UNIVERSITY OF HOHENHEIM

Spatial Data Analysis in Economics

by

Johannes Jasny

A Dissertation submitted to obtain the Doctoral Degree of Economic Sciences (Dr. oec)

Faculty of Business, Economics and Social Sciences

University of Hohenheim

Stuttgart, January 2020

This thesis has been accepted in June 2019 by the Faculty of Business, Economics and Social Sciences at the University of Hohenheim to attain the doctoral degree of economic sciences (Dr. oec).

Date of defence: 17. December 2019

Prof. Dr. Karsten Hadwich, Dean, University of Hohenheim

Prof. Dr. Alfonso Sousa-Poza, Supervisor, Institute for Household and Consumer Economics,
University of Hohenheim

Prof. Dr. Tilman Becker, Supervisor, Institute for Agricultural Markets and Marketing,
University of Hohenheim

Prof. Dr. Andreas Pyka, Dean's Representative, University of Hohenheim

Acknowledgements

First and foremost I want to thank Professor Tilman Becker for his valuable advice and supervision of my thesis. I am lucky to have an open-minded, interested and enthusiastic supervisor like him. I consider the work climate at his chair as very positive. Especially our fruitful discussions, his openness and creativity when it comes to new ideas, and his overall support were a blessing for me. Professor Becker ensured a working atmosphere where I could development of own ideas and allowed me to grow, personally and professionally. I would also like to thank my supervisor Professor Alfonso Sousa-Poza. The regular doctoral seminars and discussions with him were always very pleasant and helped me to refine my research.

I also like to thank all other professionals, who helped me to sharpen my work. First and foremost, I would like to thank Dr. Stergios Xouridas with whom I spend several lunches discussing ideas and simply enjoying each other. Furthermore I like to thank all the people who contributed to my work: Tanja Strohäker, Professor Michael Schenk, Professor Robert Jung, Dr. Fabian Wahl, and Mandy Badermann. I also like to thank my colleagues from the field of spatial econometrics and the Spatial Econometric Society, especially Professor Paulo Guimaraes and Professor James LeSage, who laid the groundwork for my work.

A special thanks to the staff at our faculty, especially our secretary Luitgard Abele who supported me in every possible way. I also thank Alexander Jelinek, Xuenan Li, Gudrun Mayer, Esin Vidin, Dr. Beate Gebhardt, Andrea Wöhr, and Marius Wuketich.

Finally and most importantly I would like to thank my family and friends, especially my mother Alla, my wife Elisa, and my son Aljoscha for your love and support.

Table of Content

List of Figures

List of Tables

1 Introduction

> *"I don't think the human race will survive the next thousand years, unless we spread into space"*
>
> *Prof. Stephen Hawking*

In 2008, the Nobel Prize in Economics was awarded to Prof. Paul Krugman for "*[…] his analysis of trade patterns and location of economic activity"*.[1] The fact that the price was awarded in the field of economic geography, a subdivision of geography with a focus on economic activity, highlights the importance of space in economic research. Today, eleven years after Krugman's award, the field of spatial data analysis has reached its maturity and became a widely used tool for economists. A review of the Econlit database for the last ten years, i.e. 2009 to 2019, shows that the number of published articles on spatial data analysis in the context of economics grew steadily. The literature review also identifies the most popular subjects that are concerned with spatial data analysis in economics. There is a total number of 644 published articles in academic journals on spatial data analysis, and in particular on spatial econometrics. The three most common subjects are studies on the size and spatial distribution of regional economic activity (113 articles), economic development, i.e. urban, rural, and regional analysis, housing and infrastructure (112 articles), regional economic activity such as growth, development, and environmental issues (110 articles).

Arbia (2011) conducts an extensive literature review on 237 papers concerning "spatial econometrics". He finds that the number of applications and fields expanded in the period from 2007 to 2011. He also identifies the most common application of spatial econometric techniques in fields. Anselin (2010) points out that the increasing number of published papers, the general acceptance of spatial statistics & spatial econometrics, and the increasing variety of applications prove that these techniques have reached their maturity and obtained general acceptance among other economic methodologies. This trend is also visible in educational text books, such as Baltagi (2013), who has implemented an extensive chapter on spatial data analysis in the latest edition of textbook on econometric analysis.

