University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 1/2021 (349)

CAUSES OF THE SPATIALLY UNEVEN OUTFLOW OF WARSAW INHABITANTS TO THE CITY'S SUBURBS: AN ECONOMIC ANALYSIS OF THE PROBLEM

Honorata Bogusz

Warsaw 2021

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS

Causes of the spatially uneven outflow of Warsaw inhabitants to the city's suburbs: an economic analysis of the problem

Honorata Bogusz

University of Warsaw, Faculty of Economic Sciences, Labfam, and Data Science Lab WNE UW
Corresponding author: h.bogusz@uw.edu.pl

Abstract: In this article I provide a quantitative analysis of suburban migration patterns in Warsaw, Poland. Basing this analysis on the extended gravity model of migration, an econometric panel model was built to identify key pulling factors for migrants who move from Warsaw to its suburbs. The role of residential lot prices and the resulting possible endogeneity are also discussed. It was confirmed that migrants choose boroughs of greater population density that have higher average relative income and more amenities, but at a smaller distance to Warsaw's city center and with lower residential lot prices relative to those in Warsaw.

Keywords: gravity model of migration, suburbanization, Mundlak terms, Correlated Random Effects

JEL codes: R23, P25, C23, C51

1. Introduction

In the last 11 years (2008-2019), the number of people living in the agglomeration of Warsaw (Warsaw and 70 suburban boroughs, as defined by the EU NUTS2 classification) increased by 8.18%, while the number of Warsaw inhabitants increased by 4.73% and the number of people living in the suburbs increased by 13.41%. This means that the ratio of those living in the suburbs to the total population in the agglomeration increased by 4.84%.[1] The outflow of Warsaw inhabitants to the city's suburbs is quite significant, having increased by 43.90% between 2008-2019. This outflow is compensated by the inflow of people from other parts of Poland, which prevents the desolation of Warsaw, as is typical for the suburbanization process (Caves, 2004). Nonetheless, due to the obvious shift of population from the city to the outskirts, suburbanization can still be tackled in the context of Warsaw. The inflow of people to the suburbs is not even, however. For example, in 2019 909 people moved to the most popular borough, while 0 migrants reported a new residence in some of the least popular municipalities. Even though suburbanization is a natural stage in the evolution of a city, unrestricted city growth can lead to a variety of formidable consequences. The negative impacts of urban sprawl include: an increase in vehicle mileage, residential energy consumption, and land use; degradation of air quality; increased usage of natural resources and thus of greenhouse gas emissions (Kahn, 2000); increased infrastructure costs and thus growing fiscal deficits (Downs, McCann & Mukherji, 2005); decline of social capital; residential segregation resulting in class and racial divisions (Duany, Plater-Zyberk & Speck, 2010); and health deterioration as a result of increased vehicle usage (Sturm & Cohen, 2004). However, during the communist period, city growth, residential mobility, and land and housing development were under tight political control. Suburbanization "in a Western sense" has been a recent phenomenon in European cities of the former Eastern Bloc. It is believed to have begun in the post-communist countries in the 90s, following the political transformation (Lisowski, 2004, Timar & Varadi, 2001, Nuissl & Rink, 2005).

The causes of metropolitan suburbanization are heavily discussed in the literature and several theories have been proposed (Mieszkowski & Mills, 1993). Kok (1999), Lisowski (2004), Murray & Szelenyi (2009), and Nuissl & Rink (2005) offered some insight about suburbanization processes in post-communist cities. Quantitative measures of suburbanization's determinants are scarce in the literature and include works by Jordan, Ross

[1] Own calculation based on population data from the Polish Statistical Office.

& Usowski (1998), Loibl (2004), and Kok (1999). In all of these three works, simple regression models were used (logit in the case of Kok (1999)) and only a few possible pulling factors were included. Surprisingly, no application of the gravity model of migration (the most well-known quantitative migration model (Poot, Alimi, Cameron, and Mare (2016)) in the context of suburban migration has been offered in the literature. Moreover, Loibl (2004) pointed out the residential lot prices as a vital attractiveness measure, but didn't include them in his analysis due to his presumption that they can be directly dependent on the demand for migrants. This matter has not yet been addressed. These two research gaps indicate the research potential of the determinants of unequal suburbanization and they are addressed in this paper. An econometric panel model spanning the years 2008-2019 and including various attractiveness measures has been built and estimated by Feasible GLS. Endogeneity, which can arise due to the included price of housing or the relative number of kindergartens, is addressed by lagging and incorporating so-called Mundlak terms (Mundlak, 1978). Finally, a measure of the price of housing relative to that in Warsaw is built on the basis of the transactional data obtained from Poland's county offices. These data include information on every transaction concerning three types of properties: flats, houses, and parcels, in every borough, in 2008-2019.[2]

