As for the research work itself, my deepest gratitude goes to D.Sc. Tanja Kirjavainen for her continuous encouragement, sound econometric advice, and fruitful colleagueship at the National Audit Office, and for suggesting the dropout prevention programme for my next performance audit topic. My ordinary e-mail discussions with our late colleague Mr Olli-Pekka Luoto were the starting point of yet another audit topic, leading to my third research article, on the Public Employment Service reform of 2013. Mr Luoto complained that his home municipality, Lieto, lost their employment office in the reform, which got me thinking.

Comment-wise, I am obliged to D.Sc. Antti Moisio for reading the draft of the audit report preceding article I and for prof. Mika Kortelainen (UTU) for commenting the draft of the audit report preceding article III. Compliments are also due to my thesis's pre-examiners D.Sc. Jani-Petri Laamanen and Ph.D. Kari Hämäläinen for their valuable comments and observations. Numerous anonymous referees of the original articles had worthwhile suggestions for improvements, too. I also thank Ms Päivi Ilves for proofreading the articles II and III before their publication.

Table of Contents

Tables

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I   Vehkasalo, V. (2018): Effectiveness of EU Regional Policy: Evidence from a Natural Experiment in Finland. *Region*, vol. 5, no. 3, 1–19.

II Vehkasalo, V. (2020): Dropout prevention in vocational education: Evidence from Finnish register data. *Nordic Journal of Vocational Education and Training*, vol. 10, no. 2, 81–105.

III Vehkasalo, V. (2020): Effects of face-to-face counselling on unemployment rate and duration: evidence from a Public Employment Service reform. *Journal for Labour Market Research*, vol. 54, no. 11, 1–14.

The original publications have been reproduced with the permission of the copyright holders.

1   Introduction

Establishing causality is one of the most challenging tasks in social sciences. In order to make informed policy decisions, it would be highly valuable to know, for instance, the causal effect of student loans on graduation rates and future incomes. But unlike e.g. in medicine, where randomised controlled trials (RCTs) are the norm, in social sciences we seldom have the opportunity or the means to conduct randomised experiments. Sometimes this would even be unethical; think of randomisation of last-resort social assistance. Almost invariably, the treated units (persons, firms, schools, municipalities, etc.) self-select themselves into treatment – the word "treatment" interpreted here very broadly – and later, the evaluator faces an observational data set plagued with an indeterminate amount of self-selection bias. For example, only students with skills and prospects above average might apply for student loans in the first place.

Underlying this bias is naturally the existence of omitted variables: if we could observe every confounding characteristic of the treated and non-treated units, estimating the causal effect of the treatment – for example student loans – would be quite straightforward. However, this kind of omniscience is not reality in the present or in the foreseeable future. And as some of these characteristics always will be non-quantifiable, or subject to considerable measurement error, we might never see the day when none of the relevant variables are omitted and the problem of self-selection bias is resolved.

Randomisation effectively removes self-selection bias, as units cannot choose whether they are treated or not. If the sample is large enough, the distributions of the other characteristics will be approximately equal between the treatment arms (treated/non-treated), except for insignificant random variation. Comparison of the outcomes after the treatment will then yield credible causal estimates of the treatment effect.

When controlled experimentation is impossible or impractical, as a next-best alternative we should then look for changes in policy circumstances that accidentally produce random-like variation in treatment exposure. Such cases can be labelled as *natural experiments*, although there is no universally accepted definition for this expression (Craig et al. 2012). These study designs have a long history in

epidemiology, beginning with Snow's (1855) famous study of the 1854 cholera outbreak in London.

Contrary to mid-nineteenth century belief that cholera was spread by "foul air", Snow demonstrated that polluted drinking water was the culprit. He collected data on cholera cases from various districts of London which were exposed to exogenous changes in water supply. Snow commented on his research design: "The experiment, too, was on the grandest scale. No fewer than three hundred thousand people of both sexes, of every age and occupation, and of every rank and station, from gentlefolks down to the very poor, were divided into two groups without their choice, and, in most cases, without their knowledge; one group being supplied with water containing the sewage of London, and, amongst it, whatever might have come from the cholera patients, the other group having water quite free from such impurity."[1]

Thus, the key difference with respect to controlled experiments is that the researcher is unable to manipulate who is treated and who is not treated. Treatment exposure is determined for example by changes in legislation, or a natural disaster, or an economic crisis. A typical example entails exogenous changes in some regions while others remain in status quo, such as in the seminal case study by Card and Krueger (1994) on the effect of minimum wage increase on employment in fast food restaurants. Using individual-level register data, it is possible to evaluate quite complex and intimate questions, such as the effect of exogenous job losses on family size (Huttunen and Kellokumpu 2016). The situation's potential as a natural experiment may not even be understood immediately, and the subjects might not realise their current partaking in an experiment, as in the cholera outbreak.

This feature of natural experiments could even be an advantage over controlled experiments. The now infamous Hawthorne lighting experiments in the Hawthorne electrical plant in Illinois during 1924–1927 claimed a significant "observer effect" was detected in the experiments. Productivity increased even when the lighting was dimmed. However, this finding was later proven false (Levitt and List 2009). Nevertheless, Levitt and List found more subtle long-run productivity effects in the original Hawthorne data. Therefore, the study subjects' knowledge of the treatment may well have a bearing on the results. RCT experiments in the field of medicine routinely employ a double-blind approach, where neither the patient nor the attending physician knows whether the patient is exposed to the treatment or a placebo medicine.

---

[1]   Snow (1855, 75).

In this summary article, we first outline the theoretical framework of the study and its objectives in Section 2. This is followed by short presentations of the research designs and the major findings of our case studies in Section 3. We then discuss the experimental validity of each case study in Section 4, and finally in Section 5 offer some concluding remarks.

2 Framework of the Study

2.1 Theoretical framework[2]

Suppose that we are interested in evaluating the effects of a binary treatment w. If $w = 1$, the observed unit (individual, firm, school, state etc.) is exposed to the treatment, and if $w = 0$, the unit is not exposed to the treatment. Outcomes with and without the treatment are denoted y_1 and y_0, respectively. The main problem of the evaluation is that we are unable to observe both y_0 and y_1; a single unit cannot be in both states simultaneously. Hence, we essentially encounter a missing data problem.

Using a random sample from the population, we aim to measure the effect of the treatment on the outcome: $y_1 - y_0$. However, as this is a unit-specific random variable, we actually need an estimate of the expected effect, or the average treatment effect (ATE):

$$(1) \qquad\qquad ATE = \mathrm{E}(y_1 - y_0).$$

ATE measures the average effect of the treatment on a randomly sampled unit from the population. A closely related measure is the average treatment effect on the treated (ATT):

$$(2) \qquad\qquad ATT = \mathrm{E}(y_1 - y_0 \mid w = 1),$$

which measures the expected effect on the treated ($w = 1$). In certain special circumstances, ATE equals ATT, but this is quite unusual in practice.

As mentioned above, the basic problem is that we cannot observe both y_0 and y_1 for the same unit. Therefore, the observed outcome, y, is a combination of non-treated and treated states:

$$(3) \qquad\qquad y = (1 - w)y_0 + wy_1 = y_0 + w(y_1 - y_0).$$

If the treatment is randomly assigned, then w and the potential outcomes (y_0, y_1) are independent. Then we can assert that ATT and ATE are identical, since:

$$(4) \qquad\qquad \mathrm{E}(y_1 - y_0 \mid w = 1) = \mathrm{E}(y_1 - y_0)$$

---

[2]   This subsection is heavily based on Wooldridge (2010, Ch. 21).

because the treatment and the potential outcomes are uncorrelated. In this case, estimation of *ATE* is straightforward, as

(5) $$\mathrm{E}(y \mid w = 1) = \mathrm{E}(y_1 \mid w = 1) = \mathrm{E}(y_1).$$

Similarly,

(6) $$\mathrm{E}(y \mid w = 0) = \mathrm{E}(y_0 \mid w = 0) = \mathrm{E}(y_0).$$

Both *ATE* and *ATT* are therefore estimated by a difference in sample means:

(7) $$ATE = ATT = \mathrm{E}(y_1) - \mathrm{E}(y_0).$$

As mentioned in the Introduction, in the social sciences, randomised treatments are usually not possible, and consequently, simple comparisons of sample means are severely biased. We therefore have to utilise more involved methods and accept the fact that our estimates contain an unmeasurable degree of bias. Nevertheless, even a biased estimate is better in policy analysis than no estimate at all, provided that the sign of the estimate is correct.