Why is there a rising interest in spatial data analysis among social scientists and why is it useful to consider geography in economic analysis? There are several answers to that question, however, data availability,

[1] See: https://www.nobelprize.org/prizes/economic-sciences/2008/summary/

rising interest in data visualisation, spatial pattern recognition, and spatial interactions as well as improved estimation techniques are among the most popular motivations. Over the last ten years, georeferenced economic and demographic data has become more available at finer aggregated levels. For instance, economic data, such as GDP or tax income, became increasingly available on regional aggregation levels such as districts and municipalities. Although this type of data had been available previously, the quality and sample sizes of the datasets have improved which is a necessary condition for meaningful spatial data analysis (Fischer and Getis, 2010). Obviously, data availability is a driving force for analytical tools such as spatial statistics, spatial econometrics, and data visualisation through maps. Especially maps have been used for centuries in order to simplify the world and understand geographical phenomena. Certainly this motivation is also true for economic issues.

Data availability increased the interaction between economists, geographers, and regional scientists who initiated the development of spatial econometrics. A less romantic but very important motivation are improvements in estimation techniques which arise from spatial data analysis. For example, studying the spatial distribution of economic activity in the context of classic growth theory illustrates why spatial techniques are necessary to obtain consistent coefficient estimates. Figure 1.1 shows the regional distribution of absolute GDP growth in German districts. It illustrates that economic activity is not completely random, but clusters in certain regions. For example, regions with high GDP growth, i.e. “hot spots” are mostly observed in North Rhine-Westphalia, Baden-Württemberg, and partly at the border to the Czech Republic. What leads to the geographical variation and clustering of GDP growth? Although the clustering of GDP can be partially explained by observable determinants from growth theory such as capital or labour supply, it cannot be fully explained by those determinants. In other words, even when these factors are implemented in the model, GDP clustering may still sustain and requires other techniques to obtain estimate consistency. One major motivation for spatial data analysis techniques is the omitted variable bias caused by existing spatial structures, which affect the values of the observations. Hence, it is essential for an applied economist who is studying the determinants of GDP, to consider the underlying spatial structure of the GDP distribution. In other words, excluding spatial structures from a georeferenced GDP growth regression will cause biased estimates (LeSage and Pace, 2009). Another motivation for spatial techniques are unobserved variables. For instance, innovation often results from regional knowledge transfers, which can be hard to quantify. Company founders in Silicon Valley benefit to a large extent from the regional knowledge transfer which makes the region particularly innovative. Regional knowledge transfer can be understood as unobserved positive spatial externalities which are not directly measurable by an independent variable.