Since the suburbanization processes happening now in Warsaw are in relatively early stages compared to Western cities, their negative consequences may be curtailed to some extent. Therefore, it is highly useful to identify factors that push migrants from Warsaw and pull them to the suburbs. Some of these factors can be directly influenced by local planning, e.g., institutional or transport amenities. My aim is to identify the features that are key pulling or pushing factors for migrants. I believe that this work sheds light not only on the local situation of Warsaw, but also contributes to understanding the bigger picture of suburbanization in post-communist cities. The remaining part of this paper is structured as follows. In order to identify possible pulling factors, the second section covers a review of suburbanization theories and the existing empirical evidence, and is followed by methodological issues and dataset description. Thereafter, empirical analysis is conducted. My conclusions complete the study.

2. Literature Review

An excellent, thorough review of the existing theories of suburbanization can be found in Mieszkowski & Mills (1993) and according to those authors, two classes of theories have been offered. The first one is "natural evolution theory" and it is based on a simple chain of events.

[2] One of the 70 boroughs (Żyrardów) defined as part of the agglomeration is excluded from this analysis due to the lack of transactional data.

The job market is originally located in the city center. To minimize the costs of commuting to the central business district (CBD), people settle in close proximity to it. As the centrally located residential units become filled up, new houses are built in further located spaces. As the quality of life in the city center deteriorates, affluent city residents prefer to move to these new, suburban settlements, as they offer more living space and newer properties. Historically, due to the high relative cost of commuting, higher social strata lived in the city centers, while the poorer people inhabited the outskirts. Today, this tendency has been reversed. The second class of theories is based on the Tiebout (1956) model and states that the given city's institutional problems are what underlie the causes of suburbanization. These problems are: high taxes, poor quality of public schools, criminality, and low quality of the living environment. People disappointed with living conditions in the city center move to the outskirts, while those who cannot afford to remain there. The population outflow fuels the worsening fiscal and social situation of the center, which in turn leads to further migration, creating a vicious circle. Mieszkowski & Mills (1993) also stress the importance of social affiliation preferences: people generally prefer to live in a neighborhood whose inhabitants evince a similar financial background. The authors state that those two classes of theories have a number of interconnections and that therefore it is challenging to distinguish them empirically. Nonetheless, both are relevant to this work, as they point out certain features of suburban municipalities that can attract migrants, e.g., higher average income.

Quantitative measures of suburbanization existing in the literature are scarce and include works by Jordan et al. (1998), Loibl (2004), and Kok (1999). Jordan et al. (1998) investigated the outflow to suburbs in 35 American metropolises in 1980-90, taking the population density gradient as the dependent variable. They intended to compare the same process in different cities and found out that suburbanization proceeds more quickly in places previously unaffected by it – namely, those of greater population growth rate where the job market is more centralized. On the other hand, the more expensive the rents in the suburbs, the fewer people settle there. Loibl (2004) offered yet another quantitative study of suburbanization patterns and adapted a multi-agent system approach to simulate different trends within urban sprawl in Vienna. The simulations for a 30-year span were compared with real observations. The author found out that a remarkable decrease of urban sprawl can be achieved by applying the right planning measures, even when the number of migrating households stays the same. Part of the simulation design was to identify "attractiveness measures" and 4 were distinguished: landscape attractiveness (measured by the forest area quota in the neighborhood); the supply of local services (access and number of attorneys); core-city availability (calculated by applying the shortest-path model to find the minimum travel time to Vienna); and residential lot prices and availability of lots information (zoned, but still vacant residential areas). Loibl (2004) tested the influence of these factors on the net number of migrants by linear regression. He identified core-city accessibility, landscape attractiveness, and the supply of services as significant pulling factors. Residential lot prices were not included in the final regression, as, according to the author, they are directly dependent on the demand of migrants. The only empirical work tackling suburbanization in post-communist countries was made by Kok (1999), who adapted life-course approach to investigate individual decisions to move out. This approach assumes that the decision to move out is an outcome of striving to satisfy individual needs and that the "trajectory of migration" is closely connected to certain life events, such as taking a new job or pursuing education. Kok (1999) based his logit model on survey data for Poland (49 cities) and Hungary (19 cities), taking an individual binary decision of moving out as the dependent variable. He found out that being married, getting a new job, purchasing property, being in age groups 18-24 and 35-39, and moving out in the years 1989-1993 have a positive influence on the probability of moving to the suburbs. These findings confirm the presumption that suburban communities are rather homogeneous and that the suburbanization process started in Poland right after the political change begun in 1989. Although some pulling and pushing factors have already been suggested, little attention has been devoted to quantitative measures of suburbanization in the existing literature. Apart from that, two research gaps have remained to be addressed – and they are filled in this article. The extended gravity model of the migration framework is used to measure the inflow to suburban municipalities. Simultaneously, information about residential lot prices is incorporated in the model.

