Прикладная эконометрика, 2023, т. 71, с. 99–127.
Applied Econometrics, 2023, v. 71, pp. 99–127.
DOI: 10.22394/1993-7601-2023-71-99-127

A. Piliuk, E. Semerikova, A. Nastansky[1]

Determinants of commuting flows in Germany

The paper studies commuting flows between German regions. Using panel data of 400 German regions from 2013 to 2019 we evaluate the effect of the wide range of indicators determining the magnitude of the commuting flows: demographic factors, indicators of the labour and real estate markets, welfare variables, social and educational system characteristics, etc. We employ the gravity model analysis with Poisson Pseudo Maximum Likelihood, allowing us to consider even the absence of commuters between regions. The novelty of the research is that the full structure of commuting flows, including the direction, is analyzed at the aggregated district level. In addition to other papers devoted to the economics of the labor market and focused mostly on individual data and selected determinants, we investigate a wide range of possible factors and conclude that the main macroeconomic factors determining both the intensity and direction of commuting flows: population, unemployment rate, cost of leasing housing and the number of companies per 10000 people. We also find that commuting flows between regions in the same land are 202% higher than between regions from different lands, and commuting flows between neighbouring regions are 414.5% higher than between regions without a common border.

Keywords: commuting; gravity models; Germany; panel data; PPML.
JEL classification: J61; R23.

1. Introduction

Nowadays stable economic growth is inextricably linked with the dynamic labour market. One of the key characteristics of efficient labour resources and high economic activity in a market economy is the mobility of the labour force (Goetz et al., 2010). Free labour mobility across the country can only be realised with a developed road and transport infrastructure that provides communication between regions. Germany occupies one of the first places in the world in terms of transport network density — this applies to both road and rail connections between regions. High road density contributes to increased labour market mobility (Sun et al., 2017), the emergence of frequent territorial movements along the home-to-work route and a socio-economic phenomenon named commuting. The conducted studies demonstrate that the lack of alternatives for work in rural areas, combined with the desire to live outside a large city, is a fundamental push and pull factor for the development of commuting (Partridge et al., 2010). The analysis of the factors affecting the direction and intensity of commuting flows can be useful for both

[1] **Piliuk, Anastasiia** — HSE University, Moscow, Russia; ivnas2011@mail.ru.
Semerikova, Elena — HSE University, Moscow, Russia; lena.sem7@mail.ru.
Nastansky, Andreas — Berlin School of Economics and Law, Berlin, Germany; andreas.nastansky@hwr-berlin.de.

the regulation of the socioeconomic and regional policy of the country and an overview of local labour markets' trends for companies, hiring employees. The effect of commuting on fertility (Huinink, Feldhaus, 2012), life satisfaction (Stutzer, Frey, 2007), health (Hansson et al., 2011; Künn-Nelen, 2016), and regional disparity (Niebuhr et al. 2012) has been investigated in similar research. It was shown that commuting has no significant impact on the wage gap reduction between East and West Germany, but it has a relationship with the unemployment disparity reduction, which corresponds with the neoclassical approach in economics (Niebuhr et al., 2012). Thus, the main labour market factor which produces such dynamics is not the regional wage level, but employment opportunities. In our study, we justify this conclusion. Some other determinants of commuting were also taken into account. In particular, Skora et al. (2020) established a relationship between the first child born and women's commuting distance reduction by 33%. Such a reduction results in a corresponding wage drop for those women with children, who have reduced their commuting distance, causing a so-called "motherhood wage gap". The commuting distance can also be influenced by housing density or vacancy ratio (Zhao, 2013). The data for Beijing 2001 has shown that high-density housing corresponds with a lower commuting distance.

The main research question of this work is about which factors affect the magnitude and intensity of commuting flows between German regions. Instead of specializing in a particular area, we investigate a wide range of possible determinants to find out the determinants of commuting in Germany. It allows us to choose the most relative variables among different economic indicators with no limits imposed by a specific domain, which was not the case before. While most researchers are focused on individual-level data, originating from surveys and census, we use more homogeneous district-level panel data to ensure sufficient and robust data for further analysis. To conduct a comprehensive analysis, it is necessary to consider simultaneously the origin and destination regions of migrants along with the dynamic structure of the migration network. Finally, the inclusion in the analysis not only existing but also zero flows between regions allows us to conclude what is the reason for the absence of commuting between them.

The purpose of this study is to identify the determinants of the intensity of commuting flows in Germany and determine the direction of their influence. To the best of our knowledge, this is the first attempt to conduct such a study for Germany. For the analysis, we employ data on German interregional commuting networks and regional socio-economic characteristics as determinants of the flows' magnitude from 2013 to 2019. The study period ends before the start of the COVID-19 pandemic. The main scientific novelty is in modelling the full structure of commuting flows, which consists of all 400 regions of Germany. The peculiarity of the model is also taking into account the indicator of belonging to the same land, which serves as a proxy for various cultural, economic, ethnic and other characteristics.

The article is structured as follows. The first section presents a literature overview on commuting and gravity models as a method for studying various economic aspects, the second discusses the main hypotheses put forward in this study. The third section explains the data, the fourth describes the econometric model. The fifth and sixth sections present the results of the study. The conclusions follow at the end of the paper. Additional reference materials are issued in the form of Appendixes to work.

2. Literature overview

In the literature, "commuting" may stand for various types of migration flows with no residence change. Namely, research on student flows (Bringolf-Isler et al., 2008; Whalen et al., 2013; Kobus et al., 2015), and urban flows (Kim et al., 2012) are all treated as "commuting". In particular, much research is focused on specific vehicle use for commuting e.g., bicycle-commuting or commuting by personal cars (Handy, Xing, 2011; Santos et al., 2013). To be precise, in our research "commuting" stands for such flows only, that are committed between residence and workplaces by workers, whose residence and working address are in different territorial units. Therefore, the literature review encompasses similar research only.

A significant part of the research on commuting is based on survey data of individuals and households, where their characteristics act as independent variables (Tkocz, Kristensen, 1994; Artis et al., 2000; Shen, 2000; Lee, McDonald, 2003). The authors note that there are significant differences in male and female commuting — men more often work in the suburbs, while women prefer to work in the centre (Tkocz, Kristensen, 1994). Moreover, research results show a tendency for decreasing time to work as the residential place gets closer to the city centre, the exception to this pattern are representatives of low-income groups, who, despite living in the centre, spend significantly more time commuting to work than representatives of more affluent groups (Shen, 2000). Östh and Lindgren (2012) conducted a study on nearly 12 million commutes over 16 years in Sweden. They studied how changes in GDP affect the distance between home and work. First, the results of the model suggest that changes in GDP have a much greater effect on commuting between large centres and their densely populated suburbs than on longer trips. Having also studied the relationship of commuting and changes in GDP with demographic characteristics, the authors came to the conclusion that changes in business activity (GDP) affect young people under the age of 24 more, but gender or ethnicity do not show clear differences in the response of commuters to changes in GDP. Concerning Germany, Viry et al. (2014) studied the relationship between commuting and the workers' career development in Germany. Cluster and regression analysis showed that labour mobility is quite high in the first career years and, conversely, decreases significantly after respondents reach the age of 35. The most interesting result was the absence of clear fluctuations in mobility at the age of greatest fertility (from 25 to 35 years), suggesting that even marriage and childbirth do not become a barrier to commuting movements.

Another area of research on commuting concerns the study of the commuting geography depending on the agglomeration structure where it occurs (Baum-Snow, 2010; Burger et al., 2011; Dauth, Haller, 2018). The authors conclude that employment decentralisation and commuting network dispersion are related to transport accessibility and road network expansion (Baum-Snow, 2010). Regarding Germany, Dauth and Haller (2018) note differences in the structure of commuting flows in single metropolitan areas, such as Berlin and Munich, and in regions with several large cities, such as the Ruhr region. The distance covered by migrants working in megapolises usually exceeds 20 km, suggesting that more than half of Berlin and Munich workers live outside of them and make daily trips from nearby areas. The situation is different in the Ruhr and the Rhein regions of several major cities (e. g. Düsseldorf, Essen, Dortmund), where long-distance commuting is rare and most people in this region commute within 5–10 kilometers.

Commuting flows in Germany have mostly been studied using network analysis or spatial models (Patuelli et al., 2007, 2010; Reggiani et al., 2011). Patuelli et al. (2010) analysed commuting networks in 1995 and 2005 based on data from 439 districts. The authors used two network analysis models (a scale-free network and a random graph model) and estimated the areas' tendency for mobility, and also calculated such characteristics of the commuting network as centralisation, the level of clustering, and the level of entropy. There were derived the concepts of external mobility of the region, which means the number of workers from the region who go to work to another area, and internal mobility of the region, which means the number of workers that this region attracts as a place of work. The study showed that over 10 years both indices have grown in all 9 types of regions (the authors divided all German regions into types, depending on their level of urbanisation and the tendency to create an agglomeration, where the most urbanised are "central cities in regions with urban agglomerations", and the least urbanised are "rural districts in regions with rural features"). The highest level of external mobility was observed in "central cities in regions with urban agglomerations" in both 1995 and 2005 (37 and 53% respectively), while the main sources of external mobility were "highly urbanized districts in regions with urban agglomerations" (39 and 45% respectively). Important is the fact that the network centralisation index has decreased over 10 years, while the network dispersion (entropy) has increased, which indicates that the commuting flows have become more distributed throughout the country, and the process of commuting is now more difficult to predict.

A spatial interaction model (SIM) was applied for migration modeling in (Patuelli et al., 2007) and (Reggiani et al., 2011). This framework takes into account a distance between them for the flow intensity estimation, which makes it very similar to the gravity model. The main focus of this research was on the network structure and the links establishment for new nodes. The results of the SIM model were compared with network analysis outcome with respect to the hypothesis that new nodes mainly establish their connection via main hubs. It is emphasised that since 1995 new hubs have been formed due to an increase in an average number of links per node. However, no other than spatial drivers were considered when estimating the commuting flows, what was considered as possible extensions by the authors. Therefore, defining drivers for commuting in Germany may contribute to the current state of research.