In a case when units self-select themselves into treatment – for instance, by voluntarily participating in a labour market training program – standard multiple regression or matching methods for cross-sectional data are valid, **if** we can ascertain that the *unconfoundedness* (or *ignorability*) assumption holds. That is, we can observe all the relevant confounders, or obtain reasonable proxies for the unmeasurable factors. If there are no omitted variables, treatment is as good as random after controlling for everything else, including person's motivation, social skills, appearance, innate ability, etc.

In practice, such cases are extremely rare, and the unconfoundedness assumption usually fails. If we can assume that the most relevant unobservable factors are time-invariant (e.g. student's innate ability; school location; person's labour market history before the period of observation), we may minimise self-selection bias with various panel data estimators, provided that we have access to panel data. But there might still be time-dependent unobservable variables which are of key importance in determining treatment exposure.

2.2   Study objectives

This dissertation is based on three separate case studies (articles I–III, presented in detail in Section 3), and the original motivation for these studies was to find out the actual causal effects of the treatments in question. The case studies were initially

carried out as a routine part of my performance audit[3] work in the National Audit Office of Finland, and the final audit reports (in Finnish) are available from the Internet website of the National Audit Office.[4]

Turning the audit reports into published research articles I–III has been a separate task, and not in any way officially connected to the work of the National Audit Office. My objectives in getting this work first published and then the results compiled in this summary article have been threefold:

- to deepen the understanding of how to use natural experiments in order to establish causal relationships, especially in social science

- to compare different research designs and possible sources of treatment exogeneity

- to assess the internal and external validity of study designs based on natural experiments.

Note that the original audit reports and the published articles I–III do not contain completely identical results. The research articles include more refined work and slightly different findings, partly based on their reviewers' comments and suggestions. Essentially, they still tell the same story.

2.3   Methodology

This summary article presents case studies where the units did not self-select themselves into treatment. Instead, exposure was determined by exogenous circumstances. There are several statistical methods for analysing exogeneous treatment exposure, including, but not limited to:

- difference-in-differences regression

- instrumental variables regression

- regression discontinuity designs.

As there are plenty of well-written textbooks detailing the steps of each method, we only outline the fundamentals of these approaches here; for the interested reader, an excellent reference is Wooldridge (2010).

Difference-in-differences methods were extensively used in our case studies. In the minimum, difference-in-differences methods require outcome observations of

---

[3]   "Performance auditing" is usually defined as an independent examination of the economy, efficiency, and effectiveness of government activities. Some dissidents would use a shorter term "evaluation".

[4]   http://www.vtv.fi. Reports no. 21/2016, 13/2018, 4/2020.

two distinct groups (treatment and control) and two time periods, before and after the treatment. We then compare the changes over time between the two groups. When longitudinal data is available – the same cross-sectional units are followed in time – we simply subtract the observation of the earlier time period from the observation of the latter time period, i.e. take the first difference. Regressing the difference on the treatment dummy yields the difference-in-differences estimate. In the case of pooled cross-sections, the "before" and "after" random samples include different units. In this case, obtaining the difference-in-differences estimate requires three variables, an indicator for the latter time period, an indicator for the treatment group, and the interaction of these two variables. The coefficient on the interaction term measures the treatment effect. The necessary condition for identification in difference-in-differences models is that of parallel outcome trends before the treatment. This can be tested using pre-treatment outcome observations.

Instrumental variables (IV) regression extends the standard cross-section analysis by introducing an auxiliary variable (an *instrument*), which causes exogeneous variation in treatment exposure. A valid instrument is highly correlated with the treatment, but uncorrelated with the outcome or the unobservable factors. For instance, if the treatment in question is administered in a certain geographic location only, the eligible unit's distance from this location could be used as an instrument. Distance might induce exogenous variation in treatment exposure after other factors are accounted for; the existence of this variation should be tested before proceeding. First-stage test permitting, we can then obtain the causal estimate with a two-stage IV regression procedure. The main challenge of IV methodology is finding credible instruments.

Regression discontinuity designs utilise an existing discontinuity in treatment eligibility. For example, those eligible for a labour market training program must have pre-program monthly income below some threshold level. We may then reasonably assume that units just below the threshold and just above the threshold have similar distributions of confounding variables, which is a testable assumption. If there are no statistically significant discrepancies, treatment is as good as random. However, in practical applications, sample size could be a problem; unless we obtain a very large sample, there might be only a few observations "just below" and "just above" the threshold level. Several statistical techniques have been developed to tackle this issue. The generalisation of estimated local treatment effects far beyond the threshold level could also prove problematic.

3   Case Studies

This summary article compiles the results from three previously published articles written by the author:

I   Vehkasalo, V. (2018): Effectiveness of EU Regional Policy: Evidence from a Natural Experiment in Finland. *Region*, vol. 5, no. 3, 1–19.

II   Vehkasalo, V. (2020): Dropout prevention in vocational education: Evidence from Finnish register data. *Nordic Journal of Vocational Education and Training*, vol. 10, no. 2, 81–105.

III   Vehkasalo, V. (2020): Effects of face-to-face counselling on unemployment rate and duration: evidence from a Public Employment Service reform. *Journal for Labour Market Research*, vol. 54, no. 11, 1–14.

Besides summarising the case studies' research designs and major findings, we also discuss the possible threats to internal and external validity in each case study and suggest topics for further research. Our cases have variation in the sense that the study questions are concerned with different fields: regional policy, educational policy, and employment policy. We have also used different units of observation and causes for random-like treatment exposure. Nevertheless, we can argue that the overarching theme of our case studies is distinct: we have evaluated public policies aiming to improve citizens' employment, productivity (through graduation), and re-employment possibilities.

Drawing from our practical experiences, we aim to contribute to the natural experiment literature in general, and also aid researchers who are contemplating similar study designs. Table 1 summarises the research designs of our case studies I–III.

Table 1.   Research designs of the case studies.

Article	Problem	Data	Unit of observation	Methods	Source of treatment exposure	Outcome variables
I	Has the EU regional policy been effective?	Panel of Finnish postal code areas, 2004–2013	Postal code area	Difference-in-differences regression, spatial regression	Change in supranational legislation	No. of jobs, unemployment rate, share of tertiary educated, income
II	Did the dropout prevention programme decrease dropouts?	Cross sections of three vocational student cohorts, 2002, 2007, 2012	Student	Difference-in-differences regression for pooled cross sections	Schools selected by the National Board of Education, based on school applications	Study completion rate, dropout rate
III	Is online counselling as effective as face-to-face employment counselling?	Panel of Finnish municipalities, 2006–2017	Municipality	Difference-in-differences regression, fixed effect panel data regression	Employment office reorganisation	Unemployment rate, unemployment duration

The following Sections 3.1–3.3 elaborate the respective research designs and summarise the major results of the articles. Throughout this summary article, the original research articles are referenced with the Roman numerals I–III.

3.1   Effectiveness of EU Regional Policy

The Regional Policy of the European Union aims to decrease income disparities between the developed EU areas – roughly the countries centred around Germany and France – and the less developed fringe areas. Regional Policy is implemented through structural funds, which account for a third of the annual EU budget. In the programming period 2007–2013, total outlays of the structural funds totalled approximately 50 billion euros per year.

The first structural fund, the European Social Fund (ESF), was established over 60 years ago, in 1957. The European Regional Development Fund (ERDF) followed in 1975, and the Cohesion Fund (CF) in 1994. In addition to these general-type funds, there are also more specialised funds, the European Agricultural Fund for Rural Development (EAFRD), and the European Maritime and Fisheries Fund (EMFF).