3. Methods

The gravity model of migration is one of the oldest and most popular analytical models for migration flows. According to this model, spatial flows of people depend positively on the size of target areas and negatively on the distance between them. In that sense, it resembles Newton's law of universal gravitation (Newton, 1687) what was first noticed by Ravenstein (1885, 1889). Poot et al. (2016) deliver a thorough description of that model and, following their article, the commonly applied form is:

$$M_{ij} = G\frac{P_i^{\alpha}P_j^{\beta}}{D_{ij}^{\gamma}},$$

where M_{ij} is the migration number of people who previously lived in area j (i) and moved to area i (j). P_i (P_j) is the population of that area at the beginning of the migration and D_{ij} is the distance between the two areas. G is the local constant and α, β, γ are parameters to be estimated. It is useful to take the logarithm of the above equation in order to express it in a common, econometric framework:

$$lnM_{ij} = \delta + \alpha\, lnP_i + \beta\, lnP_j - \gamma\, lnD_{ij} + \varepsilon_{ij},$$

where a zero-mean error term ε_{ij} was added to the equation and the constant term was replaced by the parameter δ. Several extended forms of the gravity model of migration have been proposed (Beine et al., 2016; Fan, 2005; Lowry, 1966; Millock, 2015) and a general linearized extended model can be expressed:

$$lnM_{ij} = \delta + \alpha\, lnP_i + \beta\, lnP_j - \gamma\, lnD_{ij} + \beta_m lnX_C + \beta_n X_D + \varepsilon_{ij},$$

where X_C is a vector of supplementary continuous variables, X_D – a vector of supplementary binary variables, β_m and β_n – vectors of coefficients of the supplementary variables.

I estimate the gravity model of migration equation with linear regression, considering a panel setting (69 boroughs in 11 years). In order to capture the effects of several variables constant in time, I choose the random effects specification. A formal justification for that choice (results of the Hausman test) is provided in Table 4. One assumption of the random effects model is the strict exogeneity of regressors with respect to both individual effects and transient errors. In my model, the first lags are taken of all time-varying regressors in order to account for possible simultaneity problems. In addition, I include a measure of the price of housing, which can cause endogeneity (Loibl, 2004). One other variable is suspected of being directly dependent on the number of migrants – namely, the number of kindergartens per 100,000 people. Despite a lack of evidence about this in the existing literature, it is common sense that if the number of residents of a borough increases as a result of migration, more kindergartens and schools are needed to foster the greater number of children. One way to allow for the unobserved heterogeneity to depend on some regressors linearly is the Mundlak (1978) framework (Woolridge, 2001). In this kind of setting, the individual effects may depend linearly on the time averages of some regressors, resulting in these time averages capturing the "environmental effect" of variables. In my case, it is feasible to assume global effects of the price and the number of kindergartens increasing in all boroughs in time. When the time averages of those variables are included, true coefficients of the time series can be revealed. It is customary to call this kind of model "Correlated Random Effects" and the time averages included are usually referred to as “Mundlak terms”. Hence, a time average of the number of kindergartens per 100,000 people is included in the regression equation. However, the possible endogeneity arising due to the price of housing variable is tackled differently. I do not include a crude measure of just the average price of m^2 of housing, but instead choose a relative measure, i.e., the average price of m^2 in a borough is divided by the average price in Warsaw. The reason for this inconsistency, with respect to Mundlak terms, is that there is naturally also a global effect of residential lot prices rising in Warsaw each year and it is likely that the prices in boroughs and in Warsaw rise by approximately the same or similar factor. Both time series of the average price of m^2 – in boroughs and in Warsaw – should exhibit a global effect, but not their ratio. The random effects model also requires sphericity of transient errors and homoskedasticity of individual effects. While it is plausible to expect that the boroughs are influenced by Warsaw more than by each other, they are spatial entities and can interact with one another to some extent. The diagnostics for the here constructed model is reported in Table A3 in the Appendix. It can be seen that, according to Pesaran’s test, cross-sectional dependence is present in the model, resulting in the collinearity of transient errors. One way to deal with this problem is to apply the Feasible Generalized Least Square estimator, which uses a robust variance-covariance matrix. It is known to yield estimates more asymptotically efficient that regular GLS, if the sphericity of the transient errors assumption doesn’t hold (Woolridge, 2001). Hence, it is used to estimate the equation in this article.[3]

4. Dataset Description

My dataset comprises observations for 69 suburban boroughs defined as belonging to the “Warsaw Metropolitan Region,” according to the EU NUTS2 classification, in the years 2008-2019. Hence, I have the most recent data at my disposal. The specific variables used in the model, with descriptions, are given in Table 1.

