



What factors contribute to uneven suburbanisation? Predicting the number of migrants from Warsaw to its suburbs with machine learning

Honorata Bogusz¹ · Szymon Winnicki² · Piotr Wójcik^{1,3}

Received: 18 May 2022 / Accepted: 5 September 2023 / Published online: 25 September 2023
© The Author(s) 2023

Abstract

This article investigates the spatially uneven migration from Warsaw to its suburban municipalities. We report a novel approach to modelling suburbanisation: several linear and nonlinear predictive models are applied, and Explainable Artificial Intelligence methods are used to interpret the shape of relationships between the dependent variable and the most important regressors. The support vector regression algorithm is found to yield the most accurate predictions of the number of migrants to the suburbs of Warsaw. In addition, we find that migrants choose wealthier and more urbanised municipalities that offer better institutional amenities and a shorter driving time to Warsaw's city centre.

JEL Classification R23 · P25 · C14 · C51 · C52

1 Introduction

Suburbanisation is a shift of population from central urban areas into suburbs, resulting in (sub)urban sprawl. As a consequence of the movement of households and businesses out of city centres, low-density, peripheral urban areas grow (Caves 2004). Most of the residents of metropolitan cities work within the central urban area, but choose to live in satellite communities called suburbs. These processes are most advanced in more economically developed countries. The United States is believed to be the first country wherein the majority of the population lives in the suburbs (Orfield 2011), rather than in the cities or rural areas. Urban sprawl, a direct consequence of suburbanisation, describes the unrestricted growth in many urban

✉ Honorata Bogusz
h.bogusz@uw.edu.pl

¹ Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland

² Warsaw, Poland

³ Data Science Lab, University of Warsaw, Warsaw, Poland

areas of housing, commercial development, and roads over large expanses of land—with little concern for urban planning. The negative impacts of urban sprawl include: an increase in residential energy consumption and land use; degradation of air quality, along with increased usage of natural resources and greenhouse gas emissions (Kahn 2000); increased infrastructure costs (Downs et al. 2005); decline of social capital; residential segregation resulting in class and racial divisions (Duany et al. 2010); growing fiscal deficit (Downs et al. 2005); and health deterioration (Sturm and Cohen 2004).

During communism, most countries in the Eastern Bloc were characterised by under-urbanisation, which meant that industrial growth occurred well in advance of urban growth and was sustained by rural–urban commuting (Murray and Szelenyi 2009). City growth, residential mobility, and land and housing development were under tight political control. Warsaw is a particular example of such circumstances—80–90% of the city’s buildings were destroyed during the 1944 uprising, and this paved the way for restricted communist city planning. Consequently, suburbanisation “in the Western sense” has been a recent phenomenon—it is believed to have begun in the post-communist countries in the 90s, following the political transformation (Nuissl and Rink 2005; Timar and Varadi 2001). In 2019, about 43% of the inhabitants of the Warsaw metropolitan area were living in the suburbs (GUS 2023a, b).

The causes of metropolitan suburbanisation have been heavily discussed in the literature and several theories have been offered (Mieszkowski and Mills 1993), mostly based on the situation in Western cities. Important papers offering insight on the suburbanisation processes in post-communist cities include: Kok (1999); Lisowski (2004); Murray and Szelenyi (2009); Nuissl and Rink (2005). The quantitative measures of suburbanisation determinants existing in the literature include works by Jordan et al. (1998); Kok (1999); and Loibl (2004). Jordan et al. (1998) identified several pulling factors of the target suburban areas in 35 US metropolises. Loibl (2004) identified the attractiveness measures of Vienna’s suburban areas. Kok (1999) used micro-level data to investigate the motives of individuals to move out of the city and to the suburbs in Budapest and Warsaw. 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, the gravity model of migration (the most well-known quantitative migration model (Poot et al. 2016) has not found many applications for local migration flows with the contributions by Bakens et al. (2018) and Simini et al. (2012) being notable exceptions.

Apart from the papers of Jordan et al. (1998), Loibl (2004), and Kok (1999), we are not aware of any study in which statistical techniques were used to explain the influx of migrants to suburban municipalities or in which the gravity model of migration framework was used in that specific context. Even though the gravity model of migration has recently gained popularity (Poot et al. 2016), it has been mainly used with respect to international migration (Beine et al. 2016; Belot and Ederveen 2012; Fan 2005) or, in regional science, intrastate and between-province migration (Pietrzak et al. 2013; Poot et al. 2016) with fewer applications to local or neighbourhood level flows (Bakens et al. 2018; Simini et al. 2012). Additionally, only a few possible pulling factors were identified in the context of suburban municipalities in the

previous studies. Finally, the possible nonlinearity of the relationships between the number of migrants and the pulling factors has yet to be addressed. Our study aims at filling these gaps. We use the gravity model of migration framework to predict the number of migrants choosing different suburban municipalities of Warsaw. On the other hand, we implement a much wider selection of possible pulling factors (25) than hitherto offered in the literature. When including so many regressors, it is reasonable to assume that the relationships between the dependent variable and some of the regressors might be nonlinear. In addition, one can expect interactions between various predictors. The wide selection of pulling factors and potentially nonlinear relationships, interactions, and collinearity call for the use of robust methods for such issues. Hence, we apply various predictive models that take into account potential nonlinearities without any prior assumptions of their shape. These models are capable of dealing efficiently with a large number of potential predictors that are also highly correlated. OLS is used as a simple benchmark, and it is also applied with an automated backward elimination of predictors. In addition, to explain the applied machine learning models and unhide the identified shape of relationships, the novel approach of Explainable Artificial Intelligence (XAI) is used.

The main aim of our paper is to accurately predict the number of migrants to the suburbs of Warsaw, given a large set of pulling factors identified in the existing literature. Second, since the machine learning methods we apply are capable of flexibly uncovering relationships between predictors and outcome (see Sect. 4. Methods), we facilitate our findings by presenting associations of different factors with suburban migration flows.

Because of a direct identification of pulling factors associated with the migration to the suburbs, this work can arguably contribute theoretical knowledge of the suburbanisation processes needed for designing local policies. Given that the suburbanisation process is fairly recent in Warsaw, adequate spatial planning can be executed in order to curb the above-mentioned negative consequences of urban sprawl. Hence, we believe that this work can be useful for both Warsaw and municipal authorities. Moreover, the example of Warsaw can contribute to understanding the suburbanisation patterns in post-communist cities.

The remaining part of this paper is structured as follows. The second section includes a review of suburbanisation theories and then offers existing empirical evidence in order to identify the possible pulling factors needed to build predictive models. The third section discusses dataset construction. In the fourth section we introduce methodological issues concerning different predictive models and XAI tools. The fifth section discusses empirical results including variable importance based on each model and the analysis of the relationships between the dependent variable and the most important predictors. A discussion then completes the study.

2 Review of the literature

According to Mieszkowski and Mills (1993), two classes of theories on suburbanisation have been tabled. The first is called “natural evolution theory” and is based on a simple chain of events. When employment is concentrated in the centre of a city,

residential development takes place from the inside out. To minimise commuting costs to the Central Business District (CBD), central areas are developed first and, as they become filled in, development moves to open lands in the suburbs. The older, centrally located units, built when average real incomes were lower, filter down to lower income groups. As more affluent households prefer to reside in outlying suburban areas, this natural functioning of the housing market leads to social stratification. Transportation costs further reinforce the tendency of the middle class to live in the suburbs. Historically, when the cost of moving goods and people within cities was high, and urban areas were dense and spatially small, high-income groups were located at the centre. Today, due to the relatively low costs of public and private transportation, this tendency has been reversed.

A second class of explanations of suburbanisation is a generalisation of the Tiebout model (Tiebout 1956) and stresses the fiscal and social problems of central cities: high taxes, low quality public schools and other government services, racial tensions, crime, and low environmental quality. These problems lead high-income central city residents to migrate to the suburbs, which leads to a further deterioration of the fiscal situation, and hence the quality of life in central areas, which then induces further out-migration. Social affiliation preferences also play a role in this vicious circle: people generally prefer to live in a group of similar income and social background. Hence, the suburbs are often homogeneous entities. Mieszkowski and Mills (1993) argue that these two theories have a number of interactions and thus, it is difficult to distinguish them empirically. Nonetheless, both theories identify factors that can contribute to the outflow of people from the city centre (costs of commuting to the CBD, average income, and institutional amenities in the suburbs) and are, hence, relevant to our study.

Several researchers have investigated the outflow to the suburbs in different metropolises quantitatively. Jordan et al. (1998) analysed the outflow in 35 American metropolises in 1980–1990, taking Brueckner's population density gradient (Brueckner 1987) as the dependent variable. They compared the same process in various cities and found that suburbanisation proceeds more quickly in areas of greater population growth rate and those where employment is concentrated in the city centre. Suburbanisation slows down with the decentralisation of the job market. Finally, more expensive rents in the suburbs contribute to a decrease of influx.

Loibl (2004) offers yet another quantitative study of suburbanisation patterns. They adapted a multi-agent system approach to simulate differing urban sprawl trends in Vienna. The simulation runs for a 30-year span were compared with the real observations. It was found 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. Loibl (2004) hypothesised that the pulling factors have to be examined in detail, because “polycentric growth dynamics seem to be dependent on regional attractiveness patterns within the suburban areas neighboring the core city”. Hence, part of the simulation design was to identify these “attractiveness patterns” and four were distinguished: landscape attractiveness (measured by the forest area quota in the neighbourhood); local services supply (access and number of attorneys); core-city availability (calculated by applying the shortest-path model to find the minimum travel time to Vienna); residential lot prices and

availability of information on lots. Loibl (2004) tested the influence of these factors on the net number of migrants by linear regression in two groups: high and low educated migrants (as proxies for income groups). Core-city accessibility, landscape attractiveness, services supply, and the population total in the previous period were identified as significant pulling factors.

An empirical study of suburbanisation in post-communist countries was made by Kok (1999), who examined its determinants in Poland (49 province cities) and Hungary (19 province cities) during the time of the political transition. They investigated the individual decisions to move out of the city of 4977 people (in the case of Poland), adapting Mulder's (1993) life-course approach. This approach assumes that a "trajectory of migration" is closely connected to some spheres of life, such as work and education. Through a logit model, Kok (1999) found out that the variables having significant impact on the probability of moving out are: taking new employment vacancy, obtaining one's own real estate, being married, being in the age groups 18–24 and 35–39 and making that decision between 1989 and 1993. These findings confirm the presumption that suburban communities are rather homogeneous and that the suburbanisation process started in Poland right after the political turnabout.

3 Dataset description

Our dataset consists of 70 observations corresponding to the municipalities defined as a part of the Warsaw Metropolitan Area according to the EU NUTS2 division.¹ The observations are for one year only and we use the latest data available, which means the years 2018–2020 depending on the variable. Migration flows are measured in 2019. All characteristics of municipalities which we could obtain from Statistics Poland (see Table 3 in Appendix) are lagged by one year and thus measured in 2018. The rest of the variables we used were scraped from open source databases (OpenStreetMap contributors 2015; Gratka [Lucky Strike] 2020—portal with real estate sales announcements; e-podróżnik.pl [e-traveller.pl] 2020—portal with local bus and train timetables) in early 2020. As historical data is not available in all of those databases, we only include one point of observation. We measure differences between municipalities, rather than in time, and hence, we assume that no serious changes in the variables' levels happened between these years. As prediction accuracy, rather than causal identification of effects is the main point of this contribution, reverse causality is not our main concern.

Our dependent variable is the number of migrants who previously lived in Warsaw and moved to one of the suburban municipalities. The source of the number of migrants is the registration of residence data provided by Statistics Poland on a municipality level. Figure 3 in Appendix shows a time series with migration flows between Warsaw and the suburbs and net migration. It can be seen that that number was consistently positive between 2002 and 2019, even though it fluctuated in time,

¹ NUTS2 region PL91 <https://ec.europa.eu/eurostat/documents/345175/7451602/2021-NUTS-2-map-PL.pdf> Access 22.02.2023.

noting a minimum of 3866 people in 2015 and maximum of 6804 people in 2007. Such a sizable flow to the suburbs, which is not compensated by the flow to the city, provides further justification for investigating the suburbanisation of Warsaw.

We predict the number of migrants using 25 regressors. The reason for taking into consideration as many as 25 features of municipalities (many of which are similar to each other) is to identify the ones that contribute to predicting the number of migrants most accurately. Their description and sources can be found in Table 3 in Appendix. Population and distance (travel time in from a municipality to Warsaw's city centre by car in minutes) are the standard explanatory variables of the gravity model of migration (see Sect. 4. Methods). Following the literature and accounting for data availability, we have chosen 23 supplementary measures. We included total income (total Personal Income Tax paid by the municipality inhabitants) as a measure of the wealth in a municipality and the number of unemployed as a measure of the job market condition. The numbers of votes for the ruling conservative party (Prawo i Sprawiedliwość [Law and Order]—PiS) and the socially and economically liberal opponent in 2019's parliamentary election (Koalicja Obywatelska [Civic Coalition]—KO) depict the conservatism/progressiveness of the community. The three types of municipalities (according to Polish administrative law, they can be classified as rural, urban–rural, or urban) are an alternative measure of urbanisation. Area is an alternative (to population) way to characterise municipality size. The total forest area captures the extent to which forest areas hinder habitable spaces.² Total greenery area, the number of available places in nurseries, the number of kindergartens, shops, tourist sites, leisure sites, sport sites, restaurants, and places of worship are measures of amenities. The presence of a suburban train station, as well as travel time from the municipality to the city centre with public transport depict the transport system's condition and the latter is included as an alternative measure of distance. The price per square metre (m²) of housing, as well as the number of parcels available for sale and their median price, measure the real estate market. Finally, we include the total area in the municipality inhabited by humans calculated following The World Settlement Footprint 2019 (Marconcini et al. 2020; DLR 2021). The raster underlying this method was built through an advanced classification system which exploits optical and radar satellite imagery with 10 m resolution and has been shown to be a measure of settlement/urbanisation which significantly outperforms all previously-existing layers (Marconcini et al. 2020). We include it to reflect the metropolitan area in a municipality. Since our model specification is based on the gravity model of migration (see Sect. 4. Methods), all characteristics of municipalities are included as absolute values rather than ratios. As population is one of the basic variables of the gravity model, the inclusion of relative income as total income divided by population, for example, would result in biased estimates. The same applies to other predictors such as number of unemployed, number of votes, etc.

² By Polish law, it is forbidden to settle in close proximity to national parks and one of the biggest Polish national parks (Kampinowski Park Narodowy) is located to the North-West of Warsaw, directly spanning the area of several municipalities.

Figure 4 in Appendix shows the visual representation of the number of migrants. Note that the coordinates used to create the presented maps come from OpenStreet-Map, which means, in fact, Polish Registry of Borders. Appendix Figures 5, 6, and 7 present the key variables of the gravity model of migration (population and distance), as well as total income, which is expected to have a large influence on the number of migrants. The spatially uneven outflow of migrants is plain at first glance. Visual inspection leads to the conclusion that the number of migrants seems to be higher in the municipalities of greater population that offer a shorter driving time to Warsaw and higher total income.

Summary statistics of all variables are attached in Table 4 in Appendix. The mean number of migrants was around 146 in 2019. There were municipalities where no migrants reported residence, while 909 people moved to the most popular one. The average population is around 19,214 people, but it differs greatly between municipalities (2742 min, 83,792 max). The average distance measured as driving time to the Warsaw city centre is 43 min. The closest municipality is 21 min away from Warsaw's city centre, while the farthest one is located over an hour away. Figure 8 (in Appendix) presents the correlation matrix. It can be seen that some predictors are highly correlated, which justifies the use of methods robust to correlation, such as accumulated local effects.

4 Methods

The gravity model of migration is one of the oldest and most popular analytical models of migration flows. According to this model, spatial flows of people depend positively on the economic size of target areas and negatively on the distance between them. In that sense, it resembles Newton's law of universal gravitation, as was first noticed by Ravenstein (1885; 1889). Poot et al. (2016) deliver a thorough description of the model and 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 constant that measures the proportion and α , β , γ are parameters to be estimated. It is customary to logarithm the above equation, in order to express it in a common, econometric framework. However, in the empirical part we consider both the original and log-transformed values of the dependent variable and quantitative predictors in various model specifications to check if the log transformation allows us to obtain more accurate predictions of the number of migrants.

While economic size is usually included as the number of people living in a specific place (Ramos 2016), different measures of distance were used in the existing literature, such as straight line distance (Lewer and van der Berg 2008),

railroad distance (Fan 2005), and road travel distance (Courchene 1970). As the suburbs of Warsaw are largely car-dependent, we assume that migrants choose the destination based on the time of driving to Warsaw with a car. Thus, we use road travel distance in minutes (between the centroid of the municipality and the metro station “Centrum” in Warsaw—the normative city centre) as our primary measure of distance. The distance from the central (bus or train) station in a municipality and the central metro station in Warsaw via public transport is included as an alternative measure of distance (See Sect. 3. Dataset Description and Table 3 in Appendix).

Moreover, the classic form of the gravity model includes population sizes in both the place of origin and destination, as well as the distance between them. In our case, we investigate the flow in one direction only: from Warsaw to its suburbs. In such a setting, the population of origin is constant for all suburban places of destination. Such a regressor has no variance and, thus, is of little use for prediction purposes. Hence, we do not include the population of Warsaw as a predictor. Our model can be thought of as a conditional gravity model—we predict the number of migrants to the suburbs assuming that the decision to move out has already been made and agents decide where to move taking the pulling factors into account. Several extended forms of the gravity model of migration have been proposed (Beine et al. 2016; Fan 2005; Lowry 1966). The choice of additional variables is context-specific.

In our setting, there are 25 regressors for 70 observations, and this results in an increased dimensionality and poses a challenge. Even though in such a setting the OLS estimator remains unbiased, high variance typically makes it perform very poorly (unless the matrix of observations is orthogonal) resulting in increased Mean Squared Error. This might be problematic in the case of our main aim—namely, using the model for accurate prediction of the number of migrants. Keeping insignificant variables in a linear regression model (or, more generally, redundant variables that do not add explainability of the modelled phenomenon) may introduce noise to predictions and make them less precise. Therefore, in the case of our linear OLS benchmark we apply a common general-to-specific approach (Campos et al. 2005; Hendry and Doornik 2015) based on p-values and, alternatively, on Akaike Information Criterion (AIC).

In the case of prediction accuracy as the main aim, machine learning techniques are an attractive alternative. Their potential benefits compared to statistical models also include the possibility of application in high dimension data, even if the number of predictors exceeds the number of observations. In addition, they require less work and knowledge from the statistician to design and implement, e.g. automated variable transformations with kernel functions, built-in mechanisms of variable selection (so-called regularisation), easily taking into account interactions between variables in tree-based models. However, in general the power of machine learning algorithms in recognising patterns is proportional to the size of the dataset (Kokol et al. 2022). The fewer observations for model building (training), the more perfectly models can fit our data, which is a well-known problem of overfitting (Hastie et al. 2009). Common ways of avoiding this issue, especially on small samples, relate to using relatively less complex models, with less parameters or smaller depth (for tree-based algorithms) and decreasing the impact of less important features with regularisation.

As to date no machine learning algorithms other than OLS have been applied to predict the number of people migrating from the city to the suburbs, it is not known which method yields the best predictions in this framework. As a consequence, we try a variety of methods representing various estimation approaches. To limit the risk of overfitting, we apply mainly parametric (e.g. LASSO, ridge regression) or quasi-parametric (support vector regression) models and in the case of tree-based models we restrict the maximum depth of trees (random forest and extreme gradient boosting) and use regularisation (extreme gradient boosting).

LASSO (Least Absolute Shrinkage and Selector Operator) (Tibshirani 1996) is one of several regularisation methods. It can be viewed as an extension of the ordinary least squares (OLS) model. It differs from OLS because of its cost function—it not only minimises the sum of squared residuals, but also takes into account the sum of absolute values of the parameters of the linear model (excluding the intercept) as an additional constraint. Adding such a penalty in the optimisation formula results in searching for parameters that fit the data well, but additionally are as small as possible. Parameters of less important variables will shrink towards zero, some of them will even be set to be equal to zero. At the expense of a certain bias (LASSO estimates are biased), this model often allows us to obtain more precise forecasts on the test sample (Hofmarcher et al. 2015).

What is crucial from our perspective is that LASSO can be considered a variable selection method. It can be applied even when the initial number of variables exceeds the number of observations. No a priori assumptions or selection of a subset of variables are needed. One has only to determine the optimal weight for the additional constraint, and this can be done via the validation. Ridge regression (Hastie et al. 2009) is very similar to LASSO—it differs only in the formulation of the additional constraint which takes into account the sum of the squared values of parameters (commonly called L2 regularisation). Parameters at less important variables asymptotically shrink towards zero, but in the case of ridge regression reach only zero if the weight of an additional constraint is infinite. Neither for LASSO, nor for ridge regression, are standard errors of coefficients and statistical significance calculated.

Both LASSO and ridge regression still assume a linear relationship between the dependent variable (number of migrants in our case) and its predictors. While the foregoing might be true for the standard independent variables of the gravity model of migration (population size and distance; in case of the log-transformed equation), there is no reason to expect this in terms of the additional regressors, such as measures of income or amenities. If relationships between the dependent variable and the regressors fail to be linear, a linear specification is inappropriate and may lead to incorrect inference. As the shape of the relationship is not known in advance, we use other machine learning tools that can flexibly adjust to data.

Support vector regression or SVR (Vapnik 1995) is similar to OLS and fits a hyperplane that is positioned as close to all data points as possible. However, while OLS minimises the sum of squared errors, SVR tries to fit the errors within a specified distance from the hyperplane (Smola and Schölkopf 2004). Moreover, the setup includes additional regularisation hyperparameter C , which controls how much one wants to avoid misclassifying each observation. The most important advantage of

SVR over OLS is the ability to model nonlinear relationships between variables using selected kernel functions. With them SVR applies an implicit nonlinear mapping of the matrix of predictors into a higher dimensional feature space (including nonlinear transformations of original variables and their interactions—depending on the kernel function), where it is more probable to find an appropriate hyperplane (Vapnik 1995). Thus, one can think of SVR as a process of performing a linear regression in a transformed and higher-dimensional feature space. Two widely used types of kernels are radial basis function and polynomial kernel. We apply both kernels in the empirical part of the article.

In addition, we use two common types of ensemble machine learning algorithms, combining many simple models with weak performance each (so-called weak-learners, usually decision or regression trees) into one strong model. Tree-based models allow for nonlinearities and easily take into account interactions. Random forest (RF) (Breiman 2001) is a representative of the bootstrap averaging approach (abbreviated as bagging) and consists of estimating multiple independent tree models, each trained on a different bootstrap sample of the original dataset. In addition, at each split of each tree, only a random subset of all predictors is considered. The prediction from a random forest is obtained by averaging the forecasts from individual trees. Random forests are robust to the problem of multicollinearity and can be applied to a large number of potential predictors without initial selection. In addition, they are indifferent to nonlinear interlinkages between the data. They require the tuning of two hyperparameters—the number of trees and the number of predictors considered at each split. Extreme gradient boosting (XGB) is an example of the boosting approach, which is also usually based on tree models. It builds the model in an iterative fashion at each step trying to improve the previous model by giving higher weights to observations that were not fitting well during the previous step. In addition to capturing nonlinear relationships, XGB is also capable of performing regularisation, for example by shrinkage such as in the LASSO or ridge regression.

Each of these models has various hyperparameters, the values of which have to be assumed before the optimisation starts (e.g. the number of trees in the random forest). Their optimal values can be found with the use of cross-validation (Hastie et al. 2009). To eliminate randomness from the model validation process, we use the leave-one-out cross-validation (LOOCV) procedure. In this procedure each model is estimated n (number of observations) times on the sample without the 1st, 2nd, 3rd, etc., observation, respectively. The single observation left aside is used as a validation sample (out-of-sample data) to generate prediction and assess its accuracy. Based on all n individual predictions obtained in this way, various performance metrics are reported (validation sample). In addition, we also calculate the same metrics for fitted values generated from models estimated on the total sample of 70 observations (training sample). The performance on the training sample might be used to conclude how well the particular algorithm fits to the data and explains the relationships between the number of migrants and its predictors, while performance on the validation sample shows its predictive performance on new data. To test the sensitivity of our findings with respect to the choice of validation method, we also report performance metrics calculated using the alternative tenfold cross-validation in Appendix (Table 5) in which the data is randomly divided into 10 equal-sized

parts (folds) and then in subsequent steps one of the folds is left aside as the validation sample (not used in model building, but for the assessment of predictive accuracy on new data).

On the one hand, machine learning models can flexibly adjust to the data and they are usually capable of detecting the underlying functional relationship between the variables (Hastie et al. 2009). On the other hand, the classic equation of the gravity model of migration is usually expressed in a logarithmic form before fitting it with an OLS. There is no state-of-the-art solution which would indicate if a functional transformation should be made before estimating the gravity model with nonlinear machine learning methods. As a result, we try two approaches: an equation where all variables except for dummies are log-transformed ($\log(x+1)$ due to 0 values in some variables including the dependent) and the second one without prior functional transformations. Predicted values from the log-transformed estimation are inversely transformed ($\exp(x)-1$) before calculating model performance metrics for ease of comparability with the models on non-transformed variables.

We compare the performance of all the algorithms by the common benchmarks of mean absolute error (MAE) and R2, as well as the symmetric mean absolute percentage error (SMAPE), which is a modified version of MAPE that allows for zero values (Flores 1986). In addition, we employ two error measures commonly used in the local migration literature (Cameron and Cochrane 2017; Wilson 2015): weighted (by population in municipality) symmetric mean absolute percentage error (WSMAPE) and symmetric median algebraic percentage error (SMedALPE). We assess model performance primarily by those measures. SMedALPE used as the median error reduces the impact of extreme outliers. On the other hand, WSMAPE is preferable to other measures (such as MAPE) when there is a wide range of population sizes (such as in our case with nearly three thousand people living in the least populated municipality and over 80 thousand living in the most populated one). We report all performance metrics both with regard to the training sample, as well as the validation sample.

While the outcome of linear algorithms is fairly easy to explain, the interpretability of nonlinear models poses a challenge. These structures are usually called “black box models” as the shape of the relationship between variables cannot be easily derived from them in the functional form that allows for interpretation. Statistical models allow us to fit a specific probability model of a defined form and usually require a set of assumptions, such as normal distribution of the variables. On the other hand, machine learning methods find patterns in data with a minimal set of assumptions.

Even though finding the algorithm with the best prediction accuracy is our main goal, exploring associations between the migration flow and the pulling factors, as well as their importance in the achieved prediction, can arguably lead to a better understanding of the suburbanisation process. This can be achieved with Explainable Artificial Intelligence (XAI), which is a group of (usually graphical) methods that allow us to better understand how the algorithms work and based on that explain their decisions and predictions. Multiple methods on a global (whole sample) and local (a single observation) level have been proposed and a review of them can be found in Molnar (2019) or Barredo et al. (2020). Here, we focus on a brief

description of the two methods we use: permutation feature importance (PFI) and accumulated local effects (ALE). In our case, we are interested only in interpretability on the global level, not in the performance of the model with respect to individual municipalities.

Permutation-based variable importance was first introduced by Breiman (2001) in a random forest algorithm. Further extension was done by Fisher et al. (2019), who proposed a model agnostic tool for calculating the contribution of individual features into prediction accuracy. Importance is calculated by randomly permuting the values of a particular variable, running a model on a changed dataset, and computing an increase in prediction error with the newly created learning sample as compared to the original data. The permutation is applied several times and average increase in the loss function, which quantifies the goodness-of-fit, is reported as the measure of variable importance. Therefore, permutation feature importance allows one to rank the used regressors in terms of their contribution to prediction accuracy. In the empirical part of this article, we measure the importance of features as the average increase in MAE after permutations.

Furthermore, the most commonly used black box visualisation tool is the partial dependence plot (PDP) introduced by Friedman (2001). It depicts the marginal effect of an input variable and the model outcome (*ceteris paribus*) and it is a graphical representation of predictions. For a given variable, PDP averages model predictions while keeping the feature values constant. However, this method assumes no correlation between predictors, as averaging incorporates the dependence between two features. As we showed in Sect. 3, that assumption is unrealistic in our case. Apley and Zhu (2019) proposed an extension of PDP, called accumulated local effects (ALE) that takes the correlation bias into account. ALE calculates the average predicted outcome with respect to the predictor value with slight modifications. For a given predictor, ALE calculates average changes in prediction for observations in close neighbourhood to the original one. The graphical representation is then analogous to the PDP. In this way, we plot the relationships of the most important predictors (as indicated by permutation feature importance) on the outcome variable for each of the models using ALE plots.

All calculations and visualisations presented in this paper were prepared by the authors in R software.

5 Results

We start the estimation by finding the optimal values of hyperparameters with the use of leave-one-out cross-validation (LOOCV) and grid search approach. The optimal values of hyperparameters are provided in Appendix (Table 2), both for the log-transformed and -reversed specification and the specification without prior functional assumptions. The following algorithms are considered: OLS, OLS with general-to-specific (gets) based on p-values and alternatively on AIC, LASSO, ridge regression, support vector regression (SVR) with a polynomial and radial basis kernel, random forest (RF), and Extreme Gradient Boosting (XGB). We compare the predictive performance of all models using validation sample and interpret the

Table 1 Performance of models based on leave-one-out cross-validation (LOOCV)

Variables transformation	Model	Train					Validation				
		R2	SMAPE	WMAPE	SMedALPE	MAE	R2	SMAPE	WMAPE	SMedALPE	MAE
None	OLS	0.87	0.88	0.22	0.00	47.46	0.57	1.04	0.45	-0.01	90.90
	OLS gets with AIC	0.86	0.88	0.23	0.01	49.07	0.73	0.97	0.34	0.00	66.48
	OLS gets with <i>p</i> -value	0.84	0.84	0.24	-0.01	48.32	0.73	0.90	0.33	-0.02	60.86
	LASSO	0.79	0.73	0.31	0.03	53.46	0.65	0.77	0.42	0.04	67.40
	Ridge	0.82	0.73	0.30	0.13	53.00	0.69	0.79	0.41	0.13	67.93
	SVR poly	0.80	0.62	0.26	0.00	46.67	0.68	0.71	0.45	-0.01	66.17
	SVR radial	0.78	0.65	0.35	-0.02	53.29	0.69	0.72	0.46	-0.03	66.69
	Random forest	0.93	0.46	0.24	0.15	33.30	0.50	0.72	0.59	0.30	77.45
	XGB	0.83	0.60	0.32	0.09	43.34	0.51	0.86	0.55	0.08	77.67
	OLS	0.74	0.56	0.46	-0.02	66.08	0.58	0.77	0.68	-0.04	109.12
Log reversed	OLS gets with AIC	0.65	0.58	0.54	-0.09	77.06	0.59	0.65	0.66	-0.10	95.20
	OLS gets with <i>p</i> -value	0.66	0.58	0.51	-0.07	73.04	0.59	0.65	0.61	-0.08	88.53
	LASSO	0.74	0.59	0.38	-0.13	58.02	0.66	0.70	0.46	-0.15	72.36
	Ridge	0.73	0.60	0.39	-0.12	60.14	0.65	0.70	0.48	-0.18	74.54
	SVR poly	0.80	0.55	0.31	0.16	48.85	0.70	0.64	0.43	0.15	60.05
	SVR radial	0.80	0.53	0.31	-0.08	48.22	0.70	0.61	0.39	-0.08	59.95
	Random forest	0.87	0.45	0.32	-0.01	37.63	0.55	0.70	0.62	0.02	70.92
	XGB	0.99	0.17	0.05	0.00	5.27	0.56	0.70	0.55	0.07	67.81

outcome of the most accurate one. The summary of model performance measures is presented in Table 1 for both the training and validation sample. The validation metrics summarise predictive performance and are obtained from LOOCV, while the training sample metrics are based on the fitted values from a particular algorithm applied on the whole sample with optimal values of hyperparameters. For a robustness check, performance metrics calculated using the alternative tenfold cross-validation are presented in Table 5 in Appendix. There are no differences in general conclusions using the two methods.

One can clearly see that tree-based machine learning algorithms (random forest and XGB), even with a limited depth, outperform all other approaches in fit to the training data. For models based on non-transformed data, random forest explains 93% of the variability in the number of migrants (R^2) and has the lowest SMAPE. However, when weighted by population (WSMAPE), random forest shows no advantage over simple OLS or OLS with any general-to-specific approach used. What is more, random forest systematically overestimates the number of migrants (positive SMedALPE), while OLS does (almost) a perfect job. Although machine learning models are believed to be invariant to data transformations, applying the log transformation to quantitative predictors and the dependent variable (and then reversing the transformation of the dependent for the purpose of assessing model performance) seems to generally improve their train sample results. Performance metrics are also more consistent for nonlinear algorithms (both SVRs and a random forest), with the clear exception of XGB which seems to suffer from overfitting (R^2 equal to 99% and all other metrics much lower than for all other approaches). This is somehow expected, when a complex algorithm is applied on a small sample. Therefore, based on the results on the training sample, one would conclude that simple OLS modelling is enough to explain the factors of suburban migration. However, related to the aim of this paper, we should rather focus on the results in the validation sample, which summarises the predictive performance of all algorithms. Surprisingly, also here the OLS models applied on non-transformed data perform better than the other approaches once insignificant or irrelevant variables are removed by the general-to-specific method. They achieve R^2 of predictions on the level of 73% and WSMAPE of 33–34%. The conclusions change in a log-transformed setting, common for gravity-based models. In this case the best predictive performance is obtained by both SVRs independently of the metric analysed. The tree-based algorithms fail in the task of prediction, which confirmed that their application to small samples, even after limiting their complexity, might lead to overfitting.

These findings confirm that OLS is a valid model for predicting local level migration, providing that the problem of redundant variables is addressed. However, the majority of the predictors were dropped in OLS with the general-to-specific approach (see Table 6 in Appendix). This is not an issue in SVR, which was found to yield the accurate prediction regardless of the presence of redundant features. Thus, we use SVR algorithm results to identify features that are most relevant with respect to the prediction of the number of migrants with the use of permutation feature importance (PFI).

Figure 1 presents PFI for SVR with a radial basis kernel while Figure 9 in Appendix shows the same for SVR with a polynomial kernel. The top seven identified

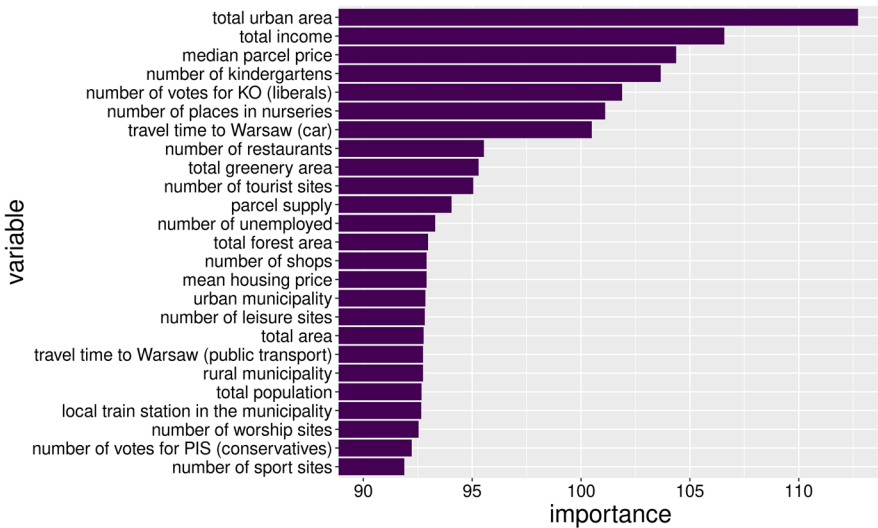


Fig. 1 Permutation feature importance based on the support vector regression model with a radial basis kernel. Note: importance measured as average increase in MAE after permutations

features are identical in both algorithms: total urban area, total income, median price of a parcel, number of kindergartens, number of votes for the liberal KO party, number of available places in nurseries, and time of driving to the Warsaw city centre. Further, accumulated local effects plots (Fig. 2) based on both SVR approaches allow us to interpret shapes of relationships between those variables and the number of migrants. All of those seven features are positively associated with the number of migrants, except for the time of driving to the city centre. Machine learning methods do not allow for a causal interpretation of those findings. Nonetheless, those associations are useful to understand patterns of suburbanisation in Warsaw. They confirm that migrants choose urbanised municipalities of good institutional amenities (kindergartens and nurseries), those located at a shorter driving distance to Warsaw, and that social affiliations play a role in suburban settlement (income depicting the wealth of a municipality, support for the socially and economically liberal party KO). The somewhat striking positive association between the number of migrants and the median parcel price is most likely a manifestation of social status as well, as the suburbs of Warsaw, especially the municipalities neighbouring Warsaw, are typified by rather affluent populations. Last, but not least, the ALE plots for the polynomial kernel in each case show simple nonlinear patterns with the slope of the relationship increasing after reaching some threshold value of the particular predictor. This is especially true in the cases of income, votes for KO, and number of kindergartens, where the ALE curves become steeper after a certain threshold, indicating a more positive relationship of those variables with the number of migrants. In our view, this increase might be driven by the wealthiest municipalities in our sample (e.g. Lesznowola, Konstancin-Jeziorna, Łomianki) which pull affluent populations and are densely populated.

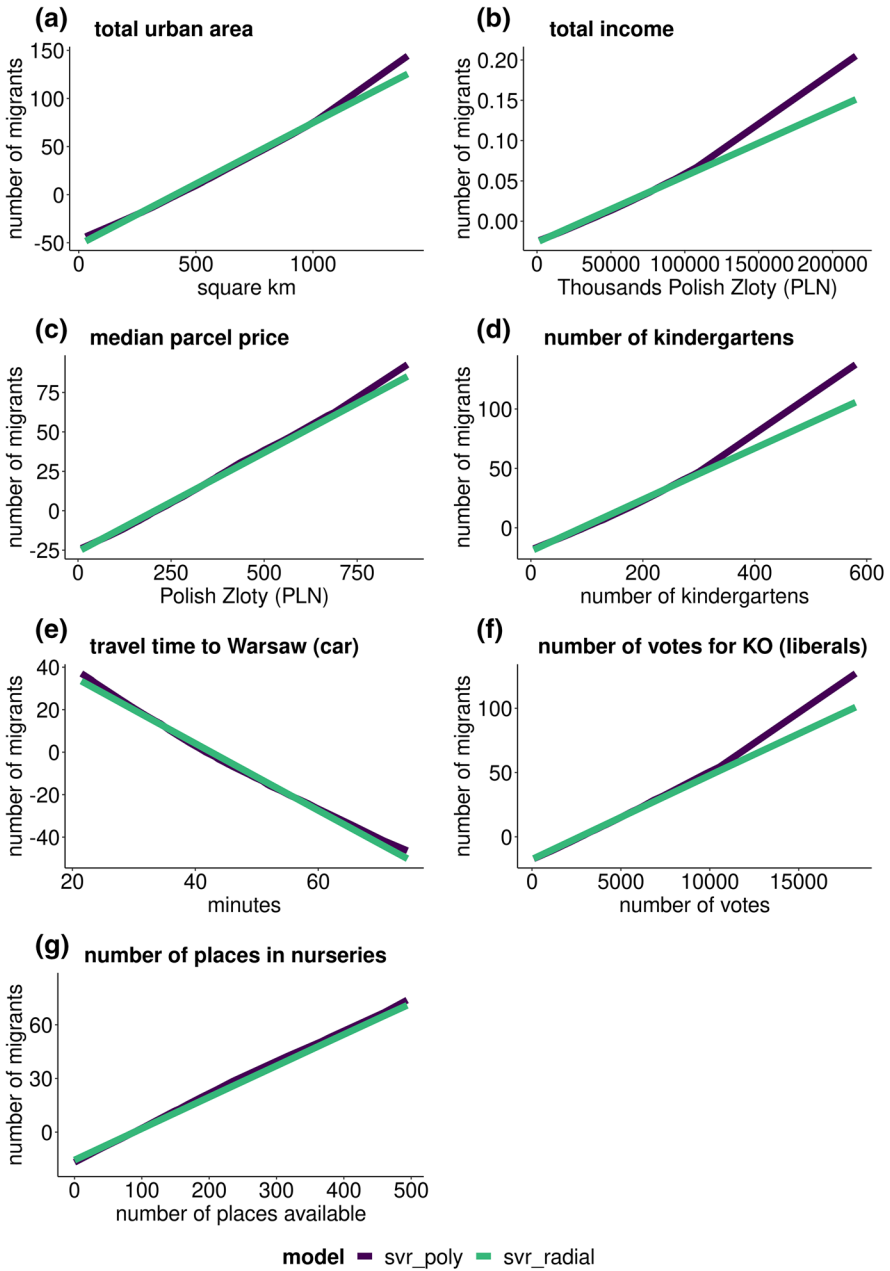


Fig. 2 Accumulated local effects plots—top seven predictors indicated by permutation feature importance for support vector regression

6 Summary and conclusions

In this paper, we investigated the phenomenon of suburbanisation in the agglomeration of Warsaw. Our primary goal was to accurately predict the number of migrants to the suburbs of Warsaw. Based on the extended gravity model of migration, we applied several predictive models and assessed their performance. Support vector regression turned out to yield the most accurate predictions in the case of the log-transformed approach of the gravity model. OLS was found as the next-best alternative, which provides justification for its wide implementation in local migration studies. Permutation-based feature importance was calculated for each chosen feature for SVR, and accumulated local effects were plotted for the top seven most important variables, allowing us to investigate associations between different pulling factors and the migration flow. We identified six pulling factors: residential area, total income, median parcel price, number of kindergartens, number of votes for the liberal KO party, and number of available places in nurseries. While income and the number of votes are likely proxies for social affiliation preferences, kindergartens and nurseries represent institutional amenities. Our findings with respect to those measures can be useful in terms of spatial planning, as measures such as the number of places in nurseries can be directly influenced by local governments.

Several limitations of our analysis need to be addressed. First, machine learning models do not allow for a causal interpretation, and hence, we cannot identify the effects of the pulling factors, but only shed light on associations using the XAI methods. Second, our analysis encompassed a wide range of possible pulling factors, most of them arguably understudied in the context of local urban migration in the previous literature (e.g. institutional daycare, political affiliations, or different kinds of amenities). Some of those variables were scraped from open source databases, such as OpenStreetMap (OpenStreetMap contributors 2015), where historical data are unavailable, and thus, we opted for including more underexplored predictors, rather than a time series. However, this strategy limits the validity of the findings for other time periods, especially exceptional ones like the Covid19 pandemic, when migration to cities slowed down in Poland (GUS 2023a, b). In addition, this approach may also introduce the issue of ‘bad controls’ (Angrist and Pischke 2009)—some variables are measured after the variable on interest (migration) and there might be reversed causality (for example between the migration and the number of shops). However, we believe that omitting potentially important predictors would also introduce a bias in the results. Finally, as an analysis of local within-metropolis migration requires sizable data collection efforts, our analysis was conducted for one city only. Since our findings may be context-specific, we cannot advertise the external validity of the obtained results with certainty. We flag those unsolved issues as important areas for further research on suburban migration.

Appendix

See Tables 2, 3, 4, 5 and 6.

See Figs. 3, 4, 5, 6, 7, 8 and 9.

Table 2 Optimal values of hyperparameters found by one-leave-out cross-validation

Model	Parameter	Without log-transformation	With log-transformation
Ridge regression	Alpha	0	0
	Lambda	54.78901	0.4552935
Lasso	Alpha	1	1
	Lambda	8.119845	0.05672426
Random forest	Number of trees	100	100
	Max number of variables considered by each split	6	11
	Min size of a node	5	10
Support vector regression with radial kernel	Sigma	1e-04	1e-04
	Cost	100	100
Support vector regression with polynomial kernel	Degree	2	2
	Scale	0.01	0.01
	Cost	1	0.5
Extreme gradient boosting	Number of rounds	100	200
	Max depth	3	5
	Eta	0.1	0.1
	Gamma	0	0
	Column sample by tree	0.5	0.7
	Min child weight	20	10
	Subsample	1	1

Table 3 Dataset description – the features of municipalities

Variable	Description	Source
Number of migrants	Number of migrants ^a	Statistics Poland
Total population	Total population (count)	Statistics Poland
Total income	Total revenue to a municipality from the Personal Income Tax, expressed in Polish Zlotys (PLN)	Statistics Poland
Number of unemployed	Number of unemployed people (registered job-seekers)	Statistics Poland
Number of votes for PIS (conservatives)	Number of votes obtained by PIS in 2019 parliamentary election	National Electoral Commission
Number of votes for KO (liberals)	Number of votes obtained by KO in 2019 parliamentary election	National Electoral Commission
Urban municipality	Municipality type—urban (dummy) ^b	Statistics Poland
Urban–rural municipality	Municipality type—urban–rural (dummy)	Statistics Poland
Rural municipality	Municipality type—rural (dummy)	Statistics Poland
Total area	Total area in ha	Statistics Poland
Total forest area	Total forest area in ha	Statistics Poland
Total greenery area	Total area of green amenities (parks, etc.) in ha	Statistics Poland
Number of kindergartens	Number of kindergartens	Statistics Poland
Number of places in nurseries	Number of places available in nurseries	Statistics Poland
Number of shops	Count of places tagged 'shop'	OpenStreetMap
Number of tourist sites	Count of places tagged 'tourist'	OpenStreetMap
Number of leisure sites	Count of places tagged 'leisure'	OpenStreetMap
Number of sport sites	Count of places tagged 'sport'	OpenStreetMap
Number of restaurants	Count of places tagged 'restaurant'	OpenStreetMap
Number of worship sites	Count of places tagged 'worship'	OpenStreetMap
Local train station in the municipality	Presence of a suburban train station (dummy)	Google Maps
Travel time to Warsaw (car)	Driving distance from a municipality to the city centre in min	Google Maps
Travel time to Warsaw (public transport)	Travel time from a municipality to the city centre with public transport in min ^c	e-podroznik.pl

Table 3 (continued)

Variable	Description	Source
Mean housing price	Price of square metre of housing in Polish Zlotys (PLN) ^d	Statistics Poland
Parcel supply	Number of parcels available for sell in a municipality, on 03.08.2020 ^e	gratka.pl
Median parcel price	Median price of m ² of a parcel, on 03.08.2020	gratka.pl
Total urban area	Total urbanised (residential) area in square kilometres	The World Settlement Footprint 2019

^aThe dependent variable: number of migrants – people who moved out of Warsaw and reported residence in one of the suburban municipalities

^bThree municipality types can be distinguished according to the classification of Polish administrative law: urban, urban–rural, and rural. Urban–rural is excluded from estimations due to perfect collinearity with the other two categories

^ce-podroznik.pl is the largest website collecting travel possibilities between places in Poland. It has information about even the smallest carriers at its disposal and was therefore chosen as the most thorough database. No information was available for 3/70 municipalities on this platform. However, public means of transport are functioning in those municipalities, even though carriers are not registered on any platform. In those cases, we called the carriers directly and asked about the travel time

^dOnly a price on a county level was available in the Statistics Poland and hence this is what we included

^egratka.pl is the largest online real estate platform and was therefore chosen as the most adequate database

Table 4 Basic summary statistics for all variables used in the analysis

Variable	Mean	Std	Min	Median	Max	Scale
Number of migrants	147.5	182.3	0	75.5	909	Number of people
Total population	19,214.8	15,549.9	2742	14,600.5	83,792	Number of people
Total income	33,274.5	34,104.5	1533.1	22,190.6	215,717.9	Thousands of Polish Zlotys (PLN)
Number of unemployed	330.5	314.5	23	210	1424	Number of people
Number of votes for PIS (conservatives)	4167.3	2865.1	747	3328.5	13,788	Number of votes
Number of votes for KO (liberals)	2803.8	2987.3	95	1889	18,197	Number of votes
Urban municipality	0.2	0.4	0	0	1	Dummy
Rural municipality	0.5	0.5	0	1	1	Dummy
Total area	8001.8	4129.6	576	8509	20,577	Dummy
Total forest area	1870.1	1608.4	0.5	1763.3	7884.7	ha
Total greenery area	55.7	63.7	2.3	24.3	286	ha
Number of kindergartens	95	98.4	5	65	580	ha
Number of places in nurseries	91.7	109.4	0	62	494	Number of places
Number of shops	140.4	170.5	1	77	781	Number of places
Number of tourist sites	53.3	153.2	0	27	1278	Number of places
Number of leisure sites	660.2	866	8	401.5	5761	Number of places
Number of sport sites	103.4	113.1	0	54.5	469	Number of places
Number of restaurants	11	13.5	0	6	64	Number of places
Number of worship sites	4.9	3.4	1	4	19	Number of places
Local train station in the municipality	0.5	0.5	0	1	1	Number of places
Travel time to Warsaw (car)	43	11.9	21.4	42.2	74.5	Dummy
Travel time to Warsaw (public transport)	46.3	22	16	43.5	105	Minutes
Mean housing price	4712.6	484.6	3889	4586	5717	Polish Zlotys (PLN)
Parcel supply	46.7	50.1	1	30	247	Number of parcels
Median parcel price	213	184.3	9	149.2	885	Polish Zlotys (PLN)
Total urban area	421.1	267.2	29.7	371.4	1404.3	Square km

Table 5 Performance of models based on tenfold cross-validation (CV10)

Variables transformation	Model	Train					Validation				
		R2	SMAPE	WSMAPE	SMedALPE	MAE	R2	SMAPE	WSMAPE	SMedALPE	MAE
None	OLS	0.87	0.88	0.22	0.00	47.46	0.60	1.04	0.43	0.06	88.87
	OLS gets with AIC	0.86	0.88	0.23	0.01	49.07	0.76	0.95	0.31	-0.02	62.44
	OLS gets with p-value	0.84	0.84	0.24	-0.01	48.32	0.75	0.87	0.33	-0.04	59.60
	LASSO	0.84	0.74	0.25	0.02	47.65	0.70	0.85	0.35	0.03	64.27
	Ridge	0.82	0.73	0.30	0.13	53.00	0.69	0.84	0.40	0.03	67.64
	SVR poly	0.97	0.58	0.10	0.07	22.03	0.75	0.79	0.31	0.05	59.78
	SVR radial	0.78	0.65	0.35	-0.02	53.29	0.69	0.77	0.45	-0.08	67.12
	Random forest	0.92	0.45	0.26	0.20	33.37	0.57	0.71	0.55	0.31	73.56
	XGB	0.83	0.65	0.32	0.14	44.03	0.56	0.83	0.51	0.17	74.59
	OLS	0.74	0.56	0.46	-0.02	66.08	0.53	0.80	0.67	0.09	125.21
Log reversed	OLS gets with AIC	0.65	0.58	0.54	-0.09	77.06	0.59	0.64	0.62	-0.09	90.55
	OLS gets with p-value	0.66	0.58	0.51	-0.07	73.04	0.59	0.65	0.57	-0.08	85.60
	LASSO	0.75	0.61	0.34	-0.11	56.31	0.68	0.69	0.41	-0.08	67.37
	Ridge	0.73	0.62	0.38	-0.14	60.17	0.66	0.70	0.43	-0.19	71.02
	SVR poly	0.80	0.55	0.31	0.16	48.85	0.68	0.64	0.45	0.12	61.91
	SVR radial	0.80	0.53	0.31	-0.08	48.22	0.65	0.64	0.44	-0.06	65.02
	Random forest	0.88	0.47	0.36	-0.02	43.05	0.58	0.68	0.61	-0.08	68.97
	XGB	0.83	0.41	0.30	-0.01	37.03	0.46	0.73	0.60	-0.07	79.59

Table 6 Coefficients of OLS models based on log-transformed variables

Model	Variable	Estimate	Std. error	VIF
All variables	(Intercept)	- 9.42	8.8	
	Total population	- 0.4	0.87	48.1
	Total income	1.07	0.75	71.6
	Number of unemployed	0.68*	0.38	11.4
	Number of votes for PIS (conservatives)	- 1.07	0.71	23
	Number of votes for KO (liberals)	0.3	0.62	56.9
	Urban municipality	1.19*	0.68	8.09
	Rural municipality	0.35	0.32	2.81
	Total area	1.26**	0.46	14.1
	Total forest area	- 0.05	0.11	3.86
	Total greenery area	0.06	0.21	6.88
	Number of kindergartens	0.22	0.3	11.4
	Number of places in nurseries	0.01	0.09	4.02
	Number of shops	- 0.07	0.19	7.68
	Number of tourist sites	- 0.33***	0.11	2.64
	Number of leisure sites	0.04	0.12	2.83
	Number of sport sites	- 0.01	0.18	7.04
	Number of restaurants	0.3	0.21	5.56
	Number of worship sites	- 0.65*	0.34	2.96
	Local train station in the municipality	0.02	0.31	2.7
	Travel time to Warsaw (car)	- 1.76**	0.73	4.48
	Travel time to Warsaw (public transport)	- 0.49	0.55	7.23
	Mean housing price	0	0	2
	Parcel supply	0.06	0.16	3.37
	Median parcel price	0	0	4.07
	Total urban area	0.11	0.34	7.13
Gets based on AIC	(Intercept)	- 14.30**	5.27	
	Total income	1.65***	0.24	9.06
	Number of unemployed	0.69**	0.27	6.86
	Number of votes for PIS (conservatives)	- 1.49***	0.46	11.8
	Urban municipality	1.08**	0.47	4.72
	Total area	1.13***	0.28	6.22
	Number of tourist sites	- 0.36***	0.08	1.75
	Number of restaurants	0.24*	0.14	2.99
	Number of worship sites	- 0.61**	0.27	2.31
	Travel time to Warsaw (car)	- 1.94***	0.53	2.91
Travel time to Warsaw (public transport)	- 0.62*	0.31	2.9	

Table 6 (continued)

Model	Variable	Estimate	Std. error	VIF
Gets based on p -value	(Intercept)	- 16.61***	5.18	
	Total income	1.81***	0.23	7.72
	Number of unemployed	0.67**	0.27	6.84
	Number of votes for pis (conservatives)	- 1.52***	0.47	11.8
	Urban municipality	1.32**	0.45	4.29
	Total area	1.23***	0.28	5.93
	Number of tourist sites	- 0.35***	0.08	1.74
	Number of worship sites	- 0.56*	0.27	2.28
	Travel time to Warsaw (car)	- 1.97***	0.54	2.91
	Travel time to Warsaw (public transport)	- 0.74**	0.31	2.76

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

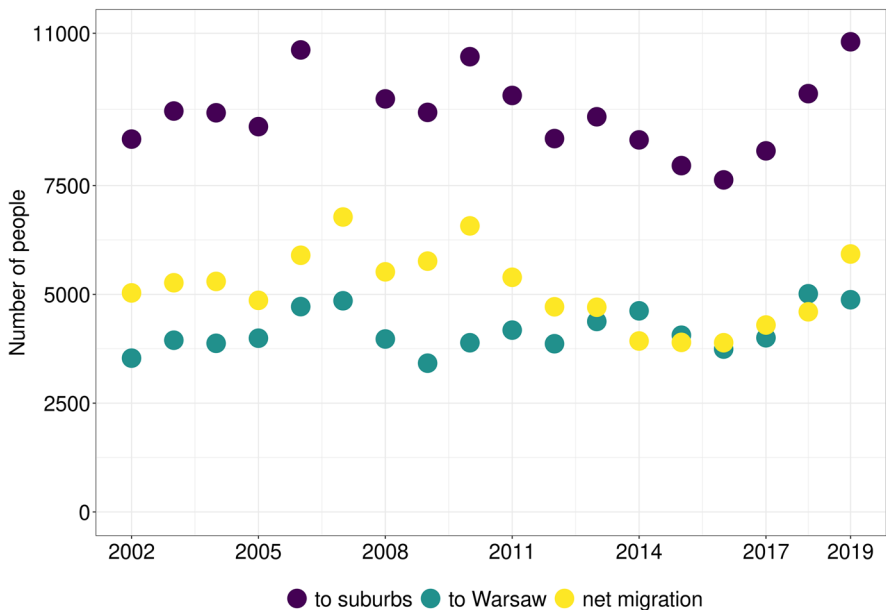


Fig. 3 Migration flows between Warsaw and the suburbs in time. Net migration represents the number of migrants from Warsaw to the suburbs net of migration in the other direction. Source of data: GUS (2023a, b)

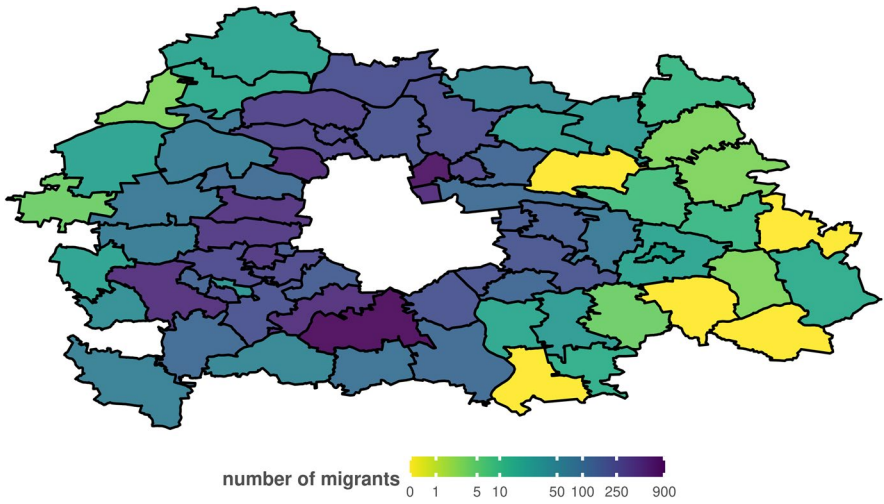


Fig. 4 Migration flows from Warsaw to suburban municipalities

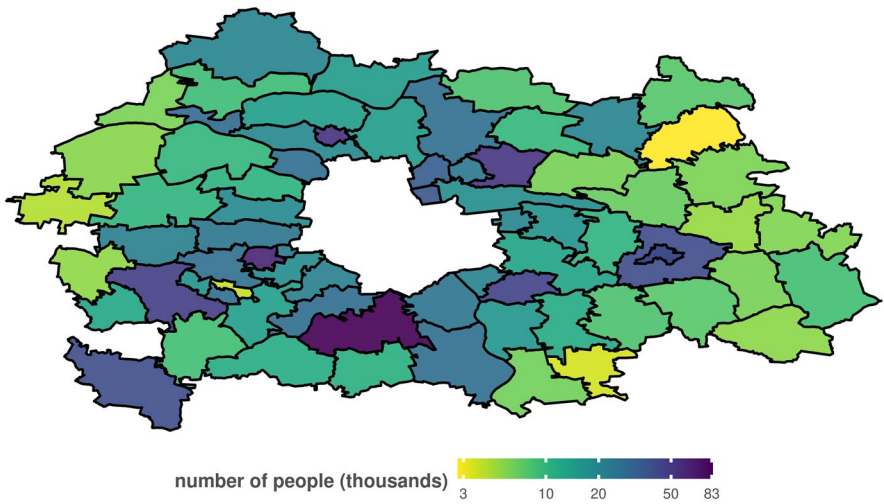


Fig. 5 Population of suburban municipalities

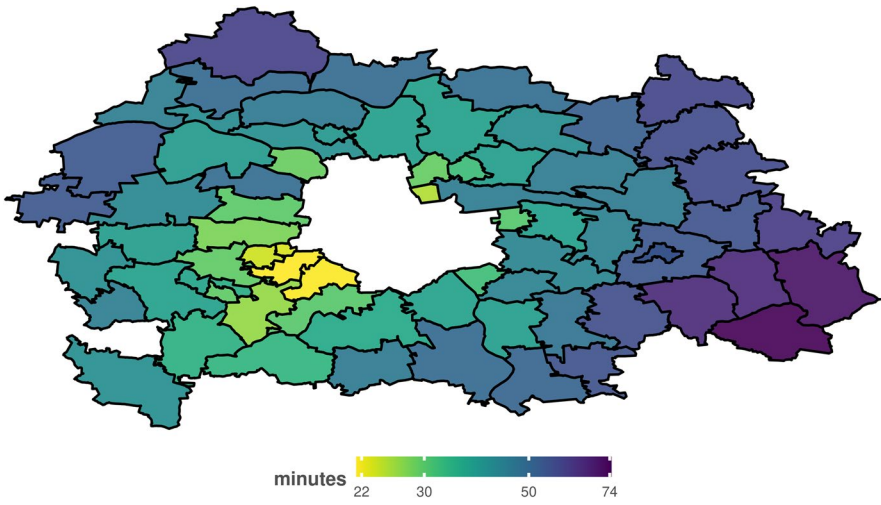


Fig. 6 Time of driving from the municipality to Warsaw city centre

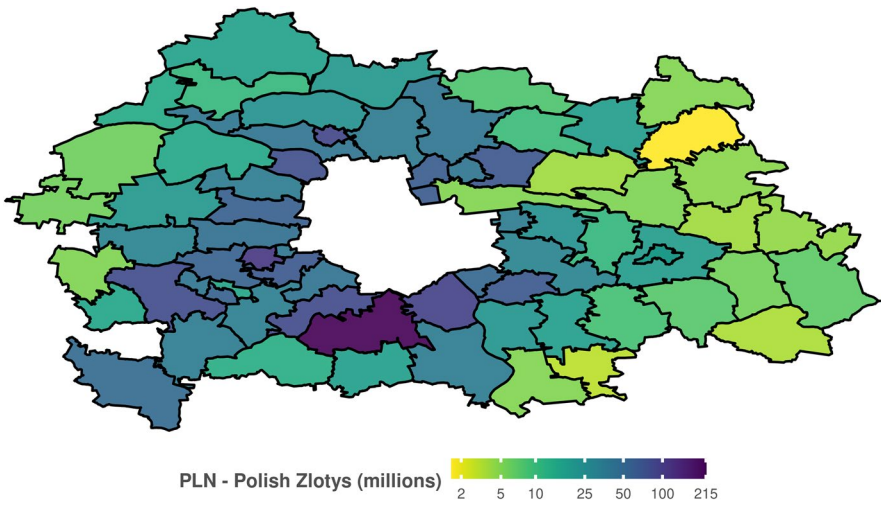


Fig. 7 Total income to the municipality from Personal Income Tax

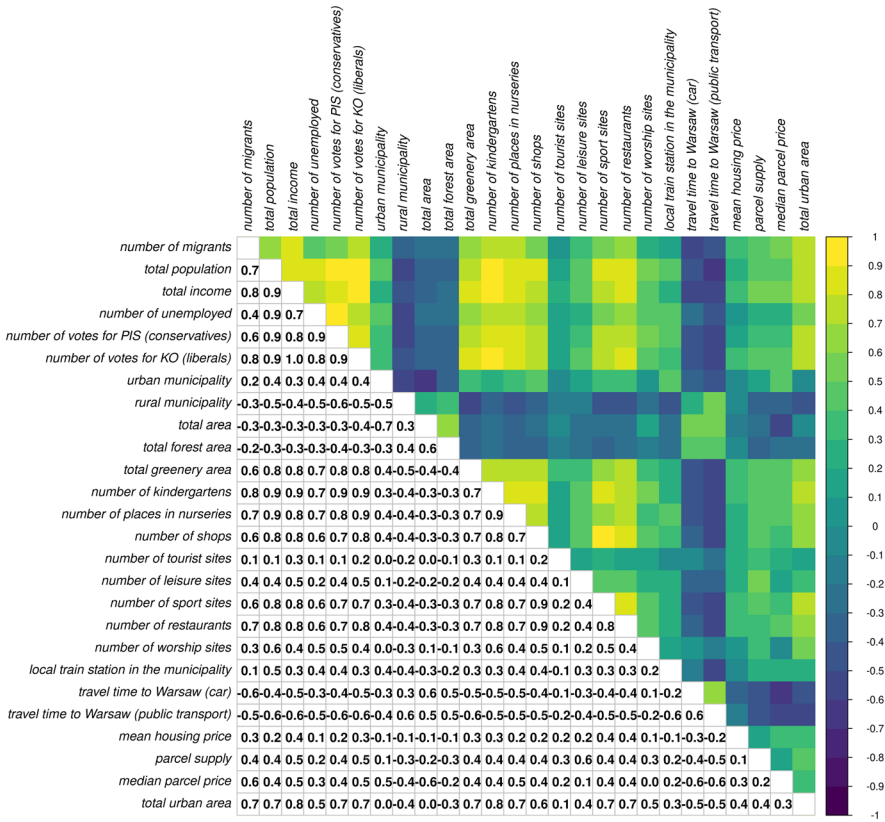


Fig. 8 Correlation plot

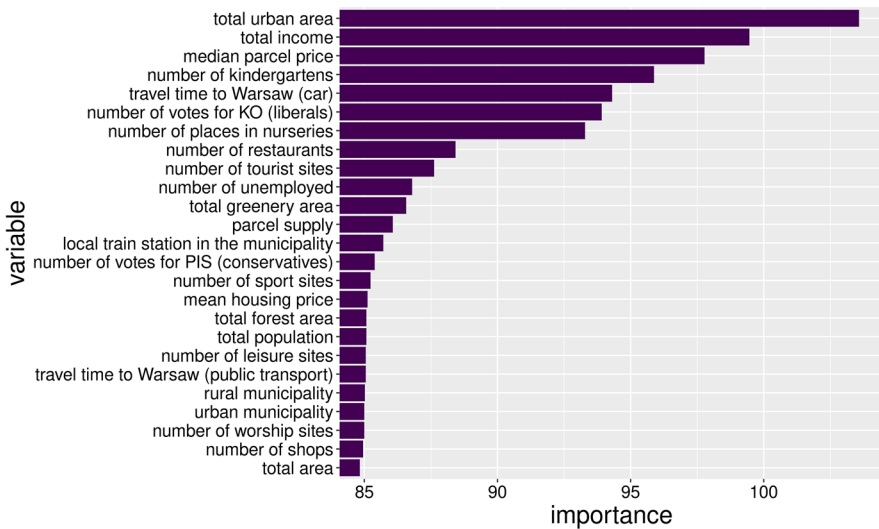


Fig. 9 Permutation feature importance based on the support vector regression model with a polynomial kernel. Note: importance measured as average increase in MAE after permutations

Acknowledgment This work was supported by the National Science Centre in Poland under Grant number 2016/21/B/HS4/00670.

Data availability The codes and datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. Map data copyrighted OpenStreetMap contributors and they are available from <https://www.openstreetmap.org>.

Declarations

Conflict of interest All authors declare that there is no conflict of interest, financial or other.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Angrist JD, Pischke JS (2009) Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton. <https://doi.org/10.2307/j.ctvc4j72>
- Apley D, Zhu J (2019) Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Ser B* 82:1059–1086. <https://doi.org/10.1111/rssb.12377>

- Bakens J, Florax RJ, Mulder P (2018) Ethnic drift and white flight: a gravity model of neighborhood formation. *J Reg Sci* 58:921–948. <https://doi.org/10.1111/jors.12390>
- BarredoArrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 58:82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Beine M, Bertoli S, Fernández-Huertas Moraga J (2016) A practitioners' guide to gravity models of international migration. *World Econ* 39:496–512. <https://doi.org/10.1111/twec.12265>
- Belot M, Ederveen S (2012) Cultural barriers in migration between OECD countries. *J Popul Econ* 25:1077–1105. <https://doi.org/10.1007/s00148-011-0356-x>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Brueckner J (1987) The structure of urban equilibria: a unified treatment of the Muth-Mills model. In: Mills ES (ed) *Handbook of regional and urban economics*. Elsevier, Amsterdam, pp 821–845
- Cameron MP, Cochrane W (2017) Using land-use modelling to statistically downscale population projections to small areas. *Australas J Reg Stud* 23:195–216
- Campos J, Ericsson NR, Hendry DF (2005) General-to-specific modeling: an overview and selected bibliography. FRB Int Finance Discuss Paper. <https://doi.org/10.2139/ssrn.791684>
- Caves R (2004) *Encyclopaedia of the city*. Routledge, London. <https://doi.org/10.4324/9780203484234>
- Courchene TJ (1970) Interprovincial migration and economic adjustment. *Can J Econ* 3:550–576. <https://doi.org/10.2307/133598>
- DLR (2021) World settlement footprint (WSF) 2019-Sentinel-1/2-Global <https://download.geoservice.dlr.de/WSF2019/#details> Accessed 14.02.2023
- Downs A, McCann B, Mukherji S (2005) *Sprawl costs: economic impacts of unchecked development*. Island Press, Washington
- Duany A, Plater-Zyberk E, Speck J (2010) *Suburban nation: the rise of sprawl and the decline of the American dream*. Macmillan, New York
- e-podróżnik.pl [e-traveller.pl] (2020) e-podróżnik.pl [e-traveller.pl] <https://e-podróżnik.pl>. Accessed 01.07.2020
- Fan CC (2005) Modelling interprovincial migration in China, 1985–2000. *Eurasian Geogr Econ* 46:165–184. <https://doi.org/10.2747/1538-7216.46.3.165>
- Fisher A, Rudin C, Dominici F (2019) All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res* 20:1–81
- Flores B (1986) A pragmatic view of accuracy measurement in forecasting. *Omega* 14:93–98. [https://doi.org/10.1016/0305-0483\(86\)90013-7](https://doi.org/10.1016/0305-0483(86)90013-7)
- Friedman J (2001) Greedy function approximation: A gradient boosting machine. *Ann Stat* 29:1189–1232
- Gratka [Lucky Strike] (2020) Gratka [Lucky Strike] <https://gratka.pl/> Accessed 01.07.2020
- GUS [Statistics Poland] (2023a) Bank Danych Lokalnych [Local Data Bank] <https://bdl.stat.gov.pl/bdl/dane/podgrup/temat> Accessed 14.02.2023a
- GUS [Statistics Poland] (2023b) Baza Demografia [Demographic Database] <https://demografia.stat.gov.pl/BazaDemografia/StartIntro.aspx> Accessed 14.02.2023b
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York
- Hendry DF, Doornik JA (2015) *General-to-specific Modeling. Empirical model discovery and theory evaluation: automatic selection methods in econometrics*. MIT Press, Cambridge.
- Hofmarcher P, Crespo Cuaresma J, Grün B, Hornik K (2015) Last night a shrinkage saved my life: economic growth, model uncertainty and correlated regressors. *J Forecast* 34:133–144. <https://doi.org/10.1002/for.2328>
- Jordan S, Ross J, Usowski K (1998) U.S. suburbanization in the 1980s. *Reg Sci Urban Econ* 28:611–627. [https://doi.org/10.1016/S0166-0462\(98\)00016-7](https://doi.org/10.1016/S0166-0462(98)00016-7)
- Kahn ME (2000) The environmental impact of suburbanization. *J Policy Anal Manage* 19:569–586. [https://doi.org/10.1002/1520-6688\(200023\)19:4%3c569::AID-PAM3%3e3.0.CO;2-P](https://doi.org/10.1002/1520-6688(200023)19:4%3c569::AID-PAM3%3e3.0.CO;2-P)
- Kok H (1999) Migration from the city to the countryside in Hungary and Poland. *GeoJournal* 49:53–62
- Kokol P, Kokol M, Zagoranski S (2022) Machine learning on small size samples: a synthetic knowledge synthesis. *Sci Prog*. <https://doi.org/10.1177/00368504211029777>
- Lewer JJ, Van den Berg H (2008) A gravity model of immigration. *Econ Lett* 99:164–167. <https://doi.org/10.1016/j.econlet.2007.06.019>

- 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>
- Loibl W (2004) Simulation of suburban migration: driving forces, socio-economic characteristics, migration behaviour and resulting land-use patterns. *Vienna Yearbook Popul Res* 2:203–226
- Lowry IS (1966) Migration and metropolitan growth: two analytical models. Chandler Pub
- Marconcini M, Metz-Marconcini A, Üreyen S et al (2020) Outlining where humans live, the World Settlement Footprint 2015. *Sci Data* 7:242. <https://doi.org/10.1038/s41597-020-00580-5>
- Mieszkowski P, Mills E (1993) The causes of metropolitan suburbanization. *J Econ Perspect* 7:135–147. <https://doi.org/10.1257/jep.7.3.135>
- Molnar C (2019) Interpretable machine learning. a guide for making black box models explainable. <https://christophm.github.io/interpretable-ml-book/>
- Mulder C (1993) Migration dynamics: a life course approach. Dissertation, University of Amsterdam
- Murray P, Szelenyi I (2009) The city in the transition to socialism. *Int J Urban Reg Res* 8:90–107. <https://doi.org/10.1111/j.1468-2427.1984.tb00415.x>
- Nuissl H, Rink D (2005) The “production” of urban sprawl in eastern Germany as a phenomenon of post-socialist transformation. *Cities* 22:123–134. <https://doi.org/10.1016/j.cities.2005.01.002>
- OpenStreetMap contributors. (2015) Planet dump [Data file from 01.07.2020]. Retrieved from <https://planet.openstreetmap.org>
- Orfield M (2011) American metropolitics: the new suburban reality. Brookings Institution Press, Washington
- Pietrzak M, Wilk J, Matusik S (2013) Gravity model as the tool for internal migration analysis in Poland in 2004–2010. Institute of Economic Research Working Papers 28/2013
- Poot J, Alimi O, Cameron M, Mare D (2016) The gravity model of migration: the successful comeback of an ageing superstar in regional science. *Investig Reg* 36:63–86
- Ramos R (2016) Gravity models: a tool for migration analysis. *IZA World Labor*. <https://doi.org/10.15185/izawol.239>
- Ravenstein E (1885) The laws of migration, part 1. *J R Stat Soc* 48:167–235. <https://doi.org/10.2307/2979181>
- Ravenstein E (1889) The laws of migration, part 2. *J R Stat Soc* 52:241–305. <https://doi.org/10.2307/2979333>
- Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. *Nature* 484:96–100. <https://doi.org/10.1038/nature10856>
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14:199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- Sturm R, Cohen D (2004) Suburban sprawl and physical and mental health. *Public Health* 118:488–496. <https://doi.org/10.1016/j.puhe.2004.02.007>
- Tibshirani R (1996) Regression shrinkage and selection via the Lasso summary. *J R Stat Soc Ser B (Methodol)* 58(1):267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Tiebout C (1956) A pure theory of local expenditures. *J Polit Econ* 64:416–424. <https://doi.org/10.1086/257839>
- Timar J, Varadi M (2001) The uneven development of suburbanization during transition in Hungary. *Eur Urban Reg Stud* 8:349–360. <https://doi.org/10.1177/096977640100800407>
- Vapnik VN (1995) The nature of statistical learning theory. Springer-Verlag, Berlin
- Wilson T (2015) New evaluations of simple models for small area population forecasts. *Popul Space Place* 21:335–353. <https://doi.org/10.1002/psp.1847>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.