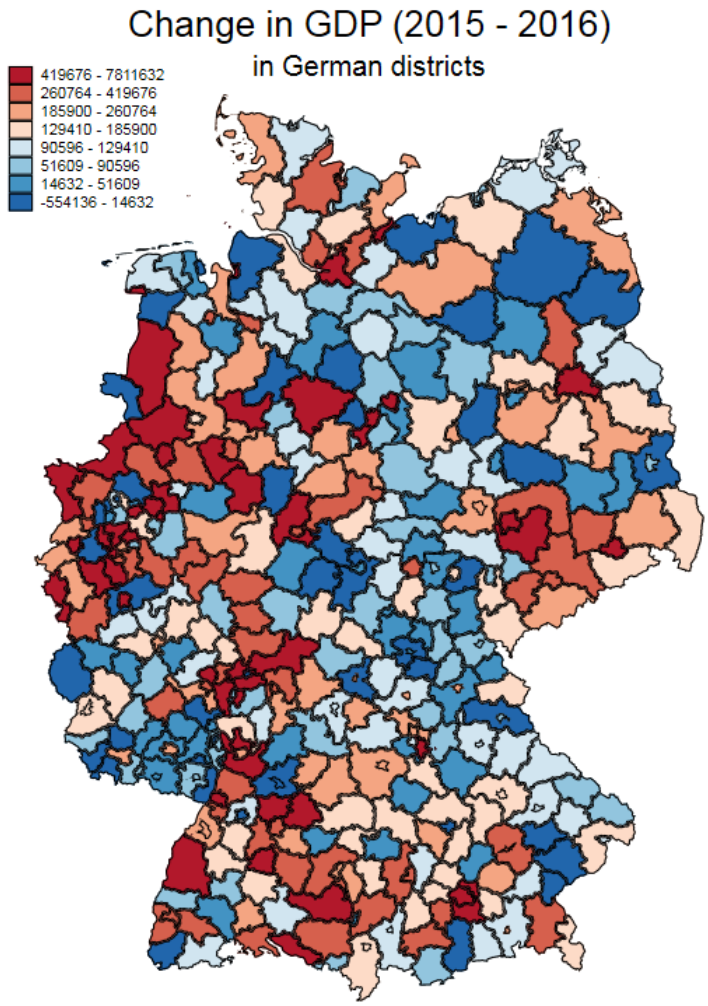

Figure 1.1: Spatial distribution of economic activity in terms of regional GDP differences

However, they can be quantified indirectly through spatial interaction terms. In other words, modelling the spatial dependence among regions with many innovative companies can be interpreted as technology transfer among those regions.

The purpose of this work is to study spatial data analysis techniques and apply those techniques on social and economic issues. There are three examples of applied spatial data analysis in this work. Article one and two study the determinants of local supply differences in the market for election gambling machines. Article three studies the influence of immigration on local voting behaviour. This work is structured as follows, the rest of chapter one contains a general introduction on spatial data analysis techniques. Chapter two studies the influence of the socioeconomic milieu on the supply of electronic gambling machines (EGM). In chapter three, we study the supply determinants and spill over effects in the EGM market. Chapter four studies the influence of immigration and other social factors on local voting behaviour in Germany. Finally, chapter five contains the discussion and some concluding remarks.

1.1 Classification of Spatial Data Analysis

Fischer and Getis (2010) provide a useful overview about different schools of thought in spatial data analysis. They distinguish between three categories of spatial data analysis, namely, (1) exploratory spatial data analysis, (2) spatial statistics, and (3) spatial econometrics.

Exploratory spatial data analysis (ESDA) has its origins in the seventies and represents an extension of Tukey-type exploratory data analysis (EDA) with georeferenced data (Tukey, 1977). Tukey suggests that exploratory data analysis is a useful tool to get a sense for patterns in the data which can be used to develop research ideas and testable hypotheses. Typical instruments of non-spatial EDA are scatterplots as a preliminary stage to correlation analysis, boxplots for outlier identification, bar charts and histograms for distribution and dispersion analysis. In contrast to EDA, the purpose of ESDA is to isolate patterns and features in spatial data and illustrate those to the analyst. Simply speaking, ESDA adds a "spatial flavour" to traditional data analysis and introduces some new tools. For instance, ESDA uses maps to display data, transforms point data into surface data, aggregates smaller surface data to larger surfaces (e.g. assign counties to the corresponding state), and illustrates changes over time in attributes of these surfaces. One key strength of ESDA is data visualisation, mostly through maps, which can motivate and facilitate an understanding of otherwise complex issues. Mapping georeferenced data can make difficult topics more accessible, illustrate spatial distributions, and make statistics overall more vibrant. A large share of ESDA is used prior to spatial statistics and spatial econometrics. Therefore, ESDA can be considered as a preparatory step to modelling. Fischer and Getis (2010) argue that ESDA and EDA represent a recent trend in research methodology. The traditional six-steps of hypothesis testing – (1) problem, (2) hypothesis, (3) sampling distribution, (4) testing, (5) results, and (6) discussion have been extended by a seventh step, namely data exploration. However, they argue that data exploration is not a step that is squeezed in between the other steps. Instead, the data exploration should accompany the six-steps at nearly all stages of the scientific process.