[3] The calculations behind the econometric modelling presented in this article were prepared in Stata16 software. Figure 1 was prepared in R software, using the ggplot2 package. Transactional data from Polish county offices were parsed using Python regular expressions.

Table 1: Description of the variables used in the analysis.

Name	Description	Source
log (migrants)	Logarithm of number of migrants who checked out of Warsaw and reported residence in one of the suburban boroughs.	Polish Statistical Office
log (popdens)	Logarithm of population density (in people/km2) in a borough.	Polish Statistical Office
log (dist)	Logarithm of distanced between the centre of Warsaw and centre of the borough, measured by a straight line.	Google Maps
log (income)	Logarithm of relative income in a borough, measured by a PIT based income per capita divided by the yearly average weighted by total population.	Polish Statistical Office
log (greenery)	Logarithm of total greenery spaces in a borough (in ha) – parks, forests etc.	Polish Statistical Office
unempl	Unemployment rate in a borough.	Polish Statistical Office
kinder	Number of kindergartens per 100 000 people in a borough.	Polish Statistical Office
mkinder	Time average of the above.	own calculation
bu	Borough type – urban (according to Polish administrative classification). Dummy.	Polish Statistical Office
bur	Borough type – urban-rural (according to Polish administrative classification). Dummy.	Polish Statistical Office
br	Borough type – rural (according to Polish administrative classification). Dummy.	Polish Statistical Office
train	Dummy indicating presence of a suburban train station.	Polish Statistical Office
averagebwaw	Average price of m^2 of housing, weighted by the number of transactions corresponding to different types of properties: flat, house or parcel, divided by the average price of m^2 of housing in Warsaw.	own calculation based on transactional data from Polish county offices and Polish Statistical Office

Source of Table: own preparation

Following the gravity model of migration framework, a logarithm of the number of migrants who checked out of Warsaw and reported residence in one of the suburban boroughs is taken as the dependent variable. In cases where no migrants moved from Warsaw to the borough, 0 is replaced by 0.5 before taking the logarithm, as described in Poot et al. (2016). Logarithms of population density (which can be used instead of population total to make it robust to the spatial sizes of boroughs) and distance are the basic regressors of that model. Furthermore, I decided to include several other variables in the analysis. The logarithm of relative income in a borough, measured by a PIT-based income per capita divided by the yearly average weighted by total population measures the relative welfare in a borough. The unemployment rate serves as a proxy for local job market conditions. The logarithm of total greenery spaces in a borough, as well as the number of kindergartens per 100,000 people, depict the influence of local amenities on the number of migrants. The presence of a suburban train station is included as a measure of the transport system between Warsaw and a borough. Two of three levels of a borough type according to Polish administrative classification are used as proxies of urbanization. Finally, I have a perfect measurement of the residential lot prices at my disposal.

Information about prices of housing available at the Polish Statistical Office are limited to the county level in years 2015-2018, which is a very rough measure for prices in boroughs. In my attempt to find a better one, I obtained real estate transactional data from Polish county offices for each of the 69 boroughs in the years 2008-2019 for the following property types: flats, houses, and parcels. Based on that parsed data, I constructed an indicator of the relative price of housing. I included observations for each borough, type, and year where the number of transactions was greater than 10. The average price of m^2 of housing was obtained by weighting average prices of property types (in each borough and each year) by the number of transactions corresponding to different types.[4] That average price was divided by the average price in Warsaw as obtained from the Polish Statistical Office for 2015-2018 and extrapolated for 2008-2019 by the use of the index of prices of housing for Warsaw for 2008-2019. Hence, a price relative to the one in Warsaw was constructed. Furthermore, no transactions were observed for

