ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Bachelor Thesis
International Economics and Business Economics

Title
The Relationship Between Transport-Related Exclusion and Commuting Patterns in Poland

Student Name: Julia Kulakowska
Student Number: 501863jk

Abstract

Since 1989, Polish public transport infrastructure has been marked by a transformation process that gradually deteriorated its quality and availability, leading to mobility-related exclusion of many Polish inhabitants. This paper magnifies on this mobility-related exclusion and the existing transport disadvantage, by investigating the relationship between public transport accessibility in the Polish communes and commuting flows – an important proxy for economic exclusion. The research is conducted using the economically acclaimed gravity model and a logit regression analysis. The commuting and economic data has been obtained from the Central Statistical Office and the Institute of Rural Development and Agriculture PAN. Given the multidimensionality of the transport disadvantage issue in Poland, the obtained associations between the commuting flows and public transport infrastructure are less straightforward than expected. However, the findings still provide a good basis for future research and policy recommendation, especially given the fact that women and low-income households are included in the analysis, as social groups especially prone to transport exclusion. Road accessibility has proven to be an important factor of the commuting flows, which is in line with the current dependency of Polish inhabitants on private car transportation.

Supervisor: Zsolt Csafordi

Table of Contents

1. Introduction

In Poland, 1989 has been marked as a significant year, as it is when the People's Republic of Poland has been transformed into the Republic of Poland, officially ending the Communist governance (Dudek, 2011). During this transformation process, the country has undergone an array of post-communist social, economic and political changes, not all of which have resulted in positive outcomes (Dubnicki, 2019). Among others, this included changes in the social policy surrounding public communication, in particular the state coach transport managed by the Motor Transport Company (PKS) and the railways transport managed by Polish State Railways (PKP).

According to Żakowska (2019), transport disadvantage in Poland currently amounts to around 14 million passengers, with over a quarter of villages having no public transport at all. The number of public transport passengers has dramatically shrinked, falling from 2.16 billion in 1989 to only 354.6 million in 2014 (Dubnicki, 2019). This was a result of depopulation of rural areas, aging population, privatization and commercialization of public transport services and other administrative changes, all of which occurred during and after the transformation period (Dubnicki, 2018; Żakowska, 2019). This transport disadvantage has led to mobility-related exclusion of many people, which in turn has resulted in limited job-seeking opportunities, social and cultural isolation, growing urban-rural inequalities and increased car usage (Dubnicki, 2019; Żakowska, 2019). Therefore, solving transport disadvantage and the consequent mobility-related exclusion, should be of the highest priority to the government of Poland.

However, mobility-related exclusion is a complex problem to solve (Dubnicki, 2019). This is because mobility-related exclusion is a highly multidisciplinary and multilateral process (Kenyon et al., 2002). Moreover, the scientific discourse regarding transport disadvantage, its link to mobility-related exclusion and the related mitigation strategies is fragmented and almost no analysis exists on a national level in Poland (Kenyon et al., 2002; Żmuda-Trzebiatowski, 2016). What is more, the solutions that have been proposed, are rarely applicable universally (Lucas, 2012). Therefore, there is a widespread consensus amongst transport exclusion researchers, that better social evaluation of transport disadvantage is needed at every level of governance (Lucas, 2012). In Poland, this is especially the case for the research of transport exclusion associated with public transportation, which has been a neglected research topic for the past decade (Ciechański, 2020; Dubnicki, 2019). This paper contributes to this research discourse by investigating the linkage between transport disadvantage due to lacking public transport infrastructure in Poland, and its impact on economic exclusion – an important determinant of transport-related exclusion (Dubnicki, 2019; Żakowska, 2019).

As already mentioned, and given the extensiveness of the topic area, this paper will narrow its focus towards the consequences of transport disadvantage on employment being a proxy for economic exclusion. Economic exclusion occurs when high monetary, time or effort costs of travel, prevent or limit access to employment facilities and thus negatively influence incomes (Żakowska, 2019). This is because commuting acts as a spatially equilibrating mechanism in the labour market and has the potential to reduce the disparities in the regional labour outcomes with regards to unemployment or wage levels (Fransen et al., 2019). However, when an individual is unable to or limited from reaching a workspace due to commuting costs, it can have severe consequences on the social and economic security of that individual, as well as their general welfare and quality of life (Fransen et al., 2019).

As such, as suggested by Eurostat (2020) report, unemployment is a key indicator to study developments in the labor markets and is an area that has been studied within the disciplines of geography, economics, or urban and rural planning (Fransen et al., 2019). The Social Exclusion Unit (2003), considers unemployment to be one of the most important consequences and drivers of transport disadvantage, highlighting the important inter-relationship between these two, that can take a geographical but also a social dimension. The former is linked to purely a combination of spatial proximity and speed, while the latter puts these two in the context of the regional labor markets. This research aims to pursue the latter, by evaluating the impacts of transport disadvantage on the employment structures in the Polish communes, while considering their social context such as their existing demographic and economic structures. This is because, among others, unemployment levels vary greatly across demographic patterns (Eurostat, 2020). For instance, certain social groups are especially prone to transport exclusion, including women, people with disabilities or low-income households (Zawojska, 2019). Herein, improved accessibility should play a major role, as it improves the matching of economic opportunities with individual’s labor needs (Fransen et al., 2019). As such, accessibility, proxied by the level of commutinbg in this research is essential to a successfully functioning labor market.

Consequently, this research will focus on answering the following research question:

What is the relationship between public transport accessibility and commuting flows in the Polish communes?

To explore the above research question, this paper focuses on a variety of sub-aspects. First, the paper aims to investigate, what determines the existence of commuting flows between the communes. Second, the research paper aims to evaluate the impact of public transport disadvantage, on the intensity of employment flows between the communes. Third, the paper aims to identify, to what extent this public transport disadvantage affects different social groups that are potentially more prone to transport exclusion. According to Eurostat (2016), demographic factors including age, gender or education level, have a great influence on the degree of unemployment. Moreover, lower-income groups or women are often excluded from car ownership and hence are more prone to suffer from transport poverty and transport disadvantage (Currie & Delbosc, 2010; Currie et al., 2010; Fransen et al., 2019).

This approach to research is in line with the recent trends in the scientific discourse regarding transport-related exclusion, which has shifted its focus from a traditional systems-based approach to transport provision, to a people-focused and needs-based social policy perspective (Lucas, 2012). The former focuses on identifying the issue of transport disadvantage itself. The latter aims to identify the consequences of transport disadvantage on the society, including the equality of opportunity to access key services, the equity of outcomes and redistributive justice (Lucas, 2012). This research is undertaking the people-based approach to transport-exclusion, by focusing on directly linking the effects of the existing transport disadvantage in the Polish communes to their employment structures.

The paper begins with an Introduction, followed by a Theoretical Framework, which reviews the historical, social and political factors that have resulted in the current state of public transportation in Poland. Moreover, this section will review the existing economic theory regarding mobility-related exclusion, especially with regards to job accessibility. Data and Methodology describe the data sources used, as well as the analytical framework chosen to obtain the statistical results necessary for the analysis. Given the availability of data on a communal level and the decision-making frameworks in Poland (Dubnicki, 2019), this paper focuses on analysing the commuting patterns between the communes. The Results section discusses the obtained results and their importance for the policymakers. Limitations & Future Extensions section discusses the limitations of the research and suggests potential improvements. Lastly, the Conclusion section summarizes the findings of the research.

2. Theoretical Framework

2.1. Administrative & Historical Context of Transport Disadvantage in Poland

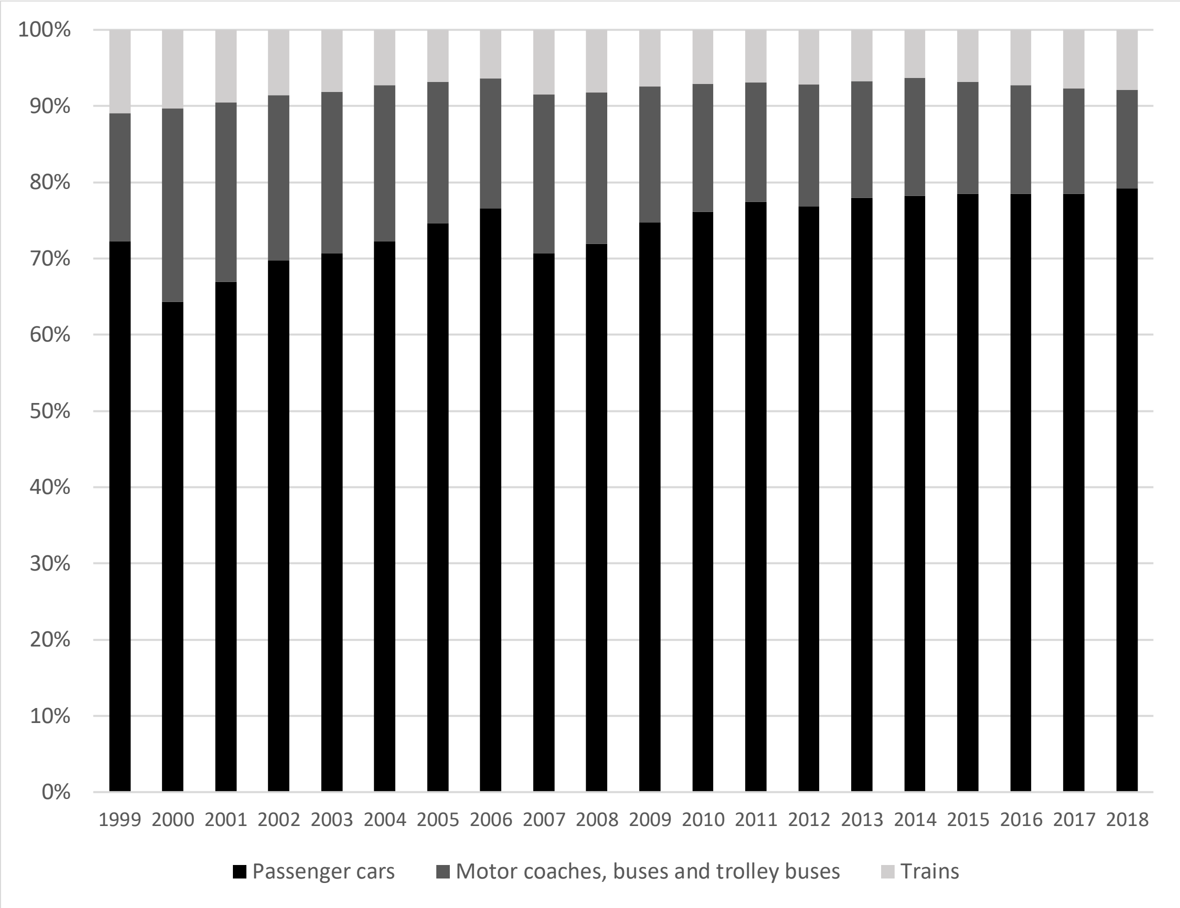
Given the unique qualities of Polish administrative structure, it is important to first define the key terms used in this analysis. The Polish regional public administration structure comprises of three levels: 1) a voivodeship also referred to as a region, 2) a powiat also referred to as a county and 3) a gmina also referred to as a commune (Mazur et al., 2018). At the central level, the government administration is composed of the Council of Ministers, the Prime Minister, individual ministries and central offices, agencies and supportive organizations (Mazur et al., 2018). This administration also operates at the voivodeships level, supported by the governors of the regions, voivodeship offices and administration. At the local level, voivodships are represented by marshals and voivodships boards. Powiats are ruled by the governors of powiat-level organization units. Lastly, gminas are ruled by mayors or city presidents and gmina-level offices and organization units. Detailed information regarding the number of communes and counties within each voivodeship can be found in Appendix A.1.1 and A.1.2.

The overview of public transportation provision history in Poland is as follows. Before 1989, railway freight and passenger transport were managed solely by the state-owned Polish State Railways (PKP) company. State coach transportation was managed by the state-owned Motor Transport Company (PKS), which provided regional and inter-city bus transit for passenger and employee lines (Dubnicki, 2019). Urban and suburban transport remained under the control of the voivodeship-based enterprises. This setting has not only led to affordable travel prices, but also to frequent travel connections between the main administrative centres and smaller towns (Dubnicki, 2019). In 1960s, railway lines started to gradually liquidate, given their high travel time in comparison to the state coach transportation, as well as the required distances to travel to the rail stations (Dubnicki, 2019). Moreover, railways were suffering from frequent traffic due to maintenance and reconstruction works.

From 1989 onwards, the Polish government started to actively seek saving opportunities (Dubnicki, 2019). Railway transportation remained under control of the state, given its strategic importance. This resulted in lagging modernization, which led to further liquidation of the railway connections, especially in the rural areas. State coach transportation, on the other hand, was much easier to restructure. As a result, PKS was split into 176 local companies, which from then onwards, were registered as commercial companies, meaning they had to compete on their attractiveness and profitability. This, however, led to disappearance of non-profitable lines, especially in the rural areas of Poland, deepening transport exclusion. Then, in 1997, the number of voivodeships in Poland was reduced from 49 to 16 (Dubnicki, 2019). This resulted in further fragmentation of the state bus transportation, as from now on, 380 counties oversaw the regional public bus transportation provision. However, in consequence, the counties primarily focused on connecting the smaller towns with the towns where their county authorities were seated. This left the responsibility of arranging the remaining intra- and inter-municipality transport to the municipalities. As a result, unprofitable inter-municipality connections started to liquidate, leading to the lack of connections between, for example, neighbouring municipalities located in different voivodeships. The education reform in 1999 has led to even more severe liquidation of public bus transportation and a massive outflow of passengers (Dubnicki, 2019). The percentage share of transport modes in individual transport between 1999 and 2018 is presented in Figure 1. From Figure 1, it can be concluded that there has been a stable increase in car passenger transport since 2000, right after the educational reform in 1999. For public bus transportation, there has been a big decline in 1999, followed by a steady decline since 2000 and reaching its lowest level in 2018. The percentage of trains has been rather stable, with a slight decrease since 2015. This is beacuse railways have remained under the constant control of the state-owned PKP throughout the years.

Figure 1

The Percentage Share of Individual Passenger Land Transport Modes in Poland in the Years 1999-2018



Note. Adapted from Eurostat (2021)

In the present, transport exclusion has first appeared on the agenda of the currently ruling Law & Justice party in 2019 (Dubnicki, 2019). In 2019, the Reform of the Law on Public Transport and the Law on a Bus Transport Development Fund were implemented (Duszan, 2020). The former aims to limit the reduction of bus networks and to create new ones, as well as to give more power to the counties to reshape public transport networks (Duszan, 2020). On January 26th, 2021, the Council of Ministers has also published a degree, according to which some PKS companies will be merged with PKP (Morawiecki, 2021). The remaining PKS companies, will be still managed by the counties, however, will now become members of the Industry Development Agency, with the aim of restructuring and share selling, as well as to facilitate their dismantling where necessary (Duszan, 2020).

2.2. Transport Exclusion Economics

The topic of transport disadvantage has gained momentum in the scientific world at the turn of late 1990s and early 2000s (Lucas, 2012). The concept of social exclusion has been adopted to transport disadvantage, to help policy makers identify the multidimensionality of the problem, the relational nature of transport disadvantage, its dynamism over time and space and the economic and social outcomes of transport disadvantage (Lucas, 2012). As such, transport-related exclusion may be defined as:

The process by which people are prevented from participating in the economic, political and social life of the community because of reduced accessibility to opportunities, services and social networks, due in whole or part to insufficient mobility in a society and environment built around the assumption of high mobility (Kenyon et al., 2002).

The concept of transport-related exclusion stems from the fact that transport is a derived demand (Rodrigue et al., 2020). This means that the demand is induced by the demand for other goods and services. This demand can be directly induced, stemming from work-related activities or shopping. Indirectly induced transportation demand is a consequence of other movements. For instance, the transport demand for road assistance is a result of the existing passenger movements, which in turn have been directly induced. Access to jobs can be treated as a direct derived demand. When denied the supply of transportation through transport disadvantage, the derived demand for transportation cannot be met, leading to unsatisfied demand for goods and services. The concept of transport-related exclusion is, however, not synonymous to transport disadvantage (Lucas, 2012). Instead, these concepts directly and indirectly interact with each other. Moreover, it is possible to be socially excluded with a good access to transportation and to be socially included despite being transport disadvantaged.

Transport-related exclusion is a result of individual factors such as age, disability, gender or race, the structure of local areas and public transport services, the condition of national structures and the labour market or cultural influences (Lucas, 2012). Moreover, certain groups are especially prone to social exclusion (Dubnicki, 2019), including the elderly, who for health reasons cannot use a car to access health centres and government institutions; students, who are too young to drive a car; women, who tend to have less driving licenses (Dubnicki, 2019); or the low-income groups, who might suffer from transport-poverty as a result of the necessity to purchase a car despite limited financial resources (Currie et al., 2010). Therefore, resolving transport-related exclusion should be of the highest priority, especially for these groups. The power to do so lies mainly within the power of the delivery agencies, including the national government of Poland in case of trains and the county governments in case of public bus transportation. Effective transport improvement policies require the removal of systemic exclusion of the above and many other social groups. To do so, the policy delivery needs to accurately recognize the abilities, skills, capacities, resources and experiences of the excluded individuals, in line with the people-based approach towards transport exclusion (Lucas, 2012).

This research focuses specifically on the aspect of job market accessibility within the area of transport-related exclusion. As Hansen (1959) defines, accessibility is a measure of how easily and how many opportunities are reachable from a given location. Apart from the degree of transport disadvantage, it is also a dependent on factors including job-housing imbalance, car dependence or urban spatial structure (Fransen et al., 2019). As such, it is a combination of land use and transport domains and is widely used for investigation and conceptualization of issues existing within those two domains (Fransen et al., 2019). Moreover, job accessibility can only be improved with a combination of improved transport supply and the adjustment and optimization of land use patterns. Given the limited scope of the research, the focus will be put on the former, with commuting being a proxy for the job market patterns. Moreover, the research focuses both on investigating public transport infrastructure and road infrastructure, given the increasing dependency of Polish commuters on cars (as presented in Figure 1). It is important to identify the two aspects, as given the negative externalities of car usage including noise of air pollution or road safety, solutions from public transport need to be provided to reduce private car transportation (Matas et al., 2010).

2.4. Hypotheses

First, the paper aims to evaluate the impact of the existing transport disadvantage in the Polish communes on the employment flows between the communes. Commuter mobility patterns have been the focus of many recent studies (Varga et al., 2018; Fransen et al., 2018) with public transport infrastructure being an important factor considered, given its power to shape population distribution and economic activity (Heuermann & Schmieder, 2018). There is a general scientific consensus that better public transport connections should enhance the set of employment options to workers outside of their residence region (Heuermann & Schmieder, 2018). This paper aims to test this claim in the context of Polish commutes, given the following hypothesis:

H1: *The accessibility to public transport services is significantly and positively associated with the employment flows between the communes***.**

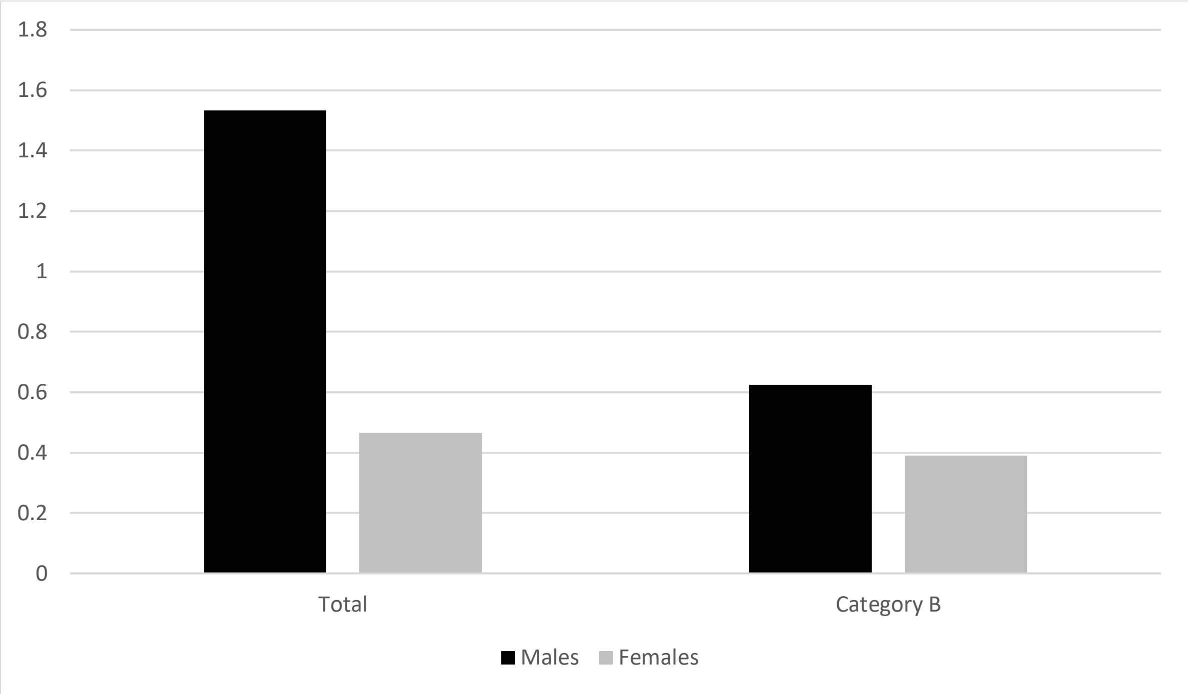
As outlined in the introduction, this research focuses on the people-based aspects surrounding the impact of transport-exclusion on employment patterns. As such, apart from investigating the factors that impact the intensity of the commuting flows, this research will also focus on investigating the relationship between public transport accessibility and disadvantaged social groups. Given limited availability of data, this research will focus only on two social groups, namely women and low-income households. Since the analysis focuses on the existing employment flows, only employed women and employed social assistance beneficiaries are considered in the study.

First, the research considers women as an important social group for the analysis, because the topic of transport exclusion of women has been an widely discussed topic in the economic literature (Matas et al., 2010). This is because their car ownership rates tend to be significantly lower than the ones of men, making them more prone to transport exclusion (Matas et al., 2010). This is also the case in Poland, where, as presented in Figure 2, women have obtained much fewer driving licenses, both when considering the total number, but also when considering the Category B driving licenses – the most popular out of the categories. On top of that, women are already much more prone to transport exclusion due to accessibility issues related to the built environment and transport infrastructure (Dobbs, 2005). What is more, as the existing literature suggests, women are already relatively more disadvantaged in terms of labour outcomes, as they suffer from lower participation rates and higher unemployment (Matas et al., 2010). Taking all these factors into consideration, this paper aims to test the following hypothesis:

H2: *Improved public transport accessibility within the communes has the highest significant positive association on the amount of employment flows in the origin communes with a high % of employed women.*

Figure 2

Amount of Driving Licenses Owned per Capita, by Gender, 2016

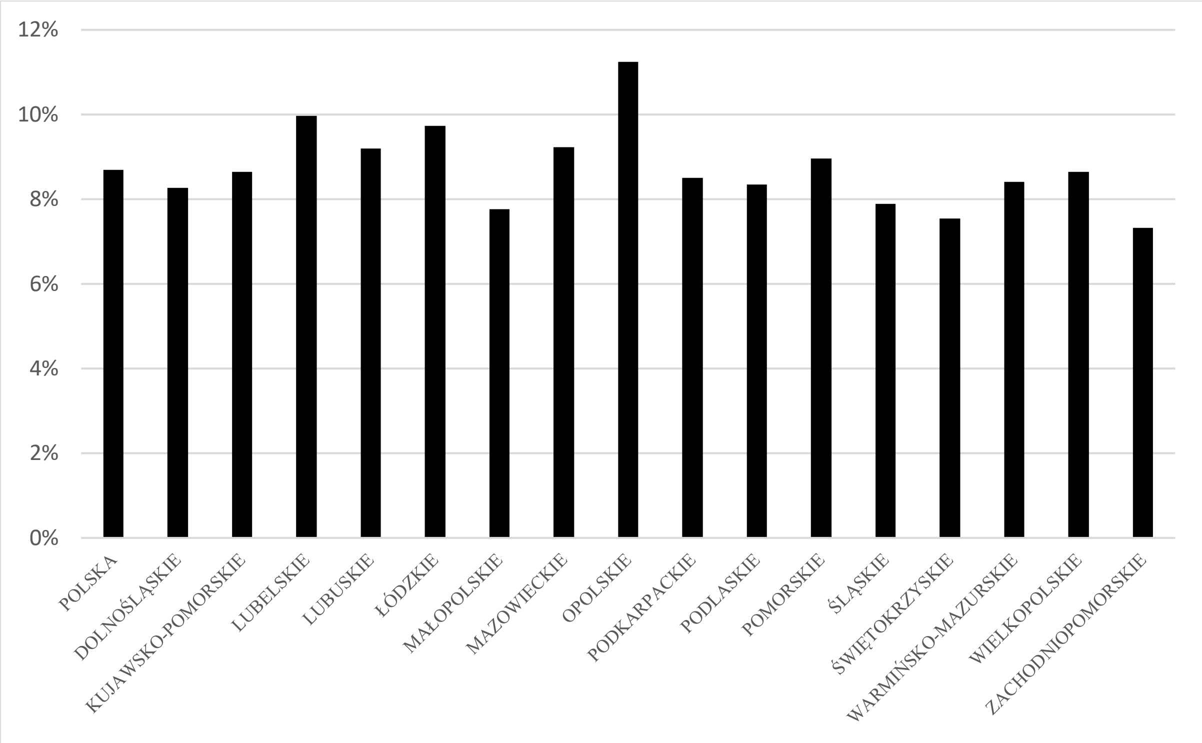


Note. The graph displays the average amount of driving licenses owned per capita, distinguishing between genders. The numbers were calculated by dividing the number of driving licenses in use in 2016, by the population numbers in the same year. The driving license data has been adapted from CEPiK (2016) and the population data has been obtained from GUS (2020).

Second, with regards to low-income communities, their probability of suffering from transport poverty increases due to lower average wages (Currie & Delbosc, 2010; Currie et al., 2010; Fransen et al., 2019). This lower average income often constrains their ability to purchase a car, in turn, bounding them to the use of public transport, which is often limited in terms of frequency and coverage (Dubnicki, 2019). As a result, it narrows their job seeking and educational opportunities, which further negatively impacts their current and future income-earning options (Kamruzzaman & Hine; 2012). What is more, as seen in Figure 3, transport expenditures in Poland are a significant share of household expenditure, ranging from around 7% to 11%. This high share of transport expenditures in the total expenditures makes the Polish low-income households especially prone to transport poverty.

Figure 3

Transport Expenditure as % of Total Monthly Household Expenditure in 2016, by Voivodeship



Note. This table represents data for 2016. Adapted from GUS (2020)

Given these factors and as supported by the empirical results from Kamruzzaman & Hine (2012), low-income and non-car owning groups tend to locate their activities in local spaces, hence, they also tend to concentrate their employment opportunities in their local surroundings. As such, improved accessibility should, in theory, stimulate the intensity of the commuting flows especially strongly for the low-income households. This leads to third hypothesis being as follows:

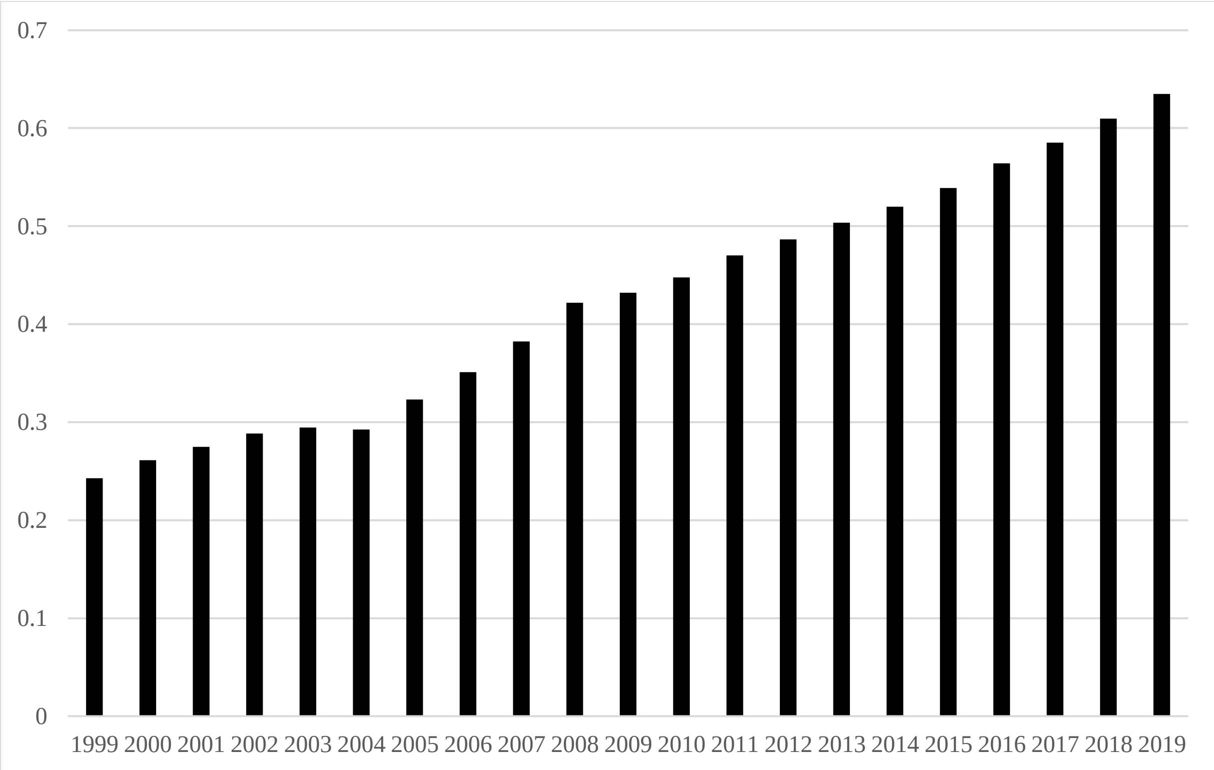
H3: *Improved public transport accessibility within the communes has the highest significant positive association on the amount of employment flows in the origin communes with a high % of households on social assistance.*

Lastly, this paper will investigate the relationship between commuting and road accessibility. This aspect is important, as the fragmentation of the public transport networks in Poland has vastly deteriorated public transport quality and has decreased the quantity of access points (Dubnicki, 2019). Such fragmentation has led to increasing car dependency, which is proxied by the rapidly increasing car ownership, as displayed in Figure 4. This topic is important from a people-based evaluation perspective, as it has increased the vulnerability to transport and economic exclusion of households and individuals that do not own cars (Matas et al., 2010). To evaluate the importance of road accessibility, this paper will test the following hypothesis:

H4: *Road accessibility is significantly and positively associated with the employment flows between the communes***.**

Figure 4

Amount of Registered Passenger Cars per Capita in Poland, 1999-2019



Note. Adapted from GUS (2020)

1. Data and Methodology

3.1 Data Selection

The cross-sectional data used for the analysis was obtained from a variety of sources. Given the limited data availability on a communal level, the variables used will refer only to the year 2016. All data, except for car ownership per inhabitant, were obtained on a communal level. First, the data regarding employment flows between the communes for 2016, has been obtained from the Central Statistical Office (GUS, 2019a). This data records the absolute number of workers travelling from a commune of living to the commune of where they are employed. The dataset only includes commuting flows that consist of a minimum of 10 workers and does not include flows within the communes. Complementary to this dataset, GUS (2019) has provided a dataset of geodetic distances (in km) between the communal centroids, the shortest distance between them via roads (in km) and the shortest travel time between them via a car (in min).

Second, the data on the amount of people employed per 1000 people within a commune, the employment division by gender, the percentage of people on social assistance due to low-income reasons, the communal population were also obtained from GUS (2020) for the year 2016. Given the fact that the data on commuting flows from 2016 only considers employed members of the population, consequently, only employed women and low-income households are considered in the subsequent variables. Lastly, the amount of spending for transport by the commune for 2016 and the income earned by a commune averaged per commune's inhabitant for 2016, were also obtained from GUS (2020).

The dummy variables regarding the percentage of villages within a commune connected to the asphalt roads and the percentage of villages having at least one train or public bus stop, were created by merging the categories displayed on the maps provided in the report of the Institute of Rural Development and Agriculture PAN (Instytut Rozwoju Wsi i Rolnictwa PAN), under the lead of Rosner & Stanny (2016). This report assesses the status of socio-economic developments in rural areas of Poland, including their spatial accessibility.

3.2. The Gravity Model

To answer the hypotheses, a gravity model is used to evaluate the impact of public transport accessibility on the flows between the communes. The gravity model stems from Isaac Netwon's *Philosophiae Naturals Principia Mathematica*, where he uses the law of universal gravitation, to claim that the gravitational force between two objects is directly proportional to the mass of each object and inversely proportional to the distance between these two objects (Yotov et al., 2017). Ever since, the gravity model has been adapted and has been part of many disciplines, including Economics, through the works of Isard (1954), Ulman (1954) or Jan Tinbergen (1962). The nature of the gravity model allows to investigate spatial interaction patterns including international trade, migration or commuting laws (Burger et al., 2009).

The model, in its elementary version, is defined as follows:

$$Flow_{ij} = A\frac{M_i^{\beta_1}M_j^{\beta_2}}{D_{ij}^{\beta_3}} \qquad (1)$$

Where:

i	denotes the commune of origin (living location);
j	denotes the destination commune (work location);
$Flow_{ij}$	is the dependent variable, representing the commuting flows between location i and j;
M_i & M_j	are the masses of locations i and j (such as population or GDP);
β_1 & β_2	are the attraction potentials of locations i and j;
$D_{i,j}$	are the travel frictions between locations i and j;
A	is a proportionality constant.

Although the model seems to be simple, it is considered to represent a realistic general equilibrium environment, which can be further augmented to simultaneously accommodate multiple economic variables (Burger et al., 2009; Yotov et al., 2017). It also must be noted that $D_{i,j}$ is not mandatorily a geographical dimension and may relate to aspects such as informational or social "distance" (Gargiulo et al., 2012). Moreover, the model can be transformed to a log-log form, yielding a testable equation which solves for non-linearity, reduces the outliers, and allows for easy elasticity-based interpretation (Burger et al., 2009).

The log-log specification, however, suffers from several limitations. First, it does not allow to control for zero-commuting flows as the zero logarithm is not defined. However, since the dataset on commuting flows described in Section 3.1. by default only includes commuting flows of a minimum of 10 people, this limitation is not relevant for this paper. Should the data be available at a more granular level, one should consider implementing the Poisson specification of the gravity model, which allows to account for non-negative continuous variables (Burger et al., 2009). It is also the most secure method to account for zero-flows in terms of theoretical and empirical justification (Burger et al., 2009).

Another limitation of the logarithmic transformation is that it might create bias in the estimation process, as the antilogarithms of the model estimates tend to underpredict large flows (Flowerdew & Aitkin, 1982). Here, again, Poisson specification of the model would be a useful alternative, as the specification estimates the variables in an absolute form, removing the antilogarithmic transformation bias.

Third, the model assumes homoscedasticity, meaning that all error terms have the same variance for all pairs of origins and destinations (Burger et al., 2009). When this assumption is not met, the efficiency and consistency of the log-linear estimators is hindered as they will cause the error term to become correlated with the covariates (Burger et al., 2009; Santos-Silvia & Tenreyro, 2006). The estimates from the Poisson regression are consistent in the presence of heteroscedasticity, and reasonably efficient especially with large samples (Burger et al., 2009). Therefore, this estimation could be a good mitigation of the heteroscedasticity issue.

Lastly, the estimation using the above gravity model estimation may suffer from endogeneity, when the explanatory variable is correlated with the error term of the model due to omitted variable bias (OVB), measurement errors or reverse causality (Baier & Bergstrand, 2007; Disidier & Head, 2008). Given the quality of the data source, the research assumes that the measurement errors do not interfere with the results. OVB is a common threat to the existing research, due to unobserved heterogeneity because of confounding variables (Baier & Bergstrand, 2007; Disidier & Head, 2008). There are several approaches to reduce OVB, such as the inclusion of control variables (Filipini & Molini, 2003). A more effective approach is to include fixed effects, which allows to account for all commune-specific, time-invariant unobserved characteristics (Agnosteva et al., 2014; Baier and Bergstrand, 2006; Egger and Nigai, 2015). However, the fixed effects approach still does not account for the time-variant unobserved characteristics, making the estimates still prone to endogeneity (Baier and Bergstrand, 2006). The third reason – reverse causality, may occur when the explanatory variable is endogenous to the commuting flows. This might be a threat to the research, as increased commuting flows between two communes, may encourage policy makers to improve public transport infrastructure and road access, to further stimulate the economic exchange and to take off the pressure of congestion. This reverse causality can be partially tackled using the lagged values of the explanatory variables as instruments, as these are unlikely to be affected by the commuting flows from 2016 (Anderson, 1979; Mathyas & Harris, 1998). However, using lagged values of explanatory variables as IVs is still not a guarantee to deal with endogeneity with full certainty (Reed, 2015). A more effective solution is to use the lagged values of the endogenous variable (Reed, 2015). However, due to the unavailability of data, this solution cannot be pursued. If endogeneity is accounted for appropriately, the gravity equation is a precise statistical and explanatory method. Else, only the associations between dependent and independent variables can be established.

To assess the hypotheses, the following gravity equation is used, which uses control variables to reduce OVB and applies robust standard errors and clustering using road distance.

The augmented model used to test hypotheses is specified as follows:

$$\begin{aligned} ln(Flow_{ij}) = {} & \beta_0 + \beta_1 public\_dummyT_i + \beta_2 publicT\_dummy_j + \beta_3 \ln(geoD_{ij}) \\ & + \beta_4 \ln(roadT_{ij}) + \beta_5 \ln(roadD_{ij}) + \beta_6 \ln(Pop_i) \\ & + \beta_7 \ln(Pop_j) + \beta_8 \ln(empl_i) + \beta_9 \ln(empl_j) + \beta_{10} \ln(Inc2016_i) \\ & + \beta_{11} \ln(Inc2016_j) \\ & + \beta_{12} \ln(expT2016_i) + \beta_{13} \ln(expT2016_j) + \beta_{14} roadA\_dummy_i \\ & + \beta_{15} roadA\_dummy_j + \beta_{16} \ln(emplGender_i) \\ & + \beta_{17} \ln(socialA_i) + \beta_{18} poviatDummy_{ij}\, \varepsilon_{ij} \end{aligned} \tag{2}$$

Where: a

i	denotes the commune of origin (living location);
j	denoted the destination commune (work location);
$ln(Flow_{ij})$	is the dependent continuous variable, representing the commuting flows between location i and j (in absolute number of commuters);

$publicT\_dummy_i$ & $publicT\_dummy_j$	are dummy variables being equal to 1 when the percentage of villages in a commune connected by at least one mode of public transport (bus or train stop), excluding school transport are between 99-100%, and 0 otherwise;
$\ln\,(geoD_{ij})$	is a continuous variable, representing the geodetic distance between the centroids of location i and j (in km);
$\ln\,(roadT_{ij})$	is a continuous variable of the average travel time by car between the centroids of communes i and j (in min);
$\ln(roadD_{ij})$	is a continuous variable, representing the distance by road between the centroids of location i and j (in km);
$\log(Pop_i)$ & $\log(Pop_j)$	are continuous variables representing the populations of locations i and j (in absolute number of people);
$\ln(empl_i)$ & $\ln\,(empl_j)$	are continuous variables representing the amount of people employed within a commune per 1000 inhabitants (in absolute numbers);
$\ln(Inc2016_i)$ & $\ln(Inc20016_i)$	are continuous variables representing the average incomes earned by the communes, averaged per commune inhabitant (in PLN);
$\ln\,(expT2016_i)$ & $\ln\,(expT2016_j)$	are continuous variables representing transport-related expenditures by the communes, averaged per inhabitant of a commune (in PLN);
$roadA\_dummy_i$ & $roadA\_dummy_j$	are dummy variables being equal to 1 when the percentage of villages within a commune with hardened (asphalt) road is between 99-100%, and 0 otherwise;
$\ln(socialA_i)$	is a continuous variable representing the % of total households on social assistance within a commune receiving social assistance due to income reasons;
$\ln\,(emplGender_i)$	is a continuous variable representing the % of employed women in the total employed population within a commune;
$powiat\_dummy_{ij}$	is a dummy variable that equals 1 when the commuting flow between commune i & j occurs within the same county;
ε_{ij}	is the unobservable error term, which is assumed to be independent and identically distributed.

In the raw dataset, the variables relating to accessibility to hardened roads and public transport were categorical variables. The % of villages within a commune with asphalt roads variable, was represented using the following categories [1: 99-100%, 2: 90-999%, 3:75-90, 4: 60-75%, 5: <60%]. However, when performing the gravity models, these categories resulted in insignificant results. As such, they were converted into a dummy variable, representing 1 if the % of villages with asphalt roads is between 99-100%, and 0 otherwise. The same has been done with the variables relating to public transport accessibility, originally representing the percentage of villages in a commune connected to at least one mode of public transport (bus or train stop), excluding school transport using the following categories [1: 99-100%, 2: 90-99%, 3: 80-90%, 4: 70-80%, 5: <70%;]. The public transport dummy is 1 if the % of villages connected is between 99-100%, and 0 otherwise. These transformations may not only improve the significance of the results but also improve the ease of interpretation and allow for interaction terms.

The descriptive statistics are given in Table1.

Table 1

Descriptive Statistics Used in Equations (1) and (2)

Variable	Obs.	Mean	Std. Dev	Min	Max
$ln(Flow_{ij})$	45,975	3.4685	1.0318	2.3026	9.0343
$ln(geoD_{ij})$	45,975	3.0500	1.1262	-1.2379	6.4438
$ln(roadT_{ij})$	45,975	3.5207	.8921	.0953	6.3534
$ln(roadD_{ij})$	45,975	3.3387	1.1004	-.6349	6.7174
$ln\ (Pop_i)$	45,975	9.4305	1.2036	5.9713	14.3774
$ln(Pop_j)$	45,975	9.9999	1.7304	6.8156	14.3774
$ln(empl_i)$	44,994	4.9506	.7289	1.0986	7.9233
$ln(empl_j)$	45,099	5.2048	.7850	2.3979	7.9233
$ln(Inc2016_i)$	29,265	8.3281	.1907	7.9451	10.7729
$ln(Inc2016_j)$	31,936	8.4337	.2838	7.9451	10.7729
$ln(expT2016_i)$	29,257	15.2428	1.8988	9.2326	22.0192
$ln(expT2016_j)$	31,929	16.3044	2.6468	9.2326	22.0192
$publicT\_dummy_i$	46,355	.1652033	.371368	0	1
$publicT\_dummy_j$	46,355	.1893647	.3918023	0	1
$roadA\_dummy_i$	46,355	.0080035	.0891043	0	1
$roadA\_dummy_j$	46,355	.0085643	.0921476	0	1

$\ln(socialA_i)$	29,265	-.4284	.1670	-1.4647	0
$ln(emplGender_i)$	45,048	-.6971	.2089	-2.4079	-.07411
$poviat\_dummy_{ij}$	45,975	.4224	.4939	0	1

Note. All numbers are rounded to 4 decimal places.

3.2. The Logit Model

The dataset of commuting flows does not contain commuting flows below 10 commuters. As such, only positive flows are included. Therefore, before proceeding with the analysis of the intensity of the flows using the gravity model, it would be insightful to determine what factors lead to having any flows at all between the communes. Moreover, it would be insightful to test whether the determinants of the existence of commuting flows are in line with the determinants of the intensity of the existing communing flows. Since generating the missing flows for all the communes proved to be too data-intense for the STATA software, the analysis will focus only on the Mazowieckie voivodeship. This voivodeship has been chosen due to the highest number of communes, amounting to 314 (see Appendix A.1.2.).

This dataset has been obtained by creating an origin-destination pair for each commune combination within the Mazowieckie region, resulting in 98,596 possible connections. Having obtained these, the population, communal income, transport expenditure, employment levels and road and transport accessibly variables were merged for each origin and destination commune. On top of that, a dummy for the commuting flows was created. Given the data availability, the commuting dummy amounts to 1 if the flow is equal to or above 10 people, and 0 otherwise.

To investigate the determinants of the commuting flows, a logit model has been chosen and specified with the commuting flow dummy as the dependent variable. This model has been chosen, as it allows to represent a binominal response through applying a generalized linear model and a maximum likelihood estimation (Powers & Xie, 2000). Logit is a linear function of the predictors with the systematic structure defined as follows:

$$logit(\pi_i) = \beta X_i \tag{3}$$

Where:

π_i	Is the linear probability;
X_i'	is a vector of the covariates;

β is a vector of the regression coefficients.

Where a β_i represents a change in log odds of the existence of a commuting flow, associated with a unit change in the j-th predictor, ceteris paribus (Powers & Xie, 2000). Thus, the probability of the existence of a commuting flow can be measured using this model. The logit model is applied to the case of Mazowieckie voivodeship, yielding the following equation:

$$logit(flow\_dummy_{ij}) = publicT\_dummy_i + publicT\_dummy_j + roadA\_dummy_i + roadA\_dummy_j + empl_i + empl_j + pop1000_i + pop1000_j + exp2016\_mln_i + exp2016\_mln_j + inc2016_i + inc2016_j + \varepsilon_{ij} \quad (4)$$

Where:

$flow\_dummy_{ij}$	is a dummy variable, being equal to 1 if there if the commuting flow between origin and destination is equal to 10 commuters or above;
$publicT\_dummy_i$ & $publicT\_dummy_j$	are dummy variables being equal to 1 when the percentage of villages in a commune connected by at least one mode of public transport (bus or train stop), excluding school transport, is between 99-100%, and 0 otherwise;
$roadA\_dummy_i$ & $roadA\_dummy_j$	are dummy variables being equal to 1 when the percentage of villages within a commune with hardened (asphalt) road is between 99-100%, and 0 otherwise;
$empl_i$ & $empl_j$	are continuous variables representing the amount of people employed within a commune per 1000 inhabitants (in absolute numbers);
$pop1000_i$ & $pop1000_j$	are continuous variables representing the populations of locations i and j (in 000s);
$exp2016\_mln_i$ & $exp2016\_mln_j$	are continuous variables representing the average incomes earned by the communes averaged per commune inhabitant (in million PLN);
$inc2016_i$ & $inc2016_j$	are continuous variables representing transport-related expenditures of the communes averaged per inhabitant of a commune (in PLN);

ε_{ij} is the unobservable error term, which is assumed to be independent and identically distributed.

The descriptive statistics for the above variables are presented in Table 2.

Table 2

Descriptive Statistics Used in Equations (3) and (4)

Variable	Obs.	Mean	Std. Dev	Min	Max
$flow\_dummy_{ij}$	98,325	.0238	.1524856	0	1
$publicT\_dummy_i$	98,325	.4548	.4979542	0	1
$publicT\_dummy_j$	98,325	.4492	.4974177	0	1
$pop1000_i$	98,325	17.9390	105.9228	1.7140	1753.9770
$pop1000_j$	97,392	17.2246	100.2308	1.7140	1753.9770
$empl_i$	98,325	129.9784	98.41996	0	511
$empl_j$	97,392	130.5821	99.07575	0	578
$inc2016_i$	98,325	4155.627	696.9849	3029.35	8417.92
$inc2016_j$	96,740	4153.129	691.3646	3029.35	8417.92
$exp2016\_mln_i$	98,325	17.1676	219.1587	.0993	3654.261
$exp2016\_mln_j$	96,740	15.8018	207.8518	.0993	3654.261
$roadA\_dummy_i$	98,325	.0096	.0977171	0	1
$roadA\_dummy_j$	98,325	.0097	.0979719	0	1

Note. All numbers are rounded to 4 decimal places.

Unfortunately, commuting frictions had to be excluded from the logit analysis, as they were obtained from the same GUS (2019) dataset as the commuting flows. As a result, they are only present for the commuting flows equal to or above 10 people. As such, they cannot be used for the binary analysis using the logit regression. The generation of the commuting frictions for the zero-flow communes is beyond the scope of this paper.

2. Results & Discussion

4.1. The Determinants of the Commuting Flows

A logit model is performed to test what determines the existence of commuting flows and whether it is in line with what determines the intensity of the existing commuting flows. This test is performed on the example of Mazowieckie voivodeship. The corresponding results can be found in Table 3.

Table 3

Logit Regression Results for Equation (4)

$flow\_dummy$	(1)	(2)	(3)
$publicT\_dummy_i$	.6113***	.6122***	.5773***
	(.0559)	(.0557)	(.0563)
$publicT\_dummy_j$	.7397***	.7475***	.6909***
	(.0564)	(.0563)	(.0571)
$pop1000_i$	.0015***		.0149***
	(.0001)		(.0012)
$pop1000_j$	.0022***		.0145***
	(.0001)		(.0010)
$empl_i$	.0027***	.0028***	.0019***
	(.0003)	(.0003)	(.0003)
$empl_j$	.0062***	.0063***	.0058***
	(.006)	(.0002)	(.0003)
$inc2016_i$	-.0002***	-.0002***	-.0002***
	(.0002)	(.0000)	(.0000)
$inc2016_j$	-.0001*	-.0000	-.0001***
	(.0000)	(.0000)	(.0000)
$expT2016\_mln_i$		.0007***	-.0063***
		(.0000)	(.0006)
$expT2016\_mln_j$		.0010***	-.0058***
		(.0001)	(.0005)
$roadA\_dummy_i$	-.3541	-.3499	-.2861
	(.2868)	(.2859)	(.2869)
$roadA\_dummy_j$	-.3817	-.3845	-.2871
	(.3227)	(.3225)	(.0000)
Constant	-5.3831***	-5.5772***	-5.2199***
	(.2210)	(.2178)	(.2239)
Observations	96,740	96,740	96,740
AIC	13705.57	13768.1	13483.05
BIC	13809.85	13872.38	13606.29
Pseudo-R^2	0.1965	0.1928	0.2098

Note. Standard errors are displayed in the parentheses. Significance level is indicated as follow: *: p<0.1, **: p<0.05, ***: p<0.01. All numbers are rounded up to 4 decimal places. AIC stands for Akaike Information Criterion and BIC stands for Bayesian Information Criterion.

The logit variables corresponding to equation (4) have been tested for multicollinearity using a correlation matrix which can be found in Appendix A2. This has been done, as multicollinearity may lead to unstable partial regression coefficients, undermining the statistical significance of the results (Allen, 1997). From the correlation matrix in Appendix A2, it may be concluded that population and expenditure of transportation are almost perfectly correlated with each other. Therefore, to avoid collinearity, one should be removed. From Table 3 it may be read that the model excluding expenditure of transport has lower AIC and BIC criteria, as well as the highest Pseudo-R^2. A such, model (1) should be used for the final interpretation of the results.

To avoid ecological fallacy, the conclusions from Table 3 are only applicable to the Mazowieckie region. However, they might still give an idea about the determinants of commuting flows. From model (1), it can be concluded that all variables are significant at 1% level, except for road accessibility dummies. Moreover, confidence intervals of the variables for the sample of no commuting flows and with commuting flows were computed and compared to test whether there exist significant differences between the variable means across the two samples. The t-test results are presented in Appendix A3. According to the analysis, all variables but road accessibility dummies have significant differences in the means between the two samples.

This insignificant result of road accessibility, which contrasts with the later gravity equation results, might be because the dummy does not consider the existence of roads but only whether they are asphalted or not. Thus, a commuter using private car transportation can still use the asphalted road, which will lower the comfort of the trip but will not make it impossible.

The remaining variables have expected results, with public transport accessibility having the strongest association with the existence of commuting flows out of all the variables. This shows that, at least in the Mazowiecki Voivodeship, public transport accessibility plays a big role in the commuting patterns. The remaining variables also have an effect, with the smallest one being for the communal income.

4.2. Testing the Hypotheses

A log-linear gravity model will be performed to test the hypotheses. However, due to possible multicollinearity between the commuting-friction-related independent variables, a baseline regression is first performed, with the corresponding results presented in Table 4.

Table 4

Baseline Gravity Equation Results

$lnFlow_{ij}$	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$publicT\_dummy_i$	-.0396***	-.0435***	-.0437***	-.0466***	-.0415***	-.0424***	-.0456***
	(.0127)	(.0126)	(.0127)	(.0126)	(.0127)	(.0126)	(.0122)
$publicT\_dummy_j$	-.0067	-.0086	-.0142***	-.0136	-.0086	-.0073	-.0125
	(.0121)	(.0121)	(.0121)	(.0121)	(.0121)	(.0121)	(.0121)
$\ln(geoD_{ij})$	-.1893***			.2051***	-.1198***		.2253***
	(.0040)			(.0305)	(.0175)		(.0310
$\ln(roadD_{ij})$		-.1996***		-.4073***		-.3193***	-.5667***
		(.0041)		(.0312)		(.0245)	(.0436)
$\ln(roadT_{ij})$			-.2369		-.0903***	.1501***	.1742***
			(.0050)		(.0217)	(.0299)	(.0303)
Constant	4.0538***	4.1438***	4.3125***	4.2132***	4.1603***	4.0145	4.0700***
	(.0142)	(.0156)	(.0193)	(.0187)	(.0285)	(.0289)	(.0294)
Observations	45,975	45,975	45,975	45,975	45,975	45,975	45,975
AIC	131340.4	131213.3	131370.3	131169.3	131325.7	131188.5	131135.7
BIC	131375.3	131248.2	131405.2	131213	131369.4	131232.2	131188.1
R^2	0.0429	0.0456	0.0423	0.0465	0.0433	0.0461	0.0473
Variance Inflation Factor (VIF)							
$publicT\_dummy_i$	1.06	1.06	1.06	1.06	1.06	1.06	1.06
$publicT\_dummy_j$	1.07	1.07	1.07	1.07	1.07	1.07	1.07
$\ln(geoD_{ij})$	1.02			52.55	17.63		53.21
$\ln(roadD_{ij})$		1.01	1.01	52.48		30.35	91.59
$\ln(roadT_{ij})$					17.59	30.32	30.70

Note. Standard errors are displayed in the parentheses. Significance level is indicated as follow: *: p<0.1, **: p<0.05, ***: p<0.01. All numbers are rounded up to 4 decimal places. AIC stands for Akaike Information Criterion and BIC stands for Bayesian Information Criterion.

According to Table 4, the variance inflation factor (VIF) is much above 10 when geodetic distance, road distance and travel time via roads variables are included in the baseline regression at the same time. This signifies that these independent variables are correlated with each other, leading to multicollinearity (Allen, 1997). This is also confirmed by the correlation table, which can be found in Appendix A4. In turn, multicollinearity may lead to unstable partial regression coefficients, undermining the statistical significance of the results. Therefore, to increase the interpretative value of the coefficients, the gravity model should not include all of them at the same time (Allen, 1997). Based on lowest VIF, AIC and BIC criteria, and the highest R^2 in the regression results from Table 4, model (2) is the most accurate to predict the gravity flows.

Based on the baseline regression results from Table 4, a gravity model will be performed with road distance being a proxy for the commuting frictions. The gravity model employs heteroscedasticity-robust standard error and clustering according to road distance, to allow for correlation of the error terms within groups defined by road distance. Such clustering should help to avoid the understatement of standard errors in the gravity models (UNESCAP, 2016). The expenditure on public transport in the origin and destination communes was excluded from the gravity model as, according to the Variable Inflation Factor (VIF) and a corresponding correlation table, the variable was multicollinear with origin and destination populations (see Appendix A5.1. & Appendix A5.2.). Excluding this variable resulted in the lowest VIFs, BIC and AIC criteria and the highest R^2 (see Appendix A5.1.).

To test the hypotheses, interaction terms between public transport accessibility and the % of people on social assistance due to income reasons and the % of female participation in the total working population were implemented, with the corresponding results visible in Table 5, model (3).

Table 5

Gravity Equation Regression Results for Equation (2)

$lnFlow_{ij}$	(1)	(2)	(3)
$publicT\_dummy_i$	-.0412**	-.0424**	-.0791
	(.0189)	(.0189)	(.0772)
$publicT\_dummy_j$	-.0452**	-.0462**	-.0466**
	(.0197)	(.0198)	(.0197)
$\ln(roadD_{ij})$	-.9840***	-.9848***	-.9849***
	(.0145)	(.0145)	(.0145)
$\ln(Pop_i)$	.3534***	.3523***	.3525***
	(.0089)	(.0090)	(.0090)
$\ln(Pop_j)$	.4255***	.4259***	.4259***
	(.0079)	(.0079)	(.0079)
$\ln(empl_i)$	-.0585***	-.0490***	-.0489***
	(.0143)	(.0159)	(.0160)
$\ln(empl_j)$	.5699***	.5742***	.5756***
	(.0155)	(.0154)	(.0154)
$\ln(Inc2016_i)$	-.0033	0.0158	.0193
	(.0453)	(.0454)	(.0454)
$\ln(Inc2016_j)$	.2637***	.2598***	.2584***
	(.0392)	(.0389)	(.0389)
$roadA\_dummy_i$	.1933***	.1833***	.1832***
	(.0650)		

		(.0653)	(.0655)
$roadA\_dummy_j$	.1828***	.1759***	.1773***
	(.0561)	(.0561)	(.0558)
ln $(emplGender_i)$		.0122	.0413
		(.0433)	(.0493)
ln $(emplGender_i)$* $publicT_i$			-.1577*
			(.0818)
ln$(socialA_i)$		.1789***	.1517***
		(.0431)	(.0470)
ln$(socialA_i)$*$publicT_i$			.1582
			(.1180)
$poviat\_dummy_{ij}$	.3953***	.3977***	.3987***
	(.0199)	(.0198)	(.0199)
Constant	-5.8246***	-5.9265***	-5.9460***
	(.4634)	(.4634)	(.4652)
Observations	21,787	21,787	21,787
AIC	49371.69	49344.43	49339.99
BIC	49475.55	49464.27	49475.81
R^2	0.5372	0.5378	0.5380

Note. Standard errors are displayed in the parentheses. Significance level is indicated as follow: *: p<0.1, **: p<0.05, ***: p<0.01. All numbers are rounded up to 4 decimal places. AIC stands for Akaike Information Criterion and BIC stands for Bayesian Information Criterion.

Overall, the coefficients of the models are rather significant, at least at the 10% significance level, with the F-test p-value being 0.0000 for all three models, meaning that the independent variables in the models are jointly significant at the 1% significance level (Frost et al., 2020). This means that the independent values used in the models improve their fit (Frost et al., 2020).

Road distance variable is significant at 1% significance level in all three models, with 1% increase in road distance, being associated with a 1% decrease in the commuting flow. This result is sensible as it is in line with the assumptions of the gravity model, that the commuting flows between two locations are inversely proportional to the distance between them (Yotov et al., 2017).

Population levels are also significant at 1% significance levels in all three models. An increase of 1% in the population of the origin communes, is associated with about 0.35% increase in the commuting flows. An increase of 1% in the population of the destination communes is associated with about 0.43% increase in the commuting flows. These coefficients are also in line with the scientific discourse, as higher population of the communes, should result in higher trip production and attraction potentials of the communes due to bigger "masses" of the locations, as outlined in the gravity model (Yotov et al., 2017).

Employment level per 1000 commune inhabitants is also significant at 1% significance levels in all three models, and in both the origin and destination communes. An increase of 1% in the employment per 1000 inhabitants in the origin communes, is associated with about 0.05%-0.06% decrease in the commuting flows. An increase of 1% in the employment levels of the destination communes is associated with about 0.57%-0.58% increase in the commuting flows. This is in accordance with the expectations, as high employment in the origin commune is a proxy for its well-functioning labour market, which should act as disincentives to commute, given the extra time and monetary costs that commuting incurs. On the other hand, high employment in the destination commune, is an incentive to commute to that commune, given its abundant labour opportunities that should boost its trip attraction potential (Gerrite, 2021).

Income earned by the origin commune is insignificant at 10% significance levels in all three models. Income earned by the destination commune is significant at 1% significance levels in all three models. An increase of 1% in the income level of the destination communes is associated with about 0.26% increase in the commuting flows. This is in accordance with the expectations, as high income earned in the destination commune is a proxy for its economic attractiveness, hence its attractiveness for the commuters. Therefore, the high income should improve the trip attraction potential of the destination communes and, subsequently, increase the commuting flows (Gerritse, 2021).

Lastly, the powiat dummy, signifying whether the commute is across a powiat border, is significant at 1% significance level in all three models. Commuting within the same powiat is associated with an increase in the commutes of about 0.40%. This is also in accordance with the expectations, as travelling within the same powiat might prove easier in terms of the geographical distance but also in terms of traffic or social and cultural distance.

4.2.1. Hypothesis 1

To recall, Hypothesis 1 has been as follows:

H1: *The accessibility to public transport services is significantly and positively associated with the employment flows between the communes***.**

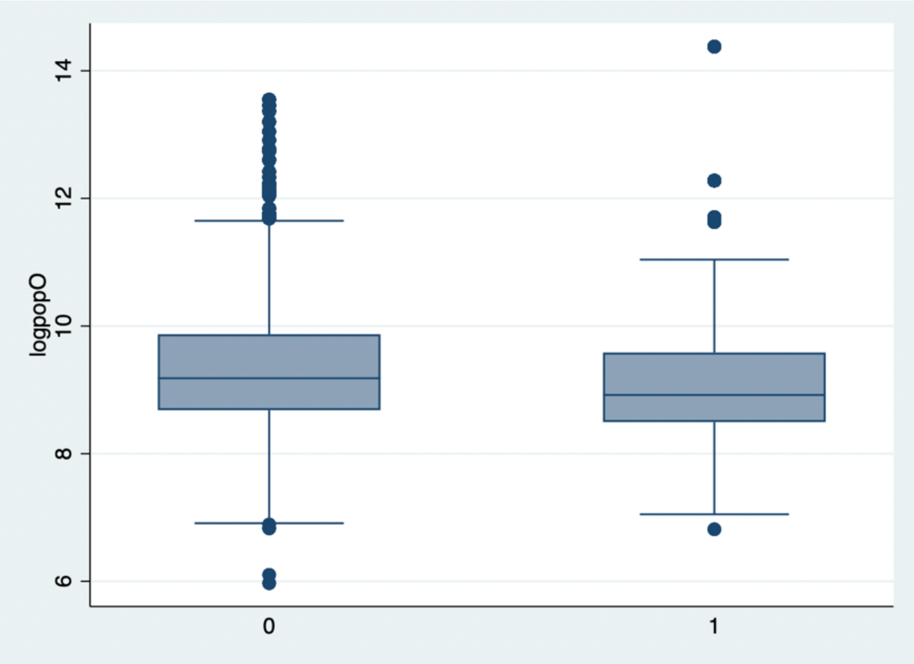
In all three models from Table 5, contrary to what Hypothesis 1 stated, the coefficients of public transport accessibility of the origin and destination communes have negative coefficients. *Perfect public transport accessibility* (i.e. when 99-100% villages within a commune are connected to public transport, represented by the public transport dummies being equal to one) in the destination communes, is associated with a decrease in the level of commutes of about -0.05%, at 5% significance level in all three models. *Perfect public transport accessibility* in the origin communes, is associated with a decrease of around -0.04% at 5% significance level in model (1) and (2) and insignificant in model (3). Therefore, Hypothesis 1 should be rejected at 5% significance levels, as majority of the public transport accessibility coefficients are negative and significant.

The negative coefficients of public transport accessibility may stem from a variety of reasons. First, the commuting flow variable is bi-directional, while the public transport accessibility dummies only refer to one commune at a time. Despite these dummy variables being the best proxies for public transport accessibility in Poland available for this research, the disparity between their uni-directionality and the bi-directionality in the commuting flow variable might have resulted in negative coefficients. To alleviate this issue, future research could include accessibility variables that involve the origin and destination communes at the same time, such as average travel time via public transport between two communes or whether there exists a direct or a transfer public transport connection between the origin and destination communes.

Second, the dummy variable for public transport accessibility differentiates only between the level of *perfect accessibility* of 99%-100% of connected villages within a commune and anything below that level. However, such level of *perfect accessibility* is hardly achievable. In this case, smaller communes might be more likely to achieve it, simply given the fact that they are more likely to have less villages that need to be connected to a bus or rail stops. In fact, as visible in Figure 5, communes with *perfect accessibility* tend to be much smaller in terms of population than the communes with *imperfect accessibility*. This small size of the communes, however, may lead to lower commuter outflow because of simply smaller population of the communes and thus a negative coefficient.

Figure 5

Boxplot of Origin Population by Public Transport Accessibility Dummy in the Origin Commune

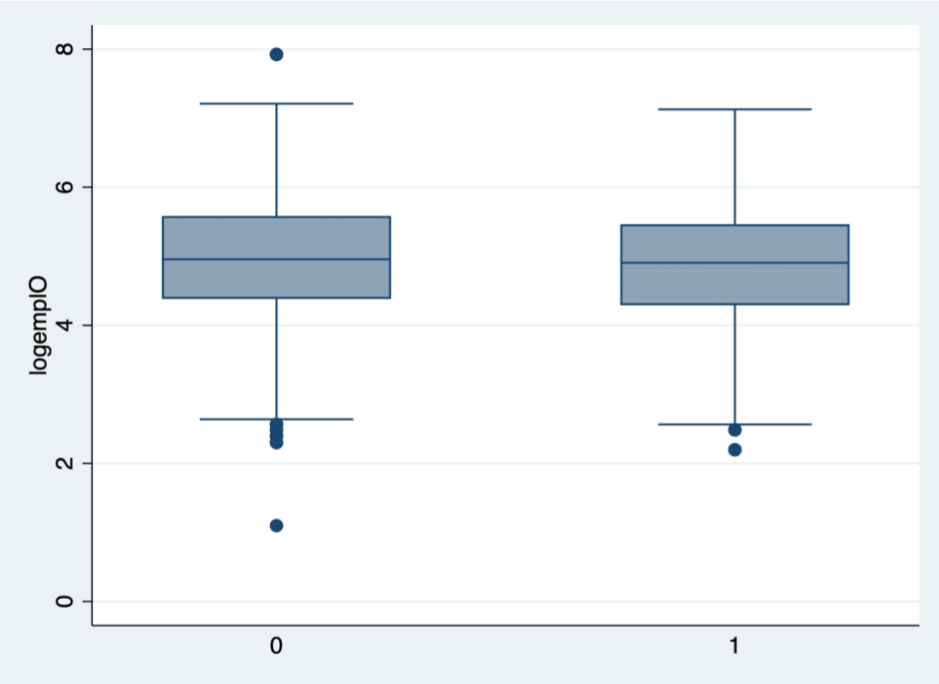


Note. When $publicT\_dummy_i = 1, \ln(Pop_i)$: M = 9.2691, SD = 1.2662; when $publicT\_dummy_i = 1, \ln(Pop_i)$: M= 9.4628, SD= 1.1881.

Another factor contributing to the negative coefficient might be the fact that the communes with perfect public transport accessibility have sufficient income to provide transportation to all villages. This good economic situation of a commune, however, might disincentivise the commuters to commute outside of their communes. This might be a possibility as, as visible in Figure 6 and Figure 7, the origin communes with perfect accessibility, also have higher employment and income levels, when compared to origin communes with imperfect accessibility.

Figure 6

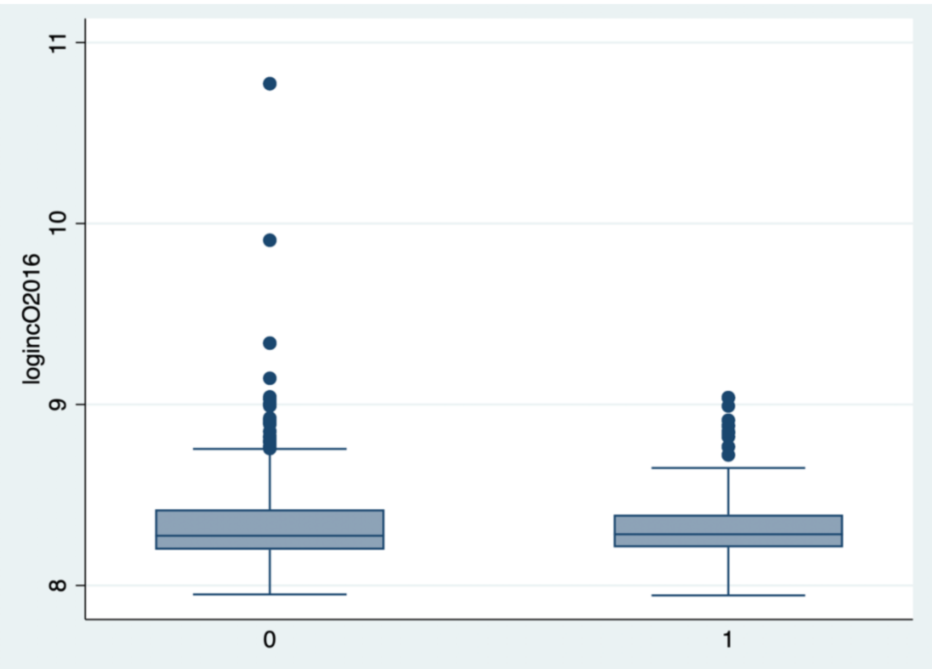
Boxplot of Origin Employment Levels by Public Transport Accessibility Dummy in the Origin Commune



Note. When $publicT\_dummy_i = 1$, $\ln(empl_i)$: M = 4.8896, SD = .7318; when $publicT\_dummy_i = 1$, $\ln(empl_i)$: M = 4.9628, SD = .7278.

Figure 6

Boxplot of Origin Income Levels by Public Transport Accessibility Dummy in the Origin Commune



Note. When $publicT\_dummy_i = 1$, $\ln(Inc2016_i)$ M = 8.3356, SD = .2123; when $publicT\_dummy_i = 1$, $\ln(Inc2016_i)$: M = 8.3266, SD = .1859

Lastly, the issue might stem from the fact that the public transport accessibility dummy, considers rail and bus transportation at the same time. However, since the structure of the two systems vastly differs as rail infrastructure is centralised, while the bus infrastructure is decentralised across many PKS companies (Dubnicki, 2019). To alleviate that, future research should seek for a measure that distinguishes between these two.

Interestingly, however, public transport has a strongly significant and positive effect on the existence of the commutes, as highlighted in the logit regression. This is an interesting insight which may suggest that *perfect public transport accessibility* may be a significant factor to consider when people switch from not commuting to commuting from one location to another. However, once commuting is present, the *perfect accessibility* does not have much association with the intensity of the flows as the existing commuters will commute either way. This might have interesting policy implications, suggesting that public transport improvement policies should focus between areas where no commuting occurs.

4.2.2 Hypothesis 2

To recall, Hypothesis 2 has been as follows:

H2: *Improved public transport accessibility within the communes has the highest significant positive association on the amount of employment flows in the origin communes with a high % of employed women.*

The variable depicting the % of women in the working population of a commune, is not significant in both model (2) and (3) at 10% significance level. However, the interaction term between this variable and public transport accessibility in the origin commune is significant at 10% significance level, meaning that there is a significant cross-over interaction between these two variables (Grace-Martin et al., 2018). When *perfect public transport accessibility* is present, an increase of 1% in share of women in the employed population of a commune, is associated with a decrease in the commuting flows of -0.16%. This contrasts with what Hypothesis 2 is stating, meaning that Hypothesis 2 must be rejected.

Hypothesis 2 stated that improved public transport would be associated with an increase in the commuting flows, when interacted with the variable regarding women employment. Such hypothesis has been drafted, assuming that women are much more prone to transport exclusion than other members of the population (Matas et al., 2010). Hence, should perfect *public transport accessibility* be achieved, it would be positively associated with the commuting in the communes where a high number of women is employed.

The interaction term was negative, however, which might stem from a variety of reasons. One reasons is highlighted by a study by CBOS (2018), which states that women are much more likely to be the caretakers of other house members, including children, animals, or people with disabilities. On average, in 16% of households, women take the sole responsibility of childcare, compared to only 2% of men. In 7% of households, only women takes care of house members with disabilities, while only 3% of men. Moreover, Polish women are on average more involved in other household responsibilities, such as cleaning, cooking or arranging administrative affairs CBOS (2018). This unequal division of household responsibilities might bound women to a work location closer to home, in comparison to men, so that they are more flexible when their help at the household is needed. As such, when the % of working women increases in a commune, the overall commuting flows between the communes decrease due to this inflexibility, despite *perfect public transport accessibility*, which is in accordance with the coefficient obtained.

Another reason for the coefficient might be the fact that the negative public transport dummy coefficients might have skewed the results of the interaction term, in contrast to the hypothesis.

4.2.3 Hypothesis 3

To recall, Hypothesis 3 has been as follows:

H3: *Improved public transport accessibility within the communes has the highest significant positive association on the amount of employment flows in the origin communes with a high % of households on social assistance.*

The coefficient of the % of people on social assistance due to income reasons is positive in model (2) and model (3) and highly significant at 1% significance level. A 1% increase in the amount of people on social assistance due to income reasons, is associated with an increase between 0.15%-0.18% in the commuting flows. The interaction term between public transport accessibility in the origin commune and social assistance is insignificant. Therefore, Hypothesis 3 cannot be rejected or accepted with certainty.

With regards to the effect of the social assistance variable itself, it contrasts with what one would assume based on the Theoretical Framework. In that section, claims from Kamruzzaman & Hine (2012) were provided, that the lowest-income groups tend to locate their activities and employment in local spaces due to limited budget to spend on transportation. This is also supported by the research of De Lima et al. (2016) or Zhu et al. (2017) who, however, focus on the urban areas and suburbs only. The disparity between the coefficient of social assistance of 0.18% and the above findings, might be because this research considers both urban and rural areas. Therefore, the commuting patterns of the lowest income groups might not be as straightforward as in the urban areas only. Moreover, high % of inhabitants of a commune on social assistance due to income reasons, might be an outcome of poor labor market in that commune, pushing the low-income employees outside of the commune despite high commuting costs. In fact, the social assistance variable is negatively correlated with the origin employment level, with a correlation equal to -0.2167 and negatively correlated with the income earned by the commune, with correlation amounting to -0.1707.

Another reason for the coefficient might be the fact that the negative public transport dummy coefficients might have skewed the results of the interaction term, in contrast to the hypothesis.

4.2.4 Hypothesis 4

To recall, Hypothesis 4 has been as follows:

H4: *Road accessibility is significantly and positively associated with the employment flows between the communes***.**

Road accessibility dummies have positive coefficients and are strongly significant at the 1% significance level. For origin communes, *perfect road accessibility* (i.e., when the percentage of villages within a commune with hardened (asphalt) road is between 99-100%,) is associated with an increase in commuting of about 0.18%-0.19%, at 1% significance in all three models. *Perfect road accessibility* in the destination communes is associated with an increase in the commutes of about 0.18%, also at 1% significance level in all three models. Therefore, Hypothesis 4 can be accepted at 1% significance level.

This strong significance and positive coefficient are probably a result of strong dependency of Polish inhabitants on private car transportation, as depicted in Figure 1 or Figure 4. As such, it is logical that better road access will improve the commuting flows between the communes.

Another interesting insight is the fact that road accessibility variables are highly significant when it comes to the intensity of commuting flows but insignificant when it comes to determinants of whether a flow exists or not, in case of the Mazowieckie voivodeship, as one may recall from the logit regression results in Table 5. This might be because the road accessibility dummy considers whether a road is asphalted, and not its existence. Therefore, the fact that a road is asphalted may highly improve the comfort of using a car, and thus increase the intensity of car usage. However, non-asphalted roads may not be enough of a disincentive for car owners to commute, which is where the insignificance of the dummies might have come from in the case of the logit model.

3. Limitations

Despite scientific and social relevance, this paper suffers from several limitations that limits its internal and external validity.

First, the commuting flow data only accounts for flows which are equal to or above 10 commuters. Moreover, no within-commune flows are provided. On one hand, this allows to apply a log-linear gravity model, since the zero-flows do not need to be accounted for. On the other, such data set only represents a subset of the Polish commuters and may limit the external validity of the research. The communal flows are also assumed to be symmetric, and they do not mention whether the commuter needs to pass through several voivodeships or make public transport transfers.

Another limitation of the log-linear model is that it might create bias in the estimation process, as the antilogarithms tend to underpredict large flows (Flowerdew & Aitkin, 1982). Moreover, the log-linear model assumes homoscedasticity, meaning that all errors terms have the same variance for all pairs of origins and destinations (Burger et al., 2009). If this assumption is not met, the efficiency and consistency of the log-linear estimators is hindered. For both cases, the Poisson specification of the model would be a useful alternative, as the specification estimates the variables in an absolute form, removing the antilogarithmic transformation bias and the estimates from the Poisson regression are consistent in the presence of heteroscedasticity, and reasonably efficient especially with large samples (Burger et al., 2009).

Another limitation of the dataset is the fact is the limited data availability on a communal level. As such, important variables from an economic perspective are excluded, such as the costs of commuting or average communal wages, rents or skills. These variables are important to determine the labour market equilibrium as they often determine the demand and supply of the local labour.

With regards to the measures of public transport and road accessibility, the drawback stems from the fact that the raw data was categorical, with irregular brackets. As such, the interpretation of the results is limited, which would not be the case if the data was continuous. Moreover, the public transport measure accounts for both train and bus connections. For more detailed research, it would be beneficial to have this measure separate for buses and trains, especially give the different administrative structures of PKS and PKP enterprises (Dubnicki, 2019).

Moreover, due to unavailability of data, the dataset is cross-sectional, which disallows the application of fixed effects in a panel data, that would improve the internal validity of the research by accounting for all time-invariant characteristics of the communes. The cross-sectionality of the data set also limits the ability to include time-related Instrumental Variables, such as the lagged commuting flows, which would improve the internal validity of the research.

With regards to the internal validity of the research, the gravity model may suffer from endogeneity due to OVB or reverse causality between public transport accessibility, road accessibility and the commuting flows, among others. As such, no causal interpretation of the results can be made and only associations can be investigated. To mitigate this OVB, as already mentioned, the fixed effects approach could be implemented to account for all time-invariant characteristics, should panel data be available (Agnosteva et al., 2014; Baier and Bergstrand, 2006; Egger and Nigai, 2015). However, this solution still does not account for time-variant omitted variables (Baier and Bergstrand, 2006). A solution to reverse causality could be to include lagged values of the explanatory variables as instruments, should such data be available, as these are unlikely to be affected by the commuting flows from 2016 (Anderson, 1979; Mathyas & Harris, 1998). However, according to Reed (2015), using lagged values of explanatory variables as IVs is still not a guarantee to deal with endogeneity with full certainty. Instead, he proposed to use the lagged values of the endogenous variable (Reed, 2020), should such data be available. Else, as already mentioned, Instrumental Variables which have a causal effect on the public transport and road accessibility, but not on the commuting flows, would be the best solution. These could be for example the transport or infrastructure policy.

The external validity is limited, given the fact that the research is not internally valid. Even if the research was internally valid, the external validity would still be limited, given the fact that the research is an empirical study for a specific country case. Therefore, to avoid ecological fallacy, no universal conclusions should be made.

With regards to the logit model, the biggest flaw is the fact that no commuting frictions can be included in it. Future extensions could include self-generated commuting frictions for the communes with zero-flows, mirroring the process conducted by GUS (2019). Moreover, since the logit model is only applicable to the Mazowieckie Voivodeship, no conclusion about the remaining communes can be made, to avoid the ecological fallacy. As such, future improvements could include expanding the model across the entire Poland, should a suitable statistical software be found. Alternatively, the logit analysis could be conducted for all 16 voivodeships.

4. Conclusion & Future Extensions

Overall, the paper serves as a good starting point to investigate the issue of transport exclusion in Poland. Through answering the research question regarding the relationship between public transport accessibility and commuting flows in the Polish commune, the research provides important insights into the issue of economic exclusion – an important topic from the social and economic policy perspective. In short, the relationship between public transport accessibility and commuting flows is complex, as it is dependent on a variety of social, economic and geographical factors which simultaneously influence each other.

Regarding the relationship between public transport accessibility and commuting, the results have been different than expected, displaying a negative coefficient, leading to the rejection of Hypothesis 1. The negative coefficient was most likely, among others, the result of the specific structure of the dummy variable, which could be improve by separating the effect between rail and bus modes and making the variable continuous.

The relationship between working women and commuting flows turned out to be less straightforward than expected, leading to the rejection of Hypothesis 2. This lack of straightforwardness is potentially a result of the multidimensional interactions that women are engaged in within a society and their economic role within it. A possible extension to further investigate the relationship between women and transport exclusion, would be to expand the model with family structures or maritial status.

The relationship between the percentage of people on social assistance due to income reasons and commuting flows also turned out to be less straightforward than expected, with insignificant interaction term leading to the impossibility of Hypothesis 3 rejection or acceptance. This lack of straightforwardness is potentially a result of the wide area of study that has been covered, including both the rural and urban areas of Poland. To improve the research in the future, a deeper analysis should be conducted on the directionality of the flows of low-income households and by considering their living location in more detail, to investigate how the commuting flows of low-income inhabitants differ for rural and urban areas.

The approval of Hypothesis 4 has provided an interesting insight into the nature of Polish commuting structures. Road accessibility was strongly significant and positively associated with the commuting flows, which shows how dependent Polish commuters are on private car transportation. Given the policy relevance of this aspect, further research should be conducted regarding possible solutions to this car dependency.

With regards to the research topic of the paper, extensions could include the investigation of the effects on other social groups, such as people with disabilities or the elderly, if such data was available per commune. Moreover, the research could dive deeper into the directionality of the commutes, by establishing the relationship between commuting between villages and cities. Lastly, the internal validity of the research should be improved, to be able to draw causal conclusions.

In conclusion, given the lack of research regarding public transport in Poland, this paper is a good starting point that establishes the essential relationships between commuting flows and public transport accessibility in Poland, as well as other important economic variables. Although it does not provide insights into causal relationships, it does help to identify the areas where such data needs to be obtained. Moreover, the insights from the presented associations have enabled to delve deeper into the complexity of the transport exclusion in Poland and can serve as a basis for future people-based research and policy recommendations.

5. References

Agnosteva, D., Anderson, J. E., & Yotov, Y. V. (2014). Intra-National Trade Costs: Measures and Aggregation. *National Bureau of Economic Research*. https://doi.org/Working Paper No. 19872.

Allen, M. P. (1997). The problem of multicollinearity. *Understanding Regression Analysis*, 176–180. https://doi.org/10.1007/978-0-585-25657-3_37

Anderson, J. (2011). The Gravity Model. *Annual Review of Economics, 3*, 133-160. doi: 10.1146/annurev-economics-111809-125114

Baier, S. L., & Bergstrand, J. H. (2005). Do Free Trade Agreements Actually Increase Members' International Trade? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.670982

Burger, M., van Oort, F., & Linders, G.-J. (2009). On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation. *Spatial Economic Analysis*, *4*(2), 167–190. https://doi.org/10.1080/17421770902834327

CBOS. (2018). *Kobiety i Mężczyźni w Domu*. https://www.cbos.pl/SPISKOM.POL/2018/K_127_18.PDF.

CEPiK. (2016). *Prawa jazdy i uprawnienia w 2016 r.* http://www.cepik.gov.pl/statystyki

Ciechański, A. (2020). *PTG 2020*, *23*(5), 34–52. https://doi.org/10.4467/2543859XPKG.20.030.13247

Currie, G. (2010). Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, *18*(1), 31–41. https://doi.org/10.1016/j.jtrangeo.2008.12.002

Currie, G., & Delbosc, A. (2010). Modelling the social and psychological impacts of transport disadvantage. *Transportation*, *37*(6), 953–966. https://doi.org/10.1007/s11116-010-9280-2

Currie, G., Richardson, T., Smyth, P., Vella-Brodrick, D., Hine, J., Lucas, K., Stanley, J., Morris, J., Kinnear, R., & Stanley, J. (2010). Investigating links between transport disadvantage, social exclusion and well-being in Melbourne – Updated results. *Research in Transportation Economics*, *29*(1), 287–295. https://doi.org/10.1016/j.retrec.2010.07.036

De Lima, J., Maia, M. L., & Lucas, K. (2016). Income vs. travel time: Why do the poorest and the richest travel fastest in northeastern Brazil? *Transportation Research Procedia*. https://doi.org/10.1016/j.trpro.2017.05.250

Disdier, A., & Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and statistics*, *90*(1), 37-48.

Dobbs, L. (2005). Wedded to the car: women, employment and the importance of private transport. *Transport Policy*, *12*(3), 266–278. https://doi.org/10.1016/j.tranpol.2005.02.004

Dubnicki, A. (2019). Transport Exclusion as a Heritage of the Post-Communist Period. The Example of Poland. *Revista De Ştiinţe Politice/Revue Des Sciences Politiques,* (62), 22-32.

Dudek, A. (2011, August 20). *The Year 1989 – The End of Communism in Poland*. ENRS. https://enrs.eu/en/articles/216-the-end-of-communism-in-poland.

Duszan, A. (2020). The Influence of Public Transport Acts in 2019 in Poland on Conflicts Concerning the Shape of the Public Transport Network. *Logistics and Transport*, *45-46*, 59–70.

Egger, P. H., & Nigai, S. (2015). Structural gravity with dummies only: Constrained ANOVA-type estimation of gravity models. *Journal of International Economics*, *97*(1), 86–99. https://doi.org/10.1016/j.jinteco.2015.05.004

Estimating the Gravity Model. (2016). *UNESCAP*. https://www.unescap.org/sites/default/files/5%20-%203.%20Estimating%20the%20Gravity%20Model_0.pdf.

Eurostat. (2020). Employment and Unemployment. https://ec.europa.eu/eurostat/web/lfs/data/main-tables.

Eurostat. (2021). Modal split of passenger transport. chart. https://appsso.eurostat.ec.europa.eu/nui/submitViewTableAction.do.

Fillipini C., & Molini V. (2003). The Determinants of East Asian trade Flows: a Gravity Equation Approach. *Journal of Asian Economics, 14,* 695-711. doi: 10.1016/j.asieco.2003.10.001

Flowerdew, R. and M. Aitkin (1982) A method of fitting the gravity model based on the Poisson distribution, Journal of Regional Science, 22, 191-202.

Fransen, K., Boussauw, K., Deruyter, G., & De Maeyer, P. (2019). The relationship between transport disadvantage and employability: Predicting long-term unemployment based on job seekers' access to suitable job openings in Flanders, Belgium. *Transportation Research Part A: Policy and Practice*, *125*, 268–279. https://doi.org/10.1016/j.tra.2018.01.023

Frost, J., B, G., Collins, K., Asfaw, G., Donkor, P., Sarah, Calvani, H., Perez, A. G., Nazia, Jk, Boba, Fernandez, N. J., Ali, S. Y., Khan, Nicho, Rashmi, Orozco-Castro, P., Stuart, C., Niklas, … Luong, D.-A. (2020, July 16). *How to Interpret the F-test of Overall Significance in Regression Analysis*. Statistics By Jim. https://statisticsbyjim.com/regression/interpret-f-test-overall-significance-regression/.

Gargiulo, F., Lenormand, M., Huet, S., & Baqueiro Espinosa, O. (2012). Commuting Network Models: Getting the Essentials. *Journal of Artificial Societies and Social Simulation*, *15*(2). https://doi.org/10.18564/jasss.1964

Grace-Martin, K., Kimberly, Diana, Yousuf, B., Manal, Brendelle, Neto, A. A., Kate, Su, F., Trickeration, Lin, P. C., Danish, Sara, Ibrahim, Yjl, Svetlana, Buru, Jane, & Karen. (2018, January 4). *Interpreting Interactions when Main Effects are Not Significant*. The Analysis Factor. https://www.theanalysisfactor.com/interactions-main-effects-not-significant/.

Gerritse, M. (2021). Lecture notes Spatial Economic Models. *Seminar UPT 20/21*.

GUS. (2019, July 8). *Oszacowanie odległości i czasu przejazdu pomiędzy wybranymi gminami w Polsce w 2016 roku*. stat.gov.pl. https://stat.gov.pl/statystyki-eksperymentalne/obszary-funkcjonalne-oraz-dostepnosc-terytorialna/oszacowanie-odleglosci-i-czasu-przejazdu-pomiedzy-wybranymi-gminami-w-polsce-w-2016-roku,11,1.html.

GUS. (2019a, May 20). *Przepływy ludności związane z zatrudnieniem w 2016 r.* stat.gov.pl. https://stat.gov.pl/obszary-tematyczne/rynek-pracy/opracowania/przeplywy-ludnosci-zwiazane-z-zatrudnieniem-w-2016-r-,20,1.html.

GUS. (2020). *Główny Urząd Statystyczny*. Strona główna. https://stat.gov.pl/.

GUS. (2021). *Główny Urząd Statystyczny / Statystyka regionalna / Jednostki terytorialne / Podział administracyjny Polski*. Strona główna. https://stat.gov.pl/statystyka-regionalna/jednostki-terytorialne/podzial-administracyjny-polski/.

Hansen, W.G., 1959. How Accessibility Shapes Land Use. Journal of the American Institute of Planners 25 (2), 73–76.

Heuermann, D., & Schmieder, J. (2018). The Effect of Infrastructure on Worker Mobility: Evidence from High-Speed Rail Expansion in Germany. https://doi.org/10.3386/w24507

Isard, W. (1954) Location theory and trade theory: short run analysis, Quarterly Journal of Economics, 68, 305-22.

Kaczorowski, J. (2019). Przedsiębiorstwa Komunikacji Samochodowej w Polsce. Rozwój, upadek i wynikające wykluczenie komunikacyjne. *V Krakowska Ogólnopolska Konferencja Naukowa Transportu "KOKONAT.* Cracow: Warsaw University of Technology

Kamruzzaman, M., & Hine, J. (2012). Analysis of rural activity spaces and transport disadvantage using a multi-method approach. *Transport Policy*, *19*(1), 105–120. https://doi.org/10.1016/j.tranpol.2011.09.007

Kenyon, S., Lyons, G., & Rafferty, J. (2002). Transport and social exclusion: investigating the possibility of promoting inclusion through virtual mobility. *Journal of Transport Geography*, *10*(3), 207–219. https://doi.org/10.1016/s0966-6923(02)00012-1

Lucas, K. (2012). Transport and social exclusion: Where are we now? *Transport Policy*, *20*, 105–113. https://doi.org/10.1016/j.tranpol.2012.01.013

Lucas, K., van Wee, B., & Maat, K. (2015). A method to evaluate equitable accessibility: combining ethical theories and accessibility-based approaches. *Transportation*, *43*(3), 473–490. https://doi.org/10.1007/s11116-015-9585-2

Matas, A., Raymond, J.-L., & Roig, J.-L. (2010). Job Accessibility and Female Employment Probability: The Cases of Barcelona and Madrid. *Urban Studies*, *47*(4), 769–787. https://doi.org/10.1177/0042098009352364

Mathyas, L., & Harris, M. (1998). The econometric of gravity models. *Melbourne Institute Working Paper*, *5*(98).

Mazur, S., Możdżeń, M., & Oramus, M. (2018). (tech.) (pp. 793–793). Brussels: European Commission: Directorate-General for Employment, Social Affairs and Inclusion.

Monte, F., Redding, S. J., & Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, *108*(12), 3855-90.

Morawiecki, M., Rozporządzenie Rady Ministrów RM-06110-16-21 (2021). https://legislacja.rcl.gov.pl/docs//3/12342853/12759851/dokument486764.pdf.

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

Powers, D. A., & Xie, Y. (2000). Logit and Probit Models for Binary Data. *Statistical Methods for Categorical Data Analysis*, 41–85. https://doi.org/10.1016/b978-012563736-7/50003-3

Rosner, A., & Stanny, M. (2016). (rep.). *Monitoring rozwoju obszarów wiejskich Etap II: Przestrzenne zróżnicowanie poziomu rozwoju społeczno-gospodarczego obszarów wiejskich*. Warsaw.

Santos-Silva, J. M. C. M., & Tenreyro, S. (2003). The Log of Gravity. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.380442

Social Exclusion Unit. (2003). Making the Connections: Final Report on Transport and Social Exclusion. Report by the Social Exclusion Unit 2003. *Proceedings of the Institution of Civil Engineers - Municipal Engineer*, *156*(2), 81–83. https://doi.org/10.1680/muen.2003.156.2.81

Tinbergen, J. (1962) Shaping the World Economy The Twentieth Century Fund, New York

Ullman, E.L. (1954), Geography as spatial interaction, in Interregional Linkages, Proceedings of the Western Committee on Regional Economic Analysis (Eds) D. Rezvan and E.S. Engelbert, UCLA, Berkeley, CA, 63-71.

Varga, L., Tóth, G., & Néda, Z. (2018). Commuting patterns: the flow and jump model and supporting data. *EPJ Data Science*, *7*(1). https://doi.org/10.1140/epjds/s13688-018-0167-3

Wikipedia Foundation. (2021, March 2). *Podział administracyjny Polski*. Wikipedia. https://pl.wikipedia.org/wiki/Podzia%C5%82_administracyjny_Polski.

Wolny, A., Ogryzek, M., & Źróbek, R. (2019). Towards Sustainable Development and Preventing Exclusions—Determining Road Accessibility at the Sub-Regional and Local Level in Rural Areas of Poland. *Sustainability*, *11*(18), 4880. https://doi.org/10.3390/su11184880

Yotov, Y., Piermartini, R., Monteiro, J.-A., & Larch, M. (2017). An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model. *World Trade Orgamnisation*. https://doi.org/10.18356/57a768e5-en

Zarząd Województwa Podlaskiego. (2018). (tech.). *Regionalny plan transportowy województwa podlaskiego na lata 2014-2020*.

Zawojska, A. (2019). Contemporary Dilemmas of Rural Development.

Żakowska, M. (2019). Wykluczenie komunikacyjne - z czego wynika i co trzeba w tej sprawie zrobić [Transport exclusion and what to do with it] "Magazyn miasta", Retrieved from: http://www.tokfm.pl/Tokfm/7,103454,24457013,wykluczenietransportowe-z-czego-wynika-i-co-trzeba-w-tej.html.

Zhu, Z., Li, Z., Liu, Y., Chen, H., & Zeng, J. (2017). The impact of urban characteristics and residents' income on commuting in China. *Transportation Research Part D: Transport and Environment*, *57*, 474–483. https://doi.org/10.1016/j.trd.2017.09.015

Żmuda-Trzebiatowski, P. (2016). Dostępność Transportowa, A Partycypacja W Aktywnościach, Ubóstwo Oraz Zagrożenie Wykluczeniem Społecznym. *Politechnika Poznańska: Zakład Systemów Transportowych*.

6. Appendix

Appendix A1.1.

Administrative Map of Poland

Note. Obtained from GUS (2021)

Appendix A1.2.

Administrative Structure of Poland

Voivodeship Name	Number of Counties	Number of Communes
dolnośląskie	26	169
kujawsko-pomorskie	19	144
lubelskie	20	213
lubuskie	12	82
łódzkie	21	177
małopolskie	19	182
mazowieckie	37	314
opolskie	11	71
podkarpackie	21	160
podlaskie	14	118
pomorskie	16	123
śląskie	17	167
świętokrzyskie	13	102
warmińsko-mazurskie	19	116
wielkopolskie	31	226
zachodniopomorskie	18	113

Note. Obtained from Wikipedia Foundation (2021)

Appendix A2

Matrix of Correlations for Variables Presented in Table 3

Variables	(1)	(2)	(3)	(4)	(5)	(1)	(7)	(8)	(9)	(10)	(11)	(12)
(1) $publicT\_dummy_i$	1.0000											
(2) $publicT\_dummy_j$	.0923	1.0000										
(3) $pop1000_i$	.1790	-.0056	1.0000									
(4) $pop1000_j$	-.0380	.3080	-.0968	1.0000								
(5) $empl_i$	.1099	-.0205	.5835	-.1571	1.0000							
(6) $empl_j$	-.0168	.1984	-.0636	.7228	-.1638	1.0000						
(7) $inc2016_i$	.2061	.0164	.7909	-.0853	.6418	-.0532	1.0000					
(8) $inc2016_j$	.0095	.3312	-.0697	.8763	-.1211	.7876	-.0371	1.0000				
(9) $expT2016\_mln_i$	.1719	-.0037	**.9976**	-.0888	.5590	-.0526	.7774	-.0626	1.0000			
(10) $expT2016\_mln_j$	-.0363	.3041	-.0960	**.9991**	-.1540	.7140	-.0840	.8708	-.0883	1.0000		
(11) $roadA\_dummy_i$	.0794	.0129	-.0223	.0119	-.0461	-.0012	-.0623	.0269	-.0197	.0119	1.0000	
(12) $roadA\_dummy_j$	-.0136	.0542	-.0116	-0.0357	.0087	-.0705	.0081	-.0647	-.0145	-.0337	-.0073	1.0000

Appendix A3

Significant Sample Mean Results

Variable Name	Observations		Confidence Intervals	
Group ($flow\_dummy_{ij}$)	0	1	0	1
$publicT\_dummy_i$	95,983	2,342	[0.4480, 0.4543]	[0.5857, 0.6252]
$publicT\_dummy_j$	95,983	2,342	[0.4421, 0.4483]	[0.5934, 0.6328]
$pop1000_i$	95,983	2,342	[15.4160, 16.5571]	[83.2785, 112.6319]
$pop1000_j$	95,050	2,342	[12.0254, 12.5836]	[194.7563, 239.0567]
$empl_i$	95,983	2,342	[128.4983, 129.7309	[160.1755, 170.5836]
$empl_j$	95,050	2,342	[126.8755, 128.0877]	[250.2654, 262.569]
$inc2016_i$	95,983	2,342	[4145.506, 4154.138]	[4348.022, 4439.066]
$inc2016_j$	95,050	2,342	[4133.182, 4141.449]	[4963.021, 5122.033]
$roadA\_dummy_i$	95,983	2,342	[0.0090, 0.0103]	[0.0051, 0.0128]
$roadA\_dummy_j$	95,983	2,342	[0.0090, 0.0102]	[0.0083, 0.0174]

Appendix A4

Matrix of Correlations for Variables Presented in Table 4

Variables	(1)	(2)	(3)	(4)	(5)
(1) $publicT\_dummy_i$	1.0000				
(2) $publicT\_dummy_j$	0.2296	1.0000			
(3) $\ln(geoD_{ij})$	-0.0161	0.1142	1.0000		
(4) $\ln(roadD_{ij})$	-0.0231	0.1058	0.9904	1.0000	
(5) $\ln(roadT_{ij})$	-0.0267	0.0995	0.9711	0.9834	1.0000

Appendix 5.1.

Baseline Gravity Equation Results for Equation (2)

$lnFlow_{ij}$	(1)	(2)	(3)
$publicT_i$	-.0392**	-.0728***	-.0412**
	(.0190)	(.0201)	(.0189)
$publicT_j$	-.0446**	-.0566***	-.0452**
	(.0196)	(.0213)	(.0197)
$\ln(roadD_{ij})$	-.9849***	-.9265***	-.9840***
	(.0146)	(.0143)	(.0145)
$\ln(Pop_i)$	.4223***		.3534***
	(.0167)		(.0089)
$\ln(Pop_j)$	.4013***		.4256***
	(.0196)		(.0079)
$\ln(empl_i)$	-.0620***	.0512***	-.0585***
	(.0142)	(.0145)	(.0143)
$\ln(empl_j)$	.5733***	.5987***	.5699***
	(.0155)	(.0167)	(.0155)
$\ln(Inc2016_i)$	.0842***	-.2105***	-.0033
	(.0495)	(.0529)	(.0453)

$\ln(Inc2016_j)$	.2303*	-.2105***	.2637
	(.0495)	(.0529)	(.0392)
$\ln(expT2016_i)$	-.0554***	.1972***	
	(.0119)	(.0067)	
$\ln(expT2016_j)$	.0205	.2895***	
	(.0143)	(.0061)	
$roadA_i$	.1857***	.2232***	.1933***
	(.0654)	(.0756)	(.0650)
$roadA_j$	.1734***	.2179***	.1828***
	(.0562)	(.0639)	(.0561)
$poviat\_dummy_{ij}$	.3991***	.2901***	
	(.0199)	(0.216)	
Constant	-4.4953***	-1.6445***	-5.825***
	(.5033)	(.5035)	(.4634)
Observations	21,776	21,776	21,787
AIC	49313.17	51250.76	49371.69
BIC	49433.00	51354.61	49475.55
R^2	0.5381	0.4951	0.5372
Variance Inflation Factor			
$publicT_i$	1.07	1.07	1.07
$publicT_j$	1.12	1.12	1.12
$\ln(roadD_{ij})$	2.54	2.35	2.54
$\ln(Pop_i)$	10.63		2.44
$\ln(Pop_j)$	23.38		3.18
$\ln(empl_i)$	2.08	1.85	2.08
$\ln(empl_j)$	2.57	2.40	2.56
$\ln(Inc2016_i)$	1.70	1.59	1.42
$\ln(Inc2016_j)$	2.80	2.38	2.14
$\ln(expT2016_i)$	10.52	2.34	
$\ln(expT2016_j)$	26.26	3.47	
$roadA\_dummy_i$	1.02	1.02	1.02
$roadA\_dummy_j$	1.02	1.01	1.01
$poviat\_dummy_{ij}$	1.69		1.69
Mean VIF	6.31	1.87	1.86

Appendix A5.2.

Matrix of Correlations for Baseline Equation (2)

	(11)	(12)	(13)	(14)
	1.0000			
	-.0220	1.0000		
	.0123	-.0195	1.0000	
	-.0227	-.2482	-.5064	1.0000

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) $publicT_i$	1.0000									
(2) $publicT_j$	.2270	1.0000								
(3) $\ln(roadD_{ij})$	-.0311	.1111	1.0000							
(4) $\ln(Pop_i)$	-.0649	-.0463	.1825	1.0000						
(5) $\ln(Pop_j)$	-.0168	.1286	.7169	-.0421	1.0000					
(6) $\ln(empl_i)$	-.0267	-.0306	.0787	.7047	-.0596	1.0000				
(7) $\ln(empl_j)$	-.0109	.0643	.6091	-.0392	.7252	-.0575	1.0000			
(8) $\ln(Inc2016_i)$	.0191	.0029	.1426	.5069	.0028	.4647	.0367	1.0000		
(9) $\ln(Inc2016_j)$	.0138	.2045	.5834	-.0466	.6546	-.0366	.6565	.0418	1.0000	
(10) $roadA\_dummy_i$	.0579	-.0068	.0258	.0083	.0067	.0023	.0109	.0558	.0195	1.0000
(11) $roadA\_dummy_j$	.0069	.0107	.0181	-.0171	.0119	-.0147	.0099	.0268	.0335	.1044
(12) $\ln(expT2016_i)$	-.0458	-.0336	.1822	**.9410**	-.0267	.6705	-.0170	.5927	-.0314	.0076
(13) $\ln(expT2016_j)$	-.0090	.1512	.7174	-.0379	**.9735**	-.0503	.7293	.0114	.7266	.0070
(14) $poviatDummy_{ij}$	.0076	-.0306	-.5434	-.2602	-.5186	-.1320	-.4962	-.2088	-.3768	-.0186