Spatial statistics has its roots in Pearson and Fisher. Their analysis was extended to a spatial framework by Whittle, Moran, Geary, Cliff and Ord. Especially Whittle (1954), Moran (1950), and Geary (1954) were pioneers in the development of autocorrelation analysis with georeferenced data. Spatial statistics is mostly concerned with problems regarding spatial entities such as points, lines, areas, or a combination of these three. Most importantly, these spatial units usually represent real world phenomena, which makes this science interesting for applied economists. For instance, point data can be used to model and map single spatial entities such as houses or firm locations. Lines are used to model the presence of roads or borders. Whereas areas can either be "real" areas, i.e. areas which are physically present and distinctive, such as forests and cities. Alternatively, areas can also have an administrative character, which means they are areas by definition rather than physical criteria (e.g. municipalities, states, countries, or common trade zones). Typical research questions in spatial statistics include conjectures about the spatial distribution of geographical entities, pattern in the spatial distribution such as clustering or dispersion, differences in the patterns, spatial relationships, and most importantly, the interaction between the defined points, lines, and areas over time and space. Besides the interaction among spatial entities, the interaction among *attributes* of the spatial entities is at the core of spatial statistics. Consider two spatial entities, Germany and France, which have GDP as a common attribute. Spatial statistics can be used to map the GDP growth at the border of these two countries and help studying growth externalities among both entities. Two of the mostly used tools from spatial statistics are spatial dependence measures and spatial weight matrices Drukker et al. (2013a, 2013b). The former measures the similarity or dissimilarity among spatial entities while the latter describes how the entities are connected with each other (e.g. neighbours). The most common spatial dependence measures are Moran's I and Geary's C statistics (Moran, 1950; Geary, 1954). Both measures are hypothesis tests for spatial randomness, which is an essential prerequisite for spatial econometric models. Spatial matrices will be discussed in section 1.2 and spatial dependence measures in section 1.3.

Spatial Econometrics is a rather new field in the history of scientific thought. The term was introduced in 1974 by Jean Paelinck at the annual meeting of the Dutch Statistical Association (Arbia, 2016). Compared to more general econometrics, spatial econometrics can be considered a rather new field. With the rise of computer technology in the eighties, followed by theories on economic geography in the nineties, and the availability of spatial big data in the first years of the new millennium, the interest in spatial econometric techniques rose steadily. Throughout the Post-Millennium years, the discipline's development had gained momentum and established in the mainstream sciences (Arbia, 2011). Traditional spatial econometric theory was formulated in the late seventies and manifests in two textbooks by Paelinck and Klaassen (1979) and Anselin (1988). Both books are still considered the "bibles" for spatial econometricians and establish some of the most commonly used tools in the field such as the spatial lag model (SLM) and the spatial error model (SEM). Both models are similar to linear regression models with the main distinction being the presence of spatial dependence. For example, the SLM model assumes a non-randomness in the distribution of the dependent variable while the SEM model assumes the same for the error term. Both models incorporate dependencies among spatial entities in different ways. In addition to the “bibles”, the book by LeSage and Pace (2009) is considered as one of the most important extension to Paclinck, Klaassen and Anselin. LeSage and Pace provide an overview on models, which incorporate more than one spatial interaction. They are proponents of the Spatial Durbin Model (SDM), which incorporates a spatial structure in the dependent variable and the independent variables simultaneously. Spatial econometrics builds upon ESDA and spatial statistics, which makes it sometimes difficult to separate the fields. For instance, ESDA is useful to visualise economic activity and detect patterns, spatial statistics helps quantifying those patterns, and spatial econometrics is ultimately used for more distinct analytics.

When it comes to applied spatial econometrics, there is vast number of literature, which incorporate those techniques into different computer programs. Fischer and Getis (2010) show a variety of applications in different programs, such as ArcGIS, SAS, R, GeoDa, Python, and others. LeSage and Pace (2009) present some MatLab procedures, Anselin and Rey (2014) provide an extensive overview on their own open-source software GeoDa, and Arbia (2014) illustrates the main procedures in R. In this work, I mostly work with the procedures by Drukker et al. (2013a, 2013b), who incorporate spatial econometric techniques in STATA. The aim of this work is to explore these techniques in economic applications, however, I will put a higher emphasis on spatial econometrics. Spatial econometric models will be discussed in section 1.3 and 1.4.