[4] However, in terms of flats, the price is related to the living space, while it corresponds to the area of a parcel for both house and parcel. This is due to the reason that no information about the living space of houses was available for about half of the boroughs in my study. The availability of details of transactions depends on the data classifying and storing system a specific county office uses and there is no standardization of such a system in Poland. In addition, I limited the area of parcels to observations greater than 200 m^2 and smaller than 2000 m^2. The reason for this distinction is that properties classified as houses or parcels are not necessarily lots habitable by private agents, but can be also service establishments. On the one hand, 200 m^2 is the minimum parcel size, where a habitable house, rather than holiday cabin can be built. On the other, 2000 m^2 is a reasonable upper boundary since private houses are rarely built on larger parcels in the suburbs of Warsaw (the average parcel size is between 1400-2000 in 90% of the boroughs in years examined) giving rise to the presumption that larger parcels function as service establishments in the vast majority.

all types or at all for some boroughs in some years. To complete the unbalanced panel, the well-known multiple imputation technique was used. Since in my setting observations for only one regressor were missing, predictive mean matching was applied. When imputing the missing values of a continuous variable, PMM may be preferable to linear regression when the normality of the underlying model is suspect (Little, 1988; Rubin, 1986). Panel models are asymptotically normal and formal normality is not required (Woolridge, 2001). As reported in Table 2 (Shapiro-Wilk normality test p-value), the variables presented here are not normal, which justifies the use of PMM. 140 of 759 observations were missing and imputed by PMM. The imputation was made before calculating the time average or price relative to the price in Warsaw.[5]

Table 2: Descriptive statistics of the variables from the dataset.

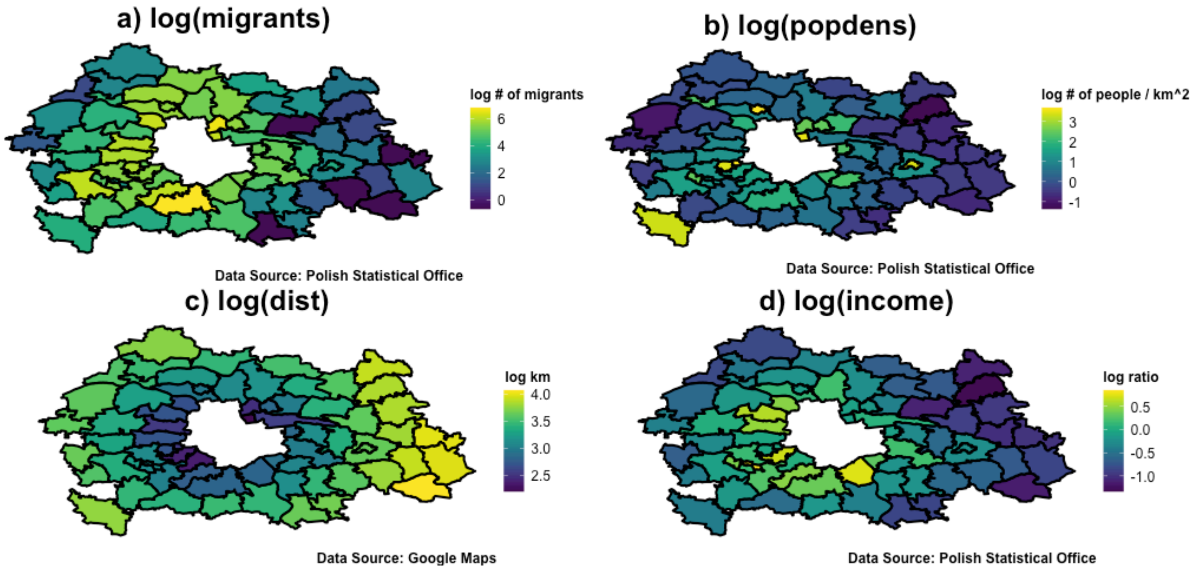
Name	**Mean**	**Std. Dev.**	**Min**	**Max**	**Shapiro-Wilk test p-value**
log(migrants)	3.90	1.62	0.41	7.03	0.00
log(popdens)	0.71	1.23	- 1.37	3.69	0.00
log(dist)	3.27	0.45	2.20	4.07	0.00
log(income)	- 0.29	0.58	- 2.33	1.26	0.01
log(greenery)	3.21	1.23	0.64	6.14	0.00
unempl	0.05	0.02	0.01	0.13	0.00
kinder	36.39	18.73	0.5	106.02	0.00
mkinder	36.49	16.14	3.09	92.53	0.00
bu	0.19	0.39	0	1	-
bur	0.28	0.45	0	1	-
br	0.54	0.50	0	1	-
train	0.51	0.50	0	1	-
averagebwaw	0.13	0.16	0.00	0.67	0.00

Source of Table: own preparation

[5] A proof that imputation doesn't change the general outcome of the estimation reported here can be made avaliable on request.

Table 2 reports the summary statistics for all of the variables, after the imputation. It can be seen that the mean of log number of migrants is equal to 3.90 and the standard deviation – 1.62. While, in some boroughs, as low as 0 migrants have reported residence, it was maximally 909 – a great disparity. Great discrepancies between boroughs in the levels of the logarithm of population density, logarithm of distance, logarithm of income, logarithm of greenery, and the number of kindergartens are also evident. It is also worth noting that the levels of variables are generally comparable, which provides reasoning for the regression models being numerically stable. Visual representation of the logarithm of number of migrants, as well as the logarithm of population density, logarithm of distance from Warsaw and the logarithm of average income in 2019, is plotted on Figure 1.

Figure 1: Visual representation of the log number of migrants, log population density, log distance and log average income in boroughs.



Source: own preparation in R based on data from Polish Statistical Office and Google Maps. The coordinates for the maps come from Open Street Map, which bases the coordinates for Poland on the Polish National Registry of Borders. The presented color scale is the default for Python – the Viridis palette. It is robust to all kinds of colorblindness, and was hence used.

It can be seen that none of the variables was evenly spread between boroughs in that year. Some of the boroughs with a greater logarithm of the number of migrants are the ones of increased logarithm of population density, smaller logarithm of distance, and greater logarithm of the average relative income, which follows the anticipated relationships between the dependent variable and the three plotted regressors. Furthermore, we expect larger greenery spaces, more kindergartens, and the presence of a suburban train station to be a pulling factor.

Following the literature (e.g, Loibl, 2004), migrants should choose municipalities of lower residential lot prices. We don't expect the local job market conditions represented by the unemployment rate to play a role in the case of suburban migration, because migrants living in the suburban ring usually still commute to work to Warsaw's core.

5. Empirical Analysis

The following equation is estimated by Feasible Generalized Least Squares:

$$log(migrants)_{it} = \beta_0 + \beta_1 \, log(popdens)_{i(t-1)} + \beta_2 \, log(dist)_{i(t-1)} + \beta_3 \, log(income)_{i(t-1)} + \beta_4 \, log(greenery)_{i(t-1)} + \beta_5 unempl_{i(t-1)} + \beta_6 kinder_{i(t-1)} + \beta_7 mkinder_i + \beta_8 bur_i + \beta_9 br_i + \beta_{10} train_i + \beta_{11} averagebwaw_{i(t-1)} + \eta_i + \varepsilon_{it},$$

where $(\beta_0, \dots, \beta_{11})$ denote the coefficients to be estimated, η_i denotes individual effects and ε_{it} - transient errors, $i = 1, \dots, 69;\ t = 2008, \dots, 2019$. The results are reported in Table 3 below.

Table 3: Model estimation results

Number of observations						**690**
Number of groups						69
Time periods						10
Wald chi2 (12)						2258.66
Prob > chi2						0.00
Dependent variable						log(migrants)
Regressor	**Coefficient**	**Std. Err.**	**z**	**P>|z|**	**95% lower**	**95% upper**
log(popdens)	0.27	0.07	3.98	0.00	0.14	0.40
log(dist)	- 1.51	0.13	- 11.86	0.00	- 1.77	- 1.26
log(income)	0.32	0.10	3.28	0.00	0.13	0.52
log(greenery)	0.28	0.05	6.12	0.00	0.19	0.36
unempl	2.30	1.69	1.36	0.18	- 1.02	5.61

kinder	- 0.01	0.00	- 2.77	0.01	- 0.02	- 0.00
mkinder	0.03	0.00	6.73	0.00	0.02	0.04
bur	0.53	0.13	4.14	0.00	0.28	0.79
br	0.50	0.14	3.62	0.00	0.23	0.78
train	0.24	0.07	3.30	0.00	0.10	0.38
averagebwaw	- 0.63	0.28	- 2.22	0.03	- 1.18	- 0.07
constant	6.63	0.48	13.86	0.00	5.70	7.57