Despite the decades-long history of the structural funds, we still lack unambiguous evidence of their effectiveness. Published studies on EU regional policy have found both positive and negative effects (and zero effects) on the eligible

regions' economic development and growth.[5] The inherent difficulty in studying the effects of regional policies empirically is that the regional policy outlays are naturally allocated to poorer EU regions. Hence, we have a situation of reverse causality, and estimating the causal effect on growth and employment is difficult or impossible.

Organising a randomised controlled trial would entail selecting a random sample of EU regions that receive structural fund support and following the effects on these regions' economic development over time. However, such a research agenda would be highly impractical on both ethical and political grounds. In the funding lottery, some of the richest regions of the EU could receive substantial support payments, which would not go unnoticed in the media (and among the EU opponents), even if the experiment was only temporary. Clearly, this is not a viable study option.

As a second-best alternative, we should look for circumstances that created random-like changes in the EU regional support allocation. One candidate for such a change occurred in Finland in 2007. For legislative reasons, parts of regions previously covered by the ERDF Programme for Western Finland were reallocated to the ERDF Programme for Northern Finland. As support intensity (euros per capita) is multifold in Northern Finland, this change created a natural experiment where some sub-regions received a windfall of support.

Behind this reallocation was the new EU legislation which was adopted four years earlier; the legal framework for the NUTS classification was established with Council Regulation (EC) No 1059/2003. The classification is laid down in Annex 1 of the Regulation. For the programming period 2007–2013, previous Regional Policy Objectives 1, 2, and 3[6] were replaced with the Objectives Regional Competitiveness and Employment and European Territorial Cooperation. Council Regulation (EC) No 1083/2006 lays down general provisions on the Structural Funds for the period 2007–2013. Recital 16 of the Preamble explains that the identification of eligible areas should be based on the NUTS classification established by Regulation (EC) No 1059/2003. Furthermore, Article 6 regulates that each Member State is required to indicate the NUTS1 or NUTS2 level regions for which it will present a programme for financing by the ERDF.

In Finland, the previous programme area allocation was based on the EU Accession Treaty of 1994, and this allocation did not comply with the NUTS classification. Finland therefore was obliged to alter the ERDF area allocation so that it matched the NUTS2 regional division. At the time, there were four NUTS2 regions – Southern, Western, Eastern and Northern Finland – and each had its own regional programme during 2007–2013. The largest reallocation took place in the region of

---

[5]   See the references in article I.
[6]   Objectives 1, 2, and 3 represent the previous tiers of EU regional policy.

Ostrobothnia in Western Finland. Four sub-regions previously included in the ERDF Programme for Western Finland were reallocated to the ERDF Programme for Northern Finland (Fig. 1). As control sub-regions we used those areas that remained in the ERDF Programme for Western Finland.
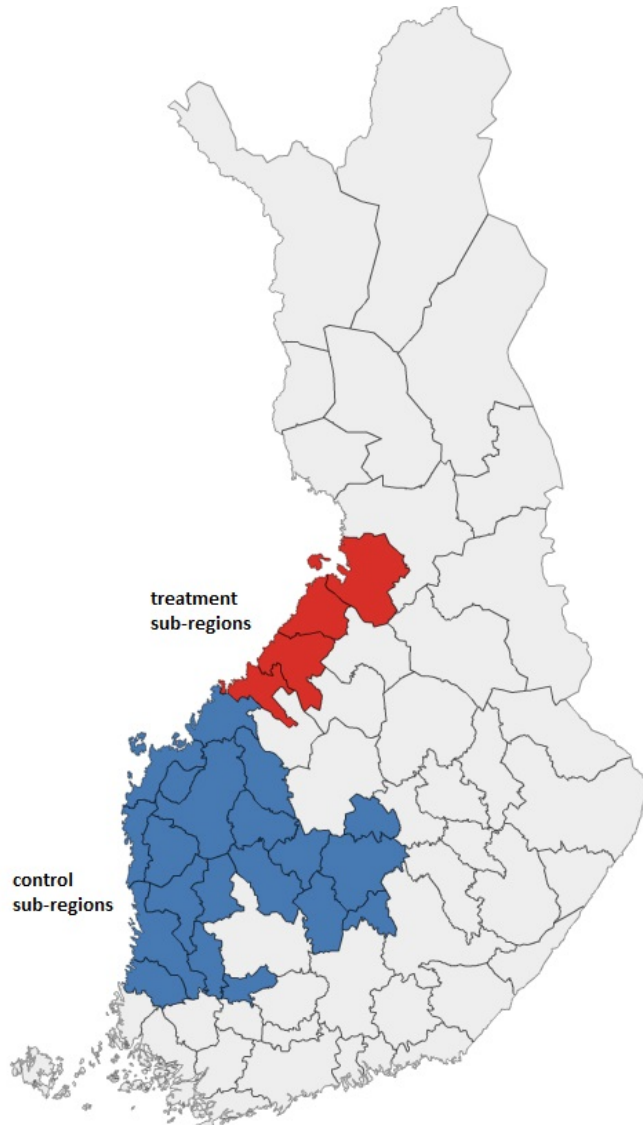
Figure 1. Treatment and control sub-regions.

As the treatment area is quite small – only four sub-regions which include twenty-two municipalities – we decided to use postal code area data in our analyses. This

allowed us to analyse 122 treatment area observations, with approximately 700 observations for the control areas. Statistical postal code areas are defined by Statistics Finland and they are based on addresses (postal codes) of firms, government offices, and inhabitants.

A further complication was that both the control and treatment sub-regions had two different support intensity levels in the previous programming period 2000–2006. The proper Objective 2 support was 33 euros/capita/year and the so-called transitional support was 9 euros/capita/year. In the new programming period 2007–2013, the treated areas had a support intensity of 70 euros/capita/year. The control areas' support intensity changed as well (Table 2, reproduced from article I).

Table 2. ERDF support intensity (EUR/capita/year) in treatment and control areas, 2000–2006 and 2007–2013.

Support category in 2000–2006	Treatment and control areas 2000–2006	Control areas 2007–2013	Treatment areas 2007–2013
Objective 2 programme areas	32.8	17.0	69.9
Transitional areas	9.4	17.0	69.9

Notes: Population data as of 1.1.2007. Source: Ministry of Economic Affairs and Employment.

Therefore, we had to split the control and treatment groups into two sub-groups, which were analysed separately. Our outcome variables were the number of jobs, the unemployment rate, the share of tertiary educated population, and the median disposable income per capita in each postal code area.[7] The selection of our outcome variables was based on the objectives of the Europe 2020 Strategy, which used the structural funds as the main policy instrument. The data covered the years 2004–2013.

Our main results in article I can be stated as follows:

- we detected a statistically significant decrease in the unemployment rate – on average, −1.6 percentage points – throughout the treatment period (2008–2013) in those sub-regions that had low support intensity (9 euros/capita) in the previous programming period 2000–2006

- in those same sub-regions, we also found positive regional policy effects on the number of jobs, but the coefficient estimates were statistically less convincing – we will take a further look at this result in Section 4

- all other coefficient estimates were statistically insignificant.

---

[7]   For summary statistics and detailed results, see article I.

It is important to understand what the last item means: doubling the support intensity from the previous levels did not have any effect on the outcome variables, i.e. unemployment, income, the number of jobs, and the share of tertiary educated inhabitants. Actual effects required a seven-fold increase in support intensity. Averaged across time periods, the estimated effect on the unemployment rate was a relative change of −0.16. The mean unemployment rate in the treated postal code areas was 9.9% before the treatment, which means that the sudden increase in ERDF support decreased the unemployment rate by approximately −1.6 percentage points in the treated areas. Besides being statistically significant, the effect was also economically significant.

Our baseline results were robust to various sensitivity checks, including spatial correlation. We therefore concluded that the EU regional policy is not totally without merit, although the long-term effects of these support measures are difficult to measure. The reason for this is that after one programming period ends, another one immediately begins, and there are no "without support" time periods in between. We discuss the measuring of the social costs and benefits of these programmes below in Section 3.4.