Another line of research for Germany is the comparison of commuting flows in the western and eastern parts of the country against the background of the 20th century (Haas, Hamann, 2008). Authors note that in addition to the commuting flows from the east to the west of the country that have increased for objective reasons, the areas located at the border between the Federal Republic of Germany and the German Democratic Republic are particularly strongly distinguished. 16 years after the German reunification, the fact of the location of these areas still greatly influenced the intensity of migration — the West German regions had a significant positive balance of commuting migrants, who in large numbers went to work from the East German territory. The process of east-west migration has stopped in the meantime and has been replaced by migration movements from north to south.

In addition, we should refer to the literature, where the gravity model is applied as mathematical methods in spatial analysis of various economic topics (export-import trade relations (Bun, Klaassen, 2002; Kimura, Lee, 2006; Baltagi et al., 2014), tourist flows (Khadaroo, Seetanah, 2008; Keum, 2010) or migration (Sawyer, 1967; Bergstrand, 1989; Kim, Cohen, 2010; Ramos, Surinach, 2017; Ahrens, Lyons, 2021)). However, we will focus on papers where authors use the method of gravity models specifically for the commuting and other types of migration.

As for the commuting, the gravity model is rarely used to analyse the behavior of commuters. However, Stefanouli and Polyzos (2017) investigate commuting between Greek regions using three traffic flow estimation models: a standard and extended gravity model, and a radiation model. The analysis showed that the extended model turned out to be better. In addition to the population it includes the GDP indicators of various sectors of the economy. The authors explain this by the fact that economic indicators (such as GDP, especially the GDP of the tertiary sector of the economy) reveal the development of the region and allow potential employees to conclude about the number of jobs in the region. The higher the GDP per capita, the more vacancies there are in that region. The radiation model also shows good results and is quite representative. In addition to the Euclidean distance between regions, it includes the travel time between regions along the main roads of Greece. Ahrens and Lyons (2021), using a gravity model for commuting in Ireland, note that there is a relationship between rental prices in major centres and commuting time. Besides, research results show that with an increase in rental prices by 10% in the largest cities, the travel time to work for commuters increases to 0.3 minutes throughout Ireland or to 1.2 minutes in the Dublin agglomeration.

Another modern method for studying commuting flows is the radiation model (Simini et al., 2012), based on the assumption that people, who are looking for a job, maximise their future income and choose the vacancy where they are offered a salary higher than in their residential place. The authors modify this model in different ways, adding the distance between home and work, as well as the transport costs of commuters (Simini et al., 2013; Varga et al., 2016). One variation of this model is the flow and jump model (FJM) proposed by Varga et al. (2018). Studying commuting travel in the USA, Hungary, and Italy, in the model indirectly was taken into account the fact that long-distance travel is less likely due to its higher cost. The authors found that commuting between home and work is more common in the United States than in Europe. Comparing the models, it turned out that the gravity model gave the same good results as the FJM model, while the standard radiation models showed a poor quality of regression fitting. This conclusion once more indicates to the advantages of gravity model in evaluation migration flows, in particular, commuting.

It is also worth noting some features of the gravity models' evaluation. Like other regression models, the gravity equation can include lagged variables to account for dynamic process. Bun and Klaassen (2002) proposed to consider the effect of trade and income lags on the dependent variable of trade flow. As a result, the authors showed the advantage of dynamic gravity models over static ones. The most important achievement of this article was the conclusion that the LSDV method is relevant for both dynamic and static gravity models, which may be important in comparing the estimates. Dynamic gravity models are more common for assessing export-import trade relations (Tham et al., 2018; Bekele, Mersha, 2019), but are also used in the analysis of migration (Bunea, 2012; Davidescu et al., 2017). Davidescu et al. (2017) found a relationship between the magnitude of migration flows from Romania to other EU countries in the current year with the migration flows of the past period. The authors explain this phenomenon by the fact that a potential immigrant has more relatives in a new country, which favors his relocation, since it simplifies the process of adaptation in a new country.

Another important issue is the estimation approach of the gravity models. Along with the standard logarithmic transformation of the gravity equation, and hence using the estimation methods traditional for panel data (pool regression, a Fixed effects model and a Random effects model), alternative estimation methods such as the Pseudo Poisson maximum likelihood method are used (Wajdi et al., 2017; Gupta et al., 2019; Vieira, Reis, 2019). Pseudo Poisson maximum likelihood estimation accounts for zero flows between regions, and secondly, unlike the FE model, it allows to estimate time-invariant variables. Hence, on the current study we rely on the PPML estimation approach.

Even though the volume of literature is considerable (Table 1 provides summary details of the studies reported in this section), there is a lack of papers in which the method of gravity model is used for investigation commuting in Germany. Among the works where a gravity model is still used to describe commuting (Stefanouli, Polyzos, 2017; Wajdi et al., 2017; Varga et al., 2018; Gupta et al., 2019; Vieira, Reis, 2019), the authors of recent studies most often use the PPML method. When working with migration data, it allows to take into account a complete sample of all possible flows between regions, and not only those that are positive, which is a significant advantage, since the accuracy of our estimates will be higher on a more complete sample. The choice of the method of estimating the gravity model will be in detail described in section 5. Finally, the application of the method to this analysis may deepen the knowledge about commuting, its causes and consequences.

Table 1. Studies analysing the commuting flows or PPML estimation of gravity models

Study	Data	Methods of analysis
Tkocz, Kristensen (1994)	1991 Statistical Yearbook of Denmark; Telephone interviews	Logit regression Linear (OLS) regression
Artis et al. (2000)	Sample of the 1991 Catalonia's Census (individual data)	Multinomial logit regression
Shen (2000)	1990 U.S. Census; Census Transportation Planning Package (CTPP)	Linear (OLS) regression
Lee, McDonald (2003)	Korean Population Census of 1995	Linear (OLS) regression
Stutzer, Frey (2007)	German Socio-Economic Panel 1985–1998	Pooled least squares regression
Patuelli et al. (2007) Reggiani et al. (2011)	Federal Employment Agency, Federal Office for Building and Regional Planning 1995 and 2004	Network analysis Spatial interaction model (SIM)
Haas, Hamann (2008)	Federal Statistical Office of Germany; Federal Agency for Labour 1995–2005	Quantitative analysis
Baum-Snow (2010)	USA Census 1960–2000	Linear (OLS) regression
Patuelli et al. (2010)	Federal Employment Agency, Federal Office for Building and Regional Planning 1995 and 2005	Network analysis Scale-free network Random graph model
Hansson et al. (2011)	Cross-sectional public health surveys of Scania, Sweden (2004, 2008)	Binary logistic regression
Burger et al. (2011)	Journey-to-work data between local authority districts from 1981 and 2001 in England and Wales. Originated from the Special Workplace Statistics (Set C) in the British Census.	A set of indices is calculated for each district to estimate its functional and morphological primacy, network density, and outward openness.
Niebuhr et al. (2012)	Employment history statistics from Institute for Employment Research, German Federal Employment Agency, BBSR, Federal Institute for Research on Building, Urban Affairs and Spatial Development 1995–2005	GMM Arellano–Bond dynamic panel models
Östh, Lindgren (2012)	The PLACE database and Statistics Sweden's (SCB) data 1990–2006	Random-effects GLS models

End of Table 1

Study	Data	Methods of analysis
Huinink, Feldhaus (2012)	Panel Analysis of Intimate Relationships and Family Dynamics 2008–2009	Multivariate probit regression Longitudinal difference model Heckman-selection probit regression
Zhao (2013)	Household interview survey conducted by a foreign housing research team in Beijing in 2001, Beijing Statistical Yearbook 2002, Beijing Transport Committee 2005	Multinomial logit regression Binary logistic regression
Viry et al. (2014)	"Job Mobilities and Family Lives in Europe" survey 2010	Cluster analysis Linear (OLS) regression Logistic regression
Künn-Nelen (2016)	British Household Panel Survey	Fixed-effects (FE) linear regression
Stefanouli, Polyzos (2017)	Census conducted by the National Statistical Service 2011	Gravity model Extended gravity model Radiation model
Wajdi et al. (2017)	2000 and 2010 Population Censuses and the 2005 Intercensal Population Survey	Gravity model, PPML
Varga et al. (2018)	CTPP 2006–2010 Census Tract Flows, 2018 USA job openings accessed, American Community Survey, Population distribution of Hungary and Italy 2011	Gravity model Radiation model Flow and jump model (FJM)
Dauth, Haller (2018)	30 percent sample of IAB Beschaftigten Historik (BeH) 2000–2014	Descriptive analytics
Gupta et al. (2019)	Trade flows of 164 developing and developed countries, 1985–2013	Gravity model FE, RE, PPML
Vieira, Reis (2019)	World Bank Indicators on Brazilian exports, 2000–2015	Gravity model, PPML
Skora et al. (2020)	German Socio-Economic Panel (SOEP) 2001–2017	Fixed-effects (FE) linear regression
Ahrens, Lyons (2021)	The Place of Work, School or College (POWSCAR) 2011–2016	Linear first-difference gravity model

3. Description of the variables

The purpose of this work is to determine the factors influencing the magnitude of the commuting flows, therefore, after reading the literature and analysing the factors that the authors usually use to study migration processes, 16 variables that reflect the characteristics of the regions were selected. They can be divided into the following groups (Karemera et al., 2000; Lauridsen, 2006; Vakulenko et al., 2011; Bunea, 2012; Zhao, 2015; Ramos, Surinach, 2017; Wajdi et al., 2017).

1) Demographic indicators (population, population density, birth rate, death rate and share of foreign population).
2) Labour market indicators (unemployment rate, average salary, number of companies in the region per 10000 people).

3) Real estate market indicators (the price of 1 m^2 of apartment housing when buying or renting it).
4) Economic development and welfare indicators (GDP per capita and the number of personal vehicles per 1000 people).
5) Other indicators of the regions' attractiveness (tourism intensity in the region, the share of people, who depends on minimum social benefits, the share of school graduates without a diploma and the price of 1 m^3 of drinking water).
6) A range of dummy variables to account for spatial and time effects.
7) Distance between regions.

Since the flow from region *i* (residential place) to region *j* (workplace) is considered as one observation, we simultaneously use these characteristics for both the residential place (in the model, such variables are denoted as "*home*") and for the workplace (they are counted as "*work*").

For each factor we expect a certain direction of influence, which is indicated in Table 2 — a positive or negative influence on the intensity of commuting flows. For most variables, such as population density, GDP per capita, average wages, number of companies per 10000 population, home purchase and rental prices, and tourism intensity, there is a negative impact in the residential region and a positive impact on intensity flows in the region of work, because all of them reflect the level of economic activity and development of the region. This means that the higher these indicators in the region where workers live, the higher the living standard and employment conditions are there, and therefore the motivation for long-distance movements of workers decreases. The opposite situation is observed with the region of work — high regional welfare indicators attract more commuting migrants.

There are several variables for which the direction of influence is not determined — this means that there are two points of view, and it is not certain which one prevails as a result.

For example, on the one hand, we assume that with a growth in the birth rate, female commuting decreases. On the other hand, based on the study by Viry et al. (2014), we know that commuting is a fairly common alternative for internal or international migration for young families. In case when wages are higher in neighbouring regions, both women and men prefer to work outside their region, while maintaining their residential place, and thus the social networks there.

Similarly, the share of school graduates who did not receive a high school diploma are those who either finished school prematurely and did not receive a diploma, which allows them to enter higher education institutions, or who were able to complete their studies, but did not pass the curriculum and are not eligible to receive a diploma. Thus, the share of school graduates who did not receive a high school diploma serves not only as the indicator of the education system, but is also as proxy for the regional crime level. It is difficult to predict what impact this indicator has in region *i*, since, on the one hand, a less favourable region pushes away the employees willing to find a job in another region. On the other hand, the growth of the share of school graduates without diploma indicates an increase in the number of young people who do not have the right to receive higher education, and, consequently, to have a prestigious well-paid job. Unskilled specialists will be less valuable in the labour market, and it will be more difficult for them to find a job and attract the attention of an employer from another region. At the same time, it is assumed that the growth of this indicator in region *j* will reduce migration flows.

A description of all paired variables and their descriptive statistics can be found in Appendix 1.

Table 2. Assumed influences of factors on commuting flows

Variable	Effect on the dependent variable *flow*	Variable	Effect on the dependent variable *flow*
distance	negative	*common_border_flg*	positive
same_land_flg	positive		
population_home	positive	*population_work*	positive
gdp_ppc_home	negative	*gdp_ppc_work*	positive
density_pop_home	negative	*density_pop_work*	positive
foreign_share_home	positive	*foreign_share_work*	not defined
livebirth_home	not defined	*livebirth_work*	not defined
death_home	positive	*death_work*	negative
earnings_home	negative	*earnings_work*	positive
unemployment_home	positive	*unemployment_work*	negative
apart_prices_home	negative	*apart_prices_work*	positive
lease_prices_home	negative	*lease_prices_work*	positive
auto_home	not defined	*auto_work*	negative
not_uni_graduate_home	not defined	*not_uni_graduate_work*	negative
water_price_home	negative	*water_price_work*	not defined
benefit_share_home	positive	*benefit_share_work*	negative
company_home	negative	*company_work*	positive
tourism_intensity_home	negative	*tourism_intensity_work*	positive

Along with the justification of the effects of the other determinants, we test the following hypotheses.

Hypothesis 1. The higher the level of income in the residential region, the lower the level of commuting, while the high level of income in the region of work contributes to more intensive movements of commuters.

Hypothesis 2. Regions that belong to the same land have more active commuting connection than regions located in different lands.

Hypothesis 3. Regions bordering each other have more intensive commuting than regions that do not have a common border.

Hypothesis 4. The cost of renting housing in outflow region has a negative effect on the commuting, whereas cost of housing in the inflow region has a positive effect.

Hypothesis 5. The number of companies per 1000 people in origin region weakens the commuting, while the growth of this indicator in the destination region enhances the magnitude of flow.

4. Data

The dependent variable in the regression models is the absolute value of commuting flows from one region to another. The main source of these data is the German Federal Labour Agency[2].

[2] https://statistik.arbeitsagentur.de/SiteGlobals/Forms/Suche/Einzelheftsuche_Formular.html; jsessionid=BD5BBE5E967E18AC7975F792510AB357?nn=24390&topic_f=beschaeftigung-sozbe-krpend.

We use commuting flow data from 2013 to 2019. The flow is the number of migrants moving from the residential place (point *i*) to the workplace (point *j*) and crossing the border of the residential place (point *i*). The important fact is that moving from *i* to *j* and from *j* to *i* are different flows, and therefore different observations.

According to the Nomenclature of Territorial Units for Statistics[3] used in the European Union, we use 3 level divisions of Germany, which corresponds to 400 administrative-territorial units, which include cities (or urban areas) and districts. In total, the data panel for 7 years consists of 1117200 flows (159600 flows for each year).

Table 3 presents the general characteristics of the migration network based on the data about the dependent variable.

Table 3. Commuting network descriptive statistics in 2013–2019

Year/ Characteristic	Number of flows greater than 0	Migration network density	Number of employees	Number of commuters	Share of commuters from the number of workers	Average flow (including zero flows)	Average flow (excluding zero flows)
2013	38303	0.24	29757461	10471415	0.352	66	273
2014	38870	0.244	30269569	10714446	0.354	67	276
2015	39794	0.249	31020998	11029108	0.356	69	277
2016	41690	0.261	31815414	11385711	0.358	71	273
2017	42641	0.267	32576261	11701601	0.359	73	274
2018	43542	0.273	33251893	12003712	0.361	75	276
2019	43704	0.274	33705175	12193622	0.362	76	279

During the period, we can observe an increase in the number of commuters, as well as an increase in network density. From 2013 to 2019 commuting flows in Germany represented approximately 25% of the maximum possible number of connections.

In 2013–2015 there was an increase in the average flow (excluding zero flows), while in 2016 it is back to the 2013 value of 273 people. The reason is a sharp increase in the number of flows in 2016, with a slight increase in the number of commuters — it means that the network has expanded significantly, and many new routes have appeared, with a small number of people moving through them so far.

We see that, on average, about 11–12 million people per year work outside their residential region. The total number of employees in 2013–2019 fluctuated from 29.7 to 33.7 million people, therefore, commuting migrants make up a third of all workers. Moreover, there is a small but steady increase in this indicator, which once again proves the fact that the phenomenon of commuting in Germany is becoming more common.

The main source of data for the independent variables is the Federal and State Statistical Office[4] STATISTIKPORTAL, which holds most of the statistical tables on commuting factors. An additional

[3] https://ec.europa.eu/eurostat/web/nuts/background.

[4] https://www.regionalstatistik.de/genesis/online?operation=previous&levelindex=0&step=0&titel=&levelid=1692881920372&acceptscookies=false.

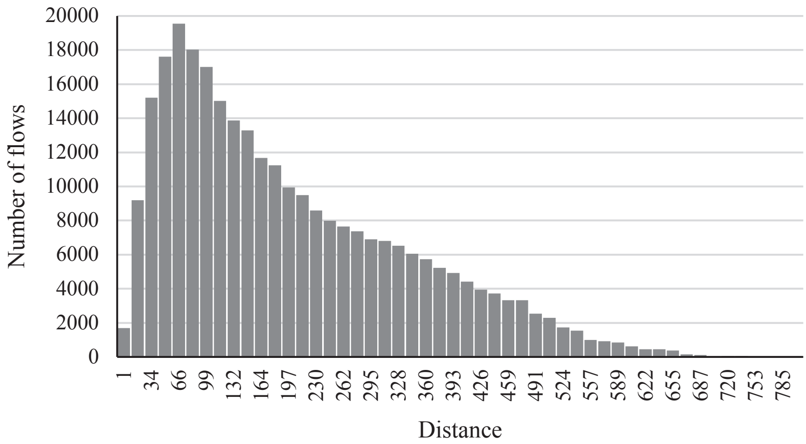

Fig 1. Distribution of distance between regions

source of information is the database of the German Federal Statistical Office DESTATIS[5]. To fill in the data gaps we used the German lands online databases, such as the Thuringian State Office for Statistics[6], and the portal Statista[7] was also used.

In the previous section we talked about paired quantitative variables, now it is worth mentioning the distance variable between regions (*distance*), as well as the dummy variables included in the model.

The distance variable is one of the key variables used in the gravity model. The calculation of it is based on the geographical coordinates of the regions' centres, such as the Euclidean distance between them. It is expected that as the distance increases, the flow between regions will decrease. On average, migrants cover 206 km, but in Fig. 1 we also see that mostly the length of routes is from 40 to 80 km. The farthest flow recorded in 2013–2019 was from Vorpommern-Rügen in the northeast of the country towards Lörrach in southwest Germany and was over 817 km. However, it should be kept in mind that it is impossible to cover such a distance every day, so this is either an example of shift or seasonal work, or an error in recording data on residential places or workplaces.

Our model also includes the following dummy variables.

- A dummy variable for whether both regions (both residential and workplace) belong to the same German land (Bundesland) — there are 16 states in total in Germany. The variable takes a value of 1 if both regions belong to the same land, and 0 otherwise. It is expected that the sign of the coefficient of this variable will be positive since the distances between regions within the same land are less than between regions from different lands, therefore commuting migrants spend less time travelling.
- Dummy variables on whether the land in which the residential region or region of work is located belongs to West or East Germany — the variable takes the value 1 if the region is in the western part of the country and 0 if in the eastern part. West Germany includes

5 Statistisches Bundesamt (2013–2019), https://www.destatis.de/DE/Home/_inhalt.html.

6 Thüringer Landesamt für Statistik (2013–2019), https://www.statistik.thueringen.de.

7 Statista (2013–2019), https://www.statista.com/.

10 of the 16 states (Schleswig-Holstein, Hamburg, Niedersachsen, Bremen, Nordrhein-Westfalen, Hessen, Rheinland-Pfalz, Baden-Württemberg, Bayern und das Saarland), in East Germany there are 6 remaining states (Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen).
- A dummy variable for the existence of a common border between the regions of residence and workplace, which takes the value 1 if the regions border each other, and 0 otherwise. It is assumed that the presence of a common border between regions will have a strong positive effect on the intensity of flows, similarly to the dummy variable on whether a region belongs to the same land.

Moreover, when evaluating the model, we control it for possible individual land effects, so we consider the fact to which land the region belongs. Similarly, accounting for time dummy variables allows us to consider possible shocks that could be observed in the economy in each of the 7 years.

5. Econometric model

The so-called gravity model used in economic and social research is an analogue of Newton's gravity law. The main idea of this model is that the intensity of the connection between two objects is directly proportional to the "importance" of these objects and inversely proportional to the distance between them. The equation of such a gravity model is as follows (1):

$$F_{ij} = G\, X_i^{\alpha} X_j^{\beta} \big/ D_{ij}^{\chi}, \tag{1}$$

where F_{ij} — the linkage intensity between objects i and j; G — constant; X_i and X_j — the magnitude or importance of the objects i and j; D_{ij} — distance between objects i and j; α, β, and χ — parameters. Traditionally, in Newton's model $\alpha = \beta = 1$, and $\chi = 2$. In the context of the study of migration in the basic gravity model X_i and X_j denoted the population of objects i and j — over time, instead of a single factor, a set of objects' characteristics began to be used as the numerator.

In our case, we use panel data working with a gravity model. Most authors use the following types of regressions to estimate such models: Pool Regression, Fixed effects model and the Random effects model. For all three types of regressions, it is required to use a logarithmic transformation for both parts of the original gravity equation (Karemera et al., 2000; Bunea, 2012). However, this type of transformation has several disadvantages. Firstly, there is no way to consider potential zero flows between regions, because it is impossible to take the logarithm of the dependent variable equal to zero, due to which a large amount of information is lost (sometimes it is more than half of the sample). Secondly, each of the three types of regressions has several drawbacks. Pool regression does not take into account data heterogeneity, which is often inherent in migration data, leading to heteroscedasticity and bias in the estimates. The FE model does not provide estimates for time-invariant variables. For our data, this is a significant disadvantage, since one of the main assumptions of the gravity model is the inverse dependence of the flow intensity on the distance between regions, and the distance variable is constant in time. Moreover, our model assumes the use of additional dummy variables, which also do not change over time. The RE model may produce inconsistent regression estimates if the basic assumption, that unobserved factors are uncorrelated with random error, does not hold.

To solve all these problems and find model estimates, we will apply the multiplicative transformation of the original equation and the Poisson Pseudo maximum likelihood method, which is often used for count data. The method requires the assumption that the conditional mean is proportional to the conditional variance. However, the data do not have to have a Poisson distribution for the estimator to be consistent. Silva and Tenreyro (2006) proved that the PPML method gives robust estimates under heteroscedasticity of observations, and it is a reliable method for estimating gravity models. Later, Yotov et al. (2016) showed that the fixed effects of such a gravity equation (2) are identical in interpretation to other estimation methods:

$$flow_{ijt} = \exp\left(\alpha + \beta \ln x'_{i(t-s)} + \gamma \ln x'_{j(t-s)} + \delta \ln(distance_{ij}) + \mu_i + \tau_j + \eta_t + \varepsilon_{ijt}\right), \ \ \varepsilon_{ijt} \sim N(\mu, \sigma^2), \quad (2)$$

where $flow_{ijt}$ — the flow of commuting migrants from region i to region j in the period t; $x'_{i(t-s)}$ — vector of characteristics of region i at the time period $t-s$ ($s = 0,1,2;$ for each s separate model is estimated); $x'_{j(t-s)}$ — vector of characteristics of region j at the time period $t-s$ ($s = 0,1,2;$ for each s separate model is estimated); $distance_{ij}$ — Euclidean distance between regions i and j; μ_i — vector of dummy variables for region i; τ_j — vector of dummy variables for region j; η_t — vector of time dummy variables; ε_{ijt} — random term. The parameters α, β, γ, δ show the effect of the factors on the dependent variable.

The interpretation of the coefficients' estimates of quantitative variables in the PPML method represents elasticity, as in other models. In order to interpret the value of dummy variables, we will use the recommendation of Yotov et al. (2016) and calculate the marginal effect as follows (3):

$$\left[\exp\left(\hat{\beta}_{dummy}\right) - 1\right] \times 100, \quad (3)$$

where $\hat{\beta}_{dummy}$ — the estimation of coefficient for the dummy variable.

It is also worth noting that earlier, in our literature overview, we have repeatedly emphasized that any type of migration needs to consider the dynamic structure. Many authors investigate the dependence of commuting not only on socio-economic factors, but also on their lags of the past period. Östh and Lindgren (2012) showed that the GDP of the previous period does not affect the commuting distance, while GDP with a lag of 2 years significantly reduces it. Due to the large number of factors selected for our model and the complexity of their interpretation, we will not take into account the current and lagged variables simultaneously — we will control the dependence of the flow on the variables of the past period by building three different models, in which the determinants without a lag, with a lag of 1 year, and with a lag of 2 years will be taken into account. In this case, we are guided by the remark: "In part, because the information is costly to acquire and requires time to decipher, migration is likely to respond with a lag to changed circumstances" (Greenwood, 1985). Thus, we believe that the factors affecting the intensity of commuting flows will show a stronger influence in subsequent periods, rather than immediately.

6. Estimation results

The results of the estimating model using the PPML method are presented in Table 4, where all 1117200 observations were included in the model:

Table 4. Parameter estimate and significance (dependent variable — *flow*)

Variables	PPML (Model 1)	PPML (Model 2)	PPML (Model 3)	PPML (Model 4)
population_home	0.861***	0.830***	0.826***	0.830***
	(0.0509)	(0.0584)	(0.0536)	(0.0565)
population_work	0.778***	0.830***	0.816***	0.816***
	(0.0484)	(0.0490)	(0.0449)	(0.0410)
gdp_ppc_home		–0.130	–0.0497	–0.0506
		(0.109)	(0.132)	(0.122)
gdp_ppc_work		0.928***	0.886***	0.959***
		(0.0725)	(0.0774)	(0.0782)
density_pop_home			–0.156***	–0.180***
			(0.0450)	(0.0595)
density_pop_work			0.0263	0.000269
			(0.0400)	(0.0444)
earnings_home		–0.0123	0.0796	0.0713
		(0.143)	(0.144)	(0.159)
earnings_work		–0.447	–0.449	–0.446
		(0.300)	(0.309)	(0.303)
unemployment_home	–0.262***	–0.152*	0.198**	–0.0340
	(0.0964)	(0.0779)	(0.0792)	(0.212)
unemployment_work	0.477***	0.243***	0.145*	0.273
	(0.0806)	(0.0867)	(0.0854)	(0.242)
apart_prices_home				0.000331
				(0.0988)
apart_prices_work				–0.226**
				(0.105)
lease_prices_home	–0.712***	–0.469***	–0.140	–0.221
	(0.154)	(0.138)	(0.152)	(0.187)
lease_prices_work	1.876***	0.610***	0.314	0.626***
	(0.130)	(0.157)	(0.200)	(0.238)
auto_home				–0.953***
				(0.270)
auto_work				–0.519**
				(0.253)
not_uni_graduate_home				0.109**
				(0.0535)
not_uni_graduate_work				0.112**
				(0.0473)
company_home	–0.339**	–0.252**	–0.240*	–0.135
	(0.137)	(0.120)	(0.130)	(0.125)
company_work	–0.306**	–0.166	–0.246**	–0.225**
	(0.129)	(0.105)	(0.109)	(0.109)
tourism_intensity_home				0.0514*

End of Table 4

Variables	PPML (Model 1)	PPML (Model 2)	PPML (Model 3)	PPML (Model 4)
				(0.0271)
tourism_intensity_work				0.0119
				(0.0299)
distance	–1.296***	–1.258***	–1.269***	–1.282***
	(0.0832)	(0.0808)	(0.0812)	(0.0749)
common_border_flg	1.592***	1.639***	1.635***	1.638***
	(0.109)	(0.112)	(0.111)	(0.102)
same_land_flg	1.106***	1.129***	1.118***	1.106***
	(0.0762)	(0.0739)	(0.0743)	(0.0688)
home_land	–0.481***	–0.462***	–0.195*	–0.103
	(0.109)	(0.112)	(0.118)	(0.133)
work_land	0.178	0.198	0.103	0.164
	(0.127)	(0.167)	(0.156)	(0.166)
Constant	–9.451***	–12.25***	–11.31***	–2.309
	(1.093)	(3.374)	(2.984)	(4.014)
Observations	1117200			
R-squared	0.564	0.575	0.565	0.577
Year FE	Yes	Yes	Yes	Yes
Home-land FE	Yes	Yes	Yes	Yes
Work-land FE	Yes	Yes	Yes	Yes
Control variables[8]	No	No	No	Yes

Note. The table represents the results of the estimation for 4 different sets of determinants. Parameter estimates obtained using PPML method (Silva, Tenreyro, 2006). Robust standard errors are presented in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

Firstly, it should be noted that if we applied standard panel data estimation methods, we would use White's standard errors to deal with heteroscedasticity, but in the PPML method they are already robust, according to Silva and Tenreyro (2006). Since we are working with a data panel in which there are many observations and a small number of time steps, the autocorrelation problem is also solved by using robust standard errors.

As for multicollinearity, firstly, it is inevitable with such many factors, and secondly, we use the same characteristics of both regions at the same time, which are highly correlated with each other. In addition, having conducted the correlation analysis, we can conclude that there is a high negative relationship between the density of personal vehicles and the unemployment rate, the share of people receiving minimal social benefits and population density. Also, the mortality rate per 10000 people and the indicators of the real estate market are correlated with many factors.

Furthermore, we proceed directly to the interpretation of the obtained estimates of the coefficients in the PPML model. We see that not for all variables the effect on the intensity of flows is significant. Regarding the factors that are most used by the authors, we have obtained results about their significance similar to the literature (Bunea, 2012; Wajdi et al., 2017). The direction

[8] Full list of estimates is presented in Appendix 2.

of the influence of population size, GDP per capita, and distance between regions on the flow also corresponds to the results obtained in similar papers (Morgenroth, 2002; Wajdi et al., 2017). We confirm our assumption that population size in both regions has a positive effect on commuting flows. Among others demographic characteristics population density in the residential region negatively affects commuting, which is explained by the tighter labour force distribution and therefore less necessity of commuting. The negative (significant at 10% level) effect of the share of the foreign population is explained by the lower intension of mobility among foreigners since they usually arrive to the region where the employer is located. The positive effect of the birth rate serves as an alternative proxy for population size.

Among the labour market indicators only the factor that reflects entrepreneurial activity in the workplace is significant: the intensity of commuting flows decreases by 2.25% with an increase in the number of companies by 10%. The factor such as wages and unemployment rates appear to be insignificant. Appendix 2 contains various specifications of the model, due to which the *earnings* variable is indifferent to commuting flows, regardless of the combination of factors in the model. The opposite situation can be observed for the *unemployment* variable. We see that in general, with almost any modification of the model and the inclusion of the most significant factors from the full model, the unemployment rate in the residential region is significant at 5 and 10% significance levels, depending on the model. In our case, we have a negative dependence — with an increase in the unemployment rate in the region of residence, commuting flows are reduced. A similar result was obtained by Morgenroth (2002) in his study of commuting patterns in Ireland. The author explains the negative effect of the unemployment rate in region i by the relative immobility of the unemployed — an increase in their number reduces the overall worker flow. Moreover, an important reason may be the lack of convenient and inexpensive public transport to commute to another region. Unemployment in the region of work is quite strongly correlated with the density of personal vehicles, therefore, when this variable is added to the model, the significance of *unemployment_work* is lost, although in specifications (1)–(3) the variable is significant at 1 and 5% levels of significance. The effect of the unemployment rate in the region of work is opposite to the region of residence — this effect can be explained by the fact that the increasing outflow of commuting migrants from the regions of work will exceed the decreasing inflow of migrants to these regions, especially if we talk about large metropolitan areas.

The real estate market is another important set of factors. The cost of both buying and renting apartment housing in the residential place does not affect the commuting flows. At the same time, an increase in the price of real estate when buying it in the region of work reduces the migration flows: the price of apartments is related to the overall economic development, so the increase in the purchase price of real estate makes the region more attractive to citizens who can afford to buy it. Consequently, some migrants buy real estate and move to live in those regions where they previously only worked — the region of residence and work now coincide, and commuting flows are reduced. Regarding the cost of housing rents, the hypothesis is confirmed, which is proved by the fact that high prices for rental properties are mostly observed in large cities, which attract intensive migration flows.

Among indicators of economic well-being, three of the four indicators are significant. The growth of economic activity (GDP per capita) in the region of work will attract more migrants. A negative impact on the flows can be observed in the number of personal cars in both regions, which confirms one of our hypotheses — for region j and proves the fact that commuting migrants are more inclined to travel by public transport and prefer to spend less time in traffic jams. It can also be added that the most likely negative effect of this factor in region *i* includes the impact on those migrants who use their cars directly to get to work. With an increase in the number of cars, the demand for parking places in the workplace increases, which contributes to an increase in prices for them. As a result, the costs of such trips become too high for labour migrants — in addition to the time spent on the road, they spend time and money on finding a parking space for their car — as their availability decreases and prices simultaneously increase, commuting to such regions reduces.

Such variables as the price of 1 m^3 of drinking water and the share of people receiving minimal social benefits turned out to be insignificant in our model. Proxy for the education system — the share of graduates without a diploma in both regions has a positive effect on commuting, which allows us to accept one of our hypotheses for region *i* — the growth in the number of young people without higher education in the region of residence contributes to the desire of qualified specialists for looking for a job in more favourable regions. At the same time, the growing share of such graduates in the workplace reduces competition for prestigious and highly paid jobs, which simplifies the search and employment process for commuter migrants with higher education. The tourist attractiveness of the region of residence is significant only at the 10% significance level and has a weak positive impact on commuting.

Our estimations confirm that the intensity of flows decreases with the increasing distance of the workplace. The distance has a strong effect on the magnitude of flows — with an increase in distance by 10%, the number of passengers decreases by 12.8%. Even higher effect has the dummy variable indicating the presence of a common border between regions — commuting flows between neighbouring regions are 414.5% higher than between regions without a common border; and the dummy variable on whether regions belong to the same land — in Germany, commuting flows between regions in the same land are 202% higher than between regions from different lands.

We also control for whether regions belong to the western or eastern part of the country. Both variables for residential place and workplace turned out to be insignificant, which suggests that there is no significant difference between the migration flows in different parts of the country.

In describing the gravity model, it was argued that previously, when applying this type of model to the estimation of migration processes, the authors used only the population size as a determinant. In our study we try to define what socio-economic characteristics determine the process of commuting in Germany. We tested several model specifications, which included different combinations of variables. To specify key socio-economic variables, we were looking at significance of both variables (both for the region of residence and work) as a priority. As a result, the final combination consisted of distance between regions and the following variables: population, unemployment rate, rent price of 1 m^2 of real estate, number of companies per 10000 people, and dummy variables on the presence of a common border between regions and on the fact of being in the same land (see Table 4, Column 1). The distance between regions is also added to it.

7. Analysis of results for models with lagged variables

Along with the analysis described above, we also build an alternative model, which is based on the values from the previous periods as independent variables. In this way, we consider the dynamic nature of migration, which we discussed in the studies' review on different types of commuting migration. Based on these additionally collected data for the previous years 2011 and 2012, we build two new models: the first takes into account all the same factors as in the standard model, but with a lag of 1 year, and the second — factors with a lag of 2 years. Thus, it is possible to track whether among our variables some indicators become significant after one or two years, and which, perhaps, lose their significance. Knowledge about the past period can have a rather strong impact on today's movements of commuters. Commuting, for example, is tightly linked to regional housing markets. In most cases, people thinking about moving to another regions take into account their previous experience, consider the changes in the region of residence and make a decision regarding the future residential and workplace. Hence, studying the dependence of commuting flows on the factors from the previous period can be more important and informative than studying it in the current period.

Table 5 compares the estimates of all three PPML models. We can conclude that the importance of some demographic factors increases over time in full model. Commuting flows significantly depend on the birth rate of the past period in the workplace. Birth rates with a lag of 2 years have a 52% stronger effect on commuting flows than birth rates in the current year. High fertility occurs in large metropolitan areas with a high population density — these are the regions where commuters from small towns and districts mainly go to work, so increase of the birth rate in region *j* leads to a higher intensity of commuting flows. On the contrary, an increase in mortality in the residential place with a lag of 2 years reduces commuting flows, which rejects the hypothesis about the rising intensity of flows due to the desire of the population to work in a more prosperous region.

Table 5. Estimates of the gravity model without a lag, with lags of 1 and 2 years were obtained using the PPML method (dependent variable — *flow*)

Variables	PPML (lag 0)	PPML (lag 1)	PPML (lag 2)
population_home	0.830***	0.831***	0.833***
	(0.0565)	(0.0564)	(0.0559)
population_work	0.816***	0.815***	0.815***
	(0.0410)	(0.0416)	(0.0408)
livebirth_home	–0.0671	–0.129	–0.200
	(0.256)	(0.262)	(0.273)
livebirth_work	0.376*	0.463*	0.570**
	(0.222)	(0.245)	(0.268)
death_home	–0.345	–0.386	–0.397*
	(0.248)	(0.245)	(0.241)
death_work	–0.152	–0.247	–0.301
	(0.229)	(0.231)	(0.232)
company_home	–0.135	–0.102	–0.0979
	(0.125)	(0.131)	(0.131)
company_work	–0.225**	–0.179*	–0.125
	(0.109)	(0.108)	(0.110)
lease_prices_home	–0.221	–0.241	–0.223
	(0.187)	(0.188)	(0.193)
lease_prices_work	0.626***	0.679***	0.767***
	(0.238)	(0.243)	(0.236)

End of Table 5

Variables	PPML (lag 0)	PPML (lag 1)	PPML (lag 2)
distance	–1.282***	–1.287***	–1.292***
	(0.0749)	(0.0736)	(0.0723)
common_border_flg	1.638***	1.633***	1.625***
	(0.102)	(0.101)	(0.0994)
same_land_flg	1.106***	1.104***	1.100***
	(0.0688)	(0.0676)	(0.0667)
home_land	–0.103	–0.673***	–0.125
	(0.133)	(0.184)	(0.131)
work_land	–0.0223	–0.0279	0.161
	(0.159)	(0.158)	(0.159)
Constant	–2.123	–0.400	–0.0928
	(4.019)	(4.111)	(4.096)
Observations		1117200	
R-squared	0.577	0.579	0.584
Year FE	Yes	Yes	Yes
Home-land FE	Yes	Yes	Yes
Work-land FE	Yes	Yes	Yes
Control variables[9]	Yes	Yes	Yes

Note. The table represents the estimates of the models with the determinants in previous periods. Here we present the main variables from the Model 1 (Table 4) and those variables, for which the dynamics of significance depending on the lag is observed. Robust standard errors in parentheses; *** — $p<0.01$, ** — $p<0.05$, * — $p<0.1$.

Interesting, that entrepreneurial activity effect on the commuting flows remains only in the short-run. The number of companies in region j has a negative impact on commuting flows in models with current value and models with a lagged value, but the significance disappears already in the model with a second lag. We can assume that information relevant for planning the place of residence or work is, first of all, based on the current situation about the number of potential employment places, the number of actively hiring companies and entrepreneurial activity. The commuters will mostly rely on real-time information about possible employers rather than look at last year's statistics when deciding whether to move or get a job. If the number of companies in a particular region is actively growing now and there are reasons to assume an increase in entrepreneurial activity, then a potential employee will rather decide to move to this region, thereby reducing commuting. Estimates of the coefficients of the remaining variables do not change much when considering factors with and without lags.

Based on the results of the analysis, we can conclude that there are independent variables among the factors that can influence the dependent variable, whose significance increases with the addition of 1 and 2 years lags, and there are those that lose it. However, we still choose the lag-free model presented in the previous chapter as the final model, because we see that the significance of one of the factors included in the final model decreases over time (*company_work*), which indicates that models with lags will be worse than models without lags of 1 or 2 years.

[9] Full list of estimates is presented in Appendix 3.

8. Conclusions

In this paper, we examined the factors of commuting in Germany using gravity models. Reviewing the studies on the gravity modelling of trade, tourism and other migration flows we identify a range of factors that are often used in this type of models. We additionally identify the methods used by the authors, especially when working with a panel data structure, and also showed the importance of considering the dynamic structure in migration and trade types of processes. In selecting factors, emphasis was placed on the maximum coverage of various sectors of the German economy and social sphere. Therefore, the final list of variables included demographic indicators, labour market and real estate market indicators, welfare variables, indicators of the social and tourist attractiveness of territories, as well as spatial and temporal effects. One of the novel variables of this research is a land indicator (Bundesland). This variable occupies cultural, ethnic, political characteristics which show the similarity of regions in one land.

Estimation of the gravity model using the PPML method allowed us to confirm the expected effect of most determinants. Empirical analysis has shown that such factors as distance, population and the cost of renting housing are significant both for the residential and workplaces, and their influence is in line with our expectations. An increase in the population in the regions of residence and work by 10% leads to growth in the flow by 8.6 and 7.8%, respectively. An increase in the distance between cities by 10% leads to a decrease in the intensity of the migration flow by 13% (see Table 4, Column 1). It is noteworthy that the result of assessing the influence of population and distance corresponds to the assumptions of the gravity model with parameters very close to the original ones: $\alpha = \beta = 1$, $\chi = 2$. According to the results of our estimation, $\alpha = 0.861$, $\beta = 0.778$, $\chi = 1.296$. At the same time, the hypothesis about the equality of α and β is not rejected at a 5% significance level.

The results regarding GDP per capita, the unemployment rate and the dummy variable for the having a common border consist with the results of Wajdi et al. (2017), who studied internal migration in Indonesia. The indicator of GDP per capita turned out to be negatively insignificant for the residential region but positively significant for the workplace. The Hypothesis 5 regarding the effect of the number of companies was partially confirmed: the intensity of commuting flows decreases by 2.25% with an increase in the number of companies by 10%. The unemployment rate, as well as the number of companies, has a significant impact in both regions, but the effect of these factors partially or completely does not coincide with our expectations, as well as Morgenroth (2002) who investigated commuting in Ireland. Such factor as wages turned out to be insignificant for both regions, which correlates with the results of Niebuhr's et al. (2012) research, that proved that the main driver of commuting process is not the difference in the level of wages in the regions of Germany, but the difference in the level of unemployment.

The final set of variables that can be used in the gravity model as characteristics of the "importance" for both regions included the following variables: population, unemployment rate, cost of renting housing and the number of companies per 10000 people. We also conclude that the cost of renting housing in the destination region has a significant positive effect on the commuting. In addition, the results confirm the location effect: the distance between regions, the presence of a common border between them, as well as the fact that the regions belong to the same land of Germany, have a strong influence on commuting flows. Commuting flows between neighbouring regions are 414.5% higher than between regions without a common border. In addition, commuting flows between regions in the same land are 202% higher than between regions from different lands.

The estimates obtained prove the applicability of gravity models with the population as a standard indicator of the regions' importance for studying such types of migration as commuting. In addition, it demonstrated the possibility of using indicators of the labour market, the real estate market or economic activity as alternative indicators of importance.

The limitations of this papers are primarily related to the availability of data for the studied period of time and for such a large number of German regions. In this regard, potential further research may be papers in which the authors will study the issue of measuring the distance between regions with other alternative approaches. It is also possible to use a wider list of indicators in each group of factors, which will make it possible to use, for example, the Principal Components Analysis to assess the complex impact of a group of factors on flows. In addition to the group of demographic characteristics and indicators of the labour and real estate markets, it will be possible to create such groups of factors that reflect the education system (primary, secondary and higher education indicators), the healthcare system (the number of hospital beds and the number of medical staff in the regions), infrastructure development (road density, green areas, transport accessibility) and the social sphere (standard of living of the population, the size of the consumption basket). Additionally, a more detailed study of the differences between commuting in East and West Germany might be performed. The analysis of the direction and intensity of commuting flows between the two territories and separately in each of them might provide a more clear view on the local labour markets. Moreover, it would be useful to analyse how the COVID-19 pandemic and the associated trend towards home offices influence commuting behaviour in Germany.

Acknowledgments. Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged.

References

Ahrens A., Lyons S. (2021). Do rising rents lead to longer commutes? A gravity model of commuting flows in Ireland. *Urban Studies*, 58 (2), 264–279. DOI: 10.1177/0042098020910698.

Artis M., Romaní J., Suriñach J. (2000). Determinants of individual commuting in Catalonia, 1986–91: Theory and empirical evidence. *Urban Studies*, 37 (8), 1431–1450. DOI: 10.1080/00420980020080191.

Baltagi B. H., Egger P. H., Pfaffermayr M. (2014). Panel data gravity models of international trade. *CESifo Working Paper Series*, No. 4616.

Baum-Snow N. (2010). Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960–2000. *American Economic Review*, 100 (2), 378–382. DOI: 10.1257/aer.100.2.378.

Bekele W. T., Mersha F. G. (2019). A dynamic panel gravity model application on the determinant factors of Ethiopia's coffee export performance. *Annals of Data Science*, 6 (4), 787–806. DOI: 10.1007/s40745-019-00198-4.

Bergstrand J. H. (1989). The generalized gravity equation, monopolistic competition, and the factor-proportions theory in international trade. *The Review of Economics and Statistics*, 71 (1), 143–153. DOI: 10.2307/1928061.

Bringolf-Isler B., Grize L., Mäder U., Ruch N., Sennhauser F. H., Braun-Fahrländer C. (2008). Personal and environmental factors associated with active commuting to school in Switzerland. *Preventive Medicine*, 46 (1), 67–73. DOI: 10.1016/j.ypmed.2007.06.015.

Bun M. J., Klaassen F. (2002). The importance of dynamics in panel gravity models of trade. https://ssrn.com/abstract=306100. DOI:10.2139/ssrn.306100.

Bunea D. (2012). Modern gravity models of internal migration. The case of Romania. *Theoretical and Applied Economics*, 4 (569), 127–144.

Burger M. J., De Goei B., Van der Laan L., Huisman F. J. (2011). Heterogeneous development of metropolitan spatial structure: Evidence from commuting patterns in English and Welsh city-regions, 1981–2001. *Cities*, 28 (2), 160–170. DOI:10.1016/j.cities.2010.11.006.

Daut W., Haller P. (2018). Berufliches pendeln zwischen wohn- und arbeitsort: Klarer trend zu längeren pendeldistanzen (No. 10/2018). IAB-Kurzbericht (in German).

Davidescu A. A. M., Strat V. A., Grosu R. M., Zgură I. D. (2017). Determinants of Romanians' migration within the European Union: Static and dynamic panel gravity approaches. *Amfiteatru Economic Journal*, 19 (46), 621–639.

Goetz S. J., Han Y., Findeis J. L., Brasier K. J. (2010). US commuting networks and economic growth: Measurement and implications for spatial policy. *Growth and Change*, 41 (2), 276–302. DOI: 10.1111/j.1468-2257.2010.00527.x.

Greenwood M. J. (1985). Human migration: Theory, models, and empirical studies. *Journal of Regional Science*, 25 (4), 521–544. DOI: 10.1111/j.1467-9787.1985.tb00321.x.

Gupta R., Gozgor G., Kaya H., Demir E. (2019). Effects of geopolitical risks on trade flows: Evidence from the gravity model. *Eurasian Economic Review*, 9 (4), 515–530.

Haas A., Hamann S. (2008). Pendeln — ein zunehmender trend, vor allem bei hochqualifizierten. Ost-West-vergleich (No. 6/2008). IAB-Kurzbericht (in German).

Handy S. L., Xing Y. (2011). Factors correlated with bicycle commuting: A study in six small US cities. *International Journal of Sustainable Transportation*, 5 (2), 91–110. DOI: 10.1080/15568310903514789.

Hansson E., Mattisson K., Björk J., Östergren P. O., Jakobsson K. (2011). Relationship between commuting and health outcomes in a cross-sectional population survey in southern Sweden. *BMC Public Health*, 11, 1–14.

Huinink J., Feldhaus M. (2012). Fertility and commuting behaviour in Germany. *Comparative Population Studies*, 37 (3–4). DOI: 10.12765/CPoS-2012-05.

Karemera D., Oguledo V. I., Davis B. (2000). A gravity model analysis of international migration to North America. *Applied Economics*, 32 (13), 1745–1755. DOI: 10.1080/000368400421093.

Keum K. (2010). Tourism flows and trade theory: A panel data analysis with the gravity model. *The Annals of Regional Science*, 44 (3), 541–557. DOI: 10.1007/s00168-008-0275-2.

Kim K., Cohen J. E. (2010). Determinants of international migration flows to and from industrialized countries: A panel data approach beyond gravity. *International Migration Review*, 44 (4), 899–932. DOI: 10.1111/j.1747-7379.2010.00830.x.

Kim C., Sang S., Chun Y., Lee W. (2012). Exploring urban commuting imbalance by jobs and gender. *Applied Geography*, 32 (2), 532–545. DOI: 10.1016/j.apgeog.2011.07.001.

Kimura F., Lee H. H. (2006). The gravity equation in international trade in services. *Review of World Economics*, 142, 92–121. DOI: 10.1007/s10290-006-0058-8.

Khadaroo J., Seetanah B. (2008). The role of transport infrastructure in international tourism development: A gravity model approach. *Tourism Management*, 29 (5), 831–840. DOI: 10.1016/j.tourman.2007.09.005.

Kobus M. B., Van Ommeren J. N., Rietveld P. (2015). Student commute time, university presence and academic achievement. *Regional Science and Urban Economics*, 52, 129–140.

Künn-Nelen A. (2016). Does commuting affect health? *Health Economics*, 25 (8), 984–1004. DOI: 10.1002/hec.3199.

Lauridsen J. (2006). Spatial autoregressively distributed lag models: Equivalent forms, estimation, and an illustrative commuting model. *The Annals of Regional Science*, 40, 297–311. DOI: 10.1007/s00168-005-0044-4.

Lee B. S., McDonald J. F. (2003). Determinants of commuting time and distance for Seoul residents: The impact of family status on the commuting of women. *Urban Studies*, 40 (7), 1283–1302. DOI: 10.1080/0042098032000084604.

Morgenroth E. L. (2002). Commuting in Ireland: An analysis of inter-county commuting flows. *ESRI Working Paper* 144. Dublin: Economic and Social Research Institute. https://www.esri.ie/publications/commuting-in-ireland-an-analysis-of-inter-county-commuting-flows.

Niebuhr A., Granato N., Haas A., Hamann S. (2012). Does labour mobility reduce disparities between regional labour markets in Germany? *Regional Studies*, 46 (7), 841–858. DOI: 10.1080/00343404.2010.532118.

Östh J., Lindgren U. (2012). Do changes in GDP influence commuting distances? A study of Swedish commuting patterns between 1990 and 2006. *Tijdschrift voor Economische en Sociale Geografie*, 103 (4), 443–456. DOI: 10.1111/j.1467-9663.2011.00697.x.

Partridge M. D., Ali K., Olfert M. R. (2010). Rural-to-urban commuting: Three degrees of integration. *Growth and Change*, 41(2), 303–335.

Patuelli R., Reggiani A., Gorman S. P., Nijkamp P., Bade F. J. (2007). Network analysis of commuting flows: A comparative static approach to German data. *Networks and Spatial Economics*, 7, 315–331. DOI: 10.1007/s11067-007-9027-6.

Patuelli R., Reggiani A., Nijkamp P., Bade F. J. (2010). The evolution of the commuting network in Germany: Spatial and connectivity patterns. *Journal of Transport and Land Use*, 2 (3/4), 5–37.

Ramos R., Surinach J. (2017). A gravity model of migration between the ENC and the EU. *Tijdschrift voor Economische en Sociale Geografie*, 108 (1), 21–35. DOI:10.1111/tesg.12195.

Reggiani A., Bucci P., Russo G. (2011). Accessibility and network structures in the German commuting. *Networks and Spatial Economics*, 11, 621–641. DOI: 10.1007/s11067-010-9149-0.

Santos G., Maoh H., Potoglou D., von Brunn T. (2013). Factors influencing modal split of commuting journeys in medium-size European cities. *Journal of Transport Geography*, 30, 127–137.

Sawyer J. A. (1967). An econometric study of international trade flows. By Hans Linnemann. Amsterdam: North-Holland Publishing Company. 1966. Pp. xiv, 234. *Canadian Journal of Economics and Political Science*, 33 (4), 633–634. DOI: 10.2307/140051.

Shen Q. (2000). Spatial and social dimensions of commuting. *Journal of the American Planning Association*, 66 (1), 68–82. DOI: 10.1080/01944360008976085.

Silva J. S., Tenreyro S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88 (4), 641–658. DOI: 10.1162/rest.88.4.641.

Simini F., González M. C., Maritan A., Barabási A. L. (2012). A universal model for mobility and migration patterns. *Nature*, 484 (7392), 96–100. DOI: 10.1038/nature10856.

Simini F., Maritan A., Néda Z. (2013). Human mobility in a continuum approach. *PloS One*, 8 (3), e60069. DOI: 10.1371/journal.pone.0060069.

Skora T., Rüger H., Stawarz N. (2020). Commuting and the motherhood wage gap: Evidence from Germany. *Sustainability*, 12 (14), 5692. DOI: 10.3390/su12145692.

Stefanouli M., Polyzos S. (2017). Gravity vs radiation model: Two approaches on commuting in Greece. *Transportation Research Procedia*, 24, 65–72. DOI: 10.1016/j.trpro.2017.05.069.

Stutzer A., Frey B. S. (2007). Commuting and life satisfaction in Germany. *Informationen zur Raumentwicklung*, 2 (3), 179–189.

Sun B., Ermagun A., Dan B. (2017). Built environmental impacts on commuting mode choice and distance: Evidence from Shanghai. *Transportation Research Part D: Transport and Environment*, 52, 441–453. DOI: 10.1016/j.trd.2016.06.001.

Tham S. Y., Goh S. K., Wong K. N., Fadhli A. (2018). Bilateral export trade, outward and inward FDI: A dynamic gravity model approach using sectoral data from Malaysia. *Emerging Markets Finance and Trade*, 54 (12), 2718–2735. DOI: 10.1080/1540496X.2017.1402176.

Tkocz Z., Kristensen G. (1994). Commuting distances and gender: A spatial urban model. *Geographical Analysis*, 26 (1), 1–14. DOI: 10.1111/j.1538-4632.1994.tb00307.x.

Vakulenko E. S., Mkrtchyan N. V., Furmanov K. K. (2011). Econometric analysis of internal migration in Russia. *Journal of Economics*, 7 (2), 21–33.

Varga L., Tóth G., Néda Z. (2016). An improved radiation model and its applicability for understanding commuting patterns in Hungary. *Regional Statistics*, 6 (2), 27–38.

Varga L., Tóth G., Néda Z. (2018). Commuting patterns: The flow and jump model and supporting data. *EPJ Data Science*, 7 (1), 1–13. DOI: 10.1140/epjds/s13688-018-0167-3.

Vieira E. R., Reis D. H. A. (2019). Determinants of Brazilian exports by levels of technological intensity: A gravity model analysis using the PPML estimator. *Journal of Business and Economics*, 10 (9), 861–879. DOI: 10.15341/jbe(2155-7950)/09.10.2019/007.

Viry G., Rüger H., Skora T. (2014). Migration and long-distance commuting histories and their links to career achievement in Germany: A sequence analysis. *Sociological Research Online*, 19 (1), 78–94. DOI: 10.5153/sro.3263.

Wajdi N., Adioetomo S. M., Mulder C. H. (2017). Gravity models of interregional migration in Indonesia. *Bulletin of Indonesian Economic Studies*, 53 (3), 309–332. DOI: 10.1080/00074918.2017.1298719.

Whalen K. E., Páez A., Carrasco J. A. (2013). Mode choice of university students commuting to school and the role of active travel. *Journal of Transport Geography*, 31, 132–142. DOI: 10.1016/j.jtrangeo.2013.06.008.

Yotov Y. V., Piermartini R., Monteiro J.-A., Larch M. (2016). *An advanced guide to trade policy analysis: The structural gravity model*. WTO. https://www.wto.org/english/res_e/booksp_e/advancedwtounctad2016_e.pdf.

Zhao P. (2013). The impact of the built environment on individual workers' commuting behavior in Beijing. *International Journal of Sustainable Transportation*, 7 (5), 389–415. DOI: 10.1080/15568318.2012.692173.

Zhao P. (2015). The determinants of the commuting burden of low-income workers: Evidence from Beijing. *Environment and Planning A: Economy and Space*, 47 (8), 1736–1755. DOI: 10.1177/0308518X15597112.

Received 24.09.2022; accepted 03.08.2023.

Appendix 1. Variables and their descriptive statistics, 2013–2019

Name	Description (i — residential place, j — workplace)	Average	Stand. deviation	Min	Max
flow	The flow of commuting migrants	275.5	1600.5	1	82206
distance	Distance between regions i and j (km)	206.9	142.9	1.1	817.9
same_land_flg	Dummy variable for belonging of regions i and j to the common land of Germany			0	1
common_border_flg	Dummy variable for the presence of a common border between regions i and j			0	1
	Demographic indicators				
population_home	Population in region i	204793	238878	34011	3669491
population_work	Population in region j				
density_pop_home	Population density in region i	529	695	36	4777
density_pop_work	Population density in region j				
foreign_share_home	Share of foreign population in region i	9.22	5.17	1	37
foreign_share_work	Share of foreign population in region j				
livebirth_home	Birthrate per 10000 people in region i	87.94	11.58	58	132
livebirth_work	Birthrate per 10000 people in region j				
death_home	Deaths per 10000 people in region i	116.78	18.38	73	178
death_work	Deaths per 10000 people in region j				
	Labor market indicators				
earnings_home	Average salary in region i	31683	4962	21920	62360
earnings_work	Average salary in region j				
unemployment_home	Unemployment rate in region i	5.64	2.66	1	15
unemployment_work	Unemployment rate in region j				
company_home	Number of companies per 10000 people in region i	73.73	19.2	33	311
company_work	Number of companies per 10000 people in region j				
	Indicators of the real estate market				
apart_prices_home	Purchase price of 1 m^2 apartment in region i	1780	835	675	7700
apart_prices_work	Purchase price of 1 m^2 apartment in region j				
lease_prices_home	Rental price of 1 m^2 apartment in region i	6.7	1.72	4	18
lease_prices_work	Rental price of 1 m^2 apartment in region j				
	Indicators of economic development and welfare				
gdp_ppc_home	GDP per capita in region i	35688	15829	14316	188453
gdp_ppc_work	GDP per capita in region j				
auto_home	Number of private cars per 1000 people in region i	576.22	70.82	204	1154
auto_work	Number of private cars per 1000 people in region j				
	Other indicators of the regions' attractiveness				
not_uni_graduate_home	Share of school graduates without certificate in region i	6.17	2.36	1	17
not_uni_graduate_work	Share of school graduates without certificate in region j				
benefit_share_home	Share of people receiving minimum social benefits in region i	8.01	3.92	1	24
benefit_share_work	Share of people receiving minimum social benefits in region j				

End of Appendix 1

Name	Description (i — residential place, j — workplace)	Average	Stand. deviation	Min	Max
tourism_intensity_home	Tourism intensity in region i	5.62	6.58	0	61
tourism_intensity_work	Tourism intensity in region j				
water_price_home	The price of 1 m^3 of drinking water in region i	1.69	0.43	0.61	3.31
water_price_work	The price of 1 m^3 of drinking water in region j				

Appendix 2. Different specifications of PPML model (dependent variable — *flow*)

Variables	Specifications						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
population_home	0.827***	0.762***	0.783***	0.826***	0.818***	0.817***	0.830***
	(0.0407)	(0.0524)	(0.0593)	(0.0548)	(0.0534)	(0.0535)	(0.0565)
population_work	1.080***	0.876***	0.899***	0.828***	0.819***	0.821***	0.816***
	(0.0445)	(0.0377)	(0.0391)	(0.0439)	(0.0423)	(0.0420)	(0.0410)
earnings_home		0.0557		0.0201	–0.0180	–0.0175	0.0713
		(0.156)		(0.160)	(0.162)	(0.163)	(0.159)
earnings_work		–0.372		–0.422	–0.423	–0.415	–0.446
		(0.341)		(0.337)	(0.334)	(0.324)	(0.303)
unemployment_home	–0.459***	–0.172*	–0.155*	–0.142	–0.215*	–0.0820	–0.0340
	(0.106)	(0.0916)	(0.0920)	(0.0873)	(0.116)	(0.226)	(0.212)
unemployment_work	0.572***	0.216**	0.193**	0.238***	0.0533	0.180	0.273
	(0.0967)	(0.0937)	(0.0824)	(0.0920)	(0.104)	(0.259)	(0.242)
company_home	–0.537***		–0.380***	–0.307**	–0.297**	–0.302**	–0.135
	(0.138)		(0.113)	(0.126)	(0.117)	(0.118)	(0.125)
company_work	0.426***		0.0130	–0.199*	–0.164	–0.171	–0.225**
	(0.115)		(0.125)	(0.116)	(0.109)	(0.109)	(0.109)
apart_prices_home				0.216**	0.177	0.168	0.000331
				(0.108)	(0.109)	(0.107)	(0.0988)
apart_prices_work				–0.0404	–0.148	–0.152	–0.226**
				(0.127)	(0.128)	(0.130)	(0.105)
lease_prices_home				–0.727***	–0.708***	–0.694***	–0.221
				(0.188)	(0.186)	(0.183)	(0.187)
lease_prices_work				0.712***	0.681***	0.684***	0.626***
				(0.237)	(0.231)	(0.232)	(0.238)
gdp_ppc_home		–0.281**	–0.210*	–0.125	–0.121	–0.103	–0.0506
		(0.110)	(0.111)	(0.107)	(0.103)	(0.109)	(0.122)
gdp_ppc_work		1.082***	1.067***	0.923***	0.960***	0.978***	0.959***
		(0.0571)	(0.0606)	(0.0712)	(0.0716)	(0.0714)	(0.0782)
auto_home					–0.378	–0.462*	–0.953***
					(0.253)	(0.240)	(0.270)
auto_work					–0.628***	–0.677***	–0.519**
					(0.211)	(0.224)	(0.253)
benefit_share_home						–0.155	0.0694

End of Appendix 2

Variables	Specifications						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
						(0.182)	(0.208)
benefit_share_work						–0.136	–0.294
						(0.211)	(0.225)
density_pop_home							–0.180***
							(0.0595)
density_pop_work							0.000269
							(0.0444)
foreign_share_home							–0.165*
							(0.0990)
foreign_share_work							0.0940
							(0.0846)
livebirth_home							–0.0671
							(0.256)
livebirth_work							0.376*
							(0.222)
death_home							–0.345
							(0.248)
death_work							–0.152
							(0.229)
not_uni_graduate_home							0.109**
							(0.0535)
not_uni_graduate_work							0.112**
							(0.0473)
water_home_year							–0.0652
							(0.0872)
water_work_year							0.129
							(0.123)
distance	–1.344***	–1.253***	–1.247***	–1.256***	–1.246***	–1.250***	–1.282***
	(0.0777)	(0.0785)	(0.0919)	(0.0788)	(0.0786)	(0.0743)	(0.0749)
common_border_flg	1.491***	1.619***	1.638***	1.645***	1.664***	1.659***	1.638***
	(0.103)	(0.110)	(0.125)	(0.109)	(0.108)	(0.103)	(0.102)
same_land_flg	1.094***	1.134***	1.146***	1.129***	1.138***	1.133***	1.106***
	(0.0714)	(0.0738)	(0.0833)	(0.0728)	(0.0727)	(0.0692)	(0.0688)
home_land	–0.632***	–0.919***	–0.482***	–0.419***	–0.424***	–0.396***	–0.103
	(0.115)	(0.180)	(0.107)	(0.120)	(0.118)	(0.130)	(0.133)
work_land	0.225	–0.122	0.0542	0.188	–0.0462	–0.0425	0.164
	(0.143)	(0.174)	(0.132)	(0.164)	(0.162)	(0.163)	(0.166)
Constant	–12.42***	–14.22***	–17.65***	–13.38***	–5.308	–4.682	–2.309
	(1.137)	(3.581)	(1.668)	(4.045)	(4.122)	(4.468)	(4.014)
Observations				1117200			
R-squared	0.543	0.573	0.544	0.572	0.571	0.577	0.577

Note. Robust standard errors in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

Appendix 3. Full table with estimates of the gravity model without a lag, with a lag of 1 year, and with a lag of 2 years (dependent variable — *flow*)

Variables	PPML	PPML (lag 1)	PPML (lag 2)
population_home	0.830***	0.831***	0.833***
	(0.0565)	(0.0564)	(0.0559)
population_work	0.816***	0.815***	0.815***
	(0.0410)	(0.0416)	(0.0408)
density_pop_home	–0.180***	–0.179***	–0.177***
	(0.0595)	(0.0581)	(0.0571)
density_pop_work	0.000269	0.00510	0.00809
	(0.0444)	(0.0435)	(0.0432)
foreign_share_home	–0.165*	–0.143	–0.123
	(0.0990)	(0.0883)	(0.0781)
foreign_share_work	0.0940	0.0904	0.0789
	(0.0846)	(0.0772)	(0.0733)
livebirth_home	–0.0671	–0.129	–0.200
	(0.256)	(0.262)	(0.273)
livebirth_work	0.376*	0.463*	0.570**
	(0.222)	(0.245)	(0.268)
death_home	–0.345	–0.386	–0.397*
	(0.248)	(0.245)	(0.241)
death_work	–0.152	–0.247	–0.301
	(0.229)	(0.231)	(0.232)
earnings_home	0.0713	0.0835	0.102
	(0.159)	(0.155)	(0.152)
earnings_work	–0.446	–0.439	–0.432
	(0.303)	(0.297)	(0.288)
unemployment_home	–0.0340	–0.0417	–0.0289
	(0.212)	(0.210)	(0.209)
unemployment_work	0.273	0.258	0.227
	(0.242)	(0.247)	(0.235)
company_home	–0.135	–0.102	–0.0979
	(0.125)	(0.131)	(0.131)
company_work	–0.225**	–0.179*	–0.125
	(0.109)	(0.108)	(0.110)
apart_prices_home	0.000331	–0.0100	–0.0247
	(0.0988)	(0.100)	(0.107)
apart_prices_work	–0.226**	–0.334***	–0.484***
	(0.105)	(0.117)	(0.132)
lease_prices_home	–0.221	–0.241	–0.223
	(0.187)	(0.188)	(0.193)
lease_prices_work	0.626***	0.679***	0.767***
	(0.238)	(0.243)	(0.236)

End of Appendix 3

Variables	PPML	PPML (lag 1)	PPML (lag 2)
gdp_ppc_home	–0.0506	–0.0530	–0.0452
	(0.122)	(0.122)	(0.121)
gdp_ppc_work	0.959***	0.940***	0.934***
	(0.0782)	(0.0792)	(0.0788)
auto_home	–0.953***	–1.042***	–1.115***
	(0.270)	(0.284)	(0.283)
auto_work	–0.519**	–0.473*	–0.482*
	(0.253)	(0.263)	(0.268)
not_uni_graduate_home	0.109**	0.133**	0.154**
	(0.0535)	(0.0572)	(0.0612)
not_uni_graduate_work	0.112**	0.123***	0.130***
	(0.0473)	(0.0464)	(0.0457)
benefit_share_home	0.0694	0.0503	0.0145
	(0.208)	(0.204)	(0.203)
benefit_share_work	–0.294	–0.279	–0.263
	(0.225)	(0.232)	(0.229)
tourism_intensity_home	0.0514*	0.0515*	0.0508*
	(0.0271)	(0.0266)	(0.0263)
tourism_intensity_work	0.0119	0.0222	0.0319
	(0.0299)	(0.0291)	(0.0283)
water_home	–0.0652	–0.0721	–0.0936
	(0.0872)	(0.0905)	(0.0965)
water_work	0.129	0.160	0.191
	(0.123)	(0.132)	(0.145)
distance	–1.282***	–1.287***	–1.292***
	(0.0749)	(0.0736)	(0.0723)
common_border_flg	1.638***	1.633***	1.625***
	(0.102)	(0.101)	(0.0994)
same_land_flg	1.106***	1.104***	1.100***
	(0.0688)	(0.0676)	(0.0667)
home_land	–0.103	–0.673***	–0.125
	(0.133)	(0.184)	(0.131)
work_land	–0.0223	–0.0279	0.161
	(0.159)	(0.158)	(0.159)
Constant	–2.123	–0.400	–0.0928
	(4.019)	(4.111)	(4.096)
Observations		1117200	
R-squared	0.577	0.579	0.584
Year FE	Yes	Yes	Yes
Home-land FE	Yes	Yes	Yes
Work-land FE	Yes	Yes	Yes

Note. Robust standard errors in parentheses; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.