1.2 Spatial Weight Matrices

One aim of spatial data analysis is the exploration of global and local patterns. A crucial aspect to study these patterns is the definition of relevant “neighbourhood”. Fischer and Wang (2011) define neighbours as spatial units which interact with each other in a meaningful way. This definition leaves some room for interpretation, because “meaningful interaction” depends on the research field itself. For example, real estate economists are often interested in spatial point data such as living houses, hospitals, factories, etc. They usually want to study the influence of local amenities such as schools, stores, train stations, and hospitals on rents and house prices. One common definition of neighbourhood for real estate economists is distance. Distance itself can be defined in two ways - first in terms of the Euclidean distance, and second, in terms of other abstract distance units (e.g. travel time or transaction costs). For instance, residents could consider a school in range of 800 metres radius as their neighbourhood because their children are more likely to visit it than another school which is five kilometres away.

Alternatively, a more abstract distance measure such as travel time could also be a reasonable definition. For example, if the school lying further away can be reached within the same time due to better public transportation infrastructure, it can also be considered as neighbourhood. On the other hand, economists, and in particular spatial econometricians, are often interested in local GDP or local unemployment variation. Whether a single household is producing more output might be not too important for an economist, but the output growth of a certain state is of key interest. As a consequence, spatial econometricians often work with administrative units, i.e. spatial area data, and require other definitions of neighbourhood. One typical question of interest for economists are spillover-effects in output growth across states. In other words: how does growth in one state affect the growth in neighbouring state? The distance definition might not be too helpful for this kind of question because states differ in shape and size which causes some inaccuracy. The "common border" definition, i.e. areas that share a common border are neighbours, proves more helpful. There are also other fields which abstract from physical distance measures entirely. For example, network analysis uses neighbourhood structures in terms of "connections" between users. Social networks, such as Facebook, consider two users as neighbours if they are connected with each other, i.e. they are "friends" within the system.

The aforementioned examples can be summarised by three types of neighbourhood definitions. That is (1) common border or contiguity, (2) the (inverse) distance definition and (3) the k-nearest neighbours definition. In spatial data analysis, the neighbourhood structure of spatial entities is nested within a spatial weights matrix W, defined as

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

Index n represents the number of observations. For instance, if we are interested in studying municipalities, then n represents the total number of municipalities in the sample. Hence, i stands for one municipality and j for another municipality. The single elements w_{ij} represent the bilateral neighbourhood relation between two entities i and j. The diagonal values of W characterise the neighbourhood relationship of each spatial entity to itself, while the off-diagonal elements represent the neighbourhood relationship to all other spatial entities. The diagonal values can be interpreted as an internal restriction, barrier or resistance within the spatial unit itself. In economics, they are usually of little interest and set to zero by assumption. However, there are some examples where the diagonal elements are unequal to zero, but they are usually considered special cases.[2] The rows and columns of W are symmetric for most applications. The symmetry of the matrix ensures that the inverse matrix W^{-1} exists, which is required for most spatial regression analysis. Symmetry also ensures that the neighbourhood structures are bilateral. For example, if Germany is neighbour to France, then France is also neighbour to Germany. Although this is a weak assumption, the spatial weights matrix is able to model unilateral neighbourhood structures as well. Further, W is often assumed to be exogenous which means that the neighbourhood structure is known and static. In most applications, an exogenous spatial weights matrix is a reasonable assumption. For instance, the borders of administrative units, such as states or municipalities, do not change frequently and do not require dynamic modelling of the spatial weight matrix. As spatial econometricians work at the scope of countries, states, or municipalities, complex neighbourhood structures and endogenous matrices are usually not required. In contrast, Gibbons and Overman (2012) argue, that the assumption of exogenous spatial weights matrix leads to biased estimates, if the true matrix is endogenous. However, there is also literature that puts emphasis on endogenous spatial weights matrices (Ahrens and Bhattacharjee, 2015).

Different neighbourhood structures, i.e. contiguity, distance, and k-nearest neighbours, are achieved by distinct formalisations of the single element w_{ij}. First, the common border or spatial contiguity is particularly useful for the analysis of economic or administrative areas. It is defined as follows

$$w_{ij} = \begin{cases} 1 & \text{if area } i \text{ shares a border with area } j \\ 0 & otherwise \end{cases}$$

The contiguity definition is binary, i.e. the values of the matrix can be either one or zero. The single element w_{ij} is equal to zero if areas i and j do not share a common border. If they share a common border, the value is equal to one and both spatial entities are considered neighbours.

Second, both areas i and j may be considered neighbours when the distance d_{ij} is less than a certain threshold, say d. This definition leads to distance-based spatial weights, formalised as

$$w_{ij} = \begin{cases} 1 & \text{if } d_{ij} < d; \text{and } d > 0 \\ 0 & \text{otherwise} \end{cases}$$

For area and line data, the distance-based definition requires centroids, i.e. the central location of each area unit. The centroids are calculated from longitude and latitude data for each area. The straight-line determines the shortest distance between two point locations, i.e. centroids, in a flat plane, whereby

[2] For example this is usually the case when intra-entity distances or barriers are modelled such as travel distance within a municipality.

longitude and latitude are treated as coordinates within the plane. If the Euclidean distance between two centroids lies within the threshold, both areas are considered neighbours.

The third popular definition of neighbourhood considers the k-nearest neighbours. It is particularly useful, when spatial entities are too far away to be modelled by a reasonable distance definition, or they don't have a common border. A typical example for these entities are islands, for example the Canary Islands belong to Spain but are rather far away and don't have a common border with Spain. Nonetheless, there is social and economic interaction between the entities. The k-nearest neighbour matrix takes the form

$$w_{ij} = \begin{cases} 1 & \text{if centroid of } j \text{ is one of the } k \text{ nearest centroids to that of } i \\ 0 & \text{otherwise} \end{cases}$$

It is evident that all three definitions of the spatial weight matrices are derived from the same spatial layout and that the choice of different spatial weights matrices can affect the outcome of the analysis. Hence, spatial statistics and spatial econometrics is evidently conditional on the choice of a spatial weight matrix. On the other hand, the choice of the spatial weight matrix should not be too overemphasised, because differences in estimates based on spatial weights matrix are also limited (Wang, Kockelmann and Wang, 2013).

1.3 Spatial Autocorrelation

Once the neighbourhood structure is defined, it can be used to study patterns in georeferenced data. An important tool and prerequisite for regression analysis is spatial autocorrelation (spatial dependence and spatial interaction are used synonymously) which determines how a variable is correlated with itself through space (auto meaning self). It helps detecting patterns, i.e. similarity or disparity, in nearby spatial entities. Fischer and Wang (2011) define spatial autocorrelation as *"the correlation among observations of a single variable […] strictly attributable to the proximity of those observations in geographic space*". In other words, spatial autocorrelation describes whether near things are more similar (or dissimilar) to each other than distant things. This definition is imbedded in Tobler's first law of geography, which states that *"everything is related to everything else, but near things are more related than distant things"* (Tobler, 1970). Measures for spatial autocorrelation work with two types of information. First, similarity among locations, i.e. proximity, and second, similarity of the values of a single variable across observations, i.e. similarity. For example, when studying GDP externalities, the first type contains information about proximity among the areal units, and the second type contains information about similarity in GDP among areal units. The central aspect of the spatial autocorrelation is the cross product between both types of information which can be described by a simple model

z_i the value of the variable of interest for region i

w_{ij} the similarity of i's and j's location (proximity)

m_{ij} the similarity of i's and j's variable values (similarity)

The measures of spatial autocorrelation are differentiated by the scale and scope of the research question. It can be distinguished between global and local measures. Global spatial autocorrelation combines both types of information, w_{ij} and m_{ij} and generates a cross product for *all* observations while local autocorrelation considers a subset of the sample. Global spatial autocorrelation is defined as

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}m_{ij}$$

Multiplying each element of matrix w_{ij} with each element of matrix m_{ij}, and summing, results in a scalar. The scalar is an average for similarity (or dissimilarity) of variable values in proximate spatial areas. In practice, the similarity of variable values m_{ij} depends on the scaling of the data. For nominal variables, m_{ij} takes the value of one if i and j are equal, and zero otherwise. For ordinal variables, the similarity of values is based on the ranking of i and j. The most common measure for global spatial autocorrelation with interval variables is Moran's I statistics. It is based on the cross-product $(z_i - \bar{z})(z_j - \bar{z})$, with $\bar{z}$ being the average of the z-values. Moran's I statistics was introduced by Moran (1950) and further developed by Cliff and Ord (1973, 1981). It takes the form

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^{n}(z_i - \bar{z})^2} \quad ; \quad i \neq j$$

The test statistics is a two-tailed test and assumes spatial randomness under the null hypothesis. If the null hypothesis sustains, there is no spatial autocorrelation, which indicates that space does not matter for the distribution of variable values. The alternative hypothesis is the existence of positive or negative spatial autocorrelation. Similar to Pearson's correlation, the value of the test statistics can either be negative or positive. The statistics ranges from $[-1; 1]$, which makes it straightforward to interpret. Positive values of Moran's I indicate that similar variable values are located in close proximity. For example, in terms of GDP, positive autocorrelation indicates, that high GDP areas neighbour other high GDP areas. Hence, positive spatial autocorrelation can be interpreted as a measure for spatial clustering of similar values of a single variable. On the other hand, negative spatial autocorrelation indicates that variable values fluctuate in neighbouring locations. This means, a high value is likely to be neighbouring a low value of the same variable and vice versa. In other words, an area with high GDP is neighbour to an area with low GDP, and vice versa. Figure 1.2 illustrates both types of spatial autocorrelation for a binary variable. The left-hand side demonstrates a concentration of similar values in the upper left corner (blue) and the bottom right corner (white).

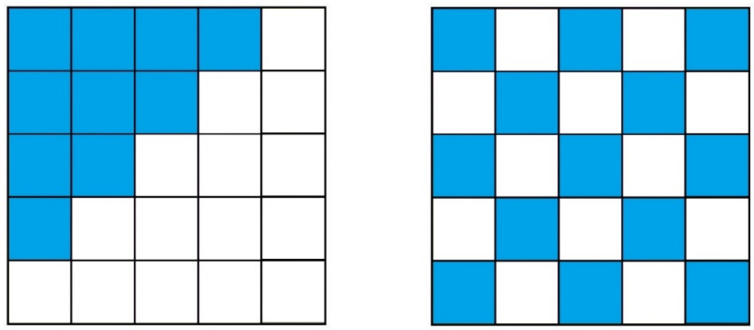

Figure 1.2: Positive spatial autocorrelation (left) and negative spatial autocorrelation (right)

The right hand side illustrates the dispersion (checkerboard pattern) of similar values in space. In practice, positive spatial autocorrelation can be interpreted as spatial clustering of similar values and negative autocorrelation indicates spatial diffusion of similar values. Figure 1.3 illustrates the unemployment rate for the working population among German districts (*Stadt- & Landkreise*). The unemployment rate is divided into eight quantiles. Regions with high unemployment rates cluster in eastern Germany and the Ruhr area in western Germany forming two “hot spots”. Low unemployment rates are mainly observed in the wealthy south of Germany and form one large “cold spot”. Moreover, the unemployment rate transitions smoothly from high rates to average to low rates, which is another indicator for positive global spatial autocorrelation. Moran’s I statistics confirms the visual inspection and quantifies the degree of similarity and proximity. Table 1.1 indicates a positive and highly significant spatial autocorrelation with a Moran’s I value of 0.623.

Table 1.1: Moran’s I statistic for the unemployment rate

$Variable$	$Moran's\ I$	$sd\ (I)$	$z-value$	$p-value$
unemployment rate	0.623 ***	0.031	20.431	0.000

****0.1% significant ; ** 1% significant; * 5% significant*

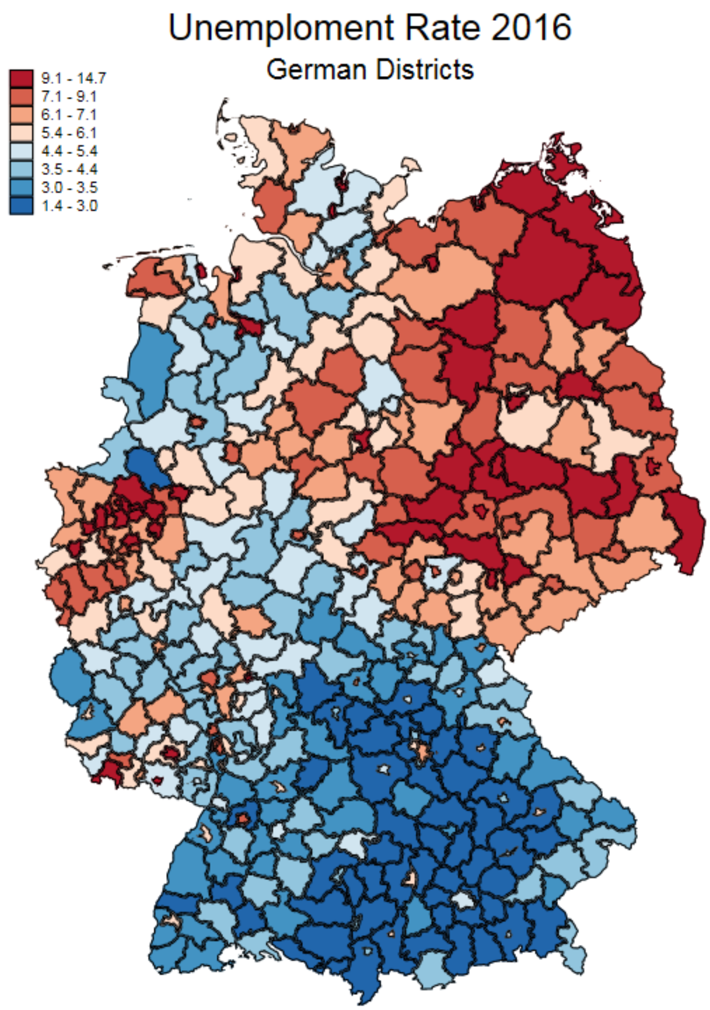

Figure 1.3: Spatial clustering in the unemployment rate among German districts for the year 2016.

Note that Moran's I considers the transition from the "hot spots" to "cold spots" because the statistic determines the scalar for the *whole* population. In contrast, local spatial autocorrelation measures, such as Geary's C, treat each hot (or cold) spot separately without necessarily considering links between them (Cliff and Ord, 1981). Strictly speaking, local spatial autocorrelation treats each cluster as an independent entity. Hence, the choice between global and local measures depends on the scope of the research question. While global spatial autocorrelation helps detecting clusters, local spatial autocorrelation is a better fitting tool to study and explain certain local clusters. Since local spatial autocorrelation measures are not the main emphasis of this work, I will further use global- and "regular" spatial autocorrelation synonymously.

1.4 Spatial Econometric Models

Exploratory spatial data analysis and spatial statistics are preliminary steps to more quantitative modelling which seeks to establish a relationship between several variables while also accounting for spatial dependence. The aim of this chapter is to provide an overview of the most common spatial econometric models for a cross-section data framework. For didactical purposes, it is useful to follow a "general to specific" approach, which means starting with the most general spatial model, i.e. the Manski model, and deriving all other models from it. The Manski model incorporates three possible types of spatial interactions (Manski, 1993; Elhorst, 2010). It incorporates (1) endogenous interactions, i.e. spatial dependence in the dependent variable, (2) exogenous interactions, i.e. spatial dependence in the independent variables, and (3) spatial interactions in the error term. The Manski model is defined as

$$Y = \lambda WY + X\beta + WX\theta + u$$

$$u = \rho Wu + \varepsilon$$

Y is an *N x 1* vector for the dependent variable where *N* is the total number of observations. It contains one observation of the dependent variable for each unit in the sample $(i = 1, \ldots, N)$. *X* is a *N x K* vector of explanatory variables with the corