

UNEMPLOYMENT AND PUBLIC TRANSPORTATION

EVIDENCE FROM HUNGARIAN MUNICIPALITIES

By

Csaba Gábor Pogonyi

Submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of

Master of Arts in Economic Policy in Global Markets

Supervisor: Professor Gábor Békés

Budapest, Hungary

2014

Abstract

This thesis analyses the relationship between regional unemployment and transportation possibilities in Hungary, with focusing on public transportation. The first question it aims to answer is whether there is a significant link between transportation possibilities and unemployment rate. The second is interested in a policy problem: if the government wants to decrease regional unemployment rates by creating new transportation linkages, which type of connection should it choose.

Models for commuting are tested on a two time-period village-city connection pair database. According to the results, there is a significant negative relationship between the transportation possibilities of a settlement and its unemployment rate. Moreover, establishing new public transportation connections towards subregion centers lowers unemployment rates the most effectively. The thesis ends with a policy recommendation that would decrease unemployment rates in the most underdeveloped regions of Hungary.

Acknowledgments

I would like to take this time to thank the Central European University for all the help they were able to provide to me in order to make this thesis possible

I would also like to thank Professor Békés, as he provided me with many great ideas and guidance whenever it was required.

Moreover, I would like to express my gratitude towards Professor János Köllő, Balázs Muraközy and Gergely Köhegyi on shaping my economic thinking and thus making it possible for me to research this important topic.

Table of Contents

1	Introduction	1
2	Context for transportation and regional unemployment in Hungary	4
2.1	<i>Transport possibilities in Hungary</i>	4
2.2	<i>Public transportation possibilities in Hungary</i>	8
2.3	<i>Public transportation quality index</i>	11
2.4	<i>Regional unemployment in Hungary</i>	14
2.5	<i>Connection between unemployment and transportation</i>	18
3	Theoretical background	22
3.1	<i>International findings on the role of commuting</i>	22
3.2	<i>Hungarian findings on commuting</i>	25
3.3	<i>Research design</i>	30
4	Data description	34
4.1	<i>Defining economic hubs</i>	37
4.2	<i>Defining catchment areas</i>	38
4.3	<i>Defining relevant periods for commuters</i>	39
4.4	<i>Settlement types in Hungary</i>	40
4.5	<i>Unemployment rate</i>	42
4.6	<i>Transport possibilities</i>	43
4.7	<i>Public transportation</i>	44
4.8	<i>Control variables</i>	45
4.8.1	<i>Individual transportation</i>	45
4.8.2	<i>Employment</i>	46
4.8.3	<i>Active population</i>	47
4.8.4	<i>Income tax paid per active population</i>	48
4.8.5	<i>Border and Austria dummies</i>	49
5	Results	49
5.1	<i>Is there a connection between public transportation and unemployment?</i>	50
5.1.1	<i>Town-level findings for 2014</i>	50
5.1.2	<i>Town-level findings for the change between 2006 and 2014</i>	57
5.2	<i>Connection-level findings from 2014</i>	61
5.3	<i>Connection-level findings for the change between 2006 and 2014</i>	67
5.3.1	<i>Which connection types are the most important?</i>	71
5.4	<i>Validity of the results</i>	75
6	Conclusions	77
7	Policy recommendation	79
8	Appendix	83
9	References	87

List of Figures

Figure 1: Average unemployment rate by settlement-type (2011, unweighted, own graph)	17
Figure 2: Scatterplot showing the connection between the unemployment rate and the number of cities in the catchment area (own graph)	19
Figure 3: Scatterplot showing the unemployment rate and the number of cities reachable by public transportation (own graph)	20
Figure 4: Scatterplot showing the unemployment rate and the PTQI (own graph)	21
Figure 5: Histograms of the Number of connections (before 8 AM and less than 1 hour) for 2006 and 2014 (own graph)	37
Figure 6: Histogram of the number of cities within the catchment area (unweighted, own graph)	44
Figure 7: Histogram of unemployment rate in Hungary for 2011 (own graph)	84
Figure 8: Histogram of the change of unemployment (own graph)	84
Figure 9: Histogram of the number of cars per active population variable (own graph)	85
Figure 10: Histogram of the employment rate for 2014 (unweighted, own graph)	85
Figure 11: Personal income tax paid per active population in 2011 (own graph)	86

List of Tables

Table 1: Settlement types in Hungary	41
Table 2: Regression output for the town-level 2014 model	52
Table 3: Regression output table for the town-level FD model	58
Table 4: Regression output table for the connection-level 2014 model	65
Table 5: Regression output for the connection-level FD model - basic specification	71
Table 6: Regression output for the connection-level FD model - advanced specifications	73
Table 7: Basic statistics of the variables used	83

List of Maps

Map 1: Vácegres and its neighborhood: defining catchment areas (own map based on OSM, 2014).....	6
Map 2: Average number of cities within 40 minutes radius (average by subregion, unweighted, own graph)	7
Map 3: Number of reachable cities by public bus connection (average by subregion, unweighted, own graph)	10
Map 4: Public Transportation Quality Index (2014, average by subregion, own graph).....	13
Map 5: Unemployment rate in Hungary for 2011 (average by subregion, unweighted, own graph)	16
Map 6: Change of unemployment between 2006 and 2014 (average by subregion, unweighted, own graph)	43
Map 7: Employment rates in Hungary (average by subregion, unweighted, own graph)	47
Map 8: Expected change in local unemployment rate due to the policy recommendatioin (own graph)	82

1 Introduction

High unemployment rate is one of the most commonly discussed policy problems in Hungary. There is an agreement about the importance of the issue; however, the solution has been creating an ever-growing debate in the sphere of policymaking. This research aims to raise awareness to an important part of the problem that has not been thoroughly researched yet: the role of public transportation.

Subsidized public transportation lowers the cost of transportation, which in turn reduces the costs of commuting for both the employer and employee. It is a classical a state subsidy: the government provides affordable transportation for its citizens, so they are able to find work in faraway places. The importance of state subsidized public transportation is the highest in low-income regions where locals cannot afford to substitute to other means of transportation (usage of car in most of the cases). Good quality public transportation enhances worker mobility, which in turn makes it cheaper for firms to hire their employees. Therefore, by subsidizing public transportation, the government helps the creation of jobs and at the same time enhances the employment of locals.

Even though this issue seems to be as current as ever, it has not been researched thoroughly for years. Even the latest analyses used evidence from 2006, and they were concerned with the existence of the link between public transportation and commuting, no policy-related question was raised.

Today, it is an important question whether the existence of the link is still significant, and if it is, how government can intervene to enhance commuting.

This thesis aims to analyze the relationship between public transportation and regional unemployment, in order to verify the validity of the hypothesis: the quality of public transportation has a significant effect on unemployment. Moreover, I will analyse the question from a policy point of view: if the government aims to reduce regional unemployment connections towards which cities are the most important?

This thesis uses a settlement-level two time-period dataset that contains Hungarian public bus company (Volán) timetables for 2006 and 2014. After defining catchment areas, I analyse public transportation possibilities in Hungary, and based on these results, a theoretical framework for commuting is applied, focusing on policy-related issues. My thesis defines a relevant catchment area for every 3111 towns in the country. A catchment area contains all relevant economic hubs for a settlement. The research finds that the widest definition can be measured as 40 minutes ride by car: it means that the catchment area of 40 minutes in most of the cases includes every important nearby economic hub. This is the definition that I use in this research

The Volán data is able to show if it was possible to get to these cities on a workday morning between 6:30 and 8:00 AM, which is the most

relevant period for commuters. I use this measure to see if there is a significant effect of public transportation on the unemployment rate of the settlement. I concentrate on village-to-city commuting, and not on suburb-to-Budapest, as here commuting patterns are fundamentally different.

The thesis finds that there is a significant effect of public transportation on unemployment. Establishing one new public transportation connection is expected to lower unemployment in settlement by .13%. Moreover, subregion capitals (kistérség) have the highest potential in decreasing regional unemployment: connecting a settlement with a previously not connected subregion capital is expected to decrease unemployment rate by .4%.

Based on my results, the government should consider putting efforts into connecting every settlement in the country with at least one subregion capital. Launching a project, which aims to provide at least one fast transportation link to the nearest subregion centre could lower local unemployment rate by 1%.

The structure of this research is the following: first, there will be an overview on Hungarian transport and public transportation possibilities and their possible effect on unemployment rates. The second chapter presents the theoretical background and previous findings on this topic internationally and in Hungary. The third chapter includes data description, whereas the fourth presents the results of the various model specifications

and the validity of the results. The fifth chapter concludes and the sixth presents a policy recommendation.

2 Context for transportation and regional unemployment in Hungary

In order to understand the relationship between public transportation and unemployment, one has to start by understanding the role of transportation in employment possibilities. This chapter aims to provide an overall view about the situation in Hungary regarding transportation possibilities, more specifically public transportation and regional unemployment.

2.1 Transport possibilities in Hungary

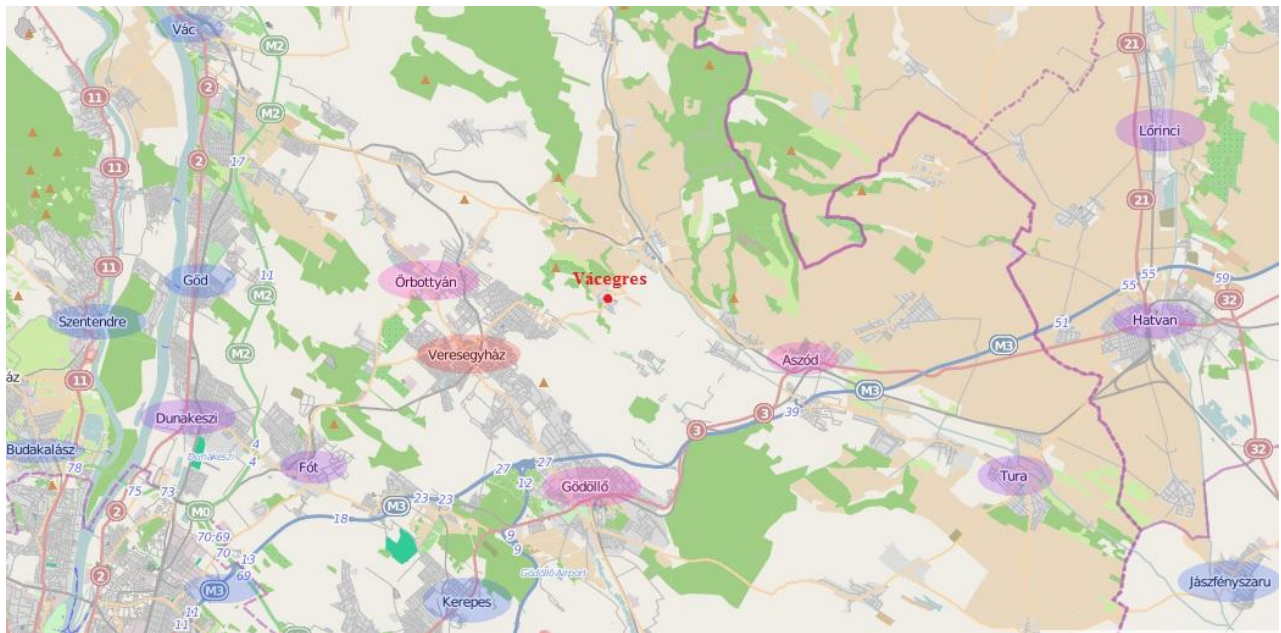
This chapter aims to provide a picture on the transport situation in Hungary (public and individual), with focusing on small town employment.

In Hungary, the most important instruments for transportation are roads. Even though the country has one of the densest railway system in Europe (Eurostat (2010)), mobility possibilities are determined by roads since they are used both by individual transportation means (car, motorcycle, etc.) and public bus companies, which have connection to every settlement in the county.

The public road system in Hungary is extensive; it is possible to reach every settlement by car. The quality is however volatile: the most important highways and roads are well-kept, but at the more backward regions of the country it is common that the maximum speed on the road is highly limited

due to geographical (mountains) and technical (potholes, etc.) reasons. This is why this study uses the average driving time as the main measure for distance instead of the length of the trip in kilometers. Average driving time is measured as if a normal car traveled from A to B abiding all laws. For example, driving from Miskolc to Debrecen (the second and the third biggest cities) the 113 km takes 63 minutes to drive, leading to an average of 108 km/h. However, from Miskolc to Bátonyterenye (small town in the mountainous North) takes 93 minutes, even though they are the same distance, leading to an average of 70 km/h. Driving time shows the distances that people encounter, this is why this research uses this measure for distance.

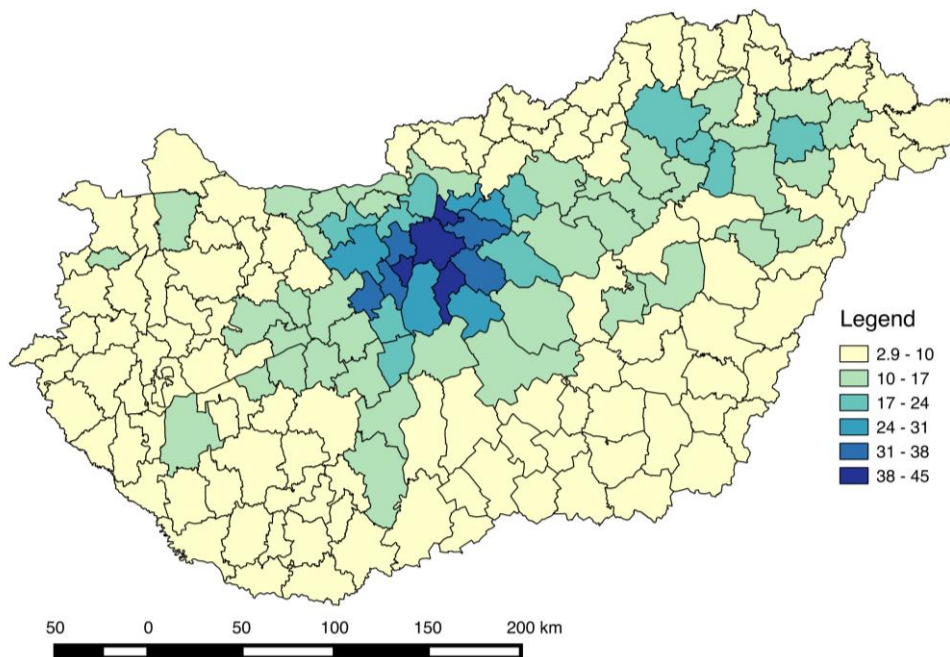
Since my thesis's interest is in regional transport possibilities, it is advisable to determine these possibilities for every town in the country. Using the ELTE TTK 2011 driving time matrix (ELTE TTK, 2014), I calculated the average driving time from every settlement to every settlement in the country. This makes it possible to define radiuses to study possible catchment areas. Map 1 shows the different radiuses for the village of Vácegres: there is only one city, Veresegyház within 10 minutes by car (indicated with red circle). There are three new cities within the 20 mins radius (rosy), five in the 30 mins (purple), and an additional 12 in the 40 mins (blue circles, not all of them are on the map).



Map 1: Vác and its neighborhood: defining catchment areas
(own map based on OSM, 2014)

Vác is above average in transport possibilities, as the country average for the 40 mins radius in 2012 is 8.99 (as opposed to Vác's 22). Map 2 shows the average number of cities within the 40 mins radius by subregions. It shows that there are huge differences in Hungary, with the center of the country being in good position. Budapest, the capital has extremely good transportation possibilities, with 45 cities around reachable within 40 mins. The reality is of course the other way around: there are 45 cities around Budapest whose inhabitants can commute everyday. These cities are in the metropolitan area of Budapest, which consists of Pest, and some parts of Fejér and Komárom-Esztergom counties. The extensive highway system around the capital created a region where daily commuting is possible from one end of the area to the other.

In addition to Budapest, the Eastern regions of Miskolc, Nyíregyháza and Debrecen are in good transport condition. Outside of this area, almost everywhere the average number of cities is below 10.



**Map 2: Average number of cities within 40 minutes radius
(average by subregion, unweighted, own graph)**

Is it a problem? If we are concerned with decreasing the unemployment rate, not necessarily. Certainly, more choices of cities where to commute decrease the risk of not finding a job, or insures against asymmetric economic shocks. However, for workers living in a small village, even one good connection to the regional economic hub can be enough to be employed. If we already have at least one connection, the intensity and the number of choices become important, however, in lots of cases in Hungary not even one connection is available.

A good example is the situation in the Cserehát mountain region, where no city is reachable by car within 30 mins. The same situation can be found in the Eastern end of the Fehérgyarmat subregion (Nagyhodos, Kishodos and Garbolc), close to the Romanian-Ukrainian border. Both of these regions are close to the border, therefore in the absence of border crossing locals can commute only towards the inside of the country. Moreover, due to first geography (mountains in Cserehát, rivers in Fehérgyarmat), roads cannot take the shortest way towards cities and are in bad condition. The result is that unemployment rate is one of the highest in country, with constant outward migration.

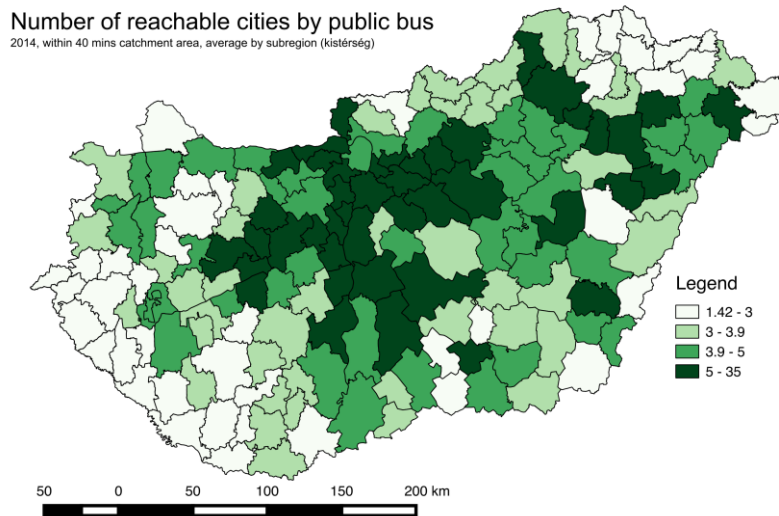
2.2 Public transportation possibilities in Hungary

The previous chapter showed the possibilities for individual transportation. However, due mainly to financial reasons, for most of the workers in the countryside, public transportation defines the real mobility possibility. The role of public transportation is crucial for those small villages, where there are a very limited number of jobs available locally, and the nearest economic hub is far away.

A typical situation can be shown through the village of Vácegres. Vácegres is in Pest county, within the subregion of Aszód with 849 inhabitants. Transportation possibilities to the subregion capital are crucial, since the place is not only a prime candidate for finding a job, but it is also a must to travel there every once in a while due to administrative affairs.

Aszód is 16 kms away, and can be reached comfortably by car within 14 mins. However, the only public transportation possibility is using the bus, which usually takes more than an hour to get there (from 51 to 71 mins), with a change in Galgamácsa, a village in-between. Moreover, if a worker finds a job in Aszód that starts at 7, the only bus arriving before that time gets there at 6:19; therefore, the worker has to wait 40 minutes until the work starts (jobs starting at 8 AM are somewhat better, with 30 minutes waiting time).

I downloaded the Hungarian public bus (Volán) timetable for May 17th, 2014 (Volán, 2014). This dataset includes every bus connection for that day in Hungary between 6:30 and 9:00 AM. It is not a problem as it was downloaded only on one day, as the Volán schedule does not change among weekdays. However, it is a very important limitation that I use only the Volán timetable; other means of transportation (train (MÁV) or ship) or not Volán city companies (Budapest (BKK), Debrecen, Miskolc, Szeged, Pécs and Kaposvár) are not included. I argue that as this research is most importantly concerned about the existence of public transportation (and its effect on unemployment), and not with the intensity of it, this data is sufficient for this task. There are no towns in the country that can only be reached by train (or ship), as opposed to Volán buses, which travel to every settlement.



Map 3: Number of reachable cities by public bus connection
(average by subregion, unweighted, own graph)

Map 3 shows the counterpart of the previous map (Map 2), it shows the number of cities reachable by public transportation (averaged for every subregion). We can see that even though the previous pattern cannot be so easily seen, it is still there: the Budapest metropolitan area and the Eastern big cities (Miskolc, Nyíregyháza and Debrecen) dominate, by having not just a good quality road system, but also an extensive public transportation system. However, this does not show the quality of the public transportation: it does not take into consideration the different importance of cities within a town's catchment area. A good connection to Budapest can mean a lot more than 5 connections to local small cities.

2.3 Public transportation quality index

I created a measure, which is able to show the quality of public transportation for every town in Hungary. It is important for this research, as the previous two measures (number of reachable cities on road and the number of reachable cities by public transportation) are not able to show the quality of public transportation as opposed to its possibilities. My thesis aims to find solutions that can decrease local unemployment rates on the medium run, without huge investment costs. The public transportation quality index shows where public transportation links are needed the most, without modifying the road network system in Hungary.

The idea is based on the famous gravity equation by Tinbergen (1962), which was later used by Persyn and Torfs (2012) to build a gravity equation for commotion flows. They claim that the most important factors that determine the commotion flow between A and B are the economic importance of the towns and their distance from each other. Based on their model, I calculated the Public transportation quality index (PTQI) the following way:

$$PTQI_i = \frac{\sum_j^n (C_j w_j)}{r_{40_i}} \quad (0.1)$$

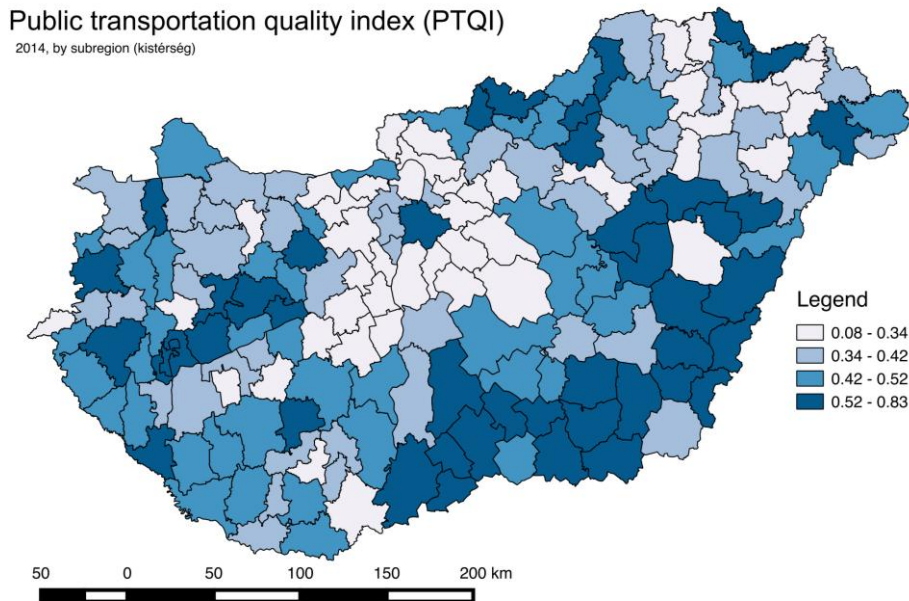
, where i denotes the town of interest, j denotes those cities that are within the 40 minutes by car radius, C is a variable that takes 1 if there is at least one relevant connection between i and j (weekday between 6:30 and 8, less

than 60 mins trip) and 0 otherwise, w is a weight that shows the economic importance of j (based on income tax paid in j), and r_{40} is the number of j -s reachable from i by car within 40 mins.

PTQI is basically the income tax weighted average of reachable cities by public bus divided by the total number of reachable cities (by car). Weighting with income tax is needed, as a bigger and richer city is more important than a smaller one. In this equation, I claimed that if there is at least one relevant connection between i and j , then j is reachable by public bus.

PTQI shows us the quality of public transportation for every town in Hungary. If it takes the value of 1, it means that all the cities within the 40 mins radius are reachable by public transportation. If it takes the value of 0, then none of the cities are reachable. If a town has a value in between, it shows how many percent of the economic hubs are actually reachable.

The value for Vácégres is 35% (which is exactly the average of the whole country), which means that 35% of all the possible economic activities are available by public transport for commuters. The other 65% are not, as there is no proper bus connection towards there. For this 65% only individual transportation can help.



Map 4: Public Transportation Quality Index
(2014, average by subregion, own graph)

Map 4 shows PTQI by subregions. We can see that what PTQI is good at is showing missed opportunities: low values can occur when there is a very important economic hub nearby, which is reachable by car, but no proper bus connections are available (this is why we see lots of white patches around Budapest and the cities of Miskolc, Nyíregyháza, Debrecen and Pécs).

In the case of the Budapest metropolitan area, we can say that this is not such an important problem, as most of the people commute anyway by car. Big road investments in the 2000s all aimed at making it easier for commuters to reach Budapest; however, they did not invest into the quality of public transportation.

Still, since people are generally richer here than anywhere else in the country, we can argue that commuters have a better chance of getting to work by car.

Therefore, the biggest problems are at those places, where the population is poor and there is no proper public transportation. If we take a look at the map, we can see that these subregions are more or less concentrated at the Northeastern part of Hungary, for example the subregion of Encs, where the previously mentioned Cserehát region is.

2.4 Regional unemployment in Hungary

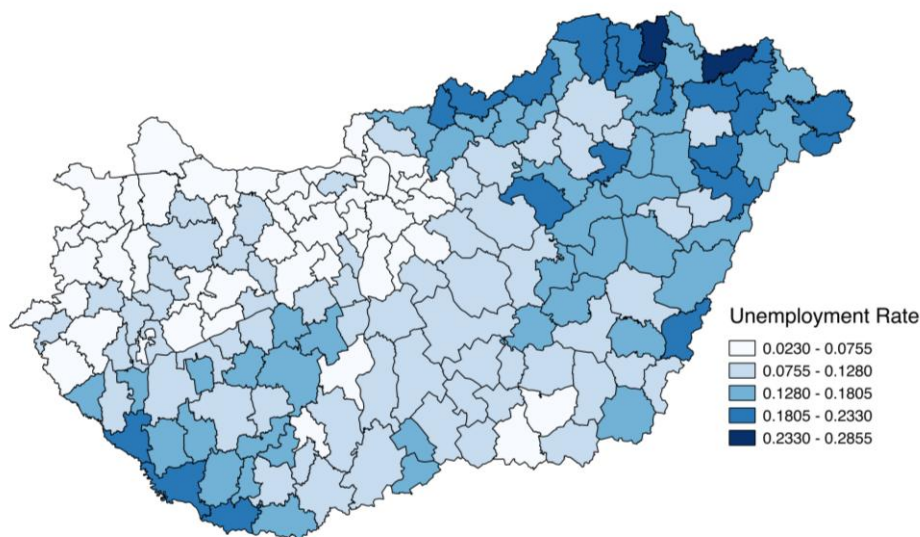
Mythesis studies unemployment on the settlement level. None of the statistical offices publish official figures on this level; therefore, I chose those variables that were published by Central Statistical Office T-Star database (ksh.hu, 2014). They publish data on the number of registered job-seekers for every settlement, and also about the number of active population (18-59 years old) in town. This definition means that we see only those who are registered job-seekers. Illegal employment is higher exactly at those places where unemployment is higher; therefore, there is an upward bias on the unemployment rate. However, there is an even more important bias that makes unemployment rate underestimated: for workers to be officially registered, they have to travel to the unemployment office, which is costly in itself; therefore, the further the unemployed individual lives from the

office, the lower is the chance of registration. According to Bartus (2011), this effect seemed to be more important both in 2006 and 2011.

Another important limitation is that the newest figures were available from 2011. As there was no other available source, I used these variables for 2014. According to the state statistical office (ksh.hu, 2014), unemployment rate fell since 2011; thus, by using the 2011 figures, I overestimate unemployment rate in my results.

I created the unemployment variable by dividing the number of registered unemployed workers by the number of active population in town (18-59 years). I accepted the classification of the state employment offices: unemployed is someone, who checks in at the employment office for unemployment benefits (KSH, 2014). In this research, all further references to unemployment are valid only for this group of people.

According to the database, the mean unemployment rate was 11% in Hungary in 2011, which is in line with the official statistics (ksh.hu, 2014). There is big variation in unemployment rate among towns in Hungary (see Figure 1 in Appendix). The highest unemployment rate was in the village of Fáj, Borsod county with over 50% rate (97 unemployed out of 187 active population). The lowest was in Harasztifalu, Vas county, just by the Austrian border with 1% (out of the 91 active workers, only 1 was unemployed). The unemployment rate of Vácegres is 10.7%, which is almost the average again.



Map 5: Unemployment rate in Hungary for 2011
(average by subregion, unweighted, own graph)

Map 5 shows the average unemployment rate by subregion in Hungary. As we can see, there is huge regional variation in the country. The Northwestern part and Budapest metropolitan area have a fairly low unemployment rate (less than 10% on average), whereas the Northeastern and some parts of the Southwest have very high rates (above 18% outside of county capitals). If we think back what we saw on map 2 and 3, we can see that exactly those places have very high rates of unemployment, where there is neither proper public transportation, nor road infrastructure. The reason why PTQI does not show these regions so clearly is because PTQI shows the quality of public transportation as opposed to the given road network. And in these regions there are huge deficiencies with the road infrastructure: complicated first geography (mountains and rivers), closeness of the border and decades of improper care for the quality of roads.

Another interesting pattern is that subregions, which have big cities behave like islands: Debrecen, Miskolc, Nyíregyháza and Szeged have significantly lower unemployment rates than their neighbouring subregions. If we take a look at Figure 2, we can see that on average, county capitals indeed seem to have a somewhat lower unemployment rate than villages or subregion capitals. Interestingly, cities that are neither county nor subregion capitals (usually small-sized) have relatively low unemployment rates.

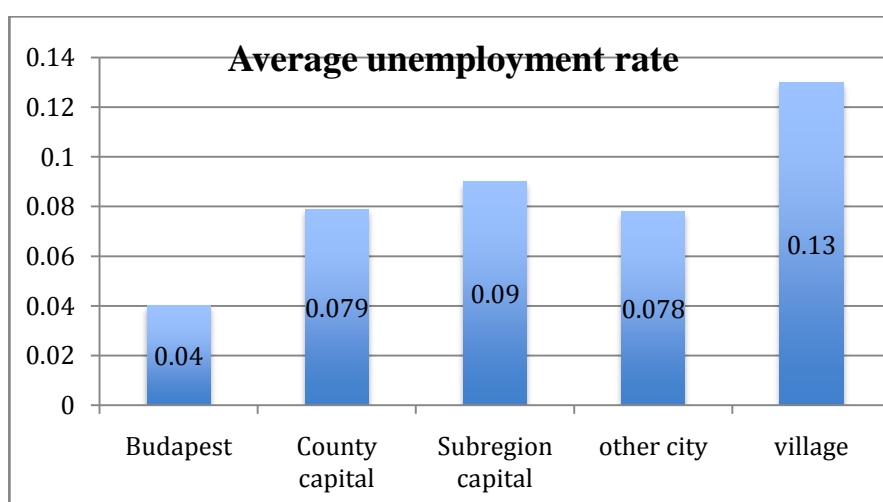


Figure 1: Average unemployment rate by settlement-type (2011, unweighted, own graph)

It is clear that there are important variations in unemployment rate among regions and also between settlement-types. What is more important, is that this variation does not seem to be random: we see high unemployment rates where there is a lack of transport infrastructure and we also see that the highest rates can be found in villages, where good transportation infrastructure is indispensable to create a link to a bigger job market.

2.5 Connection between unemployment and transportation

This chapter aims to provide a general picture about the relationship between transportation quality and unemployment. Based on our previous investigation, we suspect that there is a negative relationship between them: where transportation infrastructure is good (both road network and public transportation), unemployment rate is lower.

Figure 3 shows the relationship between the unemployment rate (vertical axis) and the number of cities reachable by car within 40 mins (horizontal axis). I also added the county averages for both variables. The elliptoid circle on the graph shows the places of towns in the Budapest metropolitan area. The overall picture is really interesting, as there seem to be a logarithmic relationship between the number of cities in the catchment area and the unemployment rate. However, if we define the Budapest metropolitan area as an outlier region (as this part of the sample is systematically different from the other side), then the whole pattern changes. If we look at only county averages, the slope completely disappears, whereas for town-level observation it becomes steep.

This means that a little improvement in the number of reachable cities can create significant decrease in the unemployment rate.

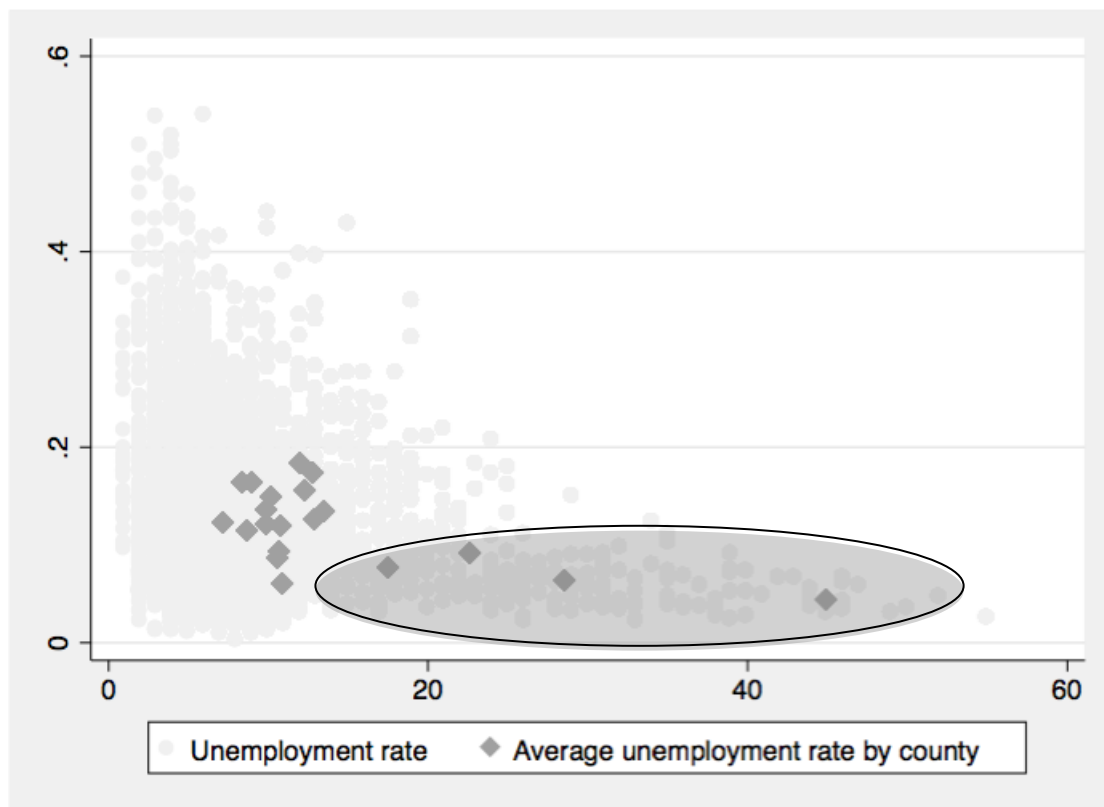


Figure 2: Scatterplot showing the connection between the unemployment rate and the number of cities in the catchment area (own graph)

Figure 4 shows the relationship between unemployment rate (vertical axis) and the number of cities reachable by public transportation (horizontal axis). Budapest was excluded from the sample as its 35 connections and 4% unemployment rate was clearly an outlier. All towns that have more than 14 connections are in Pest county; however, the county average is not so far from the other counties, so there it does not look like an outlier.

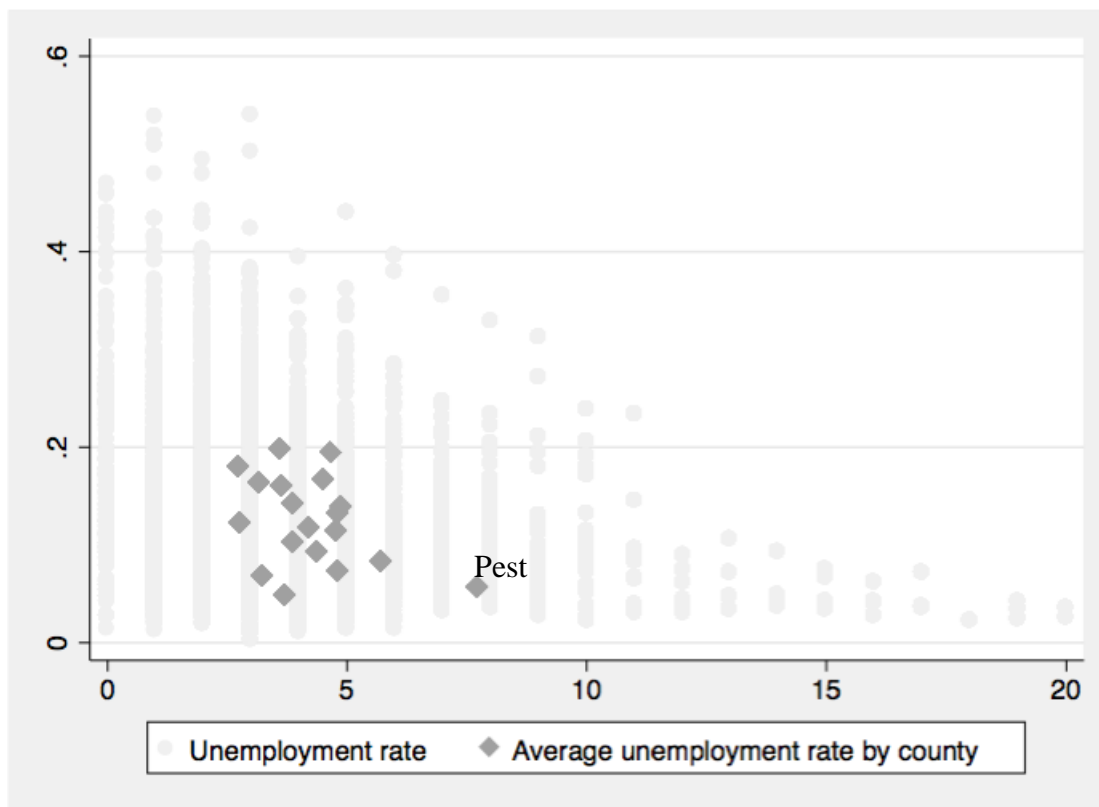


Figure 3: Scatterplot showing the unemployment rate and the number of cities reachable by public transportation (own graph)

As opposed to the previous Figure (3), here the pattern seems to be more evident: the higher is the number of reachable cities by public transportation; the lower is the unemployment rate. Moreover, the shape of the relationship looks logarithmic.

Figure 5 shows the relationship between PTQI and the unemployment rate. As we have already suspected from our previous findings, there seems to be no direct link between these two variables. Szabolcs, Borsod and Baranya counties (in the circle), where unemployment is the highest, are in the lower middle part of the PTQI distribution. It means that here there are lots of “missed opportunities”: with some public transportation investments

important economic hubs could become reachable for commuting. This is less true for Csongrád county for example, where PTQI is over 60% and the unemployment rate is relatively low.

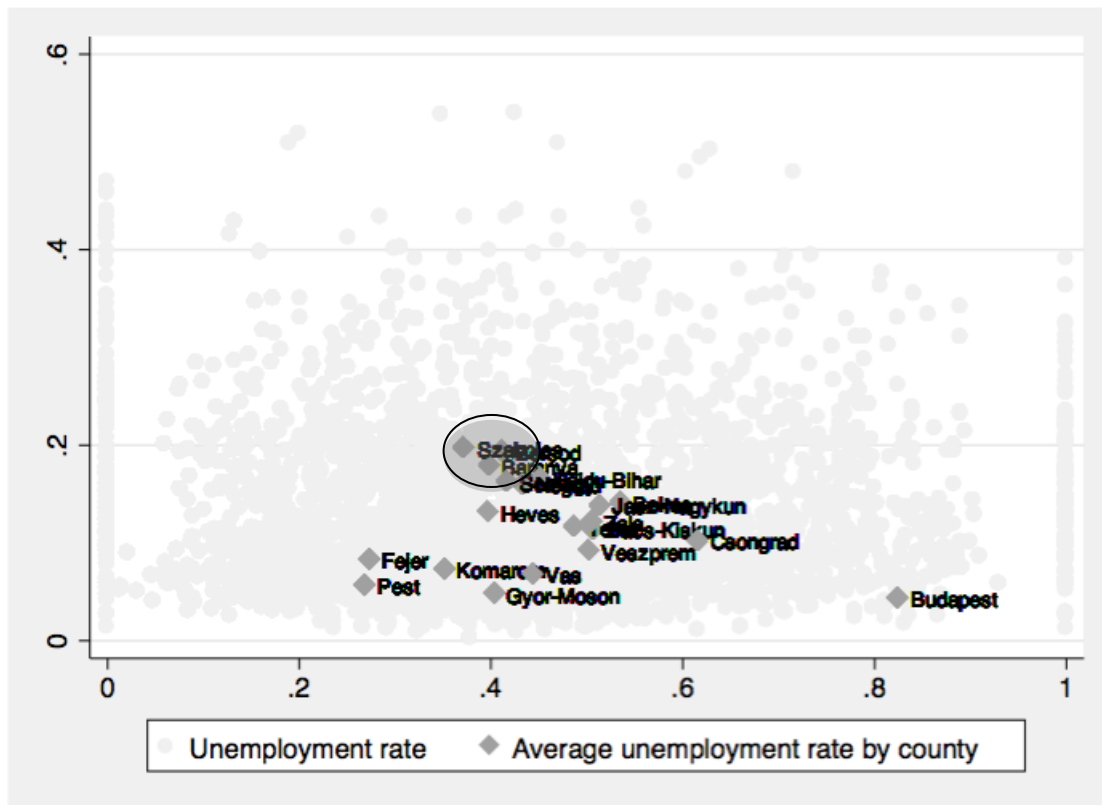


Figure 4: Scatterplot showing the unemployment rate and the PTQI (own graph)

This chapter showed that indeed the link between unemployment and transportation possibilities seems to exist. The link is rather unclear for road network (if we exclude Pest county as an outlier), whereas for public transportation possibilities a clear negative trend can be seen.

Where could these patterns come from? What are the underlying economic and/or geographical factors that shape the relationship between unemployment and transportation? The next chapter aims to build up a theory that is able to answer these questions.

3 Theoretical background

This chapter aims to provide a general picture about the possible causes and effects of mobility possibilities on regional unemployment. First, I will present the international literature about the relationship between unemployment and migration decision, further on I will focus on the Hungarian case regarding the link between commotion and unemployment. At the end of the chapter, I will present my model that is based on previous findings.

3.1 International findings on the role of commuting

If we want to understand the link between unemployment and commuting possibilities, first we have to study the migration decision of workers. This question has been studied extensively since the 1970s, as a new branch of the “economic imperialism” of Gary Becker and Jacob Mincer. Mincer’s “Family migration decisions” (1977) was the first to take into account both personal and family decisions in migration choices. These studies verified their findings on huge household-level panel data. The question of migration became important for policy purposes in the 1980s, when Pissarides & McMaster (1984) started focusing on the relationship between regional migration and unemployment in the UK. Ever since, most of the papers interested in this issue use extensive amount of empirics to back their findings and focus on actual policy solutions.

Pissarides & Wadsworth (1989) use the UK Labour Force Survey to test their modified human capital model (one which is similar to Köllő(2006)) on worker mobility. They find that unemployment can affect mobility at three levels. First, unemployed workers are more likely to move to another part of the country, as their incentive is higher than employed workers'. Second, a higher overall unemployment level lowers the probability of migration as wage differentials decrease on absolute terms. And third, households living in high unemployment regions are more likely to move to high wage regions.

If this is true then why are there still huge regional differences in Hungary? Already Pissarides & Wadsworth (1989) found the two most important factors that influence regional unemployment: workers living in high unemployment (and thus low income) regions have a harder access to capital markets and a low level of information about the job market.

These factors seem to be true not only for the U.S.; for instance Guriev and Andrienko (2004) found very similar effects for Russia between 1992 and 1999. In addition, they identified the existence of a poverty trap: migration is constrained by the lack of liquidity; therefore, an increase in income raises rather than decreases outmigration. Those who gain access to funding (the ones who are already in a better situation), use it to leave the region instead of passing on the positive effect. Their estimates show that up to a third of Russian regions are actually locked in such a poverty trap.

As already Köllő (1997) showed, Hungarian population is immobile: even though there are huge regional differences (as we have seen it in the previous chapters), we do not see high within-country migration rates. Unemployed workers living in rural areas have two typical choices outside of the migration: staying unemployed or commuting. The role of commuting possibilities is crucial in lowering regional differences and decreasing overall unemployment levels.

An interesting approach was developed by Ihlanfeldt & Sjoquist (1998). Their spatial mismatch hypothesis claims that in the ghettos of American metropolises unemployment is high, partly due to long distances from job opportunities: getting there is costly and moving there is impossible due to insufficient funds. This hypothesis is very close to Andrienko & Guriev's poverty trap theory, but it explains within-metropolitan area differences.

Regarding daily commotion, Persyn and Torfs (2012) built a model based on the famous gravity equation of Tinbergen (1962). Their approach is different from the previous researches', as their model is based on a simple spatial market structure, where commuter flows are perceived as similar to trade flows: the extent of the flow depends on the economic importance of the two endpoints (towns) and on the distance between them. They study the market distorting effect of barriers between regions in Belgium. They show that regional borders have a strong distorting effect on

commuter flows and also the direction of the border crossing has to be taken into account. They argue that border crossing (whether it is administrative or language border) works as a tax on commuting flows. Moreover, they argue that since the direction of the crossing is also accounted for, it is not the lack of infrastructure (whether road or rail) that seems to be the most important deterrent effect (as infrastructure can be used for travelling both ways). Instead, they argue that policy measures should focus on lowering administrative and other barriers (including public transportation), as these are the factors that do not work both ways. They point out that the highest gains are expected in depressed regions close to important economic hubs.

Persyn and Torfs's implications for Hungary is that we can expect the highest gains from a better public transportation in regions which already have the infrastructure to important economic hubs, however; proper public transportation link has not been yet installed. These are those regions where my PTQI measure is low (see Map 4).

3.2 Hungarian findings on commuting

Hungarian studies on this topic were mostly driven by data availability. None of the studies were able to use datasets containing individual-level data that also includes commuting variables. Still, starting from the end of the 90s, till the beginning of the 2000s, numerous studies were written about the relationship between municipality-level unemployment and the availability of urban labour markets in Hungary. Köllő (1997) points out

that the question is actually one of the oldest policy problems in the country: already in 1950, István Bibó, Ferenc Erdei and Jenő Mattyasovszky published their proposal, how to create an effective urban system in Hungary. Their objective was that every town in the country has to have the possibility to reach at least one city that is able to provide them “all the basic services of urban living” (Bibó-Mattyasovszky 1950:1 in: Köllő (1997), page 35).

Kertesi (1997) is interested in how differently regional income levels affect migration choices. He finds that unemployment levels are strongly linked with the educational level of the population and the average income level. At those regions where income level is low, already since the 1980s a strong outward migration can be seen. He found three main reasons for that: the effect of people who have already migrated, transportation possibilities and the deterrent effect of high unemployment. He finds that the highest outmigration rates can be found at places where transportation is poor and unemployment is high. This seems to back the existence of the poverty trap, introduced by Andrienko & Guriev (2004).

Moreover, Kertesi (1997) found that the reason for inward migration is not low unemployment itself, but higher local income levels. High income level mean high demand for products and services that emigrated workers can supply.

Köllő's (1997) motivation to examine the relationship between unemployment and transportation was the fact that in 1997, even within small distances there were huge regional differences in unemployment levels. We saw a similar pattern in 2014, after 17 years (see Map 5). Köllő argues that the reason is high costs of commuting. His paper focuses on the availability of transportation for unemployed workers. He argues that if cheaper and more effective commuting choices would be offered, regional disparities would lower. Moreover, due to the fact that on bigger markets frictional problems are smaller, the overall level of unemployment is expected to decrease.

Köllő finds that the “density of public transportation links” has a significant effect on unemployment. If we take two identical villages, and the first one has zero, whereas the other has 4 connections to the nearest economic hub, the unemployment rate differential is expected to be 5-6%. Moreover, he found that in the absence of public transportation, the cost of travelling by car is higher than the premium of receiving a higher salary in one of the nearest economic hubs.

Kertesi (2000) and Bartus (2003) worked on with the topic, and their papers are unique in Hungary, as they use datasets where the unit of observation is the individual. Kertesi (2000) used the 1996 microcensus data and found significant negative relationship between the cost of commuting

and the probability of it . He found that commuting costs have the biggest deterrent effect on low-educated workers.

Bartus's model (2003 and 2011), based on Köllő (1997) assumes that the reason why there is a persistent very high unemployment rate in certain regions of the country is due to high commuting costs. Here commuting costs are higher than the wage premium in the economic hub. His paper argues that previous researches miscalculated the costs of commotion. All of them assumed only an average value for costs, but they were not able to check on the individual level (due to data availability problems), whether these figures are actually true. He uses a dataset that contains survey results from 2002, for previously unemployed people who recently found a job (Köllő, 2002). The dataset contains information about the way of transportation that the worker plans to use. He found several important findings. First, almost half of the previously unemployed workers found a job at another location; therefore, they had to commute. Interestingly, only 20% of these people did not get any benefits for commuting (monthly pass, gasoline voucher, etc.). Second, the existence of a commuting cost radically lowers the probability of commuting: if the employer pays at least some benefits for commuting the worker almost surely takes the job; however, if there are no benefits, the probability of taking the job falls to 20-40%. Third, men usually got benefits if they had to travel further than 50 km. Fourth, men have a higher probability of commuting than women if there

are commuting costs. Therefore, commuting benefits have the biggest impact on women employment.

The last work published on this topic is a policy paper by Bartus (2011). According to their study, the last information about commuters is from the 2005 microcensus database. There were 1 221 000 commuters in 2005, which is the 33% percent of the more than 2.5 million workers in Hungary. In the most underdeveloped regions (Northeast Hungary and South Transdanubia), 55-62% of employees worked in another town, as opposed to the most-developed regions' 66-69%. We have to add that the workers in the most-developed regions were either commuting to Budapest (Budapest metropolitan area) or to Austria (regions close to the Austrian border). They found that even though there was a significant increase in commotion possibilities between 1994 and 2006, it did not have significant effect on unemployment. They argue that commotion possibilities improved due to the increase in minimum wage; however, this same act lowered the demand for labor, leading to no improvement on the market.

The methodology used in this paper is based on Köllő (2006), mainly due to data availability (I am using his 2006 database). He analyses the municipality-level unemployment rates and their changes between 1993 and 2001. He argues that an important endogeneity issue arises when we want to study the relationship between unemployment and transport possibilities: we cannot be sure which one has an effect on the other one. It is obvious that

transport possibilities have an effect on unemployment due to being costly; however, long-term unemployment levels (decades or centuries old) had/have an influence on where will roads be built or where will a good public transportation link be built. Köllő uses the instrument of the “percentage of Jewish population in town based on the 1941 census”: this variable has an effect on transportation possibilities (as Jews were usually living in places with good transportation), but no direct effect on current levels of unemployment.

Köllő’s most striking result is that unemployment rate differentials between integrated and remote settlements had increased since the regime change. He argues that it is not sure that by enhancing public transportation, unemployment levels would decrease (as it depends on other numerous factors); however, it is sure that it would demolish important mobility constraints that affect every local worker. He finds that only in the Cserehát region and in some parts of Somogy county can we identify such villages, where even a better infrastructure could not help, since there are no reachable economic hubs nearby.

3.3 Research design

This chapter presents the research design of this thesis. First, I present the basic one time-period settlement level model. Second, I introduce the gravity approach to mitigate the endogeneity problem by controlling for

more effect, and finally I present the final connection-level differenced model.

Due to data availability issues, instead of individual-level models, this research uses municipality and connection-level models. Based on Köllő (2006), the basic reduced-form municipality-level model is the following:

$$u_i = \beta_1 \mathbf{X}_i + \beta_2 F_i + \varepsilon_i \quad (1.1)$$

, where u_i is the unemployment rate of the settlement i , \mathbf{X} is a vector that contains control variables, and F shows how connected the settlement is within the catchment area (meaning both individual and public transportation). A catchment area includes every important nearby economic hub for every settlement.

We expect that F have a significant effect on unemployment: those settlements will have lower unemployment differentials that are more connected. As we have seen previously, there seems to be a significant difference between the existence of the infrastructure (road network) and public transportation possibilities. Therefore, F consist of two variables: one that shows how many important economic hubs are reachable by car and another one that shows how many of these hubs are actually reachable by public transportation.

The fundamental problem with this equation is that it can be heavily endogeneous: we cannot be sure that the reason why unemployment is low

is due to poor transport possibilities, or the other way around: due to historical unemployment rates (and low income levels), transportation infrastructure was neglected. One way to deal with this problem is by looking at only the change of these factors during a certain time period:

$$\Delta u_i = \beta_1 \Delta \mathbf{X}_i + \beta_2 \Delta F_i + \varepsilon_i \quad (1.2)$$

This way, we are able to partial out the “history effect”: those factors that have been affecting both unemployment and transport possibilities (like geography or former political decisions). Although we lose the potential to predict the level of unemployment, we gain the chance to test such events as a policy intervention aiming at improving transportation quality.

Persyn and Torfs (2012) show that commuting flows can be perceived as trade flows between two economic hubs. Therefore, connection-level data (one unit of observation is one origin-destination pair) can show us important details: as opposed to the town-level dataset, here the exact economic “weight” of both ends of the connection and their distance can be taken into account, which leads to a more precise estimation. In addition, one can check for heterogeneous effects of different connections: a new connection to the administrative center can have different effects than one that goes to a same-sized, but not-administrative center. Therefore, I created the following model, which incorporates Persyn and Torfs idea into Köllő’s model:

$$u_i = \beta_1 \mathbf{X}_i + \beta_2 F_j + \beta_3 I_i + \beta_4 I_j + \beta_5 D_j + \varepsilon_i \quad (1.3)$$

, where u_i is the unemployment rate of the settlement i , \mathbf{X} is a vector that contains control variables, F shows how connected the settlement is (meaning both individual and public transportation), j represent the destination city within the catchment area, I is the income of the settlement, and D is the distance between i and j . As the issue of endogeneity still holds, it is advisable to use differences again:

$$\Delta u_i = \beta_1 \Delta \mathbf{X}_i + \beta_2 \Delta U_i + \beta_3 \Delta F_i + \beta_4 \Delta I_i + \beta_5 \Delta I_j + \varepsilon_i \quad (1.4)$$

The next chapters present the data and the identification process. The first part explains the data that was used for the research, the second defines the sphere of economic hubs, catchment areas and relevant time periods for commuters, finally the fitted model is presented.

4 Data description

The first part of this chapter presents the four combined datasets and their shortcomings. I present the three important steps of the identification process: the definition of economic hubs, relevant catchment areas and the important time-period for commuters. The last part of the chapter describes all the variables used in the regressions.

The dataset was combined by four different data sources: road network data from ELTE TTK (2014), the TSTAR dataset from the Hungarian Statistical Office (KSH, 2014); I put together a dataset that contains the public bus (Volán) timetable for May 2014, and the Hungarian public bus (Volán) timetable from 2006 (MTA KTI, 2006). Below I will describe each of those data.

The first dataset contains all the distances in minutes by car for every village-city pair in Hungary. The limitation of this dataset is that it contains information only for 2011. This could be problematic, since transport possibilities could have changed due to the deterioration of existing roads or the construction of new ones. This is an important limitation; however, I argue that between 2006 and 2014 (the two studied years of the analysis) the only significant change in the Hungarian transportation systems was the construction of highways. These highways are important for those commuters who are travelling by car to the nearest county center (or

Budapest); however, for the vast majority of villages, overall transport possibilities for commuting were not affected.

The second dataset is the TSTAR, which contains general statistics about Hungary on the settlement-level. I used the years 2006 and 2011 for my thesis. As it was stated previously, there is an important bias in the figures as the 2011 data was used for 2014.

The third dataset includes the public transportation data for 2014. Based on 40-mins catchment areas, I downloaded the timetables for all the 27,350 settlement-city pairs (settlement*economic center within 40 mins). An important limitation of the dataset is that it only contains information about public bus services (Volán), and there is no information about railways and those city companies, where it is not a Volán company that provides services (Budapest, Debrecen, Miskolc, Szeged, Pécs and Kaposvár). I argue that for most of the towns in Hungary the public bus service is the main mean of transportation. Moreover, my thesis's main focus is transportation possibilities and not the actual intensity of a connection.

The fourth dataset includes public transportation data for 2006. Köllő (2006) used the same dataset and the Hungarian Academy of Sciences Economics Department was able to provide it for me. This dataset contains public bus (Volán) timetables on weekdays between 6:30 and 9:00 AM: for

every half an hour it shows the number of connections between town A and town B, and the length of the journey both in km and in minutes.

A crucial limitation of this dataset is that it includes only those connections that are less than 90 minutes by public transportation. It is a problem, since it can happen that there is an important city not far by car, however there is only a 90+ minutes bus connection available. Information about these settlement-city pairs would be important, as if we compare this timetable with the 2014 timetable, we see a huge difference between them: the number of settlement-citypairs within the 40 mins catchment area without any connection is more than double in the 2006 dataset. And most of those connections that are only present in the 2014 data are less than 90 mins long by bus.

It is unclear what could be the causes of this big difference: it could happen that indeed in 8 years service has improved this much or we can also think that there is a problem with the data. Whatever the true reason is, in order to mitigate this incompatibility issue, I decided to restrict both samples to only those connections that take less than 60 mins and arrive before 8 AM to their destination. These two restricted samples look more believable (see figure 6). Still, the number of connections is a lot higher for 2014 (“intensity of public transportation”). As my thesis’s focus is on the possibility of public transportation and not the intensity of it, this problem does not affect the end results significantly.

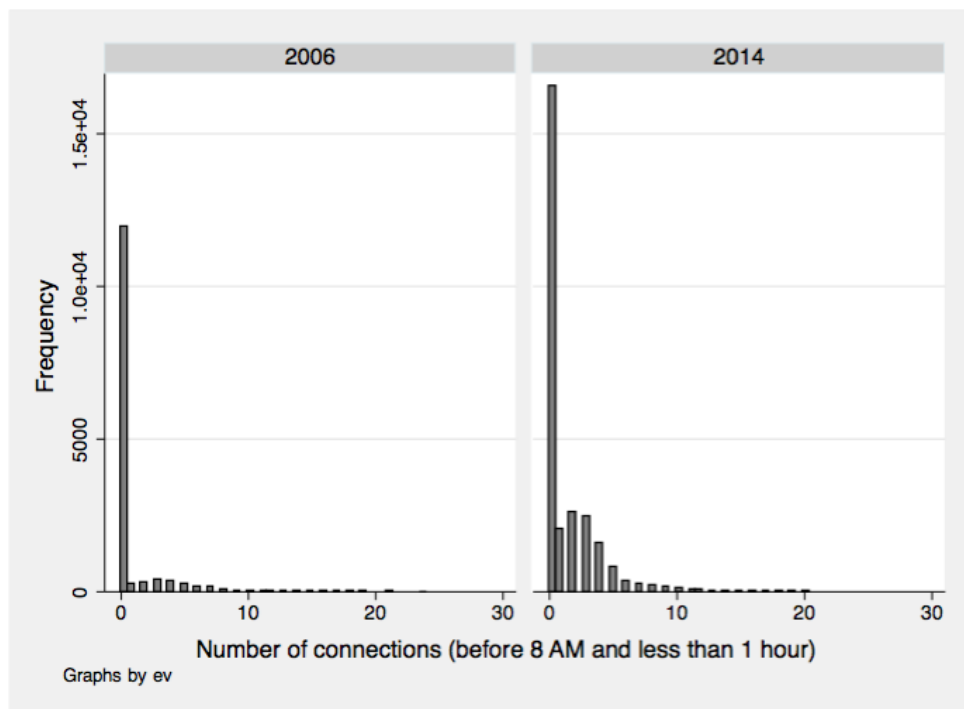


Figure 5: Histograms of the Number of connections (before 8 AM and less than 1 hour) for 2006 and 2014 (own graph)

The next three subchapters present the chosen classifications for economic hubs (cities), catchment areas and daily commuting periods. The last chapter presents the variables used for the model estimation.

4.1 Defining economic hubs

An important question is how we define relevant economic hubs. These hubs have to have significant economic activity that exposes such amount of labour demand, which cannot be satisfied with local population. In short, we are looking for towns that have at least some local industry.

Köllő (1997) defines all of those cities as economic hubs that have a labour office. In 1997, there were 170 of these towns. Kertesi(2000) uses the administrative definition of city:there were 247 cities in 2000.

This paper decided to use the administrative definition of city also.

There are villages that can have important economic effect on their neighbourhood, and there are cities that do not exert significant economic effect on their neighbourhood. Instead of using an arbitrary definition for economic hubs, I decided to stick to the simple choice of administrative classification. Since there are 346 cities in Hungary in 2014, I consider all 346 of them as economic hubs.

4.2 Defining catchment areas

Once we know which economic hubs we are interested in, we have to find for each settlement those hubs that may have significant commuting effect on workers.

Köllő (1997) defines those hubs as part of the catchment area, which are within 40 kilometers distance on road. He took into consideration only the closest four centers, as “within 40 km, finding a fifth center was very rare”. Kertesi (2000) took into consideration every hub that can be reached with maximum 4000 HUF costs for a month. The problem with Köllő’s classification is that it does not take into consideration first geography and road quality: 1 km in the mountains does not equal 1 km on the plain in terms of time or costs. Kertesi’s idea of defining an exact cost seems to be a good idea; however, almost every commuter meets a different cost: for those who travel by car it depends on the type and price of car and the price of gasoline, for those who use public transportation there are tens of

different discounts. Instead, I use the distances in minutes by car: it is simple and it controls for the quality of the road network variation.

I measured the distance by car between every settlement-city pairs based on the ELTE TTK database (2014). In practice, it means that I measured for every settlement, how many minutes does it take to travel to all the cities in the country by car. Then, it is a matter of decision to define the maximum distance that we think is relevant for the commuters in the settlement.

In this analysis, I chose this distance to be 40 minutes: longer distances would mean an almost 2 hours commute (gross travel time there and back), which is unlikely according to Bartus (2003).

4.3 Defining relevant periods for commuters

As this paper is concerned only with the transport possibilities of commuters, only those public transportation connections are relevant, which are important for commuters. We are interested in whether it is possible at all to travel to a city; therefore, looking at connections in the morning is enough: if there are no connections available, there is no public transportation link between them.

Köllő (1997) takes into consideration those bus and train connections between village and catchment cities, which are between 5:30 and 7:30 AM. If there was at least one connection, he defines the city as reachable. He argues that this way at least there is no first-order problem: if he defines a

city as unreachable by public transportation, it almost surely is. The second-order problem is not that huge either, since if there is only one connection, it arrives very early in the morning. It means that workers, who start work later, lose important time by waiting needlessly.

I accepted his classification with two changes: first, I took into consideration every connection after 6:30 that arrive before 8:00 AM. Second, due to the previously mentioned data problems, I restricted the sample to only those connections that take less than 60 minutes. I argue that connections that take longer than that are very low quality and in these cases cars become close substitutes.

The next subchapter provides descriptions for the used dataset. First, I present settlement-types in Hungary. Second, I present the independent variable used in the regressions: unemployment rate. Third, I present the two variables of interest: transportation and public transportation possibilities, and finally I present all the control variables used during the estimation process.

4.4 Settlement types in Hungary

In my thesis, there are five different settlement-classifications that are going to be used: Budapest, county capital (“megyeszékhely”), subregion capital (“kistérségi központ”), other city and village.

Table 1 contains all the important variables in settlement-type distribution for 2011 (2014 for number of bus lines). The average

unemployment rate decreases with settlement size: the average unemployment rate is the highest in villages (13.3%), and the lowest in Budapest (4.3%). Average active population is 620 people in villages, 4530 in other cities and 9028 in subregion capitals. County capitals are on average six times bigger than subregion capitals.

Table 1: Settlement types in Hungary

	Count	Unemployment average	Active population average	Income tax average	Employment rate average	Number of cars per active population average	Number of cities within catchment area average
Budapest	1	4.3%	998 918	325 858	73.1%	46.5%	45.00
county capital	18	8.1%	58 447	222 681	77.4%	47.7%	13.17
subregion capital	153	10.2%	9 028	166 093	74.8%	46.2%	12.53
other city	176	10.5%	4 530	152 445	71.6%	44.6%	17.89
village	2 769	14.3%	620	108 340	70.0%	42.4%	8.19

Average income tax paid also depends on settlement size: lowest in villages, highest in Budapest. Interestingly, this pattern breaks with average employment rate, as highest rates can be found in county capitals. Moreover, average employment rate even in subregion cities is higher than in Budapest. This might be due to tax evasion decisions: as I calculate employment rate based on the number of taxpayers, I miscalculate those workers who actually work for example in Budapest; however, pay taxes at other settlements.

The number of cars per active population is also the highest for county capitals on average. Budapest is only the second one: it can be due to

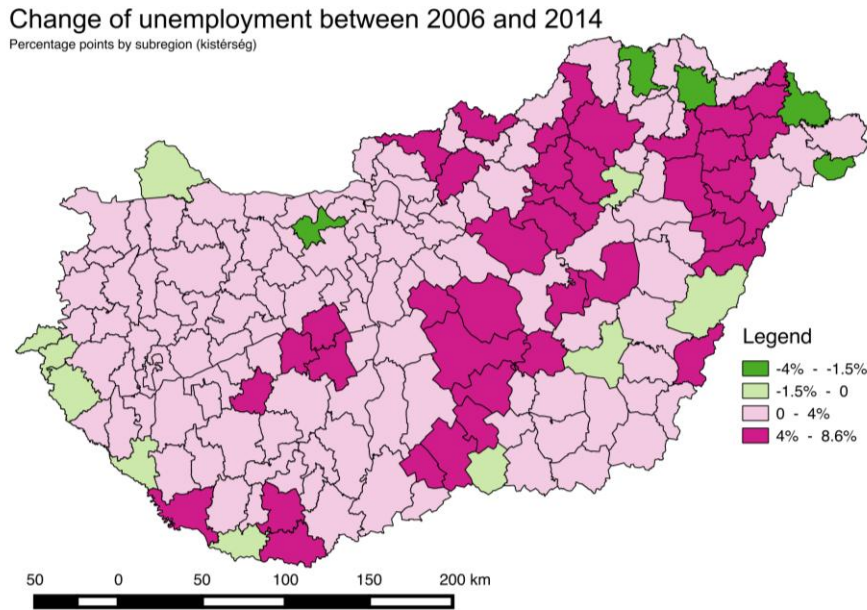
the fact that locals who work in downtown Budapest prefer using BKK services and they do not need a car.

The average number of cities within the catchment area shows an interesting pattern: “other cities” have more connections on average than county capitals or subregion capitals. This effect is mainly due to the fact that proportionally there are more “other cities” in the Budapest metropolitan area, where catchment areas are huge because of the extensive network of highways.

4.5 Unemployment rate

As it was mentioned previously, I used data from the TSTAR database to create unemployment rate by dividing the number of registered unemployed by the number of active population for every settlement. General statistics about the unemployment rate can be found in Table 2. Between 2006 and 2011; unemployment rate grew on average by 2.6 percentage points, standard deviation grew from .073 to .77 and the distribution of the change is fairly symmetrical (Figure 7 in Appendix).

There seem to be no obvious regional pattern if we take a look at Map 6. However, it is important to see that the Northwest part of the country did not experience big increase in unemployment rate, since most these big increases happened in Eastern part of Hungary. The biggest increase happened in the subregion of Záhony (+8.6%), whereas the biggest decrease occurred in the neighbouring subregion of Vásárosnamény (-4%).



Map 6: Change of unemployment between 2006 and 2014
(average by subregion, unweighted, own graph)

4.6 Transport possibilities

I have already presented transport possibilities in Hungary in chapter 2.1, here I shortly introduce the radius of 40 minutes variable that I use in my regression analysis (see Table 7 in Appendix for more details). The average number of cities within the catchment area is 9, the maximum is for Budapest (45) and there are no villages without at least 1 city in their catchment area. There are 29 villages where there is only one city in the catchment area (more than two-thirds of them are either in Borsod or in Szabolcs county).

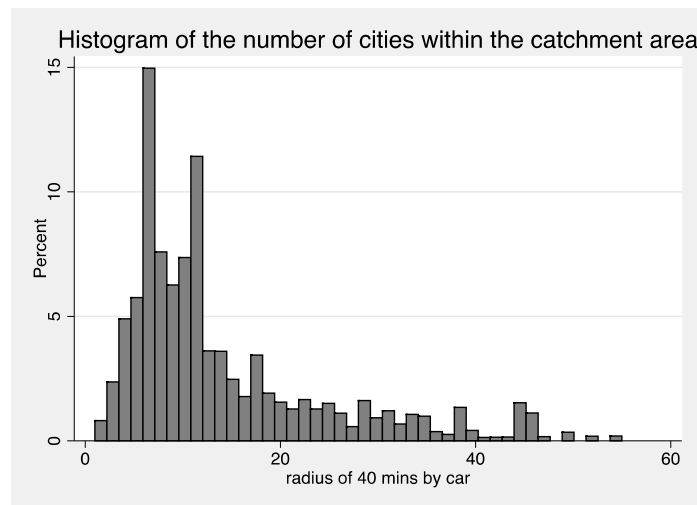


Figure 6: Histogram of the number of cities within the catchment area (unweighted, own graph)

4.7 Public transportation

In my analysis, due to the data availability problem discussed in the beginning of chapter 4, I was concerned only with the possibility of commuting from a settlement to a city within the catchment area.

The maximum number of cities reachable by public transportation has Budapest, with its 35 cities. There are 144 settlements without any public transportation link to their catchment area cities (22% of them are in Borsod, 16% in Somogy counties). On average, a settlement has 3.4 cities reachable by public transportation (see Table 2 in Appendix). If we compare this to the average number of 9 cities within the catchment area, we can see that there is a lot to improve in public transportation.

The average change of public transportation between 2006 and 2014 was 2.3, the biggest decrease was 3 (in three villages in South-Pannonia).

Only Pest county settlements went above +14 cities, the biggest increase outside of Pest happened in Sajóecseg (+13).

4.8 Control variables

This chapter shortly presents those variables, which are used in the upcoming regressions as control variables.

4.8.1 Individual transportation

In order to control for the possibility for locals to use individual means of transportation (mostly cars), an important control variable is the number of cars per active population variable. The average rate was .43 in 2014, the minimum .074 and the maximum 1.42. This 1.42 can be found in Debréte, where there were 10 cars for the 7 active people (not included in the analysis as it is clearly an outlier). The histogram of this variable can be seen in Figure 8 in Appendix.

The average change was small for this variable, only .001 between 2006 and 2014. The biggest negative change is -.35; the biggest positive is .571.

Similarly to the number of cars variable, I created a variable that is supposed to control for the effect of railway possibilities. The TSTAR database includes a variable that indicates if the settlement has a train station or not. I created a dummy variable “train station” that takes the value

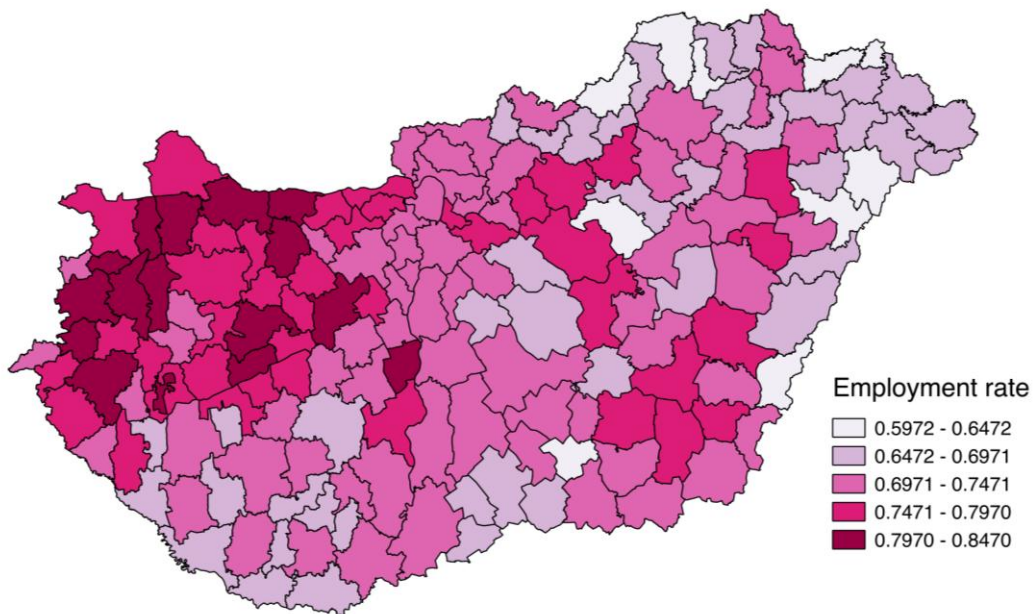
of 1 if the settlement has a train station, and 0 if not. 36% of the Hungarian settlements have a train station.

4.8.2 Employment

In my analysis, I use the employment rate as a control variable. This is needed, since it is needed to partial out the effect of employment rate due to public transportation quality: we are interested in the change of unemployment rate at a given employment level.

I call employment rate, the number of personal income tax payers divided by the local active population. This is a good approximation, as in most of the cases only those people pay income tax, which are actually officially employed. From one side, the problem of black employment arises (particularly in the underdeveloped regions): we count these people out-of-the-labour-force or unemployed (if registered at the state office). Moreover, I categorise people who are employed by public work programs as out-of-the-labour-force. Therefore, the employment variable has a strong downward bias. From the other side, however, this dataset does not contain information about the number of residents employed in the town. Instead, the number of taxpayers was used: only those people pay taxes who are employed; but there are people who are employed but do not pay taxes (most importantly household workers or illegal workers). Therefore, this proxy for employment also has an upward bias, and this one seems to be stronger.

The average employment rate of 71% is higher than the figure published by the Hungarian Statistical office by 10% (ksh.hu, 2014). As it can be seen on Map 7, there are important regional differences in the employment rate in Hungary. It was the highest in Northwest Hungary, and lowest in Borsod and Szabolcs counties.



Map 7: Employment rates in Hungary
(average by subregion, unweighted, own graph)

The average change in the employment rate between 2006 and 2014 was .052, the highest positive change was .69, whereas the negative was -.35 (both of them small villages).

4.8.3 Active population

In my research, I used the population between 18-59 as a measure for active population. Its distribution is very uneven, as Budapest, with its almost 1

million active population is an obvious outlier; however, there are also huge differences between county capitals and other cities, and the villages (see Table 1).

The average number of active population was 1908, the minimum 7 (the village of Debréte again). The change in active population was small, only -34.47 on average, with the biggest negative change in Budapest (-9910), and the biggest positive in Dunakeszi (+4340).

4.8.4 Income tax paid per active population

In order to control for the variation among settlements for wealth and economic importance, I used the personal income tax data given by the TSTAR database. I divide the total personal income tax variable with the number of active population, in order to get a measure that can be easily interpreted. A very important limitation of this variable is that it does not contain the tax paid by companies; therefore, it is only an indirect “proxy” for economic importance.

The yearly average income tax per active population was 128722 HUF, with the minimum values of 2357 HUF in Csenyété (Cserehát region, Borsod county), and the maximum of 664553 HUF in Iklánberény (Vas county, close to the Austrian border). Income tax per active population between 2006 and 2014 increased by 8652 on average, with the biggest decrease of 9816, and the biggest increase of 18934.

4.8.5 Border and Austria dummies

In order to control for the effect of borders, I created two dummy variables: the border dummy, which takes the value of 1, if the settlement is in a subregion that is adjacent to the border of Hungary, and the value of 0 otherwise. The border dummy also takes the value of 0 if the subregion is adjacent to the Austrian border, as for these settlements the Austria dummy takes the value of 1.

There are 41 subregions that are adjacent to the border, and 8 which are to Austria. For 34% of the settlements is border dummy =1, and for 6% of the settlements is the austria dummy=1.

5 Results

This chapter shows the results of the econometric models. First, I concentrate on the existence of the link between public transportation and unemployment. I start with town-level regressions for 2014, and then I move on to the first difference model in order to mitigate the previously presented selection bias. Second, I show the fitted model for connection-level 2014 dataset in order to answer the policy question: which public transportation connections are the most important in decreasing unemployment. Due to the selection bias, I present two connection-level first differenced models: one that tests the significance of a new connection on unemployment, and a second one that also shows which type of

connections are the most important. The last part of the chapter considers the validity of the results.

5.1 Is there a connection between public transportation and unemployment?

This chapter aims to test whether there is a statistically significant link between public transportation and unemployment. In order to test this, first I will use the town-level aggregated dataset, which contains observations for all Hungarian settlements in 2014. Second, I will use a dataset that includes the change of every variable between 2006 and 2014.

5.1.1 Town-level findings for 2014

Based on the equation outlined in the 3rd chapter, the basic equation tested is the following:

$$u_i = \beta_0 + \beta_1 \log(F_i) + \beta_2 \log(G_i) + \beta_3 \log(X_i) + \varepsilon_i \quad (2.1)$$

, where u_i is the unemployment rate in the settlement, F_i is the number of reachable cities within 40 minutes by car, G_i is the number of reachable cities within 40 minutes by public transportation, and X_i is a vector of control variables: employment rate, active population, income tax, the existence of train station in town, whether the subregion is by the border or by the Austrian border, and the number of cars in town. A dummy variable

for Pest county was also added to the regression to solve the lack of data for the Budapest public transportation company (BKK). The relationship between F_i and u_i and between G_i and u_i seemed logarithmic (see chapter 2.4); therefore, I am using the natural logarithmic form of these variables. I used the OLS method to estimate the equation, for 2967 observations in 2014.

The results of the regressions can be seen in Table 2. Column (1) shows the regression that includes only the variable “number of cities reachable by public transportation” in a natural logarithmic form. The coefficient shows that at settlements where the number of reachable cities is higher by 1%, unemployment rate is expected to be lower by 2.79%. This is a huge effect if we take into consideration that the average unemployment rate is 14%.

Table 2: Regression output for the town-level 2014 model

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
			Unemployment rate in 2014			
Number of cities reachable by public transportation (log)	-0.0279*** (0.00226)	-0.0142*** (0.00288)	0.000495 (0.00205)	-0.00166 (0.00209)	-0.00167 (0.00189)	-0.00211 (0.00187)
Number of cities reachable by car (log)		-0.0242*** (0.00321)	-0.000100 (0.00230)	-0.00239 (0.00234)	-0.00251 (0.00220)	0.0414*** (0.00599)
Income tax paid per active population(log)			-0.114*** (0.00208)	-0.116*** (0.00211)	-0.0632*** (0.00313)	-0.0565*** (0.00321)
Active population (log)				0.00449*** (0.000918)	-0.00193** (0.000942)	-0.00360*** (0.000956)
Train station					-0.000685 (0.00214)	-0.00105 (0.00212)
Employment rate					-0.143*** (0.0147)	-0.144*** (0.0146)
Border					0.0137*** (0.00213)	0.0111*** (0.00214)
Austria					-0.0465*** (0.00401)	-0.0430*** (0.00399)
Number of cars per active population (log)					-0.177*** (0.0121)	-0.170*** (0.0120)
C40 income tax rate (log)						-0.0408*** (0.00518)
Pest	-0.0620*** (0.00625)	-0.0434*** (0.00666)	-0.0149*** (0.00473)	-0.0167*** (0.00472)	-0.0237*** (0.00441)	-0.0150*** (0.00450)
Constant	0.170*** (0.00282)	0.202*** (0.00515)	1.451*** (0.0231)	1.452*** (0.0230)	1.060*** (0.0269)	1.474*** (0.0590)
Observations	2,961	2,961	2,961	2,961	2,961	2,961
R-squared	0.096	0.113	0.560	0.563	0.642	0.650

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The second regression includes additionally the variable “number of cities reachable by car” also in a natural logarithmic form. This variable is important, as we are interested in the effect of an additional public transportation connection at a given level of road infrastructure. Its coefficient is significantly different from zero, and it shows that one additional percent in the number of reachable cities on road lowers

unemployment rate by 2.42%, leaving all other factors constant. This effect shows that good location is important for a settlement: the more connections it has to nearby cities, the lower is the unemployment rate. The public transportation variable is still significant and negative; however, its coefficient halved.

Column (3) shows what happens with the regression if the most important control variable is inserted: income tax paid per active population in the settlement. The R-squared of the regression grew from .11 to .56 by inserting this variable. I use this local income as a proxy for the wealth of the settlement. Its coefficient shows that the richer the settlement is, the lower is the expected unemployment rate. The coefficients for the two previous variables became insignificant. It means that the effect that we have seen previously was partly due to the income effect: settlements that are wealthier, tend to have better road infrastructure and also better public transportation system. However, in turn, settlements, which have better road infrastructure and public transportation, tend to be wealthier.

This is an important endogeneity problem, which cannot be completely resolved. Regional income, road network and public transportation are all persistent over time; therefore, by analyzing their level at one moment in time one cannot identify their relationships. This is why the second part of this chapter tries to mitigate this problem by using a first difference model.

Column (4) adds the active population in natural logarithmic form to the equation. It has a significant positive coefficient, which means that the bigger is the active population at a settlement, the higher is the expected unemployment rate. Based on these results, unemployment rates are the highest in big and poor settlements.

Column (5) shows the regression where important additional control variables are inserted. A very important control variable is the local employment rate (Number of taxpayers in town/active population). By inserting this variable into the model, all other coefficients show marginal effects on unemployment at a given employment rate. It is important, as we can expect that transportation possibilities have an effect not just on unemployment, but also on employment, so it is advisable to partial this effect out. Employment's coefficient is significantly different from zero and negative: at settlements where employment is higher, we expect lower unemployment rate. The dummy variable train station takes the value 1 if there is at least one train station at the settlement, and 0 if there is none. This variable aims to control for the lack of data on train timetables in the dataset. Its coefficient is not significantly different from zero; therefore, the existence of a train connection does not seem to have an effect on unemployment rate.

The dummy variable Border takes the value of 1 if the settlement is in a subregion that is adjacent to the borders of Hungary (except for the

Austrian border), and 0 otherwise. One can suspect that the job market works differently in regions close to the border: there can be significant cross-border labor flow, which makes these job markets work differently from the overall pattern. Its coefficient is positive and significant: if a settlement is close to the border, the expected unemployment rate is higher by 1.37% on average, holding all other factors constant. The dummy variable of Austria takes the value of 1 if the settlement is in a subregion that is adjacent to the Austrian border. Here the coefficient is significant and negative: holding all other factors constant, a settlement is expected to have a lower unemployment rate if it is close to the Austrian border by 4.56%. The most likely reason for such an effect is that local workers also have an access to the extensive Austrian job market.

The number of cars per active population variable is used in a natural logarithmic form. I use this as a proxy for the local population's potential to use car for commuting. As we do not have data on the individual level, we do not know how many people use cars and how many use public transportation for commuting. In order to control at least somewhat for this deficiency, I use this variable. The coefficient is significant and negative: if there are more cars in a settlement, we can expect that local workers are relatively mobile and they have good possibilities for commuting.

Column (6) shows the results of the regression where the average income tax rate per active population of the catchment area cities is included

(in a natural logarithmic form). We can expect that if a settlement's catchment area is full of high income cities, local unemployment rate is going to be low. Workers have a good chance of finding a job with enough wage-differential to compensate for the costs of commuting. This hypothesis seems to be valid, as the coefficient is significant and negative: one additional percent of average catchment area income tax paid, lowers unemployment rate by 4% on average, holding all other factors constant.

We can see that the public transportation variable stayed insignificant after we have inserted the income tax of the settlement into the equation. As it was stated above, the most likely explanation for that is that there is an endogeneous relationship between these two variables. Local income has an effect on both public transportation (and transport infrastructure) and unemployment rate. The current research design is not capable of differentiating these effects.

In such a situation, one can think of two solutions. One solution is using an IV method by finding an instrumental variable, which has an effect only on public transportation, but has no direct effect (indirect through public transportation) on unemployment rate. Köllő (2006) chose this option and used the percent of local Jewish population in 1941 as an instrumental variable. This thesis presents another possible solution: analyzing the change of the variables. The next subchapter presents the results for this analysis.

5.1.2 Town-level findings for the change between 2006 and 2014

This chapter aims to test whether a new connection for a settlement that previously had no connection at all has a significant effect on the settlement's unemployment rate. Based on the equation outlined in the 3rd chapter, the equation fitted here is the following:

$$\Delta u_i = \beta_0 + \beta_1 K_i + \beta_2 \log(X_i) + \varepsilon_i \quad (2.2)$$

, where Δu_i is change of unemployment rate in the settlement, K_i is a dummy variable that takes the value of 1, if the settlement had no public transportation connections to its catchment cities in 2006, but had at least one in 2014. Otherwise, it takes the value of 0. X_i is a vector of control variables: change of income tax paid, change of active population, change of employment rate, change in the number of cars per active population, the change in the average tax rate of the cities within the catchment area, and the change in the number of cities reachable by public transportation. As I have data about travel times by car only for one time period (2011), I was not able to control for the change of transportation infrastructure between 2006 and 2014. However, as I have stated previously, I argue that local commuting possibilities have not changed very much during these years, as the most important investments were for highways, which have their biggest

effect on inter-city travel, and not on everyday commuting. The equation was tested on all the 3111 settlements of Hungary by an OLS method.

Table 3: Regression output table for the town-level FD model

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Change in unemployment rate between 2006 and 2014				
0 -> N no. of cities reachable by public transportation	0.00403** (0.00185)	0.00402** (0.00186)	0.00411** (0.00186)	0.00369** (0.00186)	0.00334* (0.00186)
Income tax paid per active population (log, diff)		9.90e-05 (0.00158)	-0.000484 (0.00159)	0.000277 (0.00159)	-0.0105*** (0.00352)
Active population (log, diff)			-0.0294*** (0.0101)	-0.0450*** (0.0130)	-0.0485*** (0.0130)
Employment rate (diff)				-0.0450*** (0.0106)	-0.0421*** (0.0106)
No. of cars per active population (log, diff)				0.0142 (0.0192)	0.0183 (0.0192)
C40 tax rate (log, diff)					0.0137*** (0.00397)
Average unemployment rate in subregion (diff)					
Constant	0.0209*** (0.00156)	0.0202* (0.0108)	0.0237** (0.0108)	0.0209* (0.0108)	0.00271 (0.0120)
Observations	3,111	3,111	3,111	3,111	3,111
R-squared	0.002	0.002	0.004	0.010	0.014

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The first column shows the results if only the variable of interest is regressed. We can see that in settlements, where the number of cities reachable by public transportation increased, unemployment also increased. The coefficient of the variable stays around the same in all seven specifications: in cities, where the number of cities reachable by public transportation grew from zero, the expected change of unemployment is +4%. This result has two possible interpretations: first is that settlements that acquired at least one good quality connection are worse off, as their

unemployment rate grew between 2006 and 2014. It is hard to think of such a setting; therefore, the second interpretation is more likely to be true: settlements where unemployment rate decreased, managed to improve their public transportation system. An example can be the case when a factory is built within the catchment area of the settlement, hiring workers from the region, and with the joint effort of the company and local municipalities, public transportation links are built up. As we can see, by analyzing the change of the variables we were not able to clear away the endogeneity problem completely.

The second column shows the regression where the change in the local income tax is also included. As opposed to the one time period model, here we can see that its change is not even significant. This effect shows that indeed, the “history effect” was very strong in the one time period model. According to this specification, the change in wealth did not have an effect on the change in unemployment rate.

The third column shows what happens we insert the change in the active population into the equation. The coefficient is significant and negative: in settlements, where active population grew by 1%, unemployment rate decreased by 2.9%, holding all other factors constant. Therefore, we expect high unemployment rate increase in small settlements.

The fourth column shows the regression extended with the change in employment rate and the change in the number of cars per active population.

The change in employment rate has a strong negative effect: the higher is the change in employment rate, the higher is the decrease in unemployment rate. The change in the number of cars did not have a significant result.

The fifth column shows the equation with the change in the average income tax paid in the catchment area cities (in a natural logarithmic form). According to the results, the higher is the change in the catchment income, the more grew the change in unemployment. This is interesting, as we might expect the link the other way around: settlement, whose catchment area cities “grew richer” experience an unemployment decrease. Here, we can suspect again the endogeneity problem: the change of income tax paid in the catchment area has a direct effect both on the change in the unemployment rate and also on whether there is going to be a new public transportation link built.

The last regression has the change in the average unemployment rate in the subregion as additional variable. This variable has a strong effect, as it works as a regional fixed effect variable: it partials out some of the factors that we were not able to control for. It is important that most of the coefficients of the other variables did not change much. This shows the robustness of the results. The effect of the change in the active population and in employment rate decreased somewhat, but the only big difference is that the change in the number of cars per active population became significant and positive.

Summing up the results from these specifications, we can see that there seem to have a connection between public transportation and unemployment. Due to the fact that it seems that we were not able to partial out the selection problem from the regressions, we cannot correctly predict the sign of the connection between unemployment rate and public transportation.

However, as it was elaborated on in the 3rd chapter, it is possible to study this relationship on the connection-level. It is important, as it can happen that different types of connections have different effect on unemployment. The next subchapter includes the results of these specifications.

5.2 Connection-level findings from 2014

This subchapter first presents the fitted model on the connection-level database for 2014, then it presents the results of the different specifications. Being connection-level means that for every settlement in the country I identified those cities that are within 40 minutes by car (within the “catchment area” of the settlement). The median number of cities within the catchment area is 10. I identified 27,350 connections in Hungary (for all the 3111 settlements). These connections are not symmetrical, as the public transportation system does not work symmetrically: it can happen that there is a regular bus line from city_A to city_B, however, there is no bus connection from city_B to city_A. Using this dataset, we are able to control for more

factors, namely, we can incorporate the “gravity-model” from the 3rd chapter. Moreover, we are able to answer the raised policy question: if the government decides to decrease unemployment rate by installing new public transportation links, which type of connections should it choose to achieve the highest change. In order to test this question, the research tests four types of connections: the effect of public transportation connections towards Budapest, a county capital, a subregion capital and other cities.

The fitted model for the connection-level dataset can be the following:

$$u_i = \beta_0 + \beta_1 PT_{ij} + \beta_2 \log(X_i) + \beta_3 \log(I_i) + \beta_4 \log(I_j) + \beta_5 \log(P_i) + \beta_6 \log(P_j) + \beta_7 D_{ij} + \varepsilon_i \quad (2.3)$$

, where u_i is the unemployment rate in the settlement, X_i is a set of control variables: number of cities reachable by car, train station, employment rate, border and Austria dummy and the number of cars per active population. I_i and P_i are the income tax paid and the active population in the settlement, I_j and P_j are the income tax paid and the active population in the destination city. D_j is the distance in minutes by car between i and j . It is a notable limitation that due to data availability the tax paid by local companies is not taken into account in the Income variable. Due to the catchment area classification, D 's maximum value is 40. As I had travel time data only for 2011, I estimate with the same catchment areas for 2006 and 2014.

PT_{ij} is a dummy variable, that takes the value =0 if the city j is in the catchment area, but there is no public transportation from i to j , and takes the value =1, if there is at least 1 public transportation link between them. If $PT_{ij}=1$, there are four different connection types that my thesis is concerned with: whether the connection is towards Budapest (80 links), a county capital (1,024 links towards 18 cities), a subregion capital (5,815 links towards 153 cities) or other type of cities (3,908 links towards 176 cities).

The model was estimated by an OLS method for 27,350 observations, with clustered standard errors for the 3111 settlements. A dummy variable for Pest county was added to every specification to control for the lack of data on Budapest public transportation system. The results can be seen in table 4.

Column (1) shows the regression when only the PT_{ij} dummy is present. We can see that on average, the expected unemployment rate is 0.4% lower by every public transportation connection towards cities in the catchment area. According to this result, if a settlement has three public transportation links, it is expected to have 1.2% lower unemployment rate than settlements without any public transportation link.

The second specification includes the three dummy variables that control for the different types of destinations. The connection to Budapest dummy equals 1, if the connection has public transportation link and it travels to Budapest. Therefore, the interpretation of the coefficient is the

following: if there is a public transportation connection to an “other” city, its expected effect on unemployment is shown by the coefficient of PT_{ij} . For the effect of a public transportation connection to Budapest, we have to add up the coefficient of PT_{ij} and the “connection to Budapest” dummy’s coefficient. County capital and subregion capital variables work similarly. As we can see, only the connection to Budapest dummy became significant from the three connection-type variables. It means that on average, we expect a 1.2% unemployment decrease for those settlements that have exactly one public transportation connection, and this one is towards Budapest.

Column (3) presents the model if we add the number of cities reachable by car variable. This variable here controls for the overall transportation possibilities of the settlement: its geographic position (lots or not many cities in the catchment area) and the existing road infrastructure. According to this regression, if a settlement has more cities in its catchment area (higher number of cities reachable by car), it is expected to have a lower unemployment rate, holding all other factors constant. Controlling for this effect made the coefficients of connection to a county capital and subregion capital significantly different from zero and negative. A connection to a subregional center lowers the expected unemployment rate by 0.7%, as opposed to a connection to an “other” city.

Table 4: Regression output table for the connection-level 2014 model

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
			Unemployment rate in 2014			
Public transportation	-0.00417*** (0.00119)	-0.00425*** (0.00157)	-0.00411*** (0.00144)	0.00175* (0.00103)	-0.000837 (0.000873)	-0.00289*** (0.00103)
Connection to Budapest		-0.00815*** (0.00246)	-0.00272 (0.00214)	0.00484** (0.00218)	0.00526*** (0.00169)	0.0109*** (0.00257)
Connection to a county capital		-0.00184 (0.00230)	-0.00744*** (0.00217)	-0.00402*** (0.00147)	-0.00394*** (0.00126)	0.00168 (0.00168)
Connection to a subregion capital		0.000615 (0.00154)	-0.00703*** (0.00139)	-0.00331*** (0.00100)	-0.000667 (0.000864)	0.00130 (0.000938)
Number of cities reachable by car (log)			-0.0352*** (0.00210)	0.00537*** (0.00177)	-0.00345* (0.00182)	-0.00212 (0.00179)
Income tax paid per active population (log)				-0.110*** (0.00251)	-0.0565*** (0.00353)	-0.0540*** (0.00352)
Active population (log)					-0.000578 (0.000811)	-0.00116 (0.000811)
Income tax per active population (log, destination)						-0.0130*** (0.00171)
Active population (log, destination)						0.000255 (0.000405)
Travel time by car (mins)						-8.16e-05** (3.25e-05)
Control variables	Only Pest	Only Pest	Only Pest	Only Pest	YES	YES
Constant	0.133*** (0.00173)	0.133*** (0.00173)	0.214*** (0.00477)	1.396*** (0.0278)	1.000*** (0.0297)	1.127*** (0.0318)
Observations	27,350	27,350	27,350	27,350	27,350	27,350
R-squared	0.129	0.129	0.192	0.607	0.692	0.695

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The fourth column includes the regression where I added the income tax variable. As before, here also this variable has strong explanatory power, the R-squared grew from .19 to .61. The coefficient of public transportation changed, now ceteris paribus one additional public transportation link to an

“other” city is expected to increase unemployment rate by .18%. This is due to the fact that the results seen in the previous regression were strongly connected to the fact that richer settlements have also better transportation links. Here this effect was partialled out. Still, an additional connection to a county capital or to a subregion capital lowers the expected unemployment rate ($0.18 - 0.4 = -0.22$ and $0.18 - 0.33 = -0.15$).

Column (5) shows the regression model where active population and all the control variables are included (train station, employment rate, border and Austria dummy and the number of cars per active population). We can see that these variables have some explanatory power as the R-squared grew by .08. The effect of the public transportation dummy became insignificant, together with the connection to a subregion capital dummy.

The last column (6) shows results from the model where not just all the control variables, but also the “gravity-model” is included. The coefficient of the destination city’s income is significant and negative: the higher income the destination city has, the higher is the negative effect on unemployment rate (holding all other factors constant). The active population of the destination city did not become significant; however, the distance is significant and negative: if a city is further away by 10 minutes, the existence of a public transportation towards it is expected to lower unemployment rate in the settlement by .08%, holding all other factors constant. This finding is important, as it shows that those commuting

variable costs that depend on the distance of the commute (like gasoline costs), do have a significant effect on unemployment.

The coefficient of the public transportation dummy became significant and negative; however, only the connection to Budapest dummy stayed significant. It means that according to these results, there is no significant difference between the effect of connections towards county capitals, subregion capitals or “other” cities.

Summing up the results from these specifications, we can see that the existence of a public transportation link seems to have an effect on unemployment rate. The gravity-approach showed that connections towards destination cities, which have high local income, are expected to lower unemployment rate in the settlement.

As we have stated previously, these regressions probably suffer from important endogeneity problems: public transportation has an effect on unemployment rate; however, unemployment (and local income) has an effect on public transportation. The next chapter aims to address this causality problem by using the changes between 2006 and 2014.

5.3 Connection-level findings for the change between 2006 and 2014

This chapter shows the final regressions of my thesis. I present the fitted model and then the results in two parts: first the connection-level differences

model for the existence of a link between public transportation and unemployment, and second for the different connection-types.

As it was previously mentioned, the main motivation for using the differences of the variables is that this way we can mitigate the endogeneous effect between unemployment and public transportation. In these specifications, our main interest is what happened in those settlements, where a new connection was established between 2006 and 2014. The following model was estimated on the connection-level data:

$$\Delta u_i = \beta_0 + \beta_1 \Delta \log(I_i) + \beta_2 \Delta \log(I_j) + \beta_3 \Delta \log(P_i) + \beta_4 \Delta \log(P_j) + \beta_5 \Delta \log(C_i) + \gamma PT_{ij} + \varepsilon_i \quad (2.4)$$

, where u_i is the unemployment rate in settlement i , I_i and P_i are the income tax paid and the active population in the settlement, I_j and P_j are the income tax paid and the active population in the destination city j , and C_i is the number of cars per active population for settlement i . All other previous control variables (including travel time by car) were constant in time; therefore, they are not included in the fitted model. As I had travel time data only for 2011, I estimate with the same catchment areas for 2006 and 2014. The equation was run on all the 27,350 observations, using the OLS method with clustered standard errors for all the 3111 settlements.

The regression results from can be seen in table 5. The first specification shows only the effect of a new public transportation connection on the change in unemployment rate between 2006 and 2014. This variable takes the value of 1 for all the settlement-city pairs within the catchment area, where in 2006 there was no public transportation, but in 2014 there was at least one. Therefore, in this regression it shows the expected change in the unemployment rate if a completely new connection was established between 2006 and 2014. The coefficient is significant and negative: at settlements, where there was one new public transportation established, the unemployment rate is expected to decrease by .17%. It is also has to be noted that even though the variable of interest is highly significant, the R-squared of the regression is very low (.000).

The second column shows the results if the change in income tax, active population, employment rate and number of cars are included. The coefficient of a new connection did not change significantly (-0.18%) and it still significantly different from zero. The change in income tax paid did not become significant, which shows that the income effect from the cross-section analysis was successfully partialled out by differencing the dataset. The change in active population became significant and negative: settlements where active population grew between 2006 and 2014 are expected to have a lower unemployment rate, holding all other factors constant. This effect can show the migration effect: settlements, which were

able to attract more workers were also able to attract companies and thus decrease unemployment rate. The change in the number of cars per active population became also significant and negative. This effect is also important for our topic, as we use this variable as a proxy for the local workers' possibility to commute by car. This significant and negative effect shows that those settlements were able to decrease their unemployment level, where local population had a higher chance of commuting by car, and thus finding a work outside of the settlement. The change in the employment rate did not become significant.

The third column includes additionally the income and active population of the destination city. The coefficient of the new public transportation link decreased somewhat, but it is still significantly different from zero and negative (-0.13%). Just like the change in the income of the settlement, the change in the income of the destination city became insignificant. However, the change in active population became significant and positive: connections which are leading to higher active population cities, increase the change in unemployment rate. This finding can also be interpreted with the migration effect: if there is a growing city in the catchment area of a settlement, it not just enhances commuting, but it can also enhance migration to the city. Active workers leave the settlement and move to the city. The significant negative coefficient of the settlement's change in active population coefficient backs this interpretation.

Table 5: Regression output for the connection-level FD model - basic specification

VARIABLES	(1)	(2)	(3)
	Change in unemployment rate between 2006 and 2014		
0 -> N connection	-0.00170*** (0.000636)	-0.00179*** (0.000631)	-0.00134** (0.000628)
Income tax paid (log, diff)		-0.00267 (0.00261)	-0.00217 (0.00253)
Active population (log, diff)		-0.0346*** (0.0121)	-0.0384*** (0.0128)
Number of cars per active population (log, diff)		-0.0622*** (0.0132)	-0.0624*** (0.0131)
Employment level (diff)		0.00416 (0.0177)	0.00380 (0.0168)
Income tax paid (log, diff, destination)			0.00350 (0.00332)
Active population (log, diff, destination)			0.0406*** (0.00804)
Average unemployment rate in subregion (diff)			
Constant	0.0273*** (0.000774)	0.0442** (0.0178)	0.0177 (0.0273)
Observations	27,350	27,350	27,350
R-squared	0.000	0.024	0.027

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Summing up the results from these specifications is that the installment of a new public transportation connection has an effect on the change of unemployment. Moreover, this effect is significantly negative. For further calculations, I use the specification in column (3).

5.3.1 Which connection types are the most important?

An important policy question is, which new connections have the biggest effect on the unemployment. Moreover, by controlling for these effects, we

can also increase the explanatory power of the model. The fitted model is the following:

$$\Delta u_i = \beta_0 + \beta_1 \Delta \log(P_i) + \beta_2 \Delta \log(P_j) + \beta_3 \Delta \log(P_i) + \beta_4 \Delta \log(P_j) + \beta_5 \Delta \log(C_i) + \gamma_1 PT_{Budapest_{ij}} + \gamma_2 PT_{count_{yij}} + \gamma_3 PT_{subregion_{ij}} + \gamma_4 PT_{other_{ij}} + \varepsilon_i \quad (2.5)$$

, where every variable is the same as equation 2.4, but PT_{ij} . Instead of PT_{ij} , there are four dummy variables used. $PT_{Budapest_{ij}}$ takes the value of 1, if in 2014 there was at least one public transportation link between i and j , but no connection in 2006, and the value of 0 otherwise. Similarly $PT_{count_{yij}}$ takes the value of 1 if in 2014 there was at least one public transportation link between i and j , but no connection in 2006, and takes the value of 0 otherwise. Similarly for $PT_{subregion_{ij}}$ towards subregion capital and $PT_{other_{ij}}$ towards other cities. I fitted the model using the OLS method on 27,350 observations with clustered standard errors for all the 3111 settlements.

The results of the regressions can be found in Table 6. Column (1) shows the results for the specification without control variables. The interpretation of the Budapest dummy variable is the following: if there was a public transportation connection built towards Budapest between 2006 and 2014, the expected change in the unemployment rate is -1.35% as opposed to the event when no connection was built. A connection built towards a subregion capital has the strongest negative effect: the expected

unemployment rate decrease is 4.56%. Interestingly, a connection towards a county capital increases unemployment. According to the regression, establishing a connection toward another city (not Budapest, county or subregional capital) does not have an effect on unemployment.

Table 6: Regression output for the connection-level FD model - advanced specifications

VARIABLES	(1) Change in unemployment rate between 2006 and 2014	(2)	(3)
0 -> N connection to Budapest	-0.0135* (0.00812)	-0.00521 (0.00774)	-0.00441 (0.00769)
0 -> N connection to a county capital	0.00351** (0.00146)	0.00360** (0.00145)	0.00511*** (0.00146)
0 -> N connection to a subregion capital	-0.00456*** (0.000835)	-0.00499*** (0.000836)	-0.00402*** (0.000823)
0 -> N connection to an other city	-0.000136 (0.000742)	3.58e-06 (0.000730)	-9.30e-05 (0.000732)
Income tax paid (log, diff)		-0.00261 (0.00260)	-0.00218 (0.00253)
Active population (log, diff)		-0.0352*** (0.0122)	-0.0385*** (0.0128)
Employment level (diff)		0.00405 (0.0176)	0.00378 (0.0168)
Number of cars per active population (log, diff)		-0.0624*** (0.0132)	-0.0625*** (0.0131)
Income tax paid (log, diff, destination)			0.00485 (0.00335)
Active population (log, diff, destination)			0.0387*** (0.00806)
Average unemployment rate in subregion (diff)			
Constant	0.0273*** (0.000774)	0.0438** (0.0177)	0.00859 (0.0274)
Observations	27,350	27,350	27,350
R-squared	0.002	0.026	0.029

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The second column includes the regression where the change in income tax paid, the change in active population, the change in employment rate and the change in the number of cars per active population is also included. The change in income tax variable did not become significant, nor did the change in employment rate. Change in active population and change in the number of cars per active population both became significant and negative. The Budapest connection dummy became insignificant, and the other city dummy stayed insignificant. The county capital dummy maintained its significant positive effect, whereas the subregion capital dummy stayed significantly negative.

The third column shows the results if the change in income tax and the change in active population of the destination is included. The income tax variable became insignificant, but the change in active population is significantly different from zero and positive. All of these effects are the same as we have seen it previously in Table 5. The only difference that happened with the connection dummies is that the county capital dummy's coefficient became higher (0.51%) and the subregion capital dummy's coefficient less negative (-0.4%).

Summing up the results from these specifications, it can be seen that two connection types have a significant effect on unemployment rate change: establishing a new public transportation toward a county capital is expected to raise unemployment rate, however; establishing a new

connection toward a subregion capital is expected to decrease unemployment rate. For further calculations, I use the specification in column (3). These results are useful for policy purposes; however, it is important to see exactly what effects were we able to identify.

5.4 Validity of the results

This chapter aims to provide an overall picture, which important questions were answered in this thesis, and which need further analysis.

The main research question of this thesis is whether there is a significant connection between public transportation possibilities and unemployment. Employment decisions are made by individuals; therefore, ideally one studies this question based on individual-level data. Instead, I used an aggregated dataset where the unit of observations was a settlement (or settlement-city connection), and I was concerned how different commuting possibilities affect the unemployment rates of the settlement. Even if a worker has the possibility to commute to another town and start working, there can be lots of important individual-level variation in these decisions, which I was not able to control for.

As the endogeneity and selection problems are very important issues in such a research, one has to try to mitigate them. I tried to include as many important control variables as possible, and used the difference between two time periods (2006 and 2014) for this purpose. However, it has to be stressed that I was able only to mitigate this issue. The use of more time

periods or a good instrumental variable can help to identify the effects more precisely.

Another inconvenience is caused by the public bus dataset for 2006, as its number of connections variable is a little surprising: it is unlikely that between 2006 and 2014, the average number of public bus connections for settlement-city pairs more than doubled (from 0.64 to 1.32). In order to avoid this possible error, I did not use the intensity of connections as an explanatory variable, only the existence of the connection.

I did not take into consideration that catchment areas changed between 2006 and 2014, as I had data about travel times by car only for the year 2011. The research is not concerned with villages within the catchment area as possible commuting places, and also the possibility that commuters can use bicycles or walk to their workplaces.

Only those connections were counted that arrived before 8 AM and did not take more than 60 minutes. This way, the results of the thesis are valid for those workers who start their work at the conventional 8 AM.

Hungary was struck by the global crisis in 2008. By inserting income tax variables into the regression, most of this effect is partialled out. However, the crisis hit industries differently: less the agriculture, more the banking sector. I did not control for these factors.

As it can be seen, there are important factors that my thesis was not able to control for; however, most of these deficiencies can be taken care of with further research.

6 Conclusions

In my research, I was concerned with the relationship between public transportation possibilities and regional unemployment rates in Hungary. After presenting the current situation in Hungary, I introduced the theoretical background and defined an identification strategy for the relationship between public transportation and unemployment. I tested these models on a settlement-level two time-period dataset for 2006 and 2014; using the ordinary least squares method.

The aim of my thesis was to answer two questions. The first was whether there is a significant relationship between public transportation possibilities and regional unemployment in Hungary. The second was concerned with a policy-problem: which public transportation connections should the government promote, if it aims to reduce unemployment rates most effectively.

My thesis was able to address both of these questions, and provided answer to them. According to my results, there is a significant, negative relationship between public transportation possibilities and regional

unemployment: establishing a new public transportation connection towards a city is expected to lower local unemployment rate by .13%.

If the government considers launching a policy to enhance commuting and thus decrease unemployment, it should concentrate on subregion capitals. According to my results, establishing a public transportation connection towards a subregion capital lowers unemployment rate by .4%, holding all other factors constant.

It is important to point out that there are important validity restrictions regarding the results. The most important is that the selection bias was not completely eliminated: the establishment of new public transportation connections depends on the unemployment level. This way, one cannot identify the sign of causality between transportation and unemployment. Moreover, this research is concerned only with the possibilities in transportation, not with the actual decision of commuters. In order to mitigate these deficiencies, further research is needed with broader data possibilities.

7 Policy recommendation

This chapter aims to provide a policy recommendation based on my results.

First I will present current policies promoting commuting in Hungary, then I will present the planned recommendation, and finally potential costs and benefits are going to be elaborated on.

In the case of commuting, state intervention is justified if due to some reason, there are fewer workers commuting as it would be optimal. As there are significant and persistent regional differences, this is definitely the case in Hungary. There is a spectrum of different interventions available, from setting up administrative regulations, to subventions for employers, employees or to transportation companies.

There have been three different state subsidies introduced in Hungary:

- Tax relief on employer's direct contribution to commuting costs
(monthly pass for public transportation or gasoline costs)
- Reimbursement of costs for previously unemployed workers
- Harmonization of regional public transportation timetables

Employers used the tax relief option extensively in the last decade.

According to Horváth et al (2006), 17% of all the employees were given some contribution to commuting costs in 2003 (at that time 33% of total employees were commuting). Unlike other benefits, skilled workers and unskilled workers benefitted from this allowance close to the average (17%

and 14%). According to Bartus (2011), most of the commuters received at least some kind of allowance.

Travel cost reimbursement for currently employed workers is available since 1994; however, only a couple of thousand employees have used it. The reason for that is that it requires considerable administration, and it is available only for new entrants to the labor market for 1 year.

Harmonization of timetables started to be carried in 2007, thanks to the organization of regional transportation offices. Their main task was to resolve parallelism between train and bus transportation and the harmonization of local and intercity connections. Sadly, their effectiveness has not been researched yet; however, based on my results, they achieved significant improvement.

As we have seen, between 2006 and 2014, public transportation possibilities improved in Hungary. However, regional differences in unemployment increased. If we add to these facts the results of this paper: transportation possibilities have significant effect on unemployment, and the most important connections are towards subregional capitals, we can see that 1) improvement in public transportation possibilities were not enough, and 2) they were not targeted effectively.

My thesis found that establishing one new public transportation connection is expected to lower unemployment rate in the settlement by

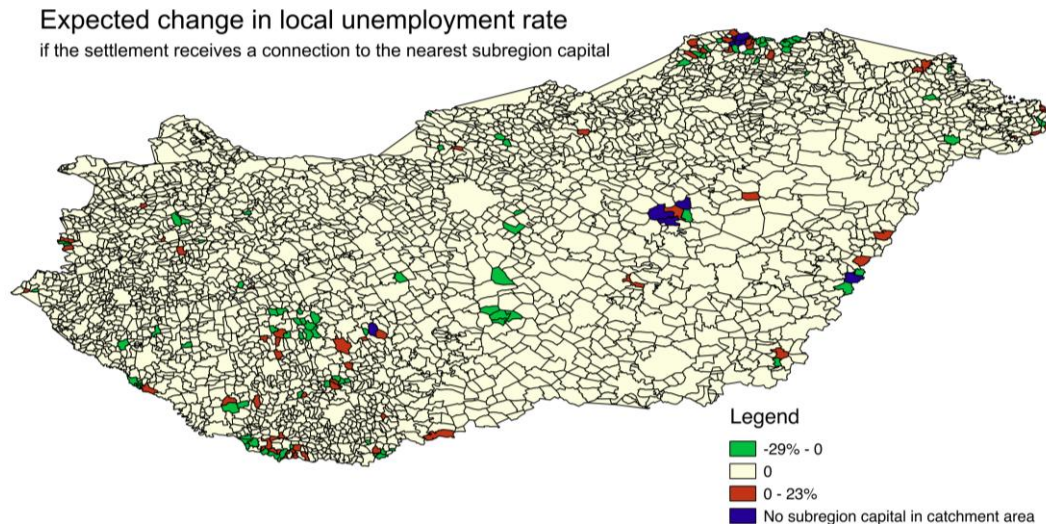
.13%. Moreover, connecting a settlement with a previously not connected subregion capital is expected to decrease unemployment rate by .4%.

Based on these results, a possible policy recommendation could be that holding all other subsidies constant, every Hungarian settlement should have at least one fast connection to a subregion capital. This is expected to increase the number of commuters; however, it does not increase outward migration. Thus regional unemployment differences may decrease.

In 2014, there were 159 settlements without a connection to a county or subregion capital (40% of them in South-Pannonia, 25% in the Cserehát region). However, 149 of them had at least either a subregion or a county capital in their catchment area. If we would connect these 149 settlements to the closest subregion or county capital (which one is closer), the expected decrease in unemployment rate would be 1% on average for these settlements. And this is only the most conservative result, as the model does not take into consideration the positive multiplicative effects of more employed workers: creating local demand for products and showing an example to fellow job-seekers.

Map 9 shows the expected change in unemployment due to this recommendation: as we can see, there are also some villages where an increase in unemployment is expected; however, for most of treated settlements a decrease in unemployment rate is expected. Those 10 settlements that have neither county, nor subregion capital in their

catchment area are also indicated: in order to improve the situation of these settlements, road infrastructure has to be upgraded.



Map 8: Expected change in local unemployment rate due to the policy recommendation (own graph)

The cost of such a policy is low: regional Volán bus services are already present in every region. What is needed is a large-scale harmonization of timetables by clinging to strict regulations that ensure quality. Such a regulation could be one that travel time is not allowed to be longer than 60 minutes and the bus should arrive to the city before 7:50 AM. Regional public transportation offices have been working on similar objectives since 2007 and it is recommended to continue this work.

Summing up, with a large-scale harmonization of bus timetables, an average decrease of 1% in unemployment rate can be expected in those areas, where public transportation quality is extremely low in Hungary.

8 Appendix

Table 7: Basic statistics of the variables used

Variable	Obs.	Mean	SD	Min	Max
Unemployment rate	3112	.13	.085	.0027	.539
Change of unemployment rate	3111	.02	.047	-.244	.285
Radius of 40 mins	3117	8.99	6.90	1	55
Number of cities reachable by public transportation	3117	3.47	2.53	0	35
Change in number of cities reachable by public transportation	3118	2.30	2.54	-3	25
Number of cars per active population	3117	.43	.106	.0748	1.42
Change of number of cars per active population	3116	.001	.061	-0.35	.571
Employment rate	3117	.703	.101	.2207	1.33
Change in employment rate	3116	.052	.097	-.45	.696
Active population	3117	1908	18706	7	998918
Change in active population	3116	-34.47	309.2	-9910	4340
Income tax paid per active population	3117	128722	59287	2357	664553
Change in income tax paid per active population	3116	8652	4671	-9816	18934
Border	3117	.344	.475	0	1
Austria	3117	.064	.244	0	1
Train station	3117	.356	.479	0	1

Figure 7: Histogram of unemployment rate in Hungary for 2011 (own graph)

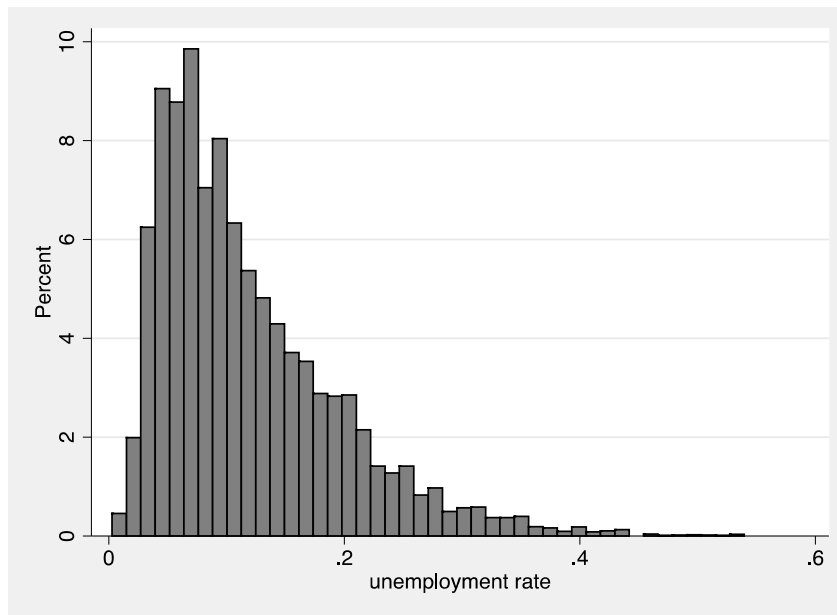


Figure 8: Histogram of the change of unemployment (own graph)

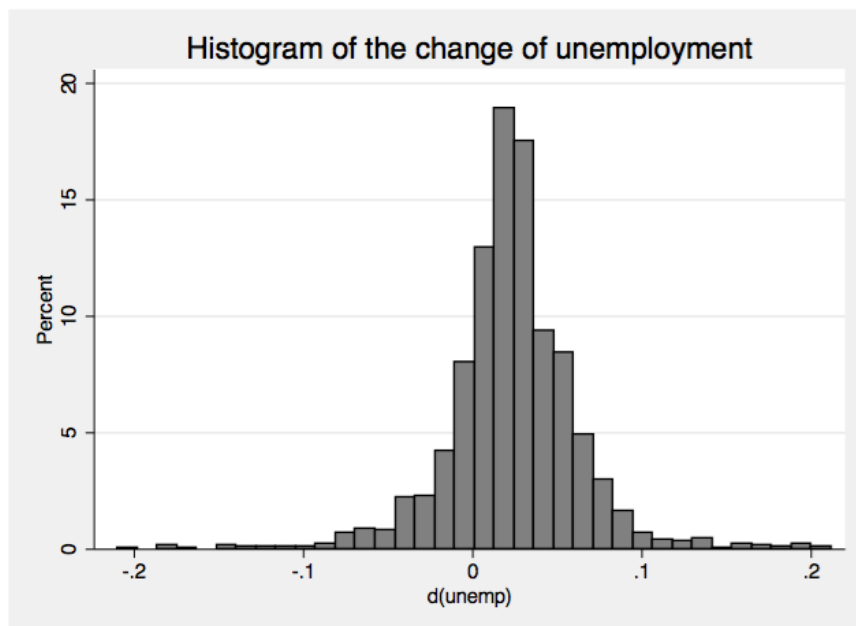


Figure 9: Histogram of the number of cars per active population variable (own graph)

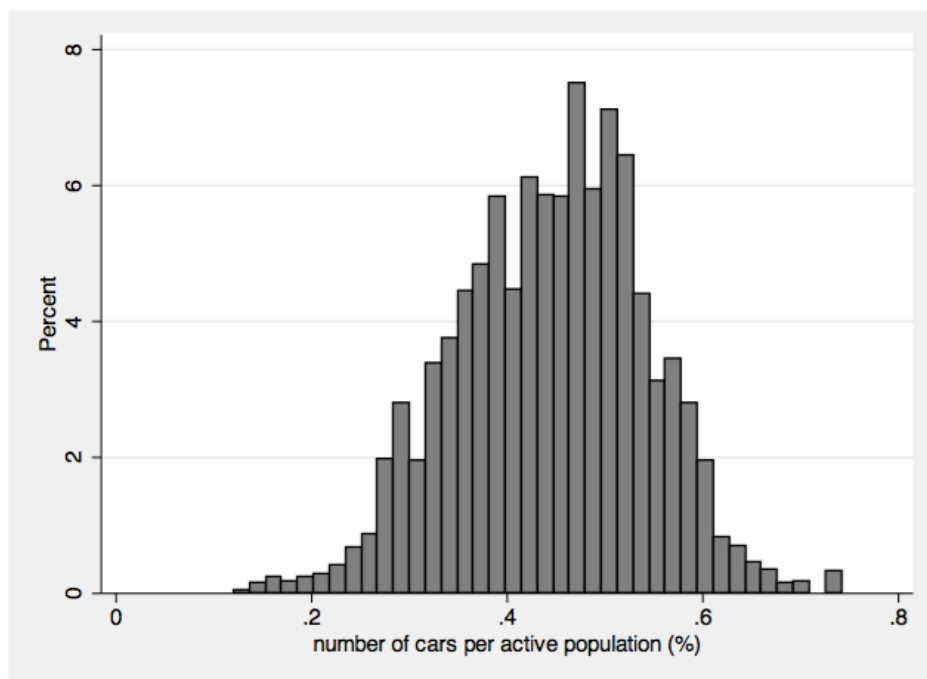


Figure 10: Histogram of the employment rate for 2014 (unweighted, own graph)

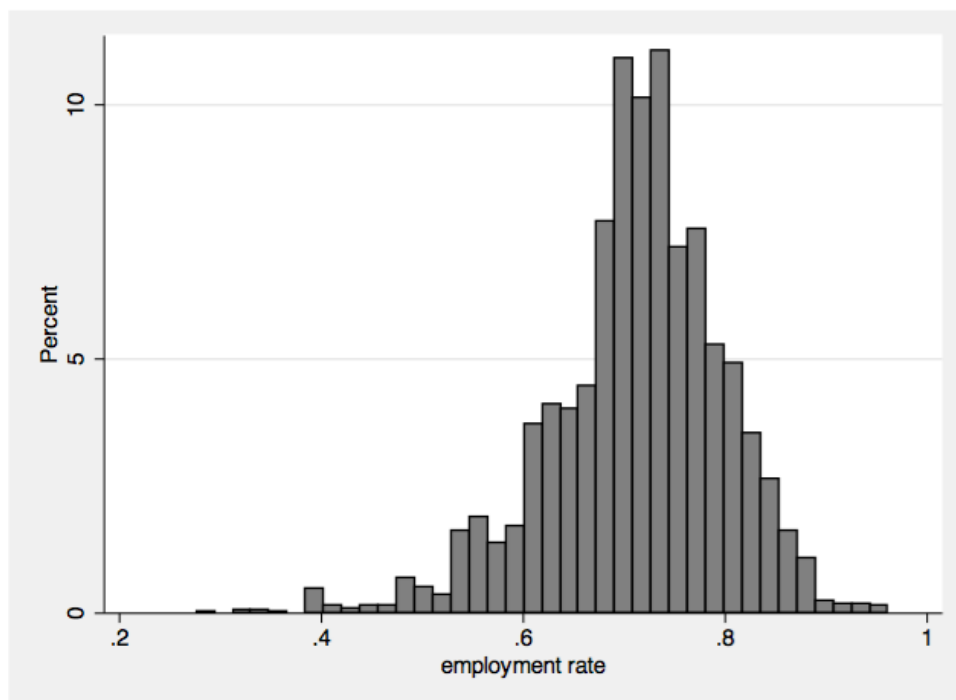
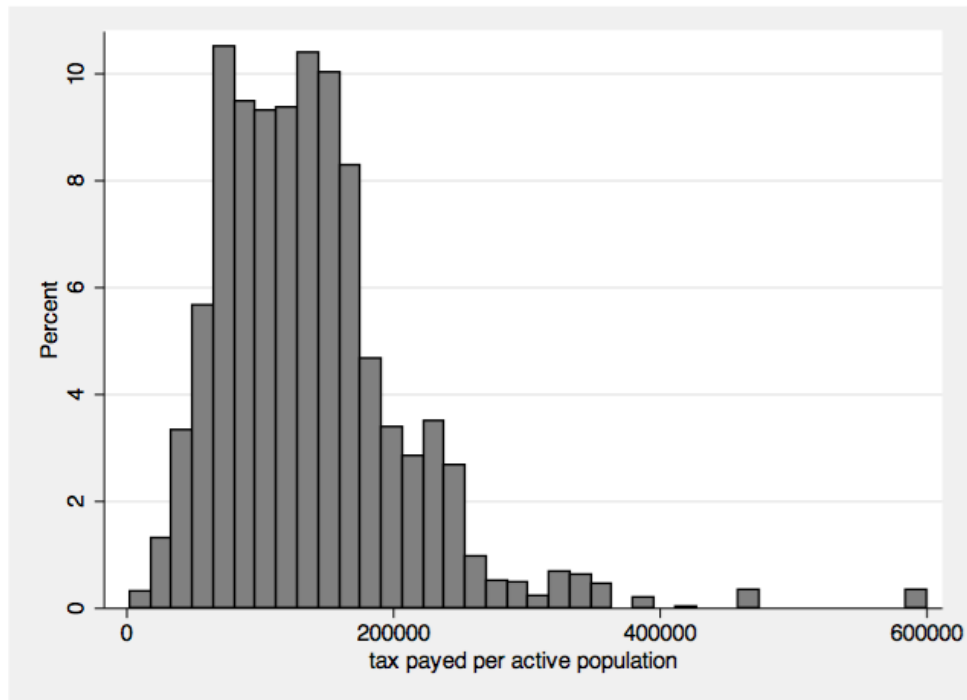


Figure 11: Personal income tax paid per active population in 2011 (own graph)



9 References

- AAandrienko, Y., & Guriev, S. (2004). Determinants of interregional mobility in Russia. *Economics of Transition*. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.0967-0750.2004.00170.x/full>
- Bartus, T. (2003). Ingázás. *Munkaerőpiaci Tükör*, 88 – 102.
- Bartus, T. (2009). Területi különbségek és ingázás. Budapest Institute Policy Paper
- Bartus, T. (2010). Commuting Time , Wages and Reimbursement of Travel Costs. *Review of Sociology*, 72–94.
- ELTE TTK. (2014). Driving time matrix for 2011.
- Eurostat. (2010). Eurostat Regional Yearbook: Transport. Retrieved June 04, 2014, from http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-HA-10-001-10/EN/KS-HA-10-001-10-EN.PDF
- Horváth, H., Hudomiet, P., & Kézdi, G. (2006). Bérek és béren kívüli juttatások. *Közelkép, Bérek, Ker*, 36 – 47.
- Ihlanfeldt, K., & Sjoquist, D. (1998). The spatial mismatch hypothesis: a review of recent studies and their implications for welfare reform. *Housing Policy Debate*. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/10511482.1998.9521321>
- Kertesi, G. (1997). A gazdasági ösztönzők hatása a népesség földrajzi mobilitására 1990 és 1994 között. *Esély*, 2, 3–32.
- Kertesi, G. (2000). Ingázás a falusi Magyarországon. Egy megoldatlan probléma. *Közgazdasági Szemle*, 47, 775 – 798.
- Köllő, J. (1997). A napi ingázás feltételei és a helyi munkanélküliség Magyarországon. *Esély*, 2, 33–61.
- Köllő, J. (2002). Az ingázási költségek szerepe a regionális munkanélküliségi különbségek fenntartásában. *Budapesti Munkagazdaságtan Füzetek*, 2.
- Köllő, J. (2006). A napi ingázás feltételei és a helyi munkanélküliség Magyarországon Újabb számítások és számpéldák. *Budapesti Munkagazdaságtan Füzetek*, 1.
- KSH. (2014). TSTAR database.
- Ksh.hu. (2014). Unemployment rate. Retrieved June 05, 2014, from http://www.ksh.hu/docs/hun/xstadat/xstadat_evkozi/e_qlf027.html
- Mincer, J. (1977). Family Migration Decisions. Retrieved from <http://www.nber.org/papers/w0199>

MTA KTI. (2006). Volán timetable.

OSM. (2014). openstreetmap.org. Map retrieved on 03/05/2014, 13:52

Persyn, D., & Torfs, W. (2012). A Gravity Equation for Commuting. *SSRN Electronic Journal*. doi:10.2139/ssrn.2285810

Pissarides, C., & McMaster, I. (1984). Regional migration, wages and unemployment: empirical evidence and implications for policy.

Pissarides, C., & Wadsworth, J. (1989). Unemployment and the inter-regional mobility of labour. *The Economic Journal*. Retrieved from <http://www.jstor.org/stable/2233768>

Tinbergen, J. (1962). Shaping the world economy; suggestions for an international economic policy. Retrieved from <http://repub.eur.nl/pub/16826/>