3.2   Dropout prevention in vocational education

A large-scale study completion programme was implemented by the Finnish National Board of Education in 2011–2014. The objective of the programme was to increase completion of vocational secondary education and decrease dropping out of vocational schools – a notorious problem during the previous decade. In some vocational schools, the yearly dropout rate exceeded ten percent. In the study completion programme, a total of 16 million euros in programme grants was allocated to participating schools. Compared to aggregate government funding for vocational education (approximately 700–800 million euros), programme resources were marginal, but when compared to the average programme grants awarded by the National Board of Education, this particular programme was quite substantial in size. In article II, we study the effects of the programme on study completion and school dropouts, using pooled cross-sections of vocational school students.

Since the schools, not the students, applied for the programme, we argue that from the viewpoint of the students, this case may also be framed as a natural experiment, provided that certain supplementary conditions are fulfilled. We take a closer look at these conditions in Section 4.

Dropout prevention measures have been widely studied, especially in the context of high schools in the United States and other developed countries (for reviews, see Wilson et al. 2011 and Hahn et al. 2015). The motivation for these measures has come from several cross-sectional studies, where dropping out of high school has

increased the probability of unemployment, poverty, and economic difficulties in general. Since dropping out is not randomly allocated, these cross-sectional estimates have been biased, but recent sibling studies have corroborated these findings (Campbell 2015).

The consensus of dropout prevention study reviews is that dropout programs have indeed succeeded in decreasing high school dropouts. However, upon closer inspection, these results may not be as solid as they seem. In the United States, Canada, and the United Kingdom, lack of high-quality register data has forced dropout researchers to use pretest-posttest survey data, where the students fill in a questionnaire before and after the preventive measure is implemented. With a 100% response rate, this popular study design produces valid and reliable longitudinal data. The problem is that usually a non-negligible number of respondents are not available for the second-wave survey. In some studies, attrition rates have been 50% or even higher.[8]

Attrition rates of even 50% are not an issue, if the subjects are missing at random; that is, the expected outcome is not correlated with unavailability for the second-wave survey. Unfortunately, several studies have shown that this is a rare occurrence. The literature suggests that missing not at random is the most likely mechanism for loss to follow-up because missing subjects tend to have different outcomes from those that remain in the programme. Depending on the associations between the expected outcome and the treatment, bias can be positive or negative. If we assume that the missing students from follow-up surveys are more likely to drop out than those that remain in the programme, our pretest-posttest estimates of programme effectiveness are biased upwards and vice versa. Therefore, a meta-analysis of those estimates also produces biased results.

To circumvent the prevalent attrition bias of earlier dropout prevention research, we decided to use high-quality register data from the student registers of Statistics Finland in our study. The disadvantage of this approach was that we were unable to identify individual students who participated in the programme, as the Finnish National Board of Education did not collect student-level participation information. Nevertheless, we were able to compare the students from the treated schools to students from the non-treated schools, without attrition. The possible measurement errors of the procedure are discussed in Section 4.

The target schedule for vocational school completion is three years. We therefore requested Statistics Finland to collect six random samples of approximately 4,000 newly enrolled students from the years 2002, 2007, and 2012, both from the treated schools and the control schools, for a total sample size of 24,000. The year 2007

---

[8]   For details, see the aforementioned review articles.

sample served as our "before the treatment" sample while the year 2012 sample served as our "after the treatment" sample. Each sample also contained information on the students' graduation – or dropping out – three years after the enrolment year. That is, the data included information whether the student enrolled in 2012 had graduated in 2015, or whether the student had dropped out by that time, or whether he or she was still enrolled as a student.

Since the dropout prevention programme was funded and implemented during 2011–2014, we assumed that the students enrolled at that particular time period would benefit the most from the programme. We then simply pooled the 2007 and 2012 samples and estimated the effect of the programme using standard difference-in-differences regression for pooled cross-sections.

Our main results in article II can be stated as follows:

- we were unable to find any evidence of the programme's effects on study completion or dropping out

- both measures improved more in the control schools than in the treated schools; prior to the treatment, there were no unobservable differences between treatment and control school students after controlling for observable factors

- likely reasons for the improvements include prolonged economic recession and tightened criteria for youth unemployment benefits.

It should be emphasised that our findings from register data are in stark contrast with earlier dropout prevention research, in which pretest-posttest survey methods were mainly used. While we recognise that our result may be due to a peculiar statistical coincidence, or poor implementation of the programme, we believe that more register-data based research is needed on this topic. As explained above, the previous positive findings of dropout programmes' effectiveness could be caused by severe attrition biases.

3.3 Effects of face-to-face counselling on unemployment rate and duration

The Internet has profoundly changed our lives in a matter of decades: our banking services, hotel reservations, and even our dating and mating habits, among other things, have undergone drastic changes after online technology has taken over. Labour markets have also been affected. In Finland, finding a job through an online search was possible already in the early 1990s, as the Public Employment Service (PES) offices' job search engine (www.mol.fi) was launched. It was one of the firsts, if not the first, public online services available in Finland.

In the year 2013, the Ministry of Economic Affairs and Employment decided to take things still further. Based on a visionary Ministry paper from the year 2010, the Ministry's goal was to create a "virtual PES office" by the year 2015. Several dozen PES offices were permanently closed, and more online services were offered instead. The unemployed clients were divided into three service categories, which had different agendas in dealing with the clients. Those in category three, i.e. the most challenging clients in need of in-depth counselling, were serviced with the bulk of the available caseworker resources. The clients in categories one and two were expected to use mainly other service channels (online or telephone services). The total PES office staff was reduced by ten percent. The main objective of the reform was to increase efficiency and productivity of the PES offices.

This major reorganisation of Public Employment Services created favourable circumstances for policy evaluation, as the affected municipalities were opposed to the reform. Permanent civil service jobs, like those of the PES office caseworkers, provide especially smaller municipalities solid and stable tax revenue, and rarely they wish to dispense with this kind of workforce voluntarily.

From the Government's point of view, the reform had potential for budget cuts: substituting online services for costly caseworkers could create significant savings in outlays. The only question is whether the effects on the jobseekers remain equal. In other words, are the new online services and the traditional face-to-face counselling services perfect substitutes in reducing unemployment? This is the topic of our case study III.

Prior research has found evidence of both positive and insignificant effects of intensive counselling efforts. Likewise, evidence on the effectiveness of Internet job search has so far been inconclusive, as this is a fairly new field of research. We could not locate studies that would have studied the aggregate (municipality-level) effects of permanent PES office closures.[9]

In this case, the "treatment" was the closure of the PES office – not necessarily a beneficial incident. The PES office closures were allocated to smaller municipalities: the average workforce in the treated municipalities was approximately half of the workforce in the control municipalities. As control municipalities we used the rest of the Finnish municipalities (= the unaffected municipalities), whether they had a PES office before the reform or not. The treated and control municipalities are depicted in Figure 2.

---

[9]   For references, see article III.
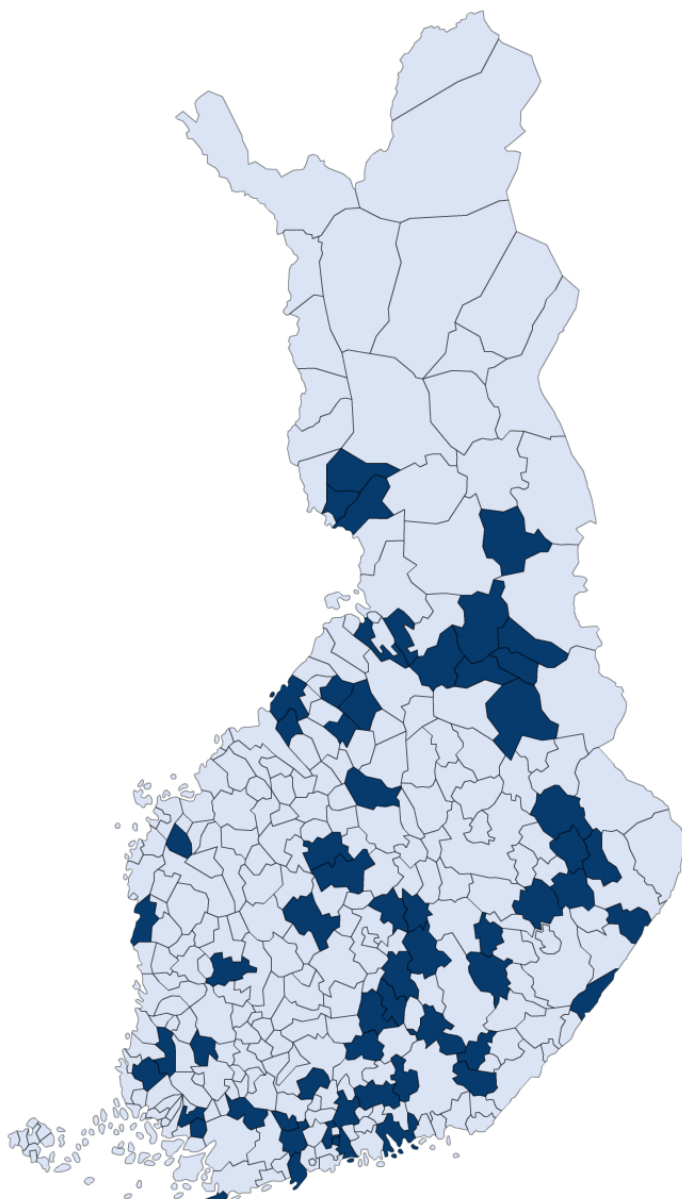
Figure 2. Treated municipalities in dark blue.

Before the reform, in 2012, the unemployment rate in the treated municipalities was 10.0%. In the control municipalities, the unemployment rate was 9.7%.[10] The

---

[10] Note that these are register-based unemployment statistics. The official unemployment statistics, which are based on a monthly survey, are not available at the municipality level.

difference of 0.3 percentage points was not statistically significant. Also, the unemployment trends were parallel before the treatment. Hence, the relative unemployment levels in the treated areas were not different – certainly not better – from the rest of the country. Retrospectively, it seems evident that the main logic in the reform was dictated by efficiency concerns: closing the smaller PES units and centralising the face-to-face counselling services into larger towns.

For our analysis, we compiled a 12-year municipality-level panel data set, covering the years 2006–2017. Our outcome variables were the unemployment rate and the average unemployment duration. Using simple two-period difference-in-differences estimators and more involved fixed effect panel data estimators, we discovered that:

- the reform increased unemployment duration by two to three weeks in the treated municipalities

- the reform also had a small, transitory effect on the unemployment rate

- the treated municipality's long distance to the nearest available PES office location increased the effect on unemployment duration.

Therefore, we found out that traditional face-to-face counselling and modern online counselling – which is largely a euphemism for self-service on the Internet – are not perfect substitutes in decreasing the length of unemployment spells. Counselling the unemployed often requires real human touch and care. Before the reform, the average unemployment duration in the affected municipalities was 37 weeks. Hence, the reform increased unemployment duration approximately by 5–10 percent – both statistically and practically significant increase. Consequently, the fiscal costs of the reform surpassed the fiscal benefits by a large margin.

3.4   Policy implications

From the point of view of policy implementation, what did we learn from our case studies I–III? Overall, the findings are not that surprising. When we increase public spending allocated to certain eligible areas, desired results can be achieved, if all goes well (article I). On the other hand, when we decrease government resources, this might lead to unexpected disadvantages in the affected areas (article III). Poorly implemented government programmes may not benefit the recipients at all (article II) and represent wasted resources.

Note that the statements above do not suggest that these or any other policies are beneficial or harmful in the social cost-benefit framework. Calculating the social costs and benefits of each of these programmes is out of the scope of this thesis. Based on the longevity of the EU regional policy, we may deduce with certainty that

the "social planner of the EU"[11] appreciates the wellbeing of the fringe areas of the EU differently from the central areas of the EU. Consequently, the planner is willing to undertake costly transfers from the developed areas to the less developed areas, year after year. Therefore, the welfare weights of EU citizens in different member states must be nonequal, which renders social cost-benefit calculations nearly impossible.[12]

---

[11]   In practice, the elected European Parliament.
[12]   With equal welfare weights, cash transfers do not increase social welfare.

4   Validity of the Case Studies

In his seminal article on natural experiments in economics, Meyer (1995) defines threats to validity as "*problems that may undermine the causal interpretations in studies*".[13] Meyer separates threats to internal validity from threats to external validity. Internal validity means that within the context of the study, we can ascertain that the observed differences in outcome variables were caused by the differences in the independent variables – and not, for instance, by omitted variables. External validity is fulfilled when effects found in a natural experiment can be generalised to different contexts and individuals. In this Section, we discuss our case study findings – summarised in Section 3 – from these perspectives.

4.1   Regional policy

Article I represents unarguably our strongest case in terms of random-like experiment. The treated sub-regions, which received a windfall of European Regional Development Fund (ERDF) support, were exposed to treatment as a result of two unrelated incidents: Finland's EU Accession Treaty of 1994 and a change in EU legislation nine years later, which defined the eligible ERDF support areas. We can assert that treatment was exogenous in this case. In fact, the treated sub-regions had slightly higher median disposable income per capita than the control sub-regions (see article I, Table 2).

However, the treated sub-regions were unfortunately quite small, and by modern standards, the number of observations (treated postal code areas) was indisputably low, 122 in total. This might also explain why we detected desirable regional policy effects only on unemployment. Differences are easier to observe when the changes are large, and the unemployment rate is easily the most volatile of our outcome variables in article I.

The limited number of observations could also explain why we were unable to find more credible evidence of the ERDF effects on the number of jobs in the transitional areas, which experienced the largest increase of ERDF support. The

---

[13]   Meyer 1995, p. 152 (italics added).

ERDF support is mainly allocated to small and medium-sized firms for investment and development purposes and to municipalities for various infrastructure projects. Hence, we should also have seen an increase in the number of jobs, as the unemployment rate decreased in the treated postal code areas. But there could be a location-based explanation behind this finding.

As an example, consider first Figure 3, which depicts the change in *ln(unemployment rate)* from 2007 to 2011 in the treated transitional areas. The mean difference in the treated areas was 0.021. Meanwhile, the mean difference in the control areas was 0.155. Hence, the coefficient on the dummy variable *treat* was −0.133 with a *p*-value of 0.01 – a highly significant estimate. Note that we only had N = 39 observations[14] from the treated postal code areas (and 200+ observations from the control areas – these are not depicted in Fig. 3).
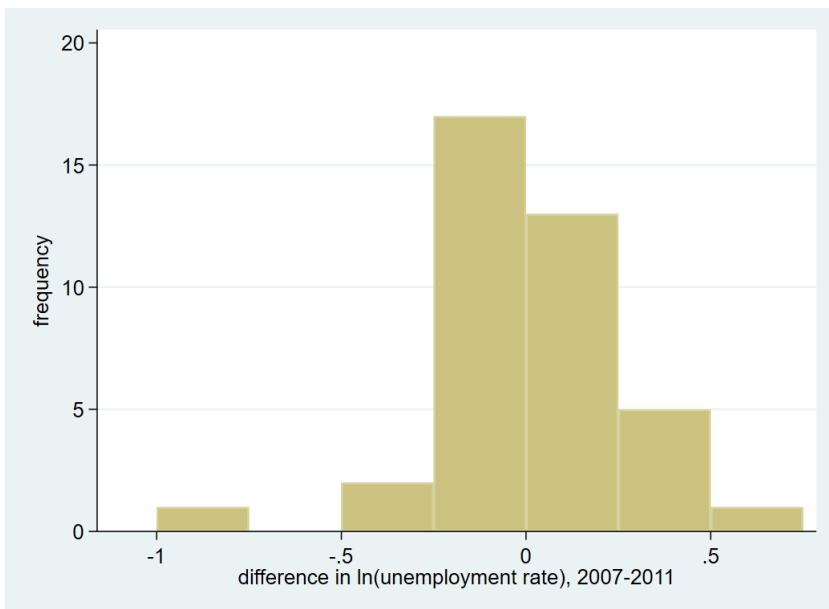

Figure 3. Difference in ln(unemployment rate) from 2007 to 2011, treated transitional areas.

Consider then Figure 4, which depicts the change in *ln(jobs)* from 2007 to 2011 in the same postal code areas. The mean difference in the treated areas was 0.108 and the mean difference in the control areas was 0.028. Hence, the coefficient on the dummy variable *treat* was 0.080 with a *p*-value of 0.145, i.e. the coefficient was not statistically

---

14   Note that our sample size is still within the bounds of a moderate sample size ($30 \leq N \leq 60$) which enables us to invoke the Central Limit Theorem (Wooldridge 2012, 783).

significant. Therefore, it seems that the job growth concentrated on just a few postal code areas, whereas the changes in the unemployment rate were more spread out.
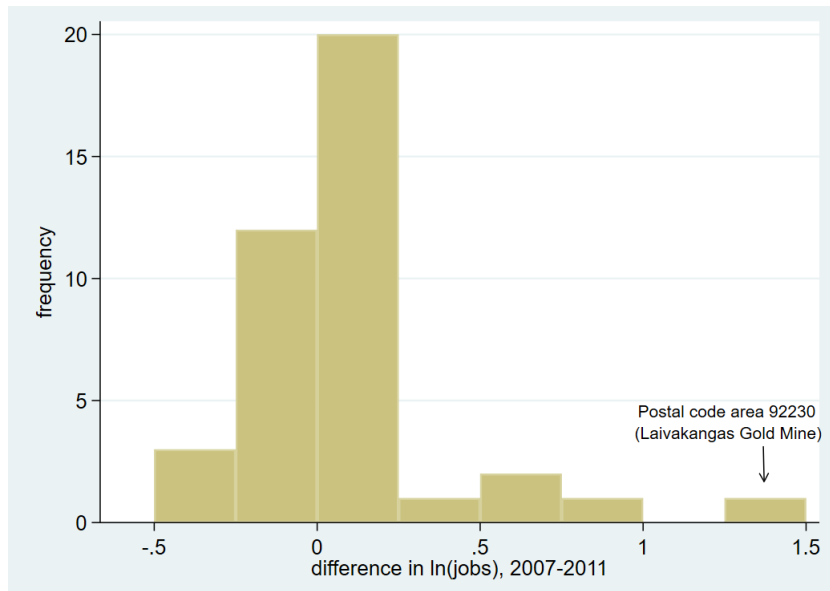
Figure 4. Difference in ln(jobs) from 2007 to 2011, treated transitional areas.

There could be a natural explanation for this: if they can avoid it, the inhabitants of any given municipality tend not to live in the industrial areas, which are usually centred around certain districts – often by deliberate municipal design. And this is mainly where the new jobs are located. But in our case, we had so few observations from those industrial postal code areas that our diff-in-diffs coefficient estimates on the policy effect on jobs were too imprecise.

In Figure 4, the largest increase in the number of jobs from 2007 to 2011 was in the postal code area 92230, where the Laivakangas Gold Mine is located. The mine is one of the largest gold mines in Europe.[15] Incidentally, the owner of the gold mine, Nordic Mines Ltd, received a support of 1,300,000 euros from the ERDF during 2009–2011 for mining infrastructure investments.[16]

Note also that those treated areas, that had higher support intensity (33 euros/capita/year) before the area reallocation, had twice the number of observations (N = 81) when compared to the transitional areas. Still we failed to find any policy effects on the outcome variables. Despite the different sample sizes, this finding

---

[15]   See for instance https://fi.wikipedia.org/wiki/Laivakankaan_kaivos (in Finnish).
[16]   For details, see https://eura2007.fi/rrtiepa.

seems reasonable: in order to achieve actual results, the policy change must be large enough in relative terms.

We argued in article I that due to our research design, our results are generalisable, at least to a degree. However, before we make any definite claims concerning the external validity of our study, we should try to replicate our findings, using settings and data from other EU member states, for instance. The only drawback is that similar random-like experiments rarely occur in EU regional policy.

It should be noted that another support area reallocation was enforced in Finland before the current 2021–2027 programming period started. The new and old NUTS2 regions are depicted in Figure 5. As a serious suggestion for further research, we should explore in the coming years whether these new support area reallocations created circumstances for usable natural experiments.
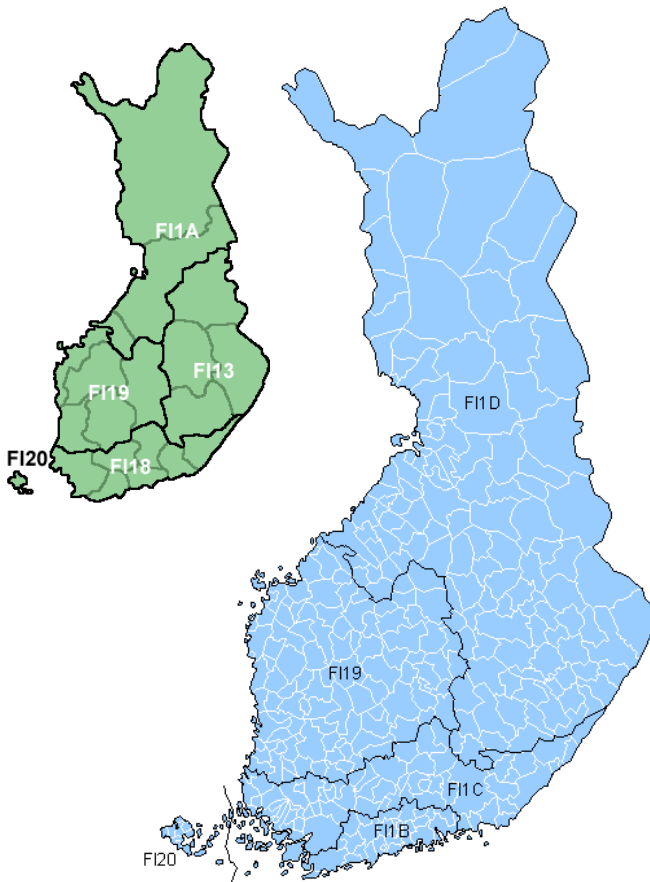
Figure 5. Finnish NUTS2 regions during the 2021–2027 programming period and during the previous programming period (small map). (Source of maps: The Ministry of Economic Affairs and Employment of Finland.)

4.2   Dropout prevention

Questions of internal validity are rightly justifiable in our second case study, article II. At the outset, our data may have suffered from measurement error in programme exposure. According to the external evaluation report (Ahola et al. 2015), approximately 85% of the funded projects declared that students were actively involved in the dropout prevention project of their school. That is, they were exposed to new teaching methods, or more intensive student counselling, or whatever was implemented in the project.

Therefore, in our year 2012 random sample, some 15% of the sampled students may not have been actively participating in their schools' dropout prevention project. But since the Finnish National Board of Education did not require the schools to collect information on the participating students, there is not much we can do to remedy this problem, except to acknowledge it. Luckily, a significant majority of the students were involved in the funded projects.

A second validity threat issue concerns selection. We argue in article II that from the viewpoint of the students, programme implementation effectively created a natural experiment. Whether the school participated in the dropout prevention programme during 2011–2014 was the decision of the National Board of Education, based on applications from voluntary schools. Hence, the schools self-selected themselves into the programme, or at least applied to the programme. Future students, enrolling in 2012, could not affect these decisions.

However, there might have existed obvious reasons why precisely these vocational schools applied for the dropout prevention programme. For instance, they could have been vocational schools where dropping out had been far more common problem than in other schools, and immediate interventions were needed. In that case, we should have observed a significant coefficient on the dummy variable *dropoutprog*, which measures the unobservable differences between the treated and control students, after controlling for their observable characteristics. In article II, Table 4 (baseline estimates), we estimated that the *dropoutprog* dummy was not statistically significant at conventional significance levels.

Thus, we found no evidence that the participating schools had problems in study completion or dropping out *ex ante*, when compared to the control schools. Also, when we looked at the average dropout rates in 2002, ten years before the program was implemented, the control schools had slightly higher three-year dropout rate (28%) than the treated schools (27%).

Another possibility which could have caused selection bias is that after the programme started, the new vocational students enrolling in 2012 selected their schools solely based on the schools' dropout programme participation, in order to benefit from the programme and to maximise their graduation possibilities. If this kind of behaviour occurred, it should have led to a large increase in new enrolments

in the participating schools. To scrutinise this prospect, we compiled data on new vocational school enrolments in the treated and control schools from 2005 to 2015, and the aggregated time series are presented in Figure 6.[17] Note that only those vocational schools that were in continuous operation during the whole time period are included in the time series.

From Figure 6 we observe that the financial crisis of 2008 increased vocational schools' appeal in the following year, most likely due to increased youth unemployment. But this was only a temporary phenomenon. In the year 2012, the year of our "after the treatment" sample, we find that the number of new students in fact decreased in the participating schools and increased in the control schools. Therefore, it seems unlikely that the mere existence of the dropout prevention programme would have influenced the nine-graders' school choice decisions in the spring of 2012.

In our opinion, this is the main argument of this dissertation, besides the original research results presented in the case studies. If certain supplementary conditions are satisfied, selection by someone else than the unit of observation itself (see Table 1 above) can create circumstances where the treatment might be considered exogenous. Note that fulfilling these conditions is strictly necessary for valid causal inference. For instance, if in our case study II the new enrolments in the treated schools would have surged in 2012, then we would have had a case where essentially the students self-selected themselves into school-specific treatment, and causal interpretation of the results would have been invalid. The situation would have been identical if we would have used school-level data in our analysis. Likewise, if only the schools prone to dropping out would have participated, treatment would have been endogenous.

---

[17]   Data source: National Board of Education.
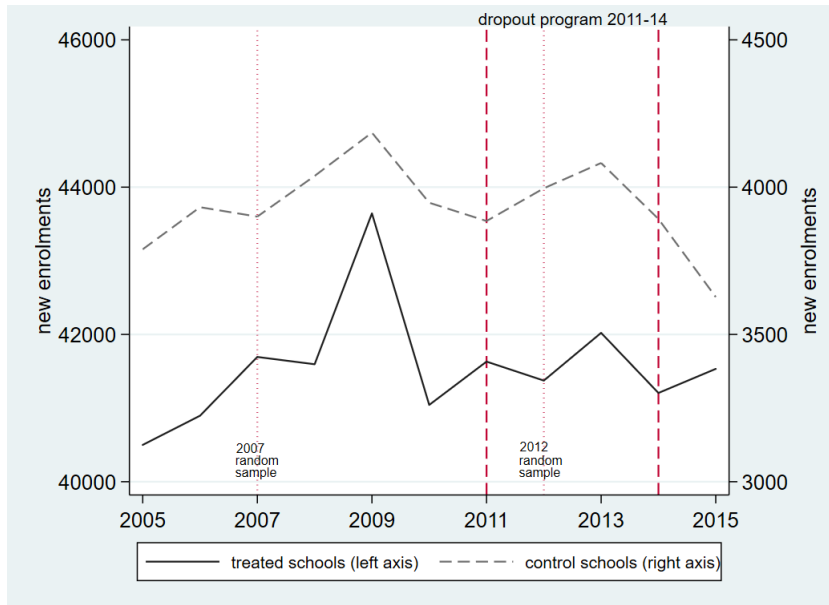
Figure 6. New enrolments in the treated schools and in the control schools, 2005–2015, excluding youth apprenticeship training and adults' competence-based qualifications.

Note that in their social science methodology handbook, Remler and Van Ryzin (2010, 432) distinguish natural experiments from quasi-experiments and argue that the latter should be defined as cases where the program or treatment is consciously implemented to produce some change in the world. Hurricanes and other natural disasters aside, we believe that there is a very fine line between these two concepts, and the main issue is the exogeneity of treatment. If treatment is endogenous, we have neither a natural nor a quasi-experiment at our hands.

Our findings in article II radically differ from earlier dropout prevention research, which is based on pretest-posttest surveys. Therefore, we should aim to replicate them with additional studies based on register data, before making any assertions on their external validity. Simultaneously, we would improve our understanding on the scale of attrition bias in survey-type study designs. But we also have to keep in mind that there are noteworthy differences in the national education systems of the Nordic countries and those of the United States, for instance.

4.3   Face-to-face counselling

In our case study article III, criteria for random-like treatment exposure are mostly fulfilled. The affected municipalities did not volunteer for the PES office closures; instead, they were steadfastly opposed to the reform. The unemployment situations

in the treated municipalities before the reform did not differ from the unemployment situations elsewhere. The main difference to randomisation was that the PES office closures were targeted at smaller municipalities rather than just being randomly allocated.

Thus, the Government's motivation was to integrate services into larger PES units and simultaneously achieve gains in efficiency and productivity. From the point of view of the affected municipalities, treatment exposure was as good as random. However, as this reorganisation was decided on a national level, the degree of treatment exogeneity is not as strong as in case study I. We cannot rule out some sort of discreet lobbying efforts by the potentially affected municipalities before the final decisions were made.

As in our first case study, our analysis in article III was hampered by small sample size: we had only 61 observations (= municipalities) in the treatment group. Unfortunately, unemployment duration data is not available at the postal code area level, so the only alternative would have been to analyse individual-level data – pooled cross-sections or longitudinal data. This could be a viable study agenda for future research on PES offices. The 2013 reform is here to stay, and Statistics Finland has a large supply of high-quality register data available, both before and after the reform. With individual-level data we could also try to estimate the effects of different service categories, which were introduced in 2013.

Small sample notwithstanding, our estimation results were remarkably robust to various sensitivity tests. Using our data as a panel data set increased the number of treated observations to N = 244 (four years, 61 municipalities). Using fixed effect estimators, we were then able to control for unobserved municipality-specific trends, for instance slow changes in population – age structure, educational levels, in-migration, out-migration, and so forth (article III, Table 4, columns 3 and 4).

Note that in estimating the so-called random trend model (article III, Table 4, columns 3 and 4) we difference the interaction terms (reform × year). We do not interact the first-differenced treatment dummy and the year indicators, as this would yield a single variable equalling 1 in the year 2014 and 0 in the other periods (and other variables consisting only of zeros). We may interpret the differenced interaction terms as measuring the 2013 change in the PES office locations for each of the treatment years (2014–2017) separately. These variables equal 1 in the year of interest and -1 in the following year and 0 otherwise. Note that this specification is also different from simply using the various lags of the first-differenced treatment indicator.

Proximity to a municipality with an existing PES office could have caused downward bias to these estimates. If a PES office is within short distance, the unemployed clients in need of face-to-face counselling can easily obtain it by

travelling to another municipality. Unfortunately, there are no PES statistics available on the number of PES office clients from another municipalities.

With our distance-based estimates in article III, we may gauge the size of the proximity bias. In Table 4, column 4, the interaction of treatment and year 2015 yielded the coefficient of 1.9 weeks and the interaction of treatment and year 2016 yielded the coefficient of 2.7 weeks for the reform's effect on unemployment duration. In Table 7, column 4, for the municipality group located the farthest from an existing PES office we obtained the coefficients of 2.3 and 3.6, respectively (for convenience, these estimates are reproduced in Table 3 below). Due to large standard errors, the differences are not statistically significant, but they are non-negligible, nevertheless.

Table 3. The reform's effect on unemployment duration, FE regression estimates, first-differenced variables. N = 3,344.

Interaction term	Baseline coefficient	Coefficient on the farthest located municipalities
treatment x year2015	1.902* (0.837)	2.336* (1.028)
treatment x year2016	2.668* (1.172)	3.633* (1.546)

Notes: Reproduced from article III, Tables 4 and 7. Municipality-clustered standard errors in parentheses. * $p \leq 0.05$.

Hence, even in random-like experiments, we have to account for the possibility of bias-inducing factors, and if possible, test for those factors' effect on the estimates. Often this requires close inspection of the relevant characteristics of the situation and detailed analysis.

We must also keep in mind that even in normal conditions, not all unemployment spells end in open market employment. According to a Finnish study based on 2012 data, approximately 50% of unemployment spells ended in employment, while some 16% ended in active labour market intervention. 20% left the workforce for studying, or retired, or had some other reason (e.g. maternity leave) for the unemployment spell ending. The remaining cases were only short breaks in unemployment, for instance short-term (1–2 weeks') temporary jobs. The data included 174,366 ended unemployment spells.[18]

The external validity of our results is debatable: we studied only one reform, in one country. We therefore refrain from making any definite propositions. Note though that in a similar study, but using individual-level data, Schiprowski (2020) found out that missing one counselling appointment increased the unemployment

---

[18]   Prime Minister's Office (2016).

spell length on average by 12 days (5 % in relative terms). However, as Schiprowski's study does not contain information on the number of PES offices in the study, it is difficult to assess the aggregate (office-level) effect of these caseworker absences.

4.4　Summary

Based on the discussions and arguments presented above, Table 4 presents a subjective ranking of our case studies in terms of their internal and external validity. Our admittedly arbitrary scale is: Good validity–Mediocre validity–Poor validity.

Table 4.　Internal and external validity of the case studies.

Case study	Internal validity	External validity
I (regional policy)	Good	Good
II (dropout prevention)	Poor/Mediocre	Mediocre
III (face-to-face counselling)	Mediocre	Mediocre

As reasoned above, case study I is our strongest study design, followed by case study III. The internal validity of our case study II – and other similar study designs – hinges on whether our study fulfils certain supplementary conditions. While the internal validity of case study II may be questionable, we still contend it is an improvement over previous dropout prevention studies, which are based on pretest-posttest survey designs.

5   Concluding Remarks

In this thesis, we have studied the effects of various Government interventions aiming to improve citizens' employment, productivity, and re-employment possibilities. In order to produce credible causal estimates of the interventions' effects, we have relied on natural experiments, or situations, where exogenous "forces of nature" have defined treatment exposure. Our case studies have dealt with regional policy, educational policy, and employment policy, and we have also used different units of observation – postal code areas, students, and municipalities. Difference-in-differences and panel data regression methods have been our major methodological tools.

This summary article has served a threefold purpose. Firstly, we compiled the main points of our case study designs and the corresponding results into a concise, comparable format, without overemphasising the econometric equations and the statistical tables. Secondly, we had to take a closer look at the internal and external validity of our case studies, and we even collected new data to assess the exogeneity of our treatments. Thirdly, we hope that our case examples contributed to the natural experiment literature in general, especially in the field of social sciences.

The basic lessons from this exercise are quite straightforward. Natural experiments, by definition, are not planned or controlled in advance, they just happen. Consequently, we are not be able to determine sample sizes, or length of treatment exposure, or treatment intensity. We must take these facts as given. In some cases, we might be able to increase sample sizes by using less aggregated data, or even individual-level data, if available. This could also strengthen the exogeneity of the treatment (as in our case study II).

However, we must bear in mind that individual-level register data adds a significant amount of complexity to data analysis, as random noise and heterogeneity in the data will increase tremendously. Grave registering errors in raw data are not uncommon either, although they are quite rare. Using more aggregated data smooths away many of these complications.

Despite small sample sizes, we were able to produce quite reasonable findings. In article I, large relative changes in ERDF support intensity lead to discernible changes in unemployment, while small changes in support did not. In article III, the

PES office closures had a larger effect on unemployment duration, as expected. We may assume that the unemployment rate is mainly determined by labour market forces, not counselling modality.[19] The estimated effects in articles I and III were also practically meaningful, not only statistically significant.

Our case study article II is pivotal in terms of whether a random-like treatment exposure occurs or not. After all, vocational schools applied for the programme, and the National Board of Education selected the participating schools. Can we truly assert that treatment exposure was exogenous to the year 2012 cohort of new students?

If our sample would only have included a cross-section of new enrolments from 2012, we would have to conclude that the unconfoundedness assumption fails; the treated students could all have been enrolled in schools that had serious problems with study completion prior to the programme. Hence, we first must obtain data from the time period(s) before the programme was implemented. Then, we must test for the existence of unobservable differences, after observable factors are accounted for. Secondly, we also need additional data in order to verify that the pre-exposure behaviour of the units of observation (in this case, students) did not change due to the programme. That is, the units did not alter their choices just because they wanted to be exposed to the treatment. Both additional conditions – no unobservable differences, no change in pre-exposure behaviour – were fulfilled in our case study II.

Discovering potential natural experiments is not something you can learn by reading an economics textbook, although this might help in framing the right questions. The key is to look for unexpected changes in policy implementation or in the relevant legislation, or even look for mistakes in policy enforcement that a hapless official has inadvertently done. The affected units may not be happy – they might even be dead, as in the London cholera outbreak – but the researcher may find solace in the fact that this was not her fault.

---

[19]   These two outcome variables are only weakly correlated in cross-sectional data (see article III, Fig. 3).

References

Ahola, S., Saikkonen, L. & Valkoja-Lähteenmäki, L. (2015): Ammatillisen koulutuksen läpäisyn tehostamisohjelma – Arviointiraportti [Programme for improving graduation rates in vocational education – Evaluation report]. National Board of Education, Helsinki. (In Finnish with an English abstract.)

Campbell, C. (2015): The socioeconomic consequences of dropping out of high school: Evidence from an analysis of siblings. *Social Science Research*, vol. 51, 108–118.

Card, D. & Krueger, A. B. (1994): Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, vol. 84, no. 4, 772–793.

Craig, P., Cooper, C., Gunnell, D., Haw, S., Lawson, K., Macintyre, S., Ogilvie, D., Petticrew, M., Reeves, B., Sutton, M. & Thompson, S. (2012): Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. *Journal of Epidemiology and Community Health*, vol. 66, 1182–1186.

Hahn, R.A., Knopf, J.A., Wilson, S.J., Truman, B.I., Milstein, B., Johnson, R.L., Fielding, J.E., Muntaner, C.J.M., Jones, C.P., Fullilove, M.T., Moss, R.D., Ueffing, E., & Hunt, P.C. (2015): Programs to increase high school completion: A community guide systematic health equity review. *American Journal of Preventive Medicine*, vol. 48, no. 5, 599–608.

Huttunen, K. & Kellokumpu, J. (2016): The Effect of Job Displacement on Couples' Fertility Decisions. *Journal of Labor Economics*, vol. 34, no. 2, 403–442.

Levitt, S. D. & List, J. A. (2009): Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. National Bureau of Economic Research, Working Paper 15016.

Meyer, B. D. (1995): Natural and Quasi-Experiments in Economics. *Journal of Business & Economic Statistics*, vol. 13, 151–161.

Prime Minister's Office (2016): Effectiveness and alternatives of employment policy. Prime Minister's Office Publications 3/2016. [Available only in Finnish: Työpolitiikan vaikuttavuus ja vaihtoehdot, VN TEAS julkaisuja 3/2016.]

Remler, D. K. & Van Ryzin, G. G. (2010): Research Methods in Practice: Strategies for Description and Causation. Sage Publications, Thousand Oaks.

Schiprowski, A. (2020): The Role of Caseworkers in Unemployment Insurance: Evidence from Unplanned Absences. *Journal of Labor Economics*, vol. 38, 1189–1225.

Snow, J. (1855): On the Mode of Communication of Cholera. 2nd Edition. John Churchill, London.

Wilson, S.J., Tanner-Smith, E.E., Lipsey, M.W., Steinka-Fry, K., & Morrison, J. (2011): Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *Campbell Systematic Reviews*, vol. 7(1), no. 8, 1–61.

Wooldridge, J. M. (2010): Econometrics of cross section and panel data. 2nd edition. MIT Press, Cambridge.

Wooldridge, J. M. (2012): Introductory Econometrics – A Modern Approach. 5th edition. South-Western, Mason.

**TURUN
YLIOPISTO**
UNIVERSITY
OF TURKU