The Impact of European Funding on Learning Outcomes of Pupils in Slovakia

Miriama Sokoláková

Vydal: ©TOPAZ
©2024
Opletalova 1603/57
11000 Praha 1
Česko

Více informací na: http://www.top-az.eu/home/

ISBN: 978-80-88350-35-4

Práce je bakalářskou prací, která byla oceněna ve 4. ročníku studentské soutěže Cena Karla Schwarzenberga. Z důvodů zachování fomátu datové analýzy je publikována v původní verzí bez úprav.

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

The Impact of European Funding on Learning Outcomes of Pupils in Slovakia

Bachelor's thesis

Author: Miriama Sokoláková

Study program: Economics and Finance

Supervisor: PhDr. Lenka Šťastná, Ph.D.

Year of defense: 2023

Declaration of Authorship

The author hereby declares that she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants Charles University permission to reproduce and distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, May 2, 2023

Miriama Sokoláková

Abstract

As a member of the European Union, Slovakia has access to various funding programs aimed at improving education and fostering economic development. This thesis examines the effectiveness of these funds in enhancing educational achievement in all relevant Slovakian primary and secondary schools. To assess the impact of European funds, the thesis employs fixed effects, between effects, and a difference-in-difference approach. We combine data collected from the National Institute of Education and Youth with socio-economic data. The analysis covers the period from 2007 to 2013, encompassing one funding cycle and allowing for a comprehensive evaluation of the long-term effects on educational achievement. The results do not reveal a correlation between European funds and the test scores of schools. However, a negative correlation was found between schools located in economically disadvantaged areas and with pupils from socially disadvantaged backgrounds. The study also identifies several challenges and areas for future research for more efficient fund allocation.

JEL Classification	C21, C23, I22, I28, R58
Keywords	school financial resources, educational achievement, educational determinants, EU funds, Slovakia, regression analysis

Title	The Impact of European Funding on Learning Outcomes of Pupils in Slovakia

Abstrakt

Jako člen Evropské unie má Slovensko přístup k různým programům financování zaměřených na zlepšení vzdělávání a podporu hospodářského rozvoje. Tato studie zkoumá účinnost těchto fondů při zlepšování vzdělávacích výsledků ve všech relevantních slovenských základních a středních školách. K posouzení dopadu evropských fondů využívá tato práce model fixních vlivů, model mezidobých vlivů a přístup "difference-in-difference". Tato práce používá data o výsledcích testů získaných z Národního institutu vzdělávání a mládeže a kombinuje socioekonomická data. Analýza pokrývá období od roku 2007 do roku 2013 zahrnující jeden cyklus financování a umožňující komplexní hodnocení dlouhodobých účinků na vzdělávací výsledky. Výsledky neodhalují souvislost mezi evropskými fondy a výsledky testů škol. Nicméně byla nalezena negativní korelace u škol nacházejících se v ekonomicky znevýhodněných oblastech a s žáky ze sociálně znevýhodněného prostředí. Studie rovněž identifikuje několik výzev a oblastí pro budoucí výzkum pro efektivnější alokace finančních prostředků.

Klasifikace JEL	C21, C23, I22, I28, R58
Klíčová slova	finančné zdroje pre školy, školské výsledky, faktory ovplyvňujúce školské výsledky, eurofondy, Slovensko, regresná analýza

Název práce	Vliv evropského financování na výsledky srovnávacích testů žáků na Slovensku

Acknowledgments

I would like to express my gratitude to PhDr. Lenka Šťastná, Ph.D.. Her guidance, feedback, and support were invaluable in shaping my bachelor's thesis into the final version. I would like to express my gratitude to my colleagues and classmates at the university, without whom I could not have imagined graduating at all. Next, I would like to thank my partner for his constant encouragement and emotional support. Last but not least, I would also like to express my sincere gratitude and appreciation to my parents and entire family for their support throughout my academic journey.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havrankova and prof. Tomas Havranek of Institute of Economic Studies, Faculty of Social Sciences, Charles University. Text fine-tuned and grammar-checked by Notion and Grammarly.

Bibliographic Record

Sokoláková, Miriama: *The Impact of European Funding on Learning Outcomes of Pupils in Slovakia.* Bachelor's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 68. Advisor: PhDr. Lenka Šťastná, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

BE Between Effects

BLUE Best Linear Unbiased Estimator

CEE Central Eastern Europe

DID Difference-in-Difference

EU European Union

FE Fixed Effects

GDP Gross Domestic Product

ICT Information and Communication Technologies

KPI Key performance indicator

MLR Multiple Linear Regression

NIVAM National Institute for Education and Youth

OLS Ordinary Least Squares

RE Random Effects

Bachelor's Thesis Proposal

Author	Miriama Sokoláková
Supervisor	PhDr. Lenka Šťastná, Ph.D.
Proposed topic	The Impact of European Funding on Learning Outcomes of Pupils in Slovakia

Research question and motivation European Cohesion Policy and Structural Funds aim to reduce regional disparities. The main objective of the 2007-2013 programming period was to accelerate the economic development of regions with Gross Domestic Product (GDP) per capita below 75% of the EU average. Slovakia, among others, was eligible to finance projects in all administrative counties within borders of 7 priority axes, including various sectors.

This thesis will investigate the Education Priority Axis under the Convergence objective, as it offers a unique opportunity to assess whether the money invested brought some comparative differences in educational outcomes of students in respective schools compared to their 'non-funded' peers. For assessing students' educational outcomes, results of nationwide educational measurement tests will be used. This data will be a relevant measure of effectivity and efficiency of European resources invested into the improvement of educational systems as PISA tests - the worldwide equivalent of this nationwide test.

This thesis aims to provide an overview of the economic and demographical factors, which influence the results of pupils in primary and secondary schools in Slovakia. Specifically, it will focus on the EU financial support over the programming period. Furthermore, we will differentiate between various types of projects - training of teachers, new technology or reconstruction of the school building.

Contribution The most significant contribution of this thesis will be the combination and analysis of data from previously uncombined sources. These are the nationwide educational tests and the details about European funding of schools in Slovakia. To the extent of our knowledge, such a study has not been conducted before.

The most similar studies, we are aware of, have been focusing on other priority axes such as unemployment and structural funds in Slovakia (Kotrč, 2020). This study has shown a positive relationship between the presence of European funds and measured Key performance indicator (KPI)s. Other studies are rather focused on more complex effects on the economic growth of the countries and are, therefore, less applicable (Dall'erba a Fang, 2015). In the case of studies concerned with explaining the educational outcomes of schools and students, they do not include European funds as one of the regressors, usually, they do not even consider school funding and investment as a regressor at all.

Methodology The National Institute for Certified Educational Measurements of the Slovak Republic will provide the educational outcomes of nationwide tests for both primary and secondary schools. Details of the Programming period 2007 - 2013 of the Cohesion Policy and Structural Funds will be acquired from the Ministry of Investments, Regional Development, and Informatization of the Slovak Republic. It will then be manually matched to specify, if and what type of funding was received by the respective educational institution. The number of educational outcomes observations is approximately 3000 schools (both primary and secondary) for the period of 11 consecutive years (2009-2019).

Other effects will be assessed as well. Examples of data available for each school are the number of students per teacher, the percentage of pupils from low-income families or the usage of interactive electronic tools during lessons. There will also be more generalized data included in the model such as the unemployment rate or average salary in the education field in the respective districts.

As can be deduced from the information above, these data will be panel data. In the case of educational outcomes observations, these are sometimes also unbalanced due to various fluctuations. Consequently, the data will be assessed through the standard analysis of panel data - fixed and random effects models. To decide, which one to use, a Hausman test will be used.

Outline

1. Introduction
2. Literature Review
3. European Cohesion Policy and Structural Funds
4. Methodology
5. Data
6. Results
7. Conclusion

Core bibliography

BOISSIERE, Maurice. Determinants of primary education outcomes in developing countries. World Bank, Independent Evaluation Group (IEG), Washington, DC, 2004.

BOUAYAD-AGHA, Salima; TURPIN, Nadine; VÉDRINE, Lionel. Fostering the development of European regions: A spatial dynamic panel data analysis of the impact of cohesion policy. Regional studies, 2013, 47.9: 1573-1593.

DALL’ERBA, Sandy; FANG, Fang. Meta-analysis of the impact of European Union Structural Funds on regional growth. Regional Studies, 2017, 51.6: 822-832.

DINU, Mihai, et al. Accessing the European funds for agriculture and rural development in Romania for the 2014-2020 period. In: Proceedings of the International Conference on Business Excellence. 2020. p. 717-727.

FIDRMUC, Jan; HULÉNYI, Martin; TUNALI, Cigdem Borke. Can money buy EU love?. European Journal of Political Economy, 2019, 60: 101804.

JACKSON, C. Kirabo; JOHNSON, Rucker C.; PERSICO, Claudia. The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. National Bureau of Economic Research, 2015.

KOTRČ, Michal. The Impact of Structural Funds and Cohesion Fund on Convergence: The Case of Slovak Districts. 2020.

MOHL, Philipp; HAGEN, Tobias. Do EU structural funds promote regional growth? New evidence from various panel data approaches. Regional Science and Urban Economics, 2010, 40.5: 353-365.

SANTOS, M. Quality of education in Argentina: determinants and distribution using Pisa 2000 test scores. Well-being and Social policy, 2007, 3.1: 69-95.

TOVAR GARCÍA, Edgar Demetrio. Determinants of educational outcomes: Analysis of the Republic of Tatarstan. Communist and Post-Communist Studies, 2014, 47.1: 39-47.

Author Supervisor

Chapter 1

Introduction

The European Union's enlargement in 2004 was the sixth but the largest to date. It increased the area of the European Union (EU) by 20% and its population by 18%, but its GDP only by 9%, while considerably decreasing GDP per capita by 8%. This expansion created many opportunities but also posed challenges for the member states and the union as a whole in addressing regional disparities. Since then, the EU has been addressing these disparities using structural and cohesion funds. Although these funds are ambitious on paper, their real, long-term effects are a topic of discussion among politicians, researchers, and even the general public. The effectiveness of these funds in addressing regional disparities and promoting economic convergence among member states has been a subject of debate. While some argue that the funds have been successful in reducing disparities, others point out that the effects have been limited and sometimes non-conclusive.

The literature on this topic does not provide a straightforward answer regarding the impact of eurofunds on convergence. The effect appears to be far from uniform, as results differ from country to country between different operational programs or programming periods. This thesis aims to assess the impact of eurofunds on the educational achievement of Slovak schools based on test scores in nationally referenced tests. This study is unique in that it combines educational achievement with additional resources granted to schools through eurofunds, a combination which has not been explored in any European country before. This may have been due to the amount of manual labour required to match projects with their respective schools.

Our dataset comprises 1409 primary schools and 657 secondary schools from Slovakia. We used national tests by National Institute of Education and Youth as a reference for educational achievement. In addition to other school characteristics and socio-economic explanatory variables, we included eurofunds from the programming period 2007 - 2013 as an independent variable. Eurofunds were split into two types: education-related and reconstruction-related funds. We conducted the analysis using three different regression models: fixed-effects, between-effects, and difference-in-difference models. This aims to assess general determinants of educational achievement, as well as what influences whether a school's test scores increase or decrease over time.

The thesis is structured as follows: Chapter 2 contains a literature review of the current state of research, which is divided into two parts. The first part examines European funds literature, while the second part looks into the impact of school resources on educational achievement more generally. Chapter 3 presents information on the European funds and the data used in our model, including specifics on the finances granted and how we matched the data to individual schools. Chapter 5 discusses other variables used in the model, with national test results as the dependent variable and school characteristics and socioeconomic variables as explanatory. Chapter 4 provides an overview of the methodology and a description of the methods used. In Chapter 6, the results of the regression models are presented and discussed. The final Chapter 7, Conclusion, summarises the findings and limitations of this thesis.

Chapter 2

Literature Review

The aim of this chapter is to assess the current state of research on the effectiveness of Structural Funds in the European Union, with a particular focus on the Central Eastern Europe (CEE) region. To do so, we thoroughly searched available papers and related literature to compare the econometric methodologies and respective results of the studies in question. This chapter will examine the limitations of these studies and will discuss the policy implications of the findings. Finally, this section will suggest how the findings can be used in the context of the bachelor thesis.

2.1 Effects of European Structural Funds

The Structural Funds and Cohesion Fund have been crucial tools for addressing regional disparities in the European Union for a significant part of its existence. The European Social Fund, the European Regional Development Fund, and the Cohesion Fund have been in place for decades, with the former dating back to 1957 and the latter to 1975 and 1994, respectively. These funds have supported a wide range of projects and initiatives. It is, therefore, understandable that many research papers have attempted to test the efficiency of allocated resources from these funds by conducting various studies. As mentioned in Dall'erba *et al.* (2009), around one hundred studies investigate European regional policies, but only a relatively small portion of them conduct proper econometric research.

Most of the papers in the field are based on the growth model proposed by Solow (1956), which introduced the theory of diminishing marginal returns to capital and exogenous technical processes. This model is one of the most important contributions to studying economic growth and continues to influence research today. This theory was later developed by the neoclassical growth model of convergence, as proposed by Barro & Sala-I-Martin (1992). It is commonly used in papers within the field. The model argues that the rate of economic growth is determined by technological progress, capital accumulation, and labour force growth.

It is commonly believed that investing more money leads to better results. However, the literature suggests that the impact of grants varies depending on factors such as grant type, allocation, and geography. Most studies on the subject focus on the effects of structural funds in Western Europe. A significant publication in this field is the meta-analysis by Dall'Erba & Fang (2017), which provides a comprehensive review of available literature and shows that the impact of grants is rather non-uniform.

Studies that focus on the geography of Central Eastern Europe are rather scarce, with exceptions like Žáček & Hrůza (2019). This study uses panel data regression analysis with fixed effects controlling for spatial effects and finds that EU funds have a positive and statistically significant impact estimated at 0.91-1.12 p.p. on the GDP growth rate of Czech regions. A similar study was conducted in Slovakia by Radvanský *et al.* (2016). The study used the counterfactual impact evaluation method and found that the EU Cohesion Policy positively impacted regional development in Slovakia. Specifically, it led to GDP growth, employment, and productivity improvements. However, the policy's effectiveness varied across regions and sectors, and some areas experienced limited positive or even negative effects. Another Poland study by Modranka (2015) used a spatial panel econometric model. The impact of the policy changes over time and across space, with the largest positive effects observed in the early years of the policy implementation. Schoenberg (2018) aimed to determine whether convergence occurred in Romania and Bulgaria. The article used a difference-in-difference analysis method and found that EU funds positively impacted GDP per capita, reduced unemployment, and decreased research and development expenditure per capita.

Recent research by Fratesi & Wishlade (2017) has shown that the impact of the Cohesion Policy is not uniform. Subsequently, in recent years, academic interest has shifted from attempts to assess its overall impact to an emphasis on the "conditioning factors" that explain when, where, and how the policy is effective. For example, according to Rodríguez-Pose & Garcilazo (2015), the effectiveness of policies can be influenced by the quality of government.

Similarly, as stated by Becker *et al.* (2013), the amount of human capital present and the quality of institutions can also impact the effectiveness of policies.

Although the list of publications can be quite extensive, to the best of the author's knowledge, there are almost no quantitative evaluations of the impacts of Structural Funds on the educational outcomes of schools in the EU. Therefore, it is necessary to examine broader examples of studies and research papers that investigate the determinants of educational outcomes in various econometric models, particularly those that include funding resources as one of the variables.

2.2 Effect of School Resources on Student Achievement

For many decades, researchers and even politicians have attempted to determine whether additional resources invested in education create additional value for money or result in poor resource management, leading to a waste of funds. One of the earliest studies that investigated the relationship between education spending and attainment in terms of labour market returns was conducted by Card & Krueger (1992). The study found that increased spending leads to higher returns to education in the labour market and a reduction in the pay gap between ethnicities.

While the intuition might suggest that more money should lead to better results, researchers are not uniform in their opinions. Another study by leading scholar Hanushek (2006) concludes that there is little consistent evidence to support that increased school resources lead to better student outcomes. Instead, the paper emphasizes the importance of teacher quality and other factors that are more difficult to measure but are likely to have a greater impact on student performance. Similar results were found among Finnish senior secondary schools by Häkkinen *et al.* (2003), and in higher institutions in Ohio and Tennessee by Hillman *et al.* (2017). Bénabou *et al.* (2009) examined additional resources that were channelled to disadvantaged schools in the education priority zones in France but did not find a relationship between those two.

On the contrary, according to a study by Hægeland *et al.* (2012), which uses the instrumental variable approach, higher revenues from local taxes on hydropower plants (that resulted in more school resources granted for schools in these particular regions) have a significantly positive effect on the achievement of students. Additionally, studies in US public schools by Jackson *et al.* (2015), the study of Michigan primary schools by Hyman *et al.* (1994), Texas schools by Kreisman & Steinberg (2019), and the study of US citizens who were tracked from kindergarten to early adulthood by Rothstein & Schanzenbach (2022) all found a positive relationship between funding and educational outcomes in various form of defining them. A recent example from Italy by Belmonte *et al.* (2020) found a positive relationship between school test scores and post-earthquake infrastructure spending for schools in certain regions in comparison to the school in regions that did not receive such funding.

A recent example is also a report by OECD (2019) that assesses the relationship between reading performance in PISA 2018 and spending on education in a particular country on national levels, as seen in Figure 2.1. The researchers found a relationship that increases but diminishes quickly. The curve itself may resemble Solow's growth model curve.

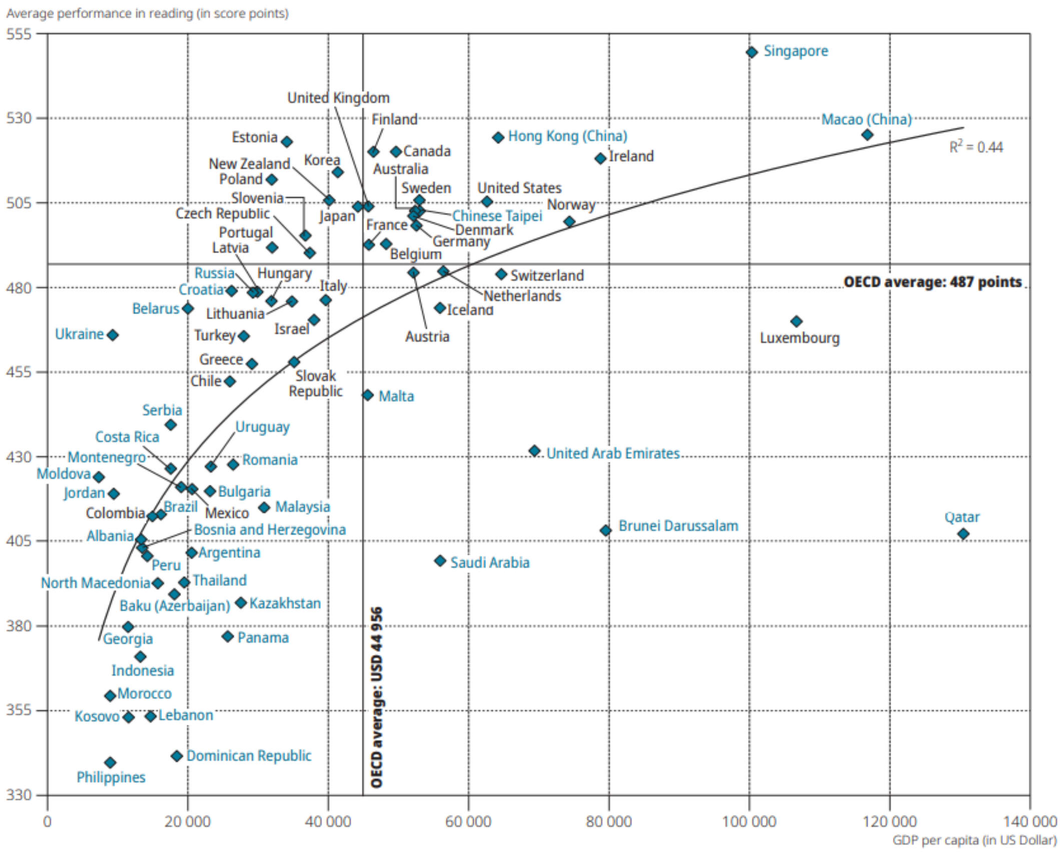


Figure 2.1: Reading performance and spending on education
Source: OECD, PISA 2018 Database, Tables I.B1.4 and B3.1.1.

It is worth noting that, to the best of the authors' knowledge, no study investigating the determinants of educational achievement in Slovakian schools has been conducted. This is a significant gap in the literature, as understanding the factors that contribute to educational achievement is crucial to implementing effective policies and interventions.

Other Determinants of Educational Achievement

It is common practice for researchers to include not only variables of interest when examining the impact of school resources on educational achievement but also additional explanatory variables. This helps the model to explain a greater proportion of the overall variability of the dependent variable, educational achievement. In this subsection, we present some commonly used explanatory variables. In this field, two main types of studies are distinguished. The first type examines the determinants of educational achievement of individuals, which is more common, while the second type examines the determinants of school performance.

In general, studies attempt to assess as many variables as possible in the dataset. Most studies use the gender and nationality of the student. Additionally, Machin *et al.* (2008) uses an interestingly detailed dataset that includes the occupation of the head of the family, as well as whether the child eats lunch at the school canteen. Häkkinen *et al.* (2003) generally include parents' education, teachers' expenditures, work during senior secondary school, or the unemployment rate in districts. Belmonte *et al.* (2020) uses school size and the share of male and native students in each school. Bénabou *et al.* (2009) goes further by adding the number of students per class, number of teachers, number of teachers per student, number of weekly hours per student, the share of young teachers, and the share of non-certified teachers. Leuven *et al.* (2007) include the socio-economic index or even the urbanization index of the area where the school is located; additionally, they include the school's religious denomination. Kreisman & Steinberg (2019) include the share of nationalities or share of the poor people living in the geographic entity to explain the individual effects of each school.

2.3 Methodology Used in the Relevant Studies

A discussion paper titled "Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide" by Schlotter *et al.* (2011) provides a comprehensive overview of the different econometric methods that can be used to evaluate education policies. The paper discusses the importance of causal inference in education research and explains how different econometric methods can help researchers address the various challenges they face. One of the paper's key contributions is its discussion of the different types of data commonly used in education research and the econometric methods best suited for each type of data. The paper also provides detailed explanations of the various econometric techniques that can be used to address common problems in education research, such as selection bias, measurement error, and endogeneity. An example of a fixed-effects approach is a study by Böhlmark & Lindahl (2008) of a voucher reform in Sweden and its effect on educational achievement. The authors found that the increase in the share of private schools in the municipality positively affected student outcomes in the short and long run. Another possibility is to use difference-in-differences models. These can identify causal effects by exploiting variations between treatment and control groups over time while controlling for time-invariant unobserved variables. For example, Machin *et al.* (2008) examine the introduction of the "literacy hour" in the United Kingdom in the 1990s. The authors found a positive effect on reading and English test results and concluded that the benefits of the literacy hour exceeded the policy's cost.

Chapter 3

European Cohesion Policy and Structural Funds

The purpose of this chapter is to examine the details and characteristics of the Cohesion Policy and Structural Funds of the European Union throughout history, as well as provide evidence of the necessity for such a policy.

Additionally, we provide a regional context relevant to Slovakia and lastly, we explain the types of education-related financed projects in Slovakia in the programming period we are going to examine and how we matched them with individual school recipients.

Necessity of Cohesion in the European Union

The fundamental economic concept behind providing additional funds to underdeveloped regions of the EU is to allow them to catch up with more developed areas through external input - financial resources. This should make convergence easier and faster. The first descendants to today's Structural Funds were established in the 1980s and have been revised several times since. The current programming period, which cover the period from 2021 to 2027, emphasize the "European Green Deal" and "digitalization" as key investment areas. This is a response to the continuing disparities in many economic indicators that exist today, such as income, employment, and education levels, which have negative effects on social cohesion and economic stability.

The Figure 3.1 below illustrates the extent of these disparities between European Union members and individual states of the United States of America.

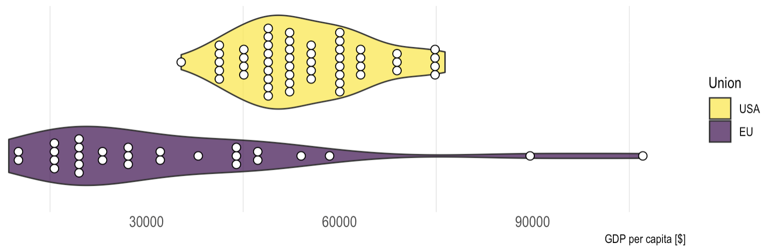


Figure 3.1: Distribution of GDP p. c. in USA and EU - 2021.
Source: Bureau of Economic Analysis and Oxford Economics.

3.1 Slovakia and EU Funds

Slovakia, as a member state of the EU since 2004, has been a recipient of European Cohesion Policy and Structural Funds. EU funds have played a crucial role in the country's development, supporting infrastructure projects, fostering innovation, promoting education and training, and enhancing environmental protection and resource efficiency.

Some examples of projects supported by the EU Funds in Slovakia during this time period include:

1. **Transport infrastructure:** Funding was provided to improve the quality and connectivity of the transport network in Slovakia, including investments in roads, rail, and public transport.

2. **Environmental protection:** Projects aimed at improving the quality of the environment in Slovakia, including investments in wastewater treatment, waste management, and air quality.

3. **Education and training:** Projects aimed at improving the quality of education and training in Slovakia, including investments in school buildings and equipment and vocational training programs.

4. **Social inclusion:** Projects to promote social inclusion and reduce poverty in Slovakia, including investments in housing, social services, and support for disadvantaged groups such as the elderly and people with disabilities.

In the programming period from 2007 to 2013, Slovakia received over 665 million euros in EU funds that were channelled into schools, according to our calculations. The extent of this can be seen on the map in the Figure 3.2.

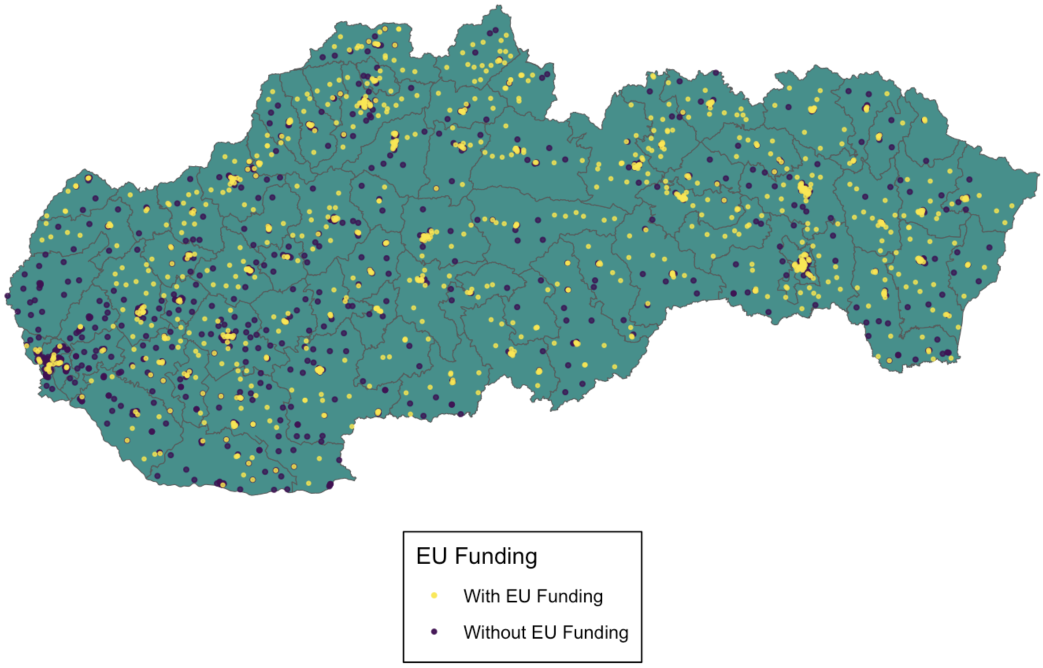


Figure 3.2: Schools granted and not granted European Funds in the programming period 2007-2013
Source: Ministry of Investment, Regional Development and Informatization of the SR

Methodology of matching EU Funds to respective schools

The uniqueness of our data set is remarkable because of the intensive manual labour that went into matching all the observations. The initial data provided by the Ministry of Investment, Regional Development and Informatization of the SR was not in a form that could be easily matched using identification codes or geocoding due to the frequent lack of clarity around the funding recipients. Despite the author's great determination, five instances of individual funding had to be deleted from the data set due to a lack of information about the recipient. A lot of effort went into compiling this data set, which could make it an interesting resource for researchers in the future.

We first filtered out all relevant variables in three types of objectives:

1. **Operational Program Education,**
2. **Regional Operational Program,**
3. **Bratislava Operational Program.**

We then subdivided the funding into two categories:

1. **Education-related funding** was allocated for educational purposes, such as specialized training for teachers, new educational resources for students, and the development of innovative teaching methodologies that foster a more dynamic learning environment. This funding was aimed at supporting the educational process and improving the quality of education for students.

2. **Reconstruction-related funding** was allocated for the reconstruction of the school building, including repairs and renovations and other infrastructure related work. This funding was not directly connected to the educational process but was aimed at ensuring that the school infrastructure was in good condition, providing a safe and conducive environment for students to learn in. This could also involve the installation of new facilities such as playgrounds, sports fields, and other recreational facilities.

We did so to be able to distinguish between effects of both, when analysing by regression models as one may have more significant effect than other. In the Figure 3.3, we can see the amount of these funds absorbed over the span of programming period.

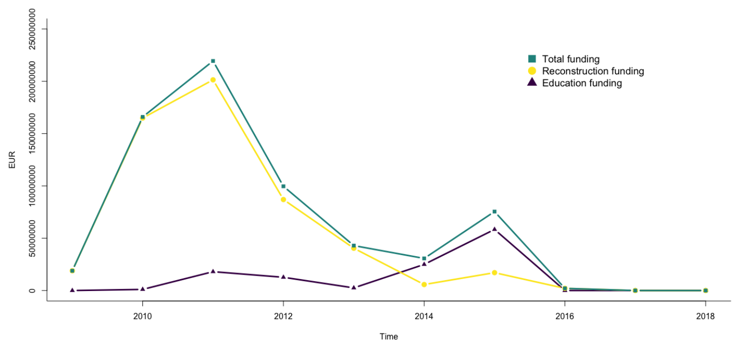


Figure 3.3: Absorption of the EU funds over time
Source: Ministry of Investment, Regional Development and Informatization of the SR

Chapter 4

Data

In this chapter, we introduce the data set and its unique characteristics. The dataset includes 1409 primary schools and 657 secondary schools from Slovakia. We only include those schools that are regular and not for pupils with special needs to prevent bias. We also only include primary schools that take part in National Institute for Education and Youth (colloquially called NIVAM) tests (primary schools with only four grades do not participate in the test, therefore, are not included in this dataset). Secondary schools were only included if there were no or a small number of missing values in the panel data (i. e. we do not include the secondary school that was established later or the secondary school was closed).

We first introduce the dependent variable – results and percentiles of individual schools from standardised NIVAM tests (Testing 9, Maturita), which every pupil in each school has to take (were the pupil not excused because of serious reasons such as long-term inability to attend a school or special condition preventing from attending). These were acquired from National Institute of Education and Youth.

We then introduce the independent variables. We discuss the descriptive characteristics of European Funding variables, which were acquired from Ministry of Investment, Regional Development and Informatization of the SR. We also provide specifications for each individual school (type, establisher, number of students, number of teachers, number of students from socially disadvantaged backgrounds...), which were acquired from Institute for Economic and Social Reforms. Regarding town-specific data, we use the percentage of the Roma population (acquired from Ministry of Interior of the Slovak Republic), and the percentage of people with higher education. Regarding district-specific data, we use the unemployment rate. Both (Roma population excluded) were acquired from Statistical Office of the Slovak Republic.

Shapefiles of Slovakia that are used in the map figures were acquired from Geodetic and Cartographic Office of Bratislava.

4.1 Dependent variable

We use the results of NIVAM tests as the dependent variable. These are norm-referenced national tests in Slovakia that are very similar to PISA tests. Their main objective is to assess the strategic competencies of students. Thus, these tests provide a good representation of educational achievement between individual schools in Slovakia.

There are three nationwide tests NIVAM is responsible for administrating:

- Testing 5 (nationwide from the school year 2015/2016) – mandatory tests from Mathematics and the Slovak language (alternatively Hungarian) of primary school pupils in 5th grade
- Testing 9 (nationwide from the school year 2004/2005) – mandatory tests from Mathematics and the Slovak language (alternatively Hungarian) of primary school pupils in 9th grade
- Maturita (nationwide from the school year 2006/2007) – mandatory test from the Slovak language (alternatively Hungarian) and optional (sometimes obligatory) tests from additional languages (English, German, Spanish or French) and Mathematics of secondary school pupils in the school-leaving grade

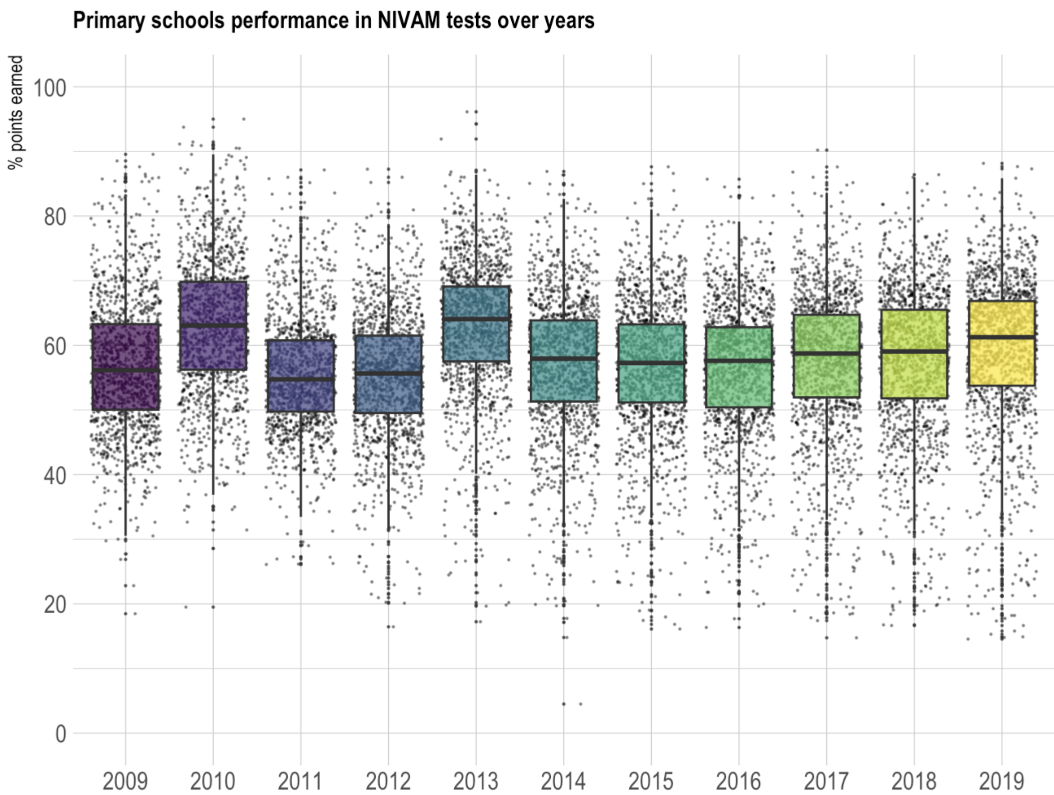
For the purpose of this bachelor's thesis, we created a special variable to measure the educational achievement of schools. For primary schools, we calculated the average scores of Testing 9 in Mathematics and Slovak (or Hungarian) language. For secondary schools, we calculated the average scores of Maturita in Slovak (or Hungarian) language and English language. Table 4.1 shows descriptive statistics and graphical representations for both primary and secondary schools.

Table 4.1: NIVAM test score results: descriptive statistics

	Primary schools						*Secondary schools*					
Year	**N**	**Min**	**Med.**	**Mean**	**Max**	σ	**N**	**Min**	**Med.**	**Mean**	**Max**	σ
2009	1376	18.5	56.1	56.8	89.5	10.2	634	27.2	57.4	56.8	86.4	9.55
2010	1374	19.5	63.1	63.0	95.0	10.5	657	27.3	56.4	57.3	88.1	9.32
2011	1377	26.1	54.7	55.6	87.1	9.0	668	28.6	57.7	58.0	87.9	10.2
2012	1374	16.4	55.6	55.3	87.2	9.9	650	30.0	53.8	54.5	81.4	9.94
2013	1396	17.2	64.1	62.9	96.1	10.2	662	30.6	55.8	56.2	85.2	9.82
2014	1400	4.5	57.9	57.1	86.9	10.7	657	15.6	56.2	56.5	89.5	11.0
2015	1402	16.1	57.3	56.6	87.6	10.4	651	19.7	46.6	46.9	83.7	11.3
2016	1400	16.4	57.6	56.5	85.7	10.3	657	23.2	49.7	50.6	81.1	11.3
2017	1400	14.8	58.7	57.5	90.2	11.2	637	27.2	51.6	51.8	85.5	9.80
2018	1409	16.6	59.0	57.8	86.4	11.3	654	26.6	50.3	51.7	81.3	10.9
2019	1409	14.6	61.2	59.3	88.2	11.8	656	21.9	46.7	49.0	83.8	12.5

Source: National Institute of Education and Youth

The Figure 4.1 demonstrates that NIVAM tests are a useful tool for differentiating between underperforming and overperforming schools. These tests have a mostly consistent median and rank schools by the average scores they get. This is a crucial aspect of the referenced tests as it helps identify areas of improvement. In addition, the test results can also be used to allocate resources and funding to schools that need it the most.



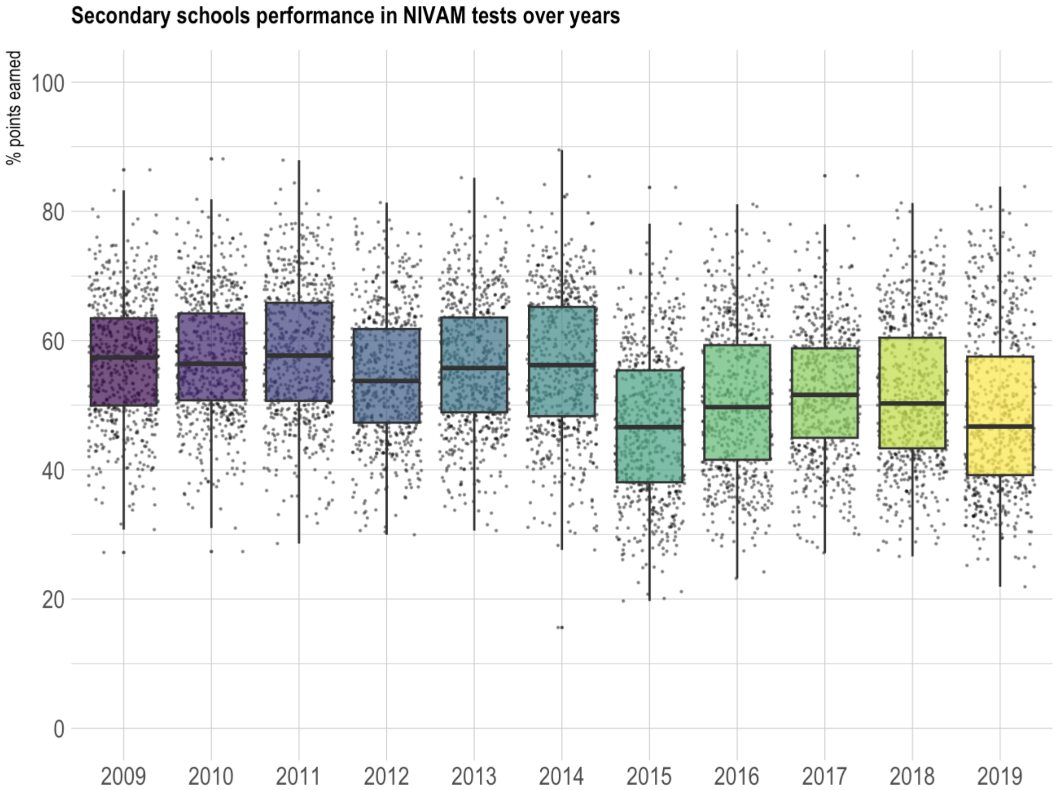


Figure 4.1: NIVAM test results between years 2009-2019
Source: National Institute of Education and Youth

While it is true that NIVAM tests are a great tool to compare schools' educational achievement, it is important to note that they are not without their limitations. One area of concern is the level of accuracy and stability of the results. To aid research, it would be beneficial to have a stable test with consistent difficulty levels. However, even within our data sample, we can see that this is not always the case. Results for individual schools can show significant fluctuations and do not follow stable trends over the years. For example, in 2013, the Figure 4.1 depicts a significant increase in test scores for primary schools. We found that this might have been due to an easier test resulting from changes in the educational curriculum in schools. That year, mathematical tests did not include questions on combinatorics, statistics, logic, and proofs, which historically were the more challenging questions. This was the main motivation for averaging individual observations in languages and mathematics, which achieved smaller year-to-year fluctuations and, consequently, smaller standard deviations.

4.2 Independent variables

4.2.1 School characteristics

As mentioned in the introduction of this chapter, we have collected various details about each school in our data set. We have subdivided them into two categories: those with numerical values and dummy variables.

Numerical variables

We include school characteristics that were available, similar to those in Belmonte *et al.* (2020) and Bénabou *et al.* (2009). These characteristics include the number of students, the number of pupils from a disadvantaged background, the number of teachers for each school, and the frequency of using Information and Communication Technologies (ICT) (Information and Communications Technology). We obtained these characteristics for each individual year, and the purpose is to determine if and how these variables affect educational achievement. For example, we want to investigate whether bigger schools are associated with worse or better results, whether ICT improves test scores over time, and if pupils from disadvantaged backgrounds influence schools' test results. However, Slovak institutions only started collecting some interesting data points - *frequency of using interactive technology* and *number of pupils from a disadvantaged background* in 2014, which will subsequently impact our possibilities with the regression model. We provide descriptive statistics in the Table 4.2.

Table 4.2: Numerical school characteristics: descriptive statistics

	Primary schools				*Secondary schools*			
Year	**Min**	**Med.**	**Mean**	**Max**	**Min**	**Med.**	**Mean**	**Max**
Number of teachers per 100 students	4.3	8.2	8.9	50.8	5.9	10.5	12.4	100.0
Number of pupils	25.7	238.3	294.5	1155.0	1.5	300.2	329.4	1144.4
Frequency of using interactive tech.	21.7	72.5	71.9	100.0	4.9	64.7	64.4	100.0
Number of pupils from disadvantaged b.	0.0	3.6	16.3	609.3	0.0	0.0	1.3	204.0

Source: Institute for Economic and Social Reforms

Dummy variables

Similar to Leuven *et al.* (2007), we further extracted descriptive information about the type of establisher, location and language of instruction and converted them into dummy variables. These are all relevant independent variables to add, as some types of students with specific characteristics will choose some schools or specific establisher will improve the results of school thanks to more funding. Descriptive statistics may be found in the Table 4.3.

Table 4.3: Dummy school characteristics: descriptive characteristics

	Primary schools	*Secondary schools*
Type of establisher	**N**	**N**
Public	1280	472
Private	29	108
Church	94	68
Location	**N**	**N**
Bratislava	76	80
District capital	432	466
Regular town	895	102
Language of instruction	**N**	**N**
Slovak	1264	500
Hungarian	121	28
Ukrainian	1	1
Bilingual	17	119

Source: Institute for Economic and Social Reforms

4.2.2 Location characteristics

Similarly to Kreisman & Steinberg (2019), we include variables such as nationality or education level of a particular location where the school is located. In the Table 4.4 we state descriptive statistics.

Table 4.4: Location characteristics: descriptive statistics

Variable	**Min**	**Median**	**Mean**	**Max**
Number of residents	0	654	1849	105468
Share of Roma population [%]	4.3	8.2	8.9	50.8
Share of citizens over 15 y.o. with university degree [%]	0.0	3.6	16.3	609.3
Unemployment [%]	3.6	10.0	11.6	28.3

Source: Statistical Office of the Slovak Republic

The **number of residents** in a particular town is a crucial piece of information that can have significant implications for both the between and Difference-in-Difference (DID) models. In terms of the between model, the number of residents is a useful indicator of the level of human capital available to the schools, as well as the level of attractiveness of the town to new teachers. The DID model, on the other hand, can benefit from changes over time, as these changes can provide valuable insights into the trends of migration and the overall desirability of the area. For instance, a decrease in the number of residents may suggest that the area is losing its appeal, whereas an increase may indicate that it is becoming more attractive to potential residents. Therefore, it is important to keep track of changes in the number of residents over time to understand better the dynamics of the town and the potential implications for its schools. We provide overview of more populated areas of Slovakia in the Figure 4.2.

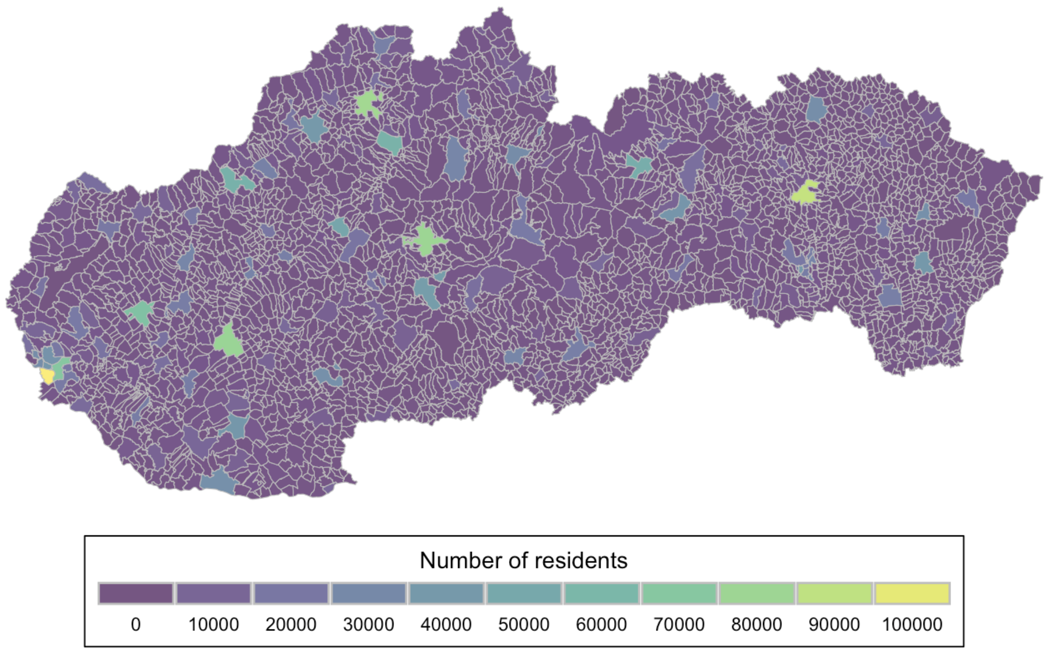


Figure 4.2: Number of residents on the town level
Source: Statistical Office of the Slovak Republic

We also include **Roma population share**, who, as a disadvantaged minority, face challenges when it comes to integrating into the education system. Many districts in Slovakia are known for having high repeat rates in the first grade of primary school. Despite various policies and programs aimed at promoting their inclusion, many Roma children still struggle with learning the language of instruction, accessing quality education, and balancing school with other responsibilities at home. This is compounded by the fact that many Roma families are themselves disadvantaged and may not have the resources or knowledge to support their children in their education. Therefore, it is not surprising that Roma children often perform poorly in school, which can lead to a cycle of low achievement and repetition. To account for this issue, we included Roma ethnicity as a control variable in our analysis. This can also be seen in the Figure 4.3.

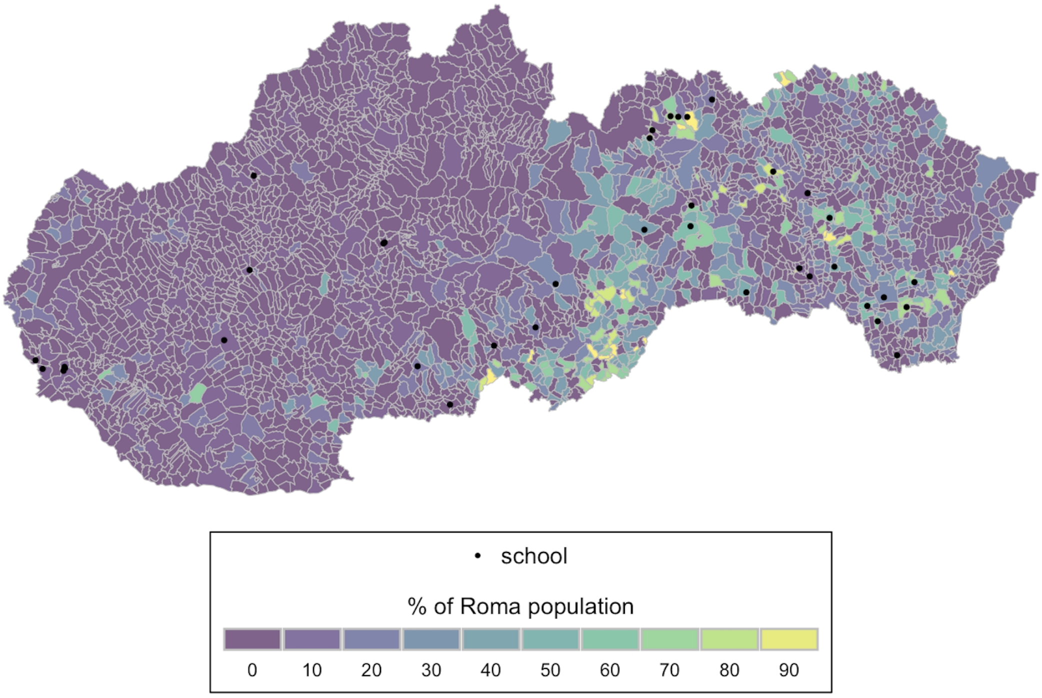


Figure 4.3: Under-performing primary and secondary schools with respect to the share of the Roma population
Source: Ministry of Interior of the Slovak Republic

Studies have shown that a student's educational level in Slovakia is often determined by the **educational level** of their parents. This is a phenomenon that has been observed in many countries around the world. While some argue that this is due to the lack of opportunities for upward mobility, others argue that it is a result of systemic inequalities in the education system. To address this issue, we are including a ratio of "over-educated residents" in a particular town. The relationship can be seen in the Figure 4.4.

After consideration, we decided to include the **unemployment rate** as an important factor to be taken into account. This is because social background plays a crucial role in determining educational achievement. By measuring the unemployment rate, we aim to capture the socioeconomic status of the environment of schools under study, which is highly relevant for understanding educational outcomes. This can also be seen in the Figure 4.5.

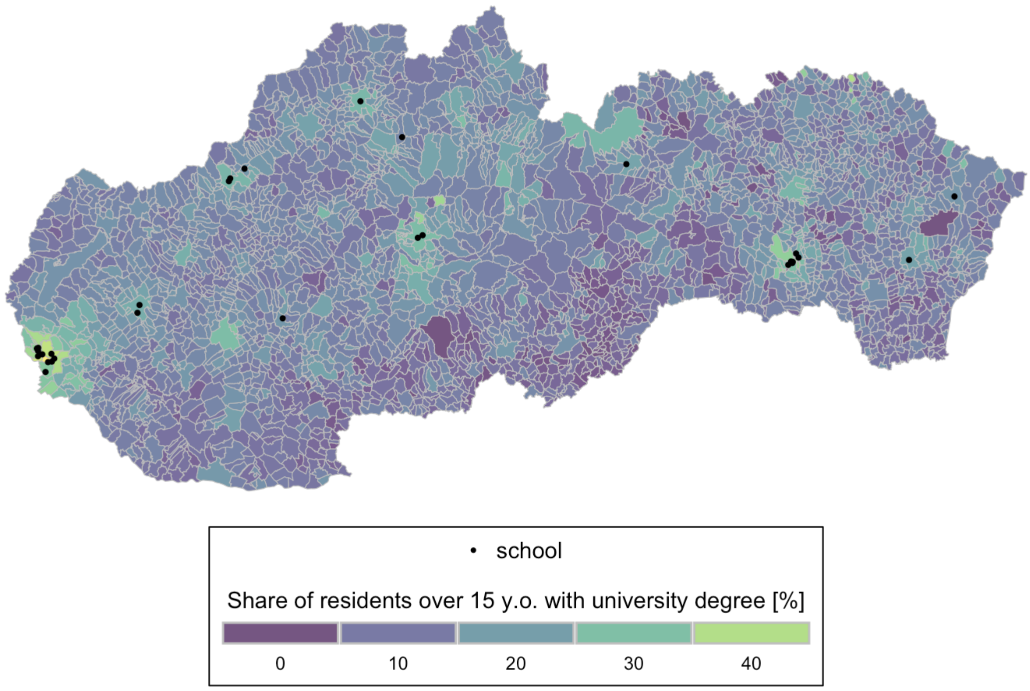


Figure 4.4: Over-performing primary and secondary schools with respect to share of residents with university degree
Source: Statistical Office of the Slovak Republic

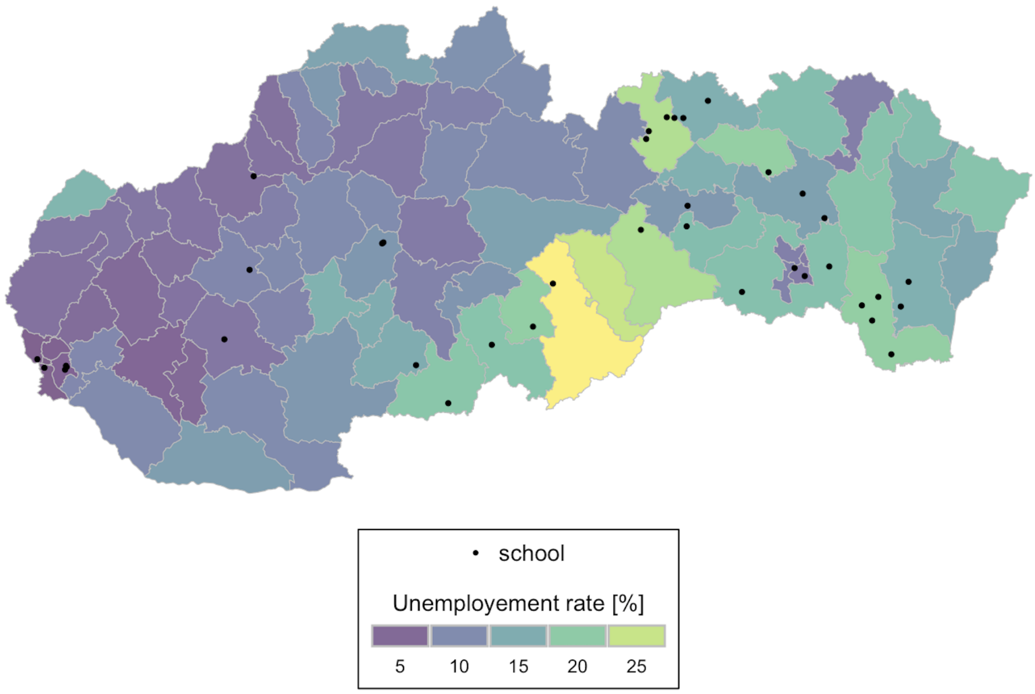


Figure 4.5: Under-performing primary and secondary schools with respect to the unemployment rate
Source: Statistical Office of the Slovak Republic

4.2.3 European Funding

Last, but not least, the European Funding variable is of main interest to us, as the whole objective of this study is to determine its significance and, particularly, impact on the educational achievement of affected schools. We describe the variable itself in the Chapter 3. In Table 4.5, we provide descriptive statistics about EU funding. Based on the data, approximately half of the projects were allocated to education-related requests and the other half to reconstruction-related requests. However, on average, reconstruction-related projects require significantly more funding.

Table 4.5: European funding: descriptive statistics

Variable	**N**	**Min**	**Med.**	**Mean**	**Max**
Reconstruction funding [th EUR]	647	201	729	835	2 960
Education funding [th EUR]	584	0.47	154	199	1 203

Source: Ministry of Investment, Regional Development and Informatization of the SR

Chapter 5

Methodology

In this chapter, we will provide a detailed account of the methods used to analyze our data set. As discussed in the previous chapter, the data set we used contained complex panel data about every primary and secondary school in the Slovak Republic, which we matched with the European funding the school received.

To begin our analysis, we implemented the Fixed Effect Model and the Random Effects Model. As the model diagnostics clearly preferred the Fixed Effect Model, we proceeded with that one. These models are a straightforward choice when dealing with panel data, as they account for the individual heterogeneity of each observation. However, we encountered some limitations with this model, which led us to explore other options.

Next, we applied Between Estimator Model, which is an extension of the Fixed Effect Model. This model averages out the time component, resulting in an equation that regresses the average scores of the schools on averages of other variables using a standard Ordinary Least Squares (OLS) estimator. Although this model provided us with more satisfactory results, we must acknowledge that it has its limitations. For instance, it does not account for changes in time and cannot estimate the direct impact of European funding on the educational achievement of students in a particular school.

Therefore, we introduced the Difference-in-Difference model to investigate the effects over time further. This model allowed us to examine how the treatment group (i.e., schools that received European funding) differed from the control group (i.e., schools that did not receive European funding) over time. By comparing the differences in changes between the two groups, we were able to isolate the effects of the treatment and estimate its impact more accurately.

Overall, our analysis has provided us with valuable insights into the relationship between European funding and educational achievement in the Slovak Republic.

5.1 Fixed effect model

The Fixed Effects Model is a method for analyzing panel data (longitudinal data) in order to control for unobserved individual heterogeneity. This method is particularly useful when dealing with unobservable variables that are constant over time but may vary across individuals, such as innate ability, motivation, or preferences (Wooldridge 2015, chap. 14).

In our case, the model should, in theory, control for each school's regional unobserved differences and other specific characteristics. For example, suppose we are interested in the effect of class size on test scores. If we do not control for individual heterogeneity, the estimated effect of class size may be biased due to unobserved differences across students, such as where the school is located. However, we can remove these biases and obtain more reliable estimates by using the Fixed Effects Model. Consider a panel data regression model that contains i schools (or cross-sectional units) and t time periods. The model can be written as:

$$y_{it} = \beta_0 + \beta_k \mathbf{X}_{it} + a_i + u_{it} \tag{5.1}$$

where y_{it} is the dependent variable for individual i at time t, $\mathbf{X}_{it}$ is a vector of k independent variables, a_i is the unobserved individual-specific effect, and u_{it} is the idiosyncratic error term.

To eliminate the unobserved individual-specific effects (a_i), we can use the within-individual (time-demeaned) transformation. For each variable, subtract the individual's mean over time:

$$y_{it} - \overline{y}_{it} = \underbrace{\beta_0 - \overline{\beta}_0}_{0} + \beta_k(\mathbf{X}_{it} - \overline{\mathbf{X}}_{it}) + \underbrace{a_i - \overline{a}_i}_{0} + u_{it} - \overline{u}_{it}. \tag{5.2}$$

Here, $\overline{y}_{it}$, $\overline{\mathbf{X}}_{it}$, and $\overline{u}_{it}$ represents the means of y_{it}, $\mathbf{X}_{it}$, and u_{it}, respectively, for individual i over the t time periods. To simplify,

$$\ddot{y}_{it} = \beta_k \ddot{\mathbf{X}}_{it} + \ddot{u}_{it} \tag{5.3}$$

where $\ddot{y}_{it}, \ddot{\mathbf{X}}_{it}$ and $\ddot{u}_{it}$ stands for transformed variables that were created as a subtraction from the original value minus the mean over all the time periods t, as shown in the Equation 5.2.

After transforming the data, we estimate the parameters using Ordinary Least Squares (OLS). The resulting estimates, β_k, will be consistent and unbiased, assuming that the independent variables $\mathbf{X_{it}}$ are uncorrelated with the time-varying error term u_{it}.

As fully listed in the Appendix A, where we provide a full list of assumptions as stated in Wooldridge (2015), we have a list of assumptions for a fixed effects model, including assumptions about the model structure, the sample, and the error terms. We will test those in the Chapter 6. These assumptions include the expected value of the idiosyncratic error being zero, the explanatory variables changing over time, and the idiosyncratic errors being uncorrelated, independently, and identically distributed, while its variance is equal to σ_u^2.

5.2 Between Estimator Model

The between estimator, also known as the "between-groups" or "cross-sectional" estimator, is a method for analysing panel data using a cross-sectional variation. The between estimator focuses on the variation between the schools rather than within each school over time. We decided to include this model so that we can show what the general pre-determinants of educational achievement are.

This estimator is not that widely used, so we fail to provide a source other than by STATA, specification for a package that is used for the analysis of panel data. The absence of a source can be explained easily, as the between-estimator model is rather a straightforward model that averages the time component, as will be shown in Equation 5.5 and uses OLS estimation.

We can obtain the regression model by averaging variables over time:

$$\overline{y}_i = \overline{\beta}_0 + \beta_k \overline{\mathbf{X}}_i + a_i + \overline{u}_i \tag{5.4}$$

where $\overline{y}_i = \frac{1}{T}\sum_t y_{it}$, $k = 1, \ldots, K$, where K is the number of explanatory variables, $\overline{\mathbf{X}}_i = \frac{1}{T}\sum_t \mathbf{X}_{it}$ and $\overline{u}_i = \frac{1}{T}\sum_t u_{it}$. The coefficients will then be estimated with the OLS estimator.

As the model does not deal with the unobservable individual-specific effect, we will include additional explanatory variables that do not change or do not change in the time period observed, and we have them only as a single observation. The final equation will therefore be:

$$\overline{y}_i = \overline{\beta}_0 + \beta_k \overline{\mathbf{X}}_i + \delta_j \mathbf{Z}_i + \overline{\epsilon}_i, \tag{5.5}$$

where $\overline{y}_i = \frac{1}{T}\sum_t y_{it}, k = 1, \ldots, K$, where K is the number of time-changing explanatory variables, $\overline{\mathbf{X}}_i = \frac{1}{T}\sum_t \mathbf{X}_{it}$, $j = 1, \ldots, J$, J is the number of time-invariant explanatory variables, $\mathbf{Z}_i$ is a vector of additional time-invariant explanatory variables and $\overline{\epsilon}_i = a_i + \frac{1}{T}\sum_t u_{it}$.

For the ordinary least squares estimator to be unbiased, several assumptions must hold; we provide them listed in Appendix A. To summarize, the assumptions that must hold for the ordinary least squares estimator to be unbiased including linearity, independence, homoskedasticity, no perfect multicollinearity, and a zero expectation for the error term. If these assumptions are met, the OLS estimator will provide unbiased estimates of the true population parameters. We will provide details about the tests conducted in Chapter 6.

5.3 Difference-in-difference model

As discussed at the beginning of this chapter, we acknowledge the limitations of the between-estimator model; specifically, it does not account for time effects. On the other hand, the Fixed Effect Model does not provide a satisfactory explanation of the variation in the dependent variable due to the limitations and fluctuations of the NIVAM tests, as illustrated in the Figure 4.1.

In theory, we want to see whether the treatment group experienced some effect that the control group did not. We will be looking at the existence or non-existence of post-treatment effects difference, as can be seen in 5.1.

We define the difference-in-difference (DID) model (Wooldridge 2001, chap. 6) as

$$y_i = \delta_0 + \delta_1 post_i + \delta_2 treatment_i + \delta_3 post_i \times treatment_i + \beta_k \mathbf{X}_i + u_i, \tag{5.6}$$

where y_i is the dependent variable, $post_i$ is the period dummy, which takes the amount of 1 when the observation is from the after-treatment period, $treatment_i$ is the treatment dummy, which takes the amount of 1 if the i-th observation was treated regardless of the period. $\mathbf{X}_i$ are additional explanatory variables that have an effect on the overall value of the dependent variable; we have to include those because, as mentioned in Wooldridge (2001), these account for the systematic different characteristics there are between individual observations. To fully assess whether the treatment group experienced some effect on the educational achievement of their students, we will be looking at $\hat{\delta}_3$ and its significance. This is visually shown in the Figure 5.1.

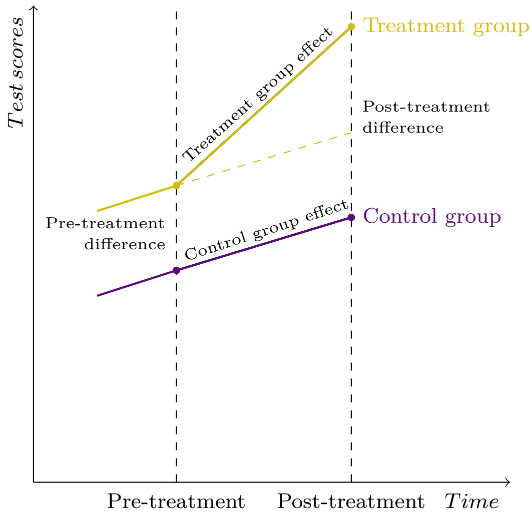


Figure 5.1: Graphical representation of DID model

We will be using the OLS estimator to estimate the equation; therefore, traditional assumptions must hold, as stated in 5.2. Nevertheless, the unbiasedness of the DID estimator requires that the treatment cannot be related to other $\mathbf{X}_i$ explanatory variables. In addition to that, we assume that the trends are the same for both control and treatment groups. We delve further into these assumptions in Chapter 6.

Chapter 6

Results

In this chapter, we will present the results of the analysis.

In the first part of this chapter, we provide an in-depth explanation of the results obtained from the fixed effect model, which is a widely used method in panel data econometrics. This model is known for its ability to account for unobserved heterogeneity among individuals or groups, thus improving the accuracy of the analysis. However, it did not work well with our data. Therefore, we will discuss the limitations and identify what the issues are.

In the second part of the chapter, we present the results of the between-estimator model analysis. This model is commonly used when the focus is on the differences between groups rather than within-group differences. We discuss the findings in detail, highlighting the strengths and limitations of the between-estimator model.

Lastly, we conclude the chapter with a Difference-in-Difference model analysis, which is a powerful tool for evaluating treatment effects. We look into the details of the analysis, discussing the assumptions and requirements for this model to be valid. We then present the results and provide a comprehensive discussion of their implications.

6.1 Fixed Effects Results

To measure the impact of EU funding on the educational achievement of the schools in time, we start our analysis by estimating a panel data model as stated below. We estimate the model individually for primary and secondary schools for data from 2014 on[1]. We provide results for two types of the $eurofonds_{it}^{educ}$ and $eurofonds_{it}^{rec}$ variables − simple lags and continuous lags. The reason for this is that we are uncertain about how these variables will affect $educresult_{it}$ exactly.

The simple lag model includes eurofunds as a single observation in the year when the school finished the project. The continuous lag model takes into account that the money invested should have an impact beyond just the year of use. Therefore, every year after the year when school finished, the project is marked with the amount of money invested. We estimate the regression model separately for primary school and for secondary school with the regression model stated in 6.1. We include 3 lags as we believe that it takes some time for the investment to influence schools' test scores fully. We also ran a regression with more than 3 lags, the results of that regression, which proved inefficient, can be found Appendix C.

$$\begin{aligned} educresult_{it} = &\beta_0 + \beta_1 pupils_{it} + \beta_2 pupilsSDB_{it} + \beta_3 ICT_{it} \\ &+ \beta_4 teachers_{it} + \beta_5 population_{it} + \beta_6 unem_{it} \\ &+ \beta_7 eurofonds_{it}^{educ} + \beta_8 eurofonds_{it+1}^{educ} \\ &+ \beta_9 eurofonds_{it+2}^{educ} + \beta_{10} eurofonds_{it+3}^{educ} \\ &+ \beta_{15} eurofonds^{rec} + \beta_{16} eurofonds_{it+1}^{rec} \\ &+ \beta_{17} eurofonds_{it+2}^{rec} + \beta_{18} eurofonds_{it+3}^{rec} + a_i + u_{it} \end{aligned} \tag{6.1}$$

First, we need to determine whether the model we selected is well-suited for the type of data we want to estimate. To do so, we run two tests − the F-test and the Hausman test - for all four models and provide the results in Table 6.1. For all four models, the Fixed Effects (FE) model is superior to the Pooled OLS model, as evidenced by the high p-value. The same can be said of the Hausman test, which also found that the FE model is superior to the Random Effects (RE) model with high p-values for all the models.

[1]This is mainly due to the fact that most of the important explanatory variables are available from the year 2014 on.

Table 6.1: Fixed Effects: diagnostics for choice of the model

	Primary schools:			*Secondary schools:*		
Test	Stat.	df	p-value	Stat.	df	p-value
Simple lags						
Pooled OLS vs FE:						
F test	7.1	6853 (1425)	2.2×10^{-16}	12.7	2406 (665)	2.2×10^{-16}
RE vs FE:						
Hausman test	730.1	8	2.2×10^{-16}	48.3	8	8.5×10^{-8}
Continuous lags						
Pooled OLS vs FE:						
F test	7.1	6853 (1425)	2.2×10^{-16}	12.7	2406 (665)	2.2×10^{-16}
RE vs FE:						
Hausman test	764.5	8	2.2×10^{-16}	102.5	8	2.2×10^{-16}

We test the assumptions of the FE estimator, as stated in Appendix A. We estimated the model using a dataset consisting of all primary and secondary schools in Slovakia. Therefore, the assumption of a random sample, as outlined in FE.2, holds. Next, we tested for the presence of multicollinearity in our dataset, as stated in FE.3. To do so, we used the "vif" package in R and found no significant correlation between explanatory variables, with the exception of the understandable correlation between lagged variables $eurofonds_{it}^{educ}$ and $eurofonds_{it}^{reconstr}$ in models with continuous lags. Since our primary interest was to estimate the overall effect of eurofunds on the dependent variable, multicollinearity may not be a significant issue as long as the model can produce accurate predictions. We keep this in mind when interpreting the results.

The original assumption FE.4 only holds when strict exogeneity holds. However, our model includes lagged variables, which automatically violates the strict exogeneity assumption. Therefore, only weak exogeneity may hold. Thus, we can achieve asymptotic unbiasedness of our estimator by having a sufficiently large sample size. Fortunately, our dataset includes 1425 primary schools and 665 secondary schools so this assumption will be satisfied.

In Table 6.2, we present the results of tests for the assumptions FE.5 and FE.6 regarding heteroskedasticity and serial correlation. We use the Breusch-Godfrey test for serial correlation and the Studentized Breusch-Pagan test for heteroskedasticity, and as can be seen, both are present. To address these issues, we use the "Arellano" robust covariance matrix specification method. Adjusted standard errors are, therefore, used in the Table 6.3 to account for the presence of serial correlation and heteroskedasticity.

Table 6.2: Fixed Effects: model diagnostics

	Primary schools:			Secondary schools:		
Test	Stat.	df	p-value	Stat.	df	p-value
Simple lags						
Breusch-Godfrey	1044.00	8	2.2×10^{-16}	586.28	8	2.2×10^{-16}
S. Breusch-Pagan	913.61	14	2.2×10^{-16}	64.72	14	1.7×10^{-8}
Continuous lags						
Breusch-Godfrey	1044.40	8	2.2×10^{-16}	590.64	8	2.2×10^{-16}
S. Breusch-Pagan	924.89	14	2.2×10^{-16}	67.63	15	5.2×10^{-9}

In conclusion, under the assumptions FE.1-4, the fixed effects estimator would be unbiased and consistent. We use robust estimation to deal with heteroskedasticity and autocorrelation; therefore, the estimator is not efficient. The estimator will not be unbiased as it includes lagged variables, which violate strict exogeneity, and only contemporaneous exogeneity can hold. Our estimator, therefore, given our large sample size, will be consistent but not Best Linear Unbiased Estimator (BLUE), as FE.5 and FE.6 are violated.

Table 6.3: Fixed Effects: Results

	Dependent variable: $educresult_i$			
	Primary schools:		*Secondary schools:*	
	(1) Simple lags	(2) Cont. lags	(3) Simple lags	(4) Cont. lags
$pupils_{it}$	0.016*** (0.004)	0.017*** (0.004)	−0.0001 (0.004)	0.0003 (0.004)
$pupilsSDB_{it}$	−0.030** (0.015)	−0.030** (0.015)	0.084 (0.049)	0.078 (0.057)
ICT_{it}	0.023*** (0.005)	0.021*** (0.005)	−0.024*** (0.007)	−0.022*** (0.007)
$teachers_{it}$	0.113 (0.110)	0.110 (0.110)	0.013 (0.019)	0.014 (0.019)
$population_{it}$	−1.01** (0.314)	−1.02** (0.311)	0.360 (0.427)	0.303 (0.426)
$unem_{it}$	−0.018 (0.039)	−0.027 (0.037)	−0.598*** (0.074)	−0.587*** (0.074)
$eurofonds^{educ}_{it}$	−0.003 (0.003)	−0.009 (0.005)	0.002 (0.008)	−0.601 (0.091)
$eurofonds^{educ}_{it+1}$	−0.007 (0.003)	0.003 (0.005)	0.008*** (0.008)	0.006*** (0.002)
$eurofonds^{educ}_{it+2}$	−0.049* (0.003)	−0.004 (0.007)	0.006*** (0.008)	−0.001 (0.002)
$eurofonds^{educ}_{it+3}$	−0.038 (0.002)	0.002 (0.003)	0.015** (0.008)	−0.002 (0.001)
$eurofonds^{rec}_{it}$	−0.002 (0.001)	−0.002 (0.001)	−0.001 (0.004)	0.005 (0.004)
$eurofonds^{rec}_{it+1}$	−0.0004 (0.001)	−0.005 (0.007)	−0.001 (0.003)	0.004 (0.006)
$eurofonds^{rec}_{it+2}$	0.005 (0.001)	−0.002 (0.008)	−0.001 (0.002)	−0.006 (0.007)
$eurofonds^{rec}_{it+3}$	−0.0002 (0.0003)	0.003 (0.005)	−0.001 (0.002)	0.002 (0.004)
Observations	8,301	8,301	3,093	3,093
R^2	0.014	0.014	0.063	0.063
Adjusted R^2	−0.193	−0.193	−0.205	−0.204
F Statistic	6.934*** (df = 14; 6861)	7.004*** (df = 14; 6861)	7.290*** (df = 14; 2413)	7.390*** (df = 14; 2413)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6.3 presents the results of the four models we estimated using fixed effects estimation. These models consider only observations from 2014 on. The two models for primary and secondary schools differ in the type of effect we assume the $eurofunds$ variable has on the dependent variable, as explained at the beginning of this section.

As shown by the results of the F test, all of the models are statistically significant at a 1% significance level. However, the models only account for 1.4% of the overall variability for primary schools and 6.3% of the overall variability for secondary schools in the $educresult$ variable. The remaining variability can be attributed to unknown and difficult-to-track effects, such as changes in management or simply cases of randomness, where some grades have better or worse-performing pupils. To ensure accurate hypothesis testing, we use robust standard errors for estimation.

The coefficient on the number of pupils is statistically significant at a 1% significance level for primary schools but insignificant for secondary schools. This means that when primary schools gain a pupil, their expected $educresult$ score increases by only 0.016 points. The relationship between these variables may indicate that student competition motivates other students to perform better and that schools with larger class sizes tend to outperform schools with smaller class sizes. Another reason might be that bigger schools are associated with more clever pupils. The relationship itself is rather weak. For example, if the school gained 100 students, it would only be expected to gain 1.6 points on the $educresult$.

Similarly, the share of pupils from socially disadvantaged backgrounds is statistically significant at a 1% significance level for primary schools but insignificant for secondary schools. When the share of socially disadvantaged students increases by 1%, the overall primary school score is expected to decrease by approximately -0.03. This follows logical reasoning, as students from socially disadvantaged backgrounds usually underperform in schools, especially in Slovakia. The insignificant effect of this variable on secondary school models may be due to the fact that not many socially disadvantaged students continue studying at the secondary school level.

The only variable that is jointly significant at a 1% significance level for both primary and secondary schools is ICT - the usage of Information and Communication Technologies in the educational process. However, the relationship differs between primary and secondary schools. In primary schools, the relationship is positive, while in secondary schools, it is negative. When the share of ICT usage increases by 1%, the overall primary school score is expected to increase by approximately 0.02, while the overall secondary school score is expected to decrease by approximately -0.02. Although it is quite straightforward to explain the positive effect of ICT usage in primary schools, the opposite effect in secondary schools is less clear. One possible explanation is that technical and specialised secondary schools may have adopted ICT more quickly, which could have created a negative effect because these schools tend to underperform.

For primary schools, an increase of 1000 citizens in the city's population had a negative effect of -1.01 at a 5% significance level. This could be due to problems with adjusting school capacities in cities with increasing populations, which negatively impacted educational results. However, this variable proved insignificant for secondary schools.

On the other hand, an increase of 1% in unemployment had a negative effect of -0.59 on the test scores of secondary schools, while this variable was insignificant for primary schools. High unemployment rates often lead to financial instability for families. This can affect students' ability to access essential resources such as books, study materials, and tutoring services, which are usually more expensive and also needed for secondary school students. As a result, their academic performance may suffer.

Now, we turn to the variables of our primary interest: the coefficients of $eurofonds_{it}^{educ}$ and $eurofonds_{it}^{reconstr}$ and their lags. Although some lags were significant, the relationship was very weak for most variables. We are unable to interpret the significant variables reliably, and we believe that if there is an effect, it would be close to 0 or not positive. Only Model (3) exhibits signs of some effect of $0.006 - 0.015$ for lags $1-3$ at a 1 to 5% significance level. That would mean the effect of the average education-related eurofund in those years on $educresult$ would be from 1.2 to 3 points.

After carefully analyzing and examining all the relevant factors and aspects, we have arrived at the conclusion that the FE model, which has a low R^2 and mostly weak relationships, is not a reliable predictor of educational achievement in this case. It is important to consider the limitations of the model, such as the lack of sufficient data before the year 2014, the limited number of explanatory variables and the complexity of the subject matter. Given these limitations, we cannot confidently draw any conclusions from the model and further research is needed to develop a more accurate and comprehensive understanding of educational achievement, as will be discussed later.

6.2 Between Effects Results

Due to the lack of accountability of the fixed effects model, to fully assess the determinants of educational achievement, we decided to include a between-effects (Between Effects (BE)) estimator that will investigate the effects of various variables on educational achievement. BE is an extension of the fixed effects estimator, and we include it as we believe it provides insightful information about schools' educational achievement in general rather than just the change over time. We, therefore, estimate 6.2 by the OLS method.

$$\begin{aligned}\overline{educresult}_i = &\overline{\beta}_0 + \beta_1\overline{pupils}_i + \beta_2\overline{pupilsSDB}_i + \beta_3\overline{ICT}_i\\ &+ \beta_4\overline{teachers}_i + \beta_5\overline{population}_i + \beta_6\overline{unem}_i\\ &+ \beta_7 eurofonds_i^{educ} + \beta_8 eurofonds_i^{rec}\\ &+ \beta_9 rompop_i + \beta_{10} university_i + \beta_{11} district_i\\ &+ \beta_{12} Bratislava_i + \beta_{13} church_i + \beta_{14} private_i\\ &+ \beta_{15} hungarian_i + \beta_{16} bilingual_i + \epsilon_i\end{aligned} \tag{6.2}$$

We test the assumptions for the OLS estimator, as listed in Appendix A. First, we assess the Multiple Linear Regression (MLR).1 assumption of a linear relationship by plotting the "Residuals vs Fitted" graph in R. In both models, the line is horizontal and without serious distinct patterns, which should indicate a linear relationship. MLR.2, a random sample assumption, holds for both models as we include a large enough sample of all primary and secondary schools in Slovakia. We checked MLR.3, which requires an expected value of 0 for the error term u_i, by plotting residuals against independent variables and finding no pattern. We checked for collinearity using the "vif" package in R and removed any variables above 2.5, so no multicollinearity should be present in our model. Next, we ran the Breusch-Pagan test for heteroskedasticity (MLR.5) and the Durbin-Watson test for serial correlation (MLR.6), and discovered both (except for secondary school model and autocorrelation), as can be seen in Table 6.4. In Table 6.5, we use adjusted standard errors that we created using the "Arellano" method of the robust covariance matrix. To finalize our assumptions testing, we created a histogram and a Q-Q plot. The histogram showed a normal distribution and the Q-Q plot showed that all the points approximately fall alongside the reference line. Therefore, we can assume normality. Under MLR.1-MLR.4, according to Wooldridge, the estimator will be unbiased, but not BLUE, as assumptions MLR.5 and MLR.6 are violated, so it is not efficient.

Table 6.4: Between Effects: model diagnostics

	Primary schools:			*Secondary schools:*		
Test	Stat.	df	p-value	Stat.	df	p-value
Durbin-Watson	1.768	–	4.0×10^{-6}	2.0497	–	0.6837
Breusch-Pagan	181.19	15	2.2×10^{-16}	35.931	15	0.0018

Table 6.5: Between Effects: Results

	Dependent variable:	
	$educresult_i$	
	(1)	(2)
	Primary schools	*Secondary schools*
$pupils_i$	0.002 (0.001)	0.003 (0.002)
$pupilsSDB_i$	−0.390*** (0.030)	−0.247* (0.226)
ICT_i	0.015 (0.011)	0.074*** (0.020)
$teachers_i$	−0.039 (0.135)	0.004 (0.058)
$population_i$	−0.001 (0.005)	−0.004 (0.008)
$unem_i$	−0.120*** (0.034)	−0.0001 (0.065)
$eurofonds_i^{educ}$	0.004** (0.002)	−0.004** (0.002)
$eurofonds_i^{rec}$	0.001** (0.0003)	0.002 (0.001)
$rompop_i$	−0.020 (0.016)	0.004 (0.034)
$university_i$	0.027 (0.021)	0.001 (0.044)
$district_i$	2.382*** (0.468)	0.929 (1.034)
$Bratislava_i$	1.745** (1.021)	2.499* (1.494)
$church_i$	1.430** (0.701)	5.049*** (0.963)
$private_i$	4.276*** (1.787)	−0.812 (0.940)
$hungarian_i$	3.484*** (0.657)	3.891** (1.836)
$bilingual_i$	2.594* (1.948)	6.760*** (0.897)
Constant	58.361*** (1.833)	45.317*** (2.301)
Observations	1,403	648
R^2	0.400	0.191
Adjusted R^2	0.392	0.188
Residual Std. Error	6.119 (df = 1387)	8.233 (df = 632)
F Statistic	61.325*** (df = 15; 1387)	9.758*** (df = 15; 632)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 6.5 shows the results of the two models estimated using between-effects estimation. The F-test results indicate that all of the models are statistically significant at a 1% significance level. Model (1) accounts for 40% of the

overall variability, while Model (2) accounts for 19.1% of the overall variability in the *educresult* variable.

The share of pupils from socially disadvantaged backgrounds has a statistically significant effect on primary and secondary school test scores, with significance levels of 1% and 10%, respectively. A 1% increase in the share of socially disadvantaged students is associated with a decrease of -0.390 in overall primary school test scores and -0.247 in overall secondary school test scores. This can be attributed to the lack of support that students from disadvantaged backgrounds receive, which leads to underperformance in the school setting.

The use of ICT is also significant for secondary schools, with a significance level of 1%. A 1% increase in the use of ICT is associated with an average increase of 0.074 points in overall test scores. This makes logical sense, as innovative education is expected to produce better results. However, it could also indirectly reflect better management, as schools with higher ICT usage are likely to have better teachers who know how to work with technology. This is contradictory to the previous FE model, where the increase in ICT was associated with a negative effect on *educresult* at secondary schools. We are unable to explain this discrepancy rationally.

Higher unemployment has a negative effect on primary school test scores, with a significance level of 1%. A 1% increase in unemployment is associated with a decrease of -0.12 points in overall primary school test scores. However, this effect is not significant for secondary schools now, contrary to the FE model. As mentioned previously, this may be a regional difference that may be an indirect indicator of the financial well-being of families of students that can influence their educational achievement.

The variables $eurofonds_i^{educ}$ and $eurofonds_i^{rec}$ are mainly informative in terms of which types of schools received funding. We observe that primary schools with marginally better performance received reconstruction-related eurofunds and education-related eurofunds, while marginally worse secondary schools received education-related eurofunds. However, the relationship is weak, so it is not necessary to delve deeper into it.

What is interesting are the significant dummy variables. For primary schools, we see that being located in the district capital leads to a likely increase in score of 2.38 at a 1% significance level, while being located in Bratislava leads to a 1.75 increase at a 5% significance level. If the school is private, the increase is 4.28 at a 1% significance level. If the school teaches in the Hungarian language, the increase is 3.45 at a 1% significance level, while being bilingual leads to an increase of 2.59 at a 10% significance level. The establisher being the church leads to an increase of 1.43 at a 5% significance level.

For secondary schools, we see that being located in Bratislava leads to a 2.49 increase at a 10% significance level. If the school teaches in the Hungarian language, the increase is 3.89 at a 5% significance level, while being bilingual leads to an increase of 6.76 at a 1% significance level. The establisher being the church leads to an increase of 5.05 at a 5% significance level. This may be due to the fact that better and more clever students tend to attend schools where the establisher is the church or where the language of instruction is bilingual. Not necessarily are those better primary or secondary schools.

We consider this model interesting because it provides insight into what the main educational determinants are for primary and secondary schools in Slovakia. It, however, has its limits, as it is not that accounting for time trends.

6.3 Difference-in-Difference Results

We finish the results section by presenting DID model we used because it is able to capture the time-variant, time-invariant and impact of eurofonds we want to assess the effect of. We use two periods in time, averaged years 2012-2014 and 2017-2019, to reduce the fluctuations of the *educresult* variable and estimate them using Equation 6.3.

$$\begin{aligned} educresult_i = {} & \beta_0 + \delta_1 post_i + \delta_2 treatment_i + \delta_3 post_i \times treatment_i \\ & + \beta_1 pupils_i + \beta_2 pupilsSDB_i + \beta_3 ICT_i \\ & + \beta_4 teachers_i + \beta_5 population_i + \beta_6 unem_i \\ & + \beta_7 rompop_i + \beta_8 university_i + \beta_9 district_i \\ & + \beta_{10} Bratislava_i + \beta_{11} church_i + \beta_{12} private_i \\ & + \beta_{13} hungarian_i + \beta_{14} bilingual_i + \epsilon_i, \end{aligned} \tag{6.3}$$

To ensure the validity of our analysis, we first test the assumptions for the OLS estimator, as listed in Appendix A. We assess the MLR.1 − MLR.4 and MLR.7 assumptions with the same method as in the previous chapter with very similar results. We also run the Breusch-Pagan test for heteroskedasticity (MLR.5) and the Durbin-Watson test for serial correlation (MLR.6), discovering both, except for the secondary school model and autocorrelation. We present the results of these tests in Table 6.6. To fix this, we use adjusted standard errors created using the "Arellano" method of the robust covariance matrix. To conclude, under MLR.1-MLR.4, according to Wooldridge, the estimator will be unbiased, but not BLUE, as assumptions MLR.5 and MLR.6 are violated, so it is not efficient.

Table 6.6: Difference-in-Difference: model diagnostics

	Primary schools:			*Secondary schools:*		
Test	Stat.	df	p-value	Stat.	df	p-value
Education-related funding						
Durbin-Watson	1.66	–	2.2×10^{-16}	2.10	–	0.9336
Breusch-Pagan	240.8	17	2.2×10^{-16}	61.3	17	2.4×10^{-6}
Reconstruction-related funding						
Durbin-Watson	1.67	–	2.2×10^{-16}	2.09	–	0.9226
Breusch-Pagan	256.31	17	2.2×10^{-16}	62.29	17	4.3×10^{-7}

Table 6.7 displays the results of the four models that were estimated using difference-in-difference estimation. The F-test results indicate that all of the models are statistically significant at a 1% significance level. Models (1) and (2) together account for 34% of the overall variability, while Models (3) and (4) account for 21% of the overall variability in the *educresult* variable.

We observe a decrease in the *educresult* variable in both primary and secondary schools for both model types, with significance levels of 1%. Specifically, the decrease is around -4 for primary schools and -6.4 for secondary schools. This may be a result of a slight decrease in the requirements of the educational curriculum over the years but unchanged tests. Additionally, in general, primary schools that received eurofunds achieved 1.1 better results than those that did not, with a significance level of 1%. This does not violate our assumption of control and treatment groups as long as the schools are similar to each other in terms of characteristics. This requirement was checked by comparing means of dependent and independent variables before the treatment, and they are not significantly different. We assume that this assumption holds.

We did not find a statistically significant coefficient for the variable $treatment \times post$. Therefore, we believe that the effect of eurofunds on educational achievement is not present. We will discuss why this is the case and examine possible limitations at the end of this chapter.

The number of pupils has a statistically significant impact on test scores in both primary and secondary school models. While it was insignificant for BE model, here, the effect is present, and we really cannot explain this discrepancy. The significance levels are 1% and 10%, respectively. Adding one more pupil is associated with an increase in the school's score by approximately 0.04 - 0.05 points in primary school and 0.03 points in secondary school. This result is logical because larger schools have more resources to support their students and can achieve better results on national tests, or better students naturally choose to attend bigger schools.

Additionally, the share of pupils from socially disadvantaged backgrounds has a significant effect on primary school test scores, with a significance level of 1%. A 1% increase in the share of socially disadvantaged students is expected to result in a decrease of approximately -0.318 in overall primary school test scores for both models. This can be attributed to the lack of support that students from disadvantaged backgrounds receive, which leads to underperformance in the school setting. This is aligned with findings from previous sections, where $pupils_{SDB}$ is almost always significant.

According to the analysis, the use of Information and Communication Technologies (ICT) has a significant impact on both primary and secondary schools, with a significance level of 10% and 1%, respectively. A 1% increase in the use of ICT is associated with an average increase of 0.015 - 0.016 points in overall test scores for primary schools and 0.064 - 0.065 points for secondary schools. These findings support the notion that innovative education leads to better outcomes and are aligned with the BE model and partly with the FE model (apart from the secondary schools model where the ICT was associated with a negative coefficient).

Higher unemployment has a negative effect on primary school test scores, with a significance level of 1%. Specifically, a 1% increase in unemployment is associated with a decrease of -0.112 points in overall primary school test scores. However, this effect is not significant for secondary schools, and this is aligned with findings from BE model.

Additionally, the educational achievement of a school is influenced by the Roma population share in the city. For every additional 1% increase in the Roma population share, the educational achievement of the school decreases by -0.023, with a significance level of 10%.

Table 6.7: Difference-in-Difference: Results

	Dependent variable: $educresult_i$			
	Primary schools:		*Secondary schools:*	
	(1) Education-related	(2) Reconstruction-related	(3) Education-related	(4) Reconstruction-related
$post_i$	−4.111*** (0.505)	−3.907*** (0.918)	−6.363*** (0.917)	−6.526*** (0.855)
$treatment_i$	1.095** (0.465)	1.113** (0.422)	−0.451 (0.774)	1.429 (1.184)
$post_i \times treatment_i$	−0.534 (0.695)	−0.696 (0.591)	−0.589 (1.078)	−1.981 (1.705)
$pupils_i$	0.005*** (0.001)	0.004*** (0.001)	0.003** (0.002)	0.003* (0.002)
$pupilsSDB_i$	−0.318*** (0.017)	−0.316*** (0.017)	−0.243 (0.166)	−0.239 (0.173)
ICT_i	0.015* (0.008)	0.016* (0.009)	0.064*** (0.015)	0.063*** (0.015)
$teachers_i$	0.061 (0.091)	0.075 (0.092)	−0.014 (0.062)	−0.009 (0.062)
$population_i$	0.006 (0.007)	0.006 (0.007)	−0.001 (0.014)	−0.001 (0.014)
$unem_i$	−0.112*** (0.040)	−0.107*** (0.063)	0.002 (0.063)	−0.014 (0.062)
$rompop_i$	−0.023* (0.016)	−0.025* (0.015)	0.023 (0.027)	0.021 (0.025)
$university_i$	0.026 (0.026)	0.024 (0.027)	0.013 (0.045)	0.012 (0.046)
$district_i$	2.508*** (0.394)	2.740*** (0.406)	0.531 (0.848)	0.550 (0.851)
$Bratislava_i$	1.885** (0.771)	2.194*** (0.790)	2.272* (1.183)	2.489** (1.188)
$church_i$	1.708*** (0.654)	1.809*** (0.656)	5.310*** (0.805)	5.421*** (0.798)
$private_i$	1.010 (1.613)	1.131 (1.612)	0.109 (0.797)	0.118 (0.800)
$hungarian_i$	2.861*** (0.620)	2.871*** (0.621)	5.533*** (1.653)	5.828*** (1.649)
$bilingual_i$	3.203** (1.883)	3.043** (1.895)	7.101*** (0.702)	7.158*** (0.700)
Constant	59.162*** (1.354)	59.162*** (1.881)	48.704*** (1.880)	48.612*** (1.875)
Observations	2,738	2,738	1,136	1,136
R^2	0.335	0.335	0.213	0.212
Adjusted R^2	0.331	0.331	0.201	0.200
Residual Std. Error	7.674 (df = 2720)	7.672 (df = 2720)	8.768 (df = 1118)	8.770 (df = 1118)
F Statistic	80.487*** (df = 17; 2720)	80.607*** (df = 17; 2720)	17.781*** (df = 17; 1118)	17.743*** (df = 17; 1118)

Note: *p<0.1; **p<0.05; ***p<0.01

What is interesting are the significant dummy variables. For primary schools, we see that being located in the district capital leads to a likely increase in score of 2.51 - 2.74 at a 1% significance level, while being located in Bratislava leads to a 1.885 - 2.194 increase at a 1% significance level. If the school teaches in the Hungarian language, the increase is 2.86 - 2.87 at a 5% significance level, while being bilingual leads to an increase of 3.04 - 3.2 at a 5% significance level. The establisher being the church leads to an increase of 1.7 - 1.8 at a 1% significance level.

For secondary schools, we see that being located in Bratislava leads to a 2.272 - 2.489 increase at a 10% and 5% significance level for different models. If the school teaches in the Hungarian language, the increase is 5.533 - 5.828 at a 1% significance level, while being bilingual leads to an increase of 7.101 - 7.158 at a 1% significance level. The establisher being the church leads to an increase of 5.31 - 5.42 at a 1% significance level.

6.4 Discussion

As previously mentioned, none of the models indicates a significant effect of eurofunds on the educational achievement of Slovak students. The only exception is the FE model (3), but its low R^2 value limits its accountability. Upon careful consideration, there are several reasons that may be the root cause of the insignificance of these variables. Firstly, we believe that our observation time may not have been long enough. For the FE model, we used data only from 2014 to 2019, which was cut short due to the COVID-19 epidemic and the cancellation of the NIVAM test. With the DID model, we were only allowed to use a 2-year gap between the before and post-treatment period, which may have influenced our results.

Another possibility is that the effect is simply not present. As seen in the Figure 3.2, a large proportion of schools received funding, which suggests that resources may not have been used efficiently and thus may not have had an effect on overall educational achievement. It's possible that due to the availability of a lot of financial resources, poor-quality applications were accepted, resulting in unnecessary spending that did not have much of an impact.

Another way to improve our model would be to make the educational achievement variable, *educresult*, more complex and collect more information about the schools. This would help eliminate fluctuations in the dependent variable over time. Additionally, an interesting path to consider would be to give this research an international setting, estimating the effect of EU funds on educational achievement (PISA or TIMMS tests) as an increase in spending on education per student and comparing it to countries without such funding.

Other findings about determinants of educational achievement, such as the significance of specific explanatory variables, align with previous literature, such as in Kreisman & Steinberg (2019). Specifically, larger schools located in district capitals or major cities, such as Bratislava, tend to have better academic results than smaller schools in rural areas. Additionally, schools with a higher rate of ICT use have been in most cases shown to perform better. On the other hand, schools located in regions with a higher rate of unemployment and more socially disadvantaged pupils are associated with worse academic outcomes. This highlights the importance of addressing the root causes of socioeconomic disparities in order to improve education for all students.

Chapter 7

Conclusion

This bachelor's thesis examines the impact of education-related EU funds that were channelled into Slovak primary and secondary schools during the 2007-2013 programming period. This time frame was selected to allow for the effects to settle and assess schools' test scores in subsequent years. We use test scores from national standardised tests and incorporate other explanatory variables related to the characteristics of the school and its location. We divide EU funding into two categories: education-related funding and reconstruction-related funding. The former includes building maintenance to ensure that students can learn in a safe and comfortable environment. The latter mostly aims to improve teacher training or ensure specific curriculum-related tools are available for students. We first used the Fixed Effects method of estimation for panel data. However, the results were not satisfactory. Therefore, we moved on to apply the Between Effects Estimator, which is an extension of the Fixed Effect method and averages out the time component. Lastly, to assess both the time effects and time-invariant explanatory variables, we used the Difference-in-Difference model.

After conducting a thorough analysis, it has been determined that there is no concrete evidence to suggest that receiving EU funds has a direct relationship with educational achievement. While some lags related to education funding may show positive significance in the Fixed Effects Model, the model's low R^2 value casts doubt on the reliability of these results. Furthermore, the DID model has also similarly found no relationship between test scores and the additional resources provided to schools through these funds, indicating that there may be other factors at play which are influencing educational achievement in these regions.

While the relationship with EU funds was absent, we were able to assess other determinants of test scores that were significant in multiple models and are associated with having an impact on the test score. The most impactful across models was the number of primary school pupils from a disadvantaged background, with an effect of -0.316 to -0.39 for an additional 1% increase in the share of these pupils. Both primary and secondary schools being located in Bratislava are associated with an influence of 1.75 to 2.5, while the primary schools being located in the district capital are associated with 2.38 to 2.74. Similarly, a higher unemployment rate by 1% is associated with a -0.112 - -0.12 decrease in test scores for primary schools. Interestingly, the frequency of usage of interactive and communication technologies in educational processes, while mostly positive, was not uniform and varied across different models.

This thesis makes a valuable contribution to the existing research on the impact of EU funds. No similar study has been conducted in Slovakia or within the area of practice of studies on EU funds. Despite the lack of a direct relationship between European funding and educational outcomes in schools, it is important for researchers to continue exploring and investigating the potential benefits of EU funds on education. Additionally, conditioning factors that contribute to the success of these funds should be examined further. This is important because Structural Funds and Cohesion Policy are not to end and are financially more ambitious than previous programming periods.

Bibliography

Barro, R. J. & X. Sala-I-Martin (1992): “Convergence.” *Journal of Political Economy* **100(2)**: pp. 223–251.

Becker, B. S. O., P. H. Egger, & M. V. Ehrlich (2013): “Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects.” *American Economic Journal: Economic Policy* **5(4)**: pp. 29–77.

Belmonte, A., V. Bove, G. DÂ€™Inverno, & M. Modica (2020): “School infrastructure spending and educational outcomes: Evidence from the 2012 earthquake in northern italy.” *Economics of Education Review* **75**: p. 101951.

Bénabou, R., F. Kramarz, & C. Prost (2009): “The French zones d’éducation prioritaire: Much ado about nothing?” *Economics of Education Review* **28**: pp. 345–356.

Böhlmark, A. & M. Lindahl (2008): “Does School Privatization Improve Educational Achievement? Evidence from Sweden’s Voucher Reform.” *IZA Discussion Papers 3691*, Institute of Labor Economics (IZA).

Card, D. & A. B. Krueger (1992): “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.” *The Journal of Political Economy* **100(1)**: pp. 1–40.

Dall’Erba, S. & F. Fang (2017): “Meta-analysis of the impact of European Union Structural Funds on regional growth.” *Regional Studies* **51(6)**: pp. 822–832.

Dall’erba, S., R. Guillain, J. Le Gallo *et al.* (2009): “Impact of structural funds on regional growth: How to reconsider a 9-year-old black box.” *Région et Développement* **30**: pp. 77–99.

Fratesi, U. & F. G. Wishlade (2017): "The impact of european cohesion policy in different contexts."

Geodetic and Cartographic Office of Bratislava (2022): "Shapefiles of Slovakia." `https://www.geoportal.sk/en/zbgis/download/` [Accessed on 2023-03-02].

Hægeland, T., O. Raaum, & K. G. Salvanes (2012): "Pennies from heaven? Using exogenous tax variation to identify effects of school resources on pupil achievement." *Economics of Education Review* **31(5)**: pp. 601–614.

Häkkinen, I., T. Kirjavainen, & R. Uusitalo (2003): "School resources and student achievement revisited: new evidence from panel data." *Economics of Education Review* **22(3)**: pp. 329–335.

Hanushek, E. A. (2006): "School Resources." *Handbook of the Economics of Education* **2**: pp. 865–908.

Hillman, N. W., A. Hicklin Fryar, & V. Crespín-Trujillo (2017): "Evaluating the Impact of Performance Funding in Ohio and Tennessee." *American Educational Research Journal* **55(1)**: pp. 144–170.

Hyman, J., S. Dynarski, J. Bound, B. Jacob, J. Smith, E. Brunner, P. Courant, E. Dillon, T. Downes, R. Garlick, M. Gideon, M. Kapustin, G. Nelson, & S. Ross (1994): "Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment." *American Economic Journal: Economic Policy* **9(4)**: pp. 256–280.

Institute for Economic and Social Reforms (2022): "Ranking for primary and secondary schools." `https://skoly.ineko.sk/metodika/` [Accessed on 2023-01-14].

Jackson, C. K., R. C. Johnson, & C. Persico (2015): " The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms." *The Quarterly Journal of Economics* **131(1)**: pp. 157–218.

Kreisman, D. & M. P. Steinberg (2019): "The Effect of Increased Funding on Student Achievement: Evidence from Texas's Small District Adjustment." *Journal of Public Economics* **176**: pp. 118–141.

Leuven, E., M. Lindahl, H. Oosterbeek, D. Webbink, E. Leuven, M. Lindahl, H. Oosterbeek, & D. Webbink (2007): "The Effect of Extra Funding for Disadvantaged Pupils on Achievement." *The Review of Economics and Statistics* **89(4)**: pp. 721–736.

Machin, S., S. McNally, S. Machin, & S. McNally (2008): "The Literacy Hour." *Journal of Public Economics* **92(5-6)**: pp. 1441–1462.

Ministry of Interior of the Slovak Republic (2019): "Atlas of Roma Communities 2019." `https://www.minv.sk/?atlas-romskych-komunit-2019` [Accessed on 2023-02-04].

Ministry of Investment, Regional Development and Informatization of the SR (2021): "Lists of Granted Eurofunds." `https://www.mirri.gov.sk` [Accessed on 2023-03-03].

Modranka, E. (2015): "Spatial dependencies in the absorption of funds from Regional Operational Programmes on NUTS 3 regional level in Poland." *Journal of Economics and Management* **19(1)**: pp. 108–122.

National Institute of Education and Youth (2022): "Dataportal." `https://vysledky.nucem.sk/` [Accessed on 2023-01-03].

OECD (2019): "Reading performance and spending on education | PISA 2018 Results (Volume I) : What Students Know and Can Do." `https://www.oecd-ilibrary.org/education/reading-performance-and-spending-on-education_7aa618c1-en` [Accessed on 2022-12-12].

Radvanský, M., K. Frank, I. Lichner, & T. Miklošovič (2016): "Impact of Cohesion Policy on Regional Development of Slovakia Ex-post Assessment of National Strategic Reference Framework 2007-2013." *Technical report*, Institute of Economic Research SAS.

Rodríguez-Pose, A. & E. Garcilazo (2015): "Quality of Government and the Returns of Investment: Examining the Impact of Cohesion Expenditure in European Regions." *Regional Studies* **49(8)**: pp. 1274–1290.

Rothstein, J. & D. W. Schanzenbach (2022): "Does money still matter? attainment and earnings effects of post-1990 school finance reforms." *Journal of Labor Economics* **40(S1)**: pp. S141–S178.

Schlotter, M., G. Schwerdt, & L. Woessmann (2011): "Econometric Methods for Causal Evaluation of Education Policies and Practices: A Non-Technical Guide." *Education Economics* **19(2)**: pp. 109–137.

Schoenberg, A. (2018): "Regional Disparities in Europe: An Assessment of the Impact of the 2007-2013 Funding Programme on Convergence in Romania and Bulgaria." *CESifo Forum* **19(01)**: pp. 32–36.

Solow, R. M. (1956): "A Contribution to the Theory of Economic Growth." *The Quarterly Journal of Economics* **70(1)**: pp. 65–94.

STATA (2019): "Xtreg - Fixed-, between-, and random-effects and population-averaged linear models." `https://www.stata.com/manuals/xtxtreg.pdf` [Accessed on 2023-04-04].

Statistical Office of the Slovak Republic (2021): "Statistical Office of the Slovak Republic." `https://slovak.statistics.sk` [Accessed on 2023-02-14].

Wooldridge, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

Wooldridge, J. M. (2015): *Introductory Econometrics: A Modern Approach*. Toronto: Nelson Education.

Žáček, J. & F. Hrůza (2019): "The Impact of EU Funds on Regional Economic Growth of the Czech Republic." *Finance a úvěr - Czech Journal of Economics and Finance* **69(1)**: p. 11.

Appendix A

Assumptions

A.1 Assumptions for Fixed Effect Model

We provide a list of assumptions for the fixed effect estimator as summarized in Wooldridge (2015, chap. 14):

Assumption FE.1
For each i, the model is $y_{it} = \beta_0 + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}$, where the β_j are the parameters to estimate and $t = 1, \ldots, T$.

Assumption FE.2
We have a random sample in the cross-sectional dimension.

Assumption FE.3
For each t, the expected value of the idiosyncratic error given the explanatory variables in all time periods and the unobserved effect is zero $E(u_{it}|X_i, a_i) = 0$.

Assumption FE.4
Each explanatory variable changes over time (for at least some i), and there are no perfect linear relationships among the explanatory variables.

Assumption FE.5
For all $t = 1, \ldots, T$: $Var(u_{it}|X_i, a_i) = Var(u_i t) = \sigma_u^2$.

Assumption FE.6
For all $t \neq s$, the idiosyncratic errors are uncorrelated (conditional on all explanatory variables and a_i): $Cov(u_{it}, u_{is}|X_i, a_i) = 0$.

Assumption FE.7
Conditional on X_i and a_i, the u_{it} are independent and identically distributed IID$(0, \sigma_u^2)$.

A.2 Assumptions for BE Model and DID Model

We provide a list of assumptions for OLS estimator as summarized in Wooldridge (2015, chap. 14):

Assumption MLR.1

The model in the population can be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i$$

where the $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown parameters of interest, and u is an unobservable random error or random disturbance term.

Assumption MLR.2

We have a random sample of n observations from the population model described above.

Assumption MLR.3

The error u has an expected value of zero, given any values of the independent variables. In other words,

$$E(u_i|X_i) = 0.$$

Assumption MLR.4

In the sample (and therefore in the population), none of the independent variables is constant, and there are no exact linear relationships among the independent variables.

Assumption MLR.5 - Homoskedasticity

$$Var(u_i|X_i) = \sigma^2.$$

Assumption MLR.6 - No serial correlation

$$Cov(u_t, u_s|X) = 0$$

for any given $t \neq s$.

Assumption MLR.7 - Normality

The residuals are independent and identically distributed.

Appendix B

Specification for models used in the analysis

To avoid confusion, we provide a full specification of the model used in this bachelor's thesis.

Fixed Effect Model

$$\begin{aligned} educresult_{it} = &\beta_0 + \beta_1 pupils_{it} + \beta_2 pupilsSDB_{it} + \beta_3 ICT_{it} \\ &+ \beta_4 teachers_{it} + \beta_5 population_{it} + \beta_6 unem_{it} \\ &+ \beta_7 eurofonds^{educ}_{it} + \beta_8 eurofonds^{educ}_{it+1} \\ &+ \beta_9 eurofonds^{educ}_{it+2} + \beta_{10} eurofonds^{educ}_{it+3} \\ &+ \beta_{15} eurofonds^{rec} + \beta_{16} eurofonds^{rec}_{it+1} \\ &+ \beta_{17} eurofonds^{rec}_{it+2} + \beta_{18} eurofonds^{rec}_{it+3} + a_i + u_{it}, \end{aligned} \tag{B.1}$$

where $t = 2014, \ldots, 2019$, denotes the year and the i subscript denotes the school. The variable $educresult_{it}$ denotes the results of the school in the NIVAM tests, $pupils_{it}$ denotes the number of pupils, $pupilsSDB_{it}$ denotes the averaged percentage of pupils from a socially disadvantaged background in respect to the overall number of pupils, ICT_{it} denotes the usage of Information and Communication Technologies in the educational process, $teachers_{it}$ denotes the number of teachers per 100 students,$population_i$ denotes the number of citizens in thousands living in the town where i-th school is located, $unem_{it}$ denotes unemployment in the district where the i-th school is located, $eurofonds^{educ}_{it}$ denotes the amount of education-related EU funds used in the i-th school in thousands of EUR and $eurofonds^{rec}_{it}$ denotes the amount of reconstruction-related EU funds used in the i-th school in thousands of EUR. [1] Subsequently, $eurofonds^{educ}_{it+1...3}$ and $eurofonds^{rec}_{it+1...3}$ are the first to seventh lags, cumulative lags or cumulative effects of education and reconstruction-related EU funding in thousands of euros, respectively.

Between Effect Model

$$\begin{aligned}
\overline{educresult_i} = &\overline{\beta}_0 + \beta_1\overline{pupils}_i + \beta_2\overline{pupilsSDB}_i + \beta_3\overline{ICT}_i \\
&+ \beta_4\overline{teachers}_i + \beta_5\overline{population}_i + \beta_6\overline{unem}_i \\
&+ \beta_7 eurofonds^{educ}_i + \beta_8 eurofonds^{rec}_i \\
&+ \beta_9 rompop_i + \beta_{10} university_i \\
&+ \beta_{11} district_i + \beta_{12} Bratislava_i + \beta_{13} church_i \\
&+ \beta_{14} private_i + \beta_{15} hungarian_i + \beta_{16} bilingual_i + \epsilon_i
\end{aligned} \tag{B.2}$$

where the variable $\overline{educresult_i}$ denotes averaged [2] results of the school in the NIVAM tests, $\overline{pupils_i}$ denotes the averaged number of pupils, $\overline{pupilsSDB_i}$ denotes the averaged percentage of pupils from a socially disadvantaged background in respect to the overall number of pupils, $\overline{ICT}_i$ denotes the averaged usage of Information and Communication Technologies in the educational process, $\overline{teachers_i}$ denotes the averaged number of teachers per 100 students, $population_i$ denotes number of citizens living in the town where i-th school is located, $\overline{unem}_i$ denotes average unemployment in the district where the i-th school is located, $eurofonds^{educ}_{it}$ denotes the amount of education-related EU funds used in the i-th school in thousands of EUR, $eurofonds^{rec}_{it}$ denotes the amount of reconstruction-related EU funds used in the i-th school in thousands of EUR, $rompop_i$ denotes the percentage of Roma population in the town where i-th school is located and $university_i$ denotes the percentage of the population with university-degree in the town where i-th school is located. Subsequently, we add dummy variables which take the value of 1 if: $district_i$ for schools located in capitals of districts (Bratislava excluded), $Bratislava_i$ for schools located in Bratislava, $church_i$ for schools with the church as the establisher, $private_i$ for schools with private establisher, $hungarian_i$ for schools with the Hungarian language of instruction, $bilingual_i$ for schools with the bilingual language of instruction.

[1] We use a year when the school stopped receiving EU funding as year 0. We decided to take the last year of receiving the EU Funds because the custom usually is to use most of the money towards the end of the absorption period.

[2] We averaged all the available years between 2009-2019 for all the variables.

Difference-in-Difference Model

$$
\begin{aligned}
educresult_i = &\beta_0 + \delta_1 post_i + \delta_2 treatment_i + \delta_3 post_i \times treatment_i \\
&+ \beta_1 pupils_i + \beta_2 pupilsSDB_i + \beta_3 ICT_i \\
&+ \beta_4 teachers_i + \beta_5 population_i + \beta_6 unem_i \\
&+ \beta_7 rompop_i + \beta_8 university_i \\
&+ \beta_9 district_i + \beta_{10} Bratislava_i + \beta_{11} church_i \\
&+ \beta_{12} private_i + \beta_{13} hungarian_i + \beta_{14} bilingual_i + \epsilon_i,
\end{aligned}
\tag{B.3}
$$

where the variable $educresult_i$ denotes averaged [3] results of the school in the NIVAM tests, $post_i$ is the period dummy, which takes the amount of 1 when the observation is from the post-treatment period, $treatment_i$ is the treatment dummy, which takes the amount of 1 if the i-th observation was treated regardless of the period, $pupils_i$ denotes the averaged number of pupils, $pupilsSDB_i$ denotes the averaged percentage of pupils from a socially disadvantaged background with respect to the overall number of pupils, ICT_i denotes the averaged usage of Information and Communication Technologies in the educational process, $teachers_i$ denotes the averaged number of teachers per 100 students, $population_i$ denotes number of citizens living in the town where i-th school is located, $unem_i$ denotes average unemployment in the district where the i-th school is located, $rompop_i$ denotes the percentage of Roma population in the town where i-th school is located and $university_i$ denotes the percentage of the population with university-degree in the town where i-th school is located. Subsequently, we add dummy variables which take the value of 1 if: $district_i$ for schools located in capitals of districts (Bratislava excluded), $Bratislava_i$ for schools located in Bratislava, $church_i$ for schools with the church as the establisher, $private_i$ for schools with private establisher, $hungarian_i$ for schools with the Hungarian language of instruction, $bilingual_i$ for schools with the bilingual language of instruction. We estimate the model individually for reconstruction-related and education-related funding.

[3]We averaged years between 2012-2014 to create the pre-treatment observation and 2017-2019 for the post-treatment group. This was done for all the variables.

Appendix C

Results for additional regression models

Table C.1: Fixed Effects with multiple lags: results

	Dependent variable:			
	$educresult_i$			
	Primary schools:		*Secondary schools:*	
	(1) Simple lags	(2) Continuous lags	(3) Simple lags	(4) Continuous lags
$pupils_{it}$	0.016*** (0.004)	0.016*** (0.004)	0.0003 (0.004)	0.0003 (0.004)
$pupilsSDB_{it}$	−0.035*** (0.015)	−0.032*** (0.015)	0.074 (0.049)	0.078 (0.057)
ICT_{it}	0.019*** (0.005)	0.021*** (0.005)	−0.023*** (0.007)	−0.022*** (0.007)
$teachers_{it}$	0.109 (0.109)	0.113 (0.109)	0.014 (0.019)	0.014 (0.019)
$population_{it}$	−0.944*** (0.313)	−0.941*** (0.313)	0.332 (0.427)	0.303 (0.426)
$unem_{it}$	0.012 (0.042)	−0.016 (0.039)	−0.584*** (0.074)	−0.587*** (0.074)
$eurofonds^{educ}_{it}$	−0.009* (0.004)	−0.006 (0.005)	0.012 (0.008)	−0.601 (0.091)
$eurofonds^{educ}_{it+1}$	−0.008* (0.005)	0.002 (0.005)	0.018** (0.008)	0.006*** (0.002)
$eurofonds^{educ}_{it+2}$	−0.012** (0.004)	−0.003 (0.007)	0.016** (0.008)	−0.001 (0.002)
$eurofonds^{educ}_{it+3}$	−0.011** (0.004)	−0.002 (0.004)	0.015* (0.008)	−0.002 (0.001)
$eurofonds^{educ}_{it+4}$	−0.010** (0.004)	0.005 (0.004)	0.009 (0.008)	−0.005*** (0.001)
$eurofonds^{educ}_{it+5}$	−0.009* (0.004)	0.002 (0.004)	0.019** (0.008)	0.009** (0.004)
$eurofonds^{educ}_{it+6}$	−0.004 (0.006)	0.002 (0.009)	0.015* (0.008)	0.001 (0.005)
$eurofonds^{educ}_{it+7}$	−0.003 (0.005)	0.011 (0.008)	0.010 (0.008)	−0.009 (0.005)
$eurofonds^{reconstr}_{it}$	−0.003 (0.001)	−0.005** (0.001)	0.001 (0.004)	0.005 (0.004)
$eurofonds^{reconstr}_{it+1}$	−0.002 (0.001)	−0.006 (0.007)	0.0003 (0.003)	0.004 (0.006)
$eurofonds^{reconstr}_{it+2}$	−0.001 (0.001)	−0.001 (0.008)	−0.0004 (0.002)	−0.006 (0.007)
$eurofonds^{reconstr}_{it+3}$	−0.001** (0.001)	0.008 (0.005)	0.002 (0.002)	0.002 (0.004)
$eurofonds^{reconstr}_{it+4}$	−0.001** (0.0004)	−0.007 (0.005)	0.001 (0.002)	−0.003 (0.003)
$eurofonds^{reconstr}_{it+5}$	−0.002*** (0.0005)	−0.001 (0.006)	−0.0001 (0.001)	0.006 (0.005)
$eurofonds^{reconstr}_{it+6}$	−0.001*** (0.0004)	0.004 (0.011)	0.002 (0.002)	−0.024* (0.007)
$eurofonds^{reconstr}_{it+7}$	−0.001** (0.0004)	−0.011 (0.010)	0.002 (0.002)	0.006 (0.010)
Observations	8,301	8,301	3,093	3,093
R^2	0.017	0.016	0.063	0.063
Adjusted R^2	−0.191	−0.192	−0.205	−0.204
F Statistic	5.396***	5.133***	7.290***	7.390***
	(df = 22; 6853)	(df = 22; 6853)	(df = 22; 2405)	(df = 22; 2405)

Note: *p<0.1; **p<0.05; ***p<0.01