Source of Table: own preparation

The results of the estimation are consistent with the prior expectations. All regressors are significant at 5% tolerance level, except for the unemployment rate (in all three cases). Since the dependent variable and the majority of the regressors are logarithmed, the coefficients can be interpreted as elasticities or semi-elasticities. Population density is a pulling factor – if it rises by 1%, the number of migrants increases by 27%, as indicated by the coefficient by that variable in the model estimations. On the other hand, distance is a pushing factor – the number of migrants decreases by 151% with a 1% increase in distance. With the highest absolute value of the coefficient, it is also the strongest factor of all included in the analysis. These findings about population density and distance are in line with the gravity model of migration assumptions and existing empirical evidence. Further on, the number of migrants increases by about 32% with a 1% increase in relative income and by about 28% with a 1% increase in total greenery areas. The unemployment rate turned insignificant in the regressions, which is in agreement with my initial guess. Most likely, migrants who move out of Warsaw's inner city still commute to work there, and thus are indifferent to the local employment situation. Both the time average of the number of kindergartens per 100,000 people and its time series turned out significant. While the coefficient of the time average is slightly positive (0.03), the coefficient of the time series is slightly negative (-0.01). This indicates that there is a positive global effect of the average number of kindergartens increasing in all boroughs in time. However, after this decomposition, the influence of the number of kindergartens per 100,000 on the number of migrants turned slightly negative, almost 0, but was still statistically significant. A probable explanation of such an ambiguous result is that migrants do not necessarily choose to place their offspring in kindergartens located near the place of residence, but rather close to their work place in Warsaw. Further on, two out of three levels of borough type were included in the analysis (urban-rural and rural). As indicated by the coefficients by these variables, the number of migrants increases about 53% if a borough is of urban-rural type, taking the urban type as a benchmark and about 50% if a borough is of rural type in reference to the urban-rural type. The presence of a suburban train station is a strong pulling factor as well – the number of migrants increases by 24% if such a station is located in a borough, illustrating the preference for better connected municipalities. Finally, the number of migrants decreases by 63% with an increase of the ratio of the average price in a borough to the average price in Warsaw by 1. This result is in conformity with the expectation that lower residential lot prices in suburbs than in the core city are a strong incentive to move out, despite the distance. As a form of robustness check, in Table 4 I report p-values of standard diagnostics tests for panel model (Woolridge, 2001). With the p-value of the Breusch Pagan Lagrange Multiplier test for individual effects being 0, I conclude that individual effects can be distinguished in the model, and thus, simple pooling would lead to incorrect inference. Furthermore, it can be seen that cross-sectional dependence is present in the model, but, as already mentioned, this problem was tackled with the use of Feasible GLS, robust to such misspecifications. The individual effects are normal, but the residuals are not. Yet, panel models are asymptotically normal (Woolridge, 2001). Finally, there is no autocorrelation in the model. Having checked the diagnostics and tackled the endogeneity, I conclude that the estimation method presented here is valid and leads to correct statistical inference.

Table 4: P-value of different diagnostic tests for the estimated model.

Hausman	**Breusch Pagan Lagrange Multiplier test for individual effects**	**Pesaran's test for cross-sectional dependence**	**Jarque-Bera test for normality of panel residuals**	**Jarque-Bera test for normality of panel individual effects**	**Woolridge test for autocorrelation in panel data**
0.29	0.00	0.00	0.00	0.27	0.27

Source of Table: own preparation

6. Conclusions

In this paper, determinants of the suburbanization of Warsaw were investigated. The aim was to find the features of boroughs which are key pulling factors for migrants. The contribution of this work is two-fold: it applies the gravity model of migration to explaining suburbanization and assesses the role of residential lot prices, having a perfect measurement of prices at its disposal.

My findings are in line with the previous research on the topic. It was confirmed that migrants choose boroughs of greater population density, higher average relative income, more total greenery spaces, and those classified as urban-rural and rural. Greater distance and residential lot prices relative to prices in Warsaw are pushing factors. Migrants are indifferent to the number of kindergartens (as their offspring likely attend kindergartens close to parents' work, rather than to their home) and the local job market, as in the majority they likely commute to work in Warsaw. Some of the regressors used in the analysis here are institutional amenities, e.g., total greenery spaces, which can be directly influenced by local planning. Thus, this work can be valuable for both Warsaw and local authorities.

Several follow-ups to this work are possible. First of all, it can be useful to include a wider range of borough features and to use methods capable of variable selection. Further on, different strategies of dealing with endogeneity can be applied to account for the residential lot prices being directly dependent on the demand for migrants. Mundlak terms were used here due to the lack of a good external instrument. However, if such an instrument can be obtained, 3SLS method could be used instead of feasible GLS with Mundlak terms. Finally, running comparative analysis of the same process in different cities in Poland and in the wider region could lead to deeper insight into the determinants of suburbanization.

7. Acknowledgements

I am thankful to Łukasz Byra, PhD, from the Faculty of Economics Sciences, University of Warsaw for supervising my Bachelor thesis, which was the starting point for my work on the topic of suburbanization. Further on, he also helped me with obtaining transactional data from Polish county offices.

References

Beine, M., Bertoli, S., & Fernandez-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration. *The World Economy*, 39(4), 496-512. https://doi.org/10.1111/twec.12265

Caves, R. W. (2004). *Encyclopaedia of the City*. Taylor and Francis. https://doi.org/10.4324/9780203484234

Downs, A., McCann, B., Mukherji, S. (2005). *Sprawl Costs: Economic Impacts of Unchecked Development.* Island Press.

Duany, A., Plater-Zyberk, E., Speck, J. (2010). *Suburban nation: the rise of sprawl and the decline of the American Dream.* Farrar, Straus and Giroux.

Fan, C. (2005). Modelling interprovincial migration in china, 1985-2000. *Eurasian Geography and Economics*, 46(3), 165-184. https://doi.org/10.2747/1538-7216.46.3.165

Jordan, S., Ross, J.P., Usowski, K.G. (1998). U.S. suburbanization in the 1980s. *Regional Science and Urban Economics.* 28(5), 611-627. https://doi.org/10.1016/S0166-0462(98)00016-7

Kahn, M. E. (2000). The environmental impact of suburbanisation. *Journal of Policy Analysis and Management.* 19(4), 569-586. https://doi.org/10.1002/1520-6688(200023)19:4<569::AID-PAM3>3.0.CO;2-P

Kok, H. (1999). Migration from the city to the countryside in Hungary and Poland. *GeoJournal*. 49(1), 53-63.

Lisowski, A. (2004). Social aspects of the suburbanisation stage in the agglomeration of Warsaw. *Dela*, 21, 531-541. https://doi.org/10.4312/dela.21.531-541

Little, R. J. A. (1988) A Test of Missing Completely at Random for Multivariate Data with Missing Values, *Journal of the American Statistical Association*, 83(404), 1198-1202. https://doi.org/10.1080/01621459.1988.10478722

Loibl, W. (2004). Simulation of suburban migration: driving forces, socio-economic characteristics, migration behaviour and resulting land-use patterns. *Vienna Yearbook of Population Research, Vienna Institute of Demography (VID) of the Austrian Academy of Sciences in Vienna,* 2(1), 203-226.

Lowry, I. S. (1966). *Migration and metropolitan growth: two analytical models.* Chandler Pub.

Mieszkowski, P., Mills, E.S. (1993). The Causes of Metropolitan Suburbanization. *Journal of Economic Perspectives*, 7(3), 135-147. https://doi.org/10.1257/jep.7.3.135

Millock, K. (2015, 11). Migration and environment. *Annual Review of Resource Economics*, 7, 35-60. https://doi.org/10.1146/annurev-resource-100814-125031

Mundlak, Y. (1978). On Pooling Time Series and Cross Section Data. *Econometrica*, 46(1), 69-85. https://doi.org/10.2307/1913646

Murray, P., Szelenyi, I. (2009). The City in the Transition to Socialism. *International Journal of Urban and Regional Research,* 8, 90-107. https://doi.org/10.1111/j.1468-2427.1984.tb00415.x

Newton, I. (1687). *Philosophiæ naturalis principia mathematica*. Benjamin Motte.

Nuissl, H., Rink, D. (2005). The "Production" of urban sprawl in Eastern Germany as a phenomenon of post-socialist transformation. *Cities*, 22(2), 123-134. https://doi.org/10.1016/j.cities.2005.01.002

Poot, J., Alimi, O., Cameron, M., & Mar ́e, D. (2016). The gravity model of migration: The successful comeback of an ageing superstar in regional science. *Investigaciones Regionales – Journal of Regional Research,* 36, 63-86.

Ravenstein, E. (1885). The laws of migration, part 1. *Journal of the Royal Statistical Society*, 48(2), 167-235. https://doi.org/10.2307/2979181

Ravenstein, E. (1889). The laws of migration, part 2. *Journal of the Royal Statistical Society*, 52(2), 241-305. https://doi.org/10.2307/2979333

Rubin, D.B., Schenker, N. (1986) Multiple Imputation for Interval Estimation from Simple Random Samples with Ignorable Nonresponse. *Journal of the American Statistical Association*, 81(394), 366-374. https://doi.org/10.1080/01621459.1986.10478280

Sturm, R., Cohen, D.A. (2004). Suburban Sprawl and Physical and Mental Health. *Public Health,* 118(7), 488-496. https://doi.org/10.1016/j.puhe.2004.02.007

Tiebout, C. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy,* 64(5), 416-424. https://doi.org/10.1086/257839

Timar, J., Varadi, M. (2001). The Uneven Development of Suburbanization during Transition in Hungary. *European Urban and Regional Studies*, 8(4), 349-360. https://doi.org/10.1177/096977640100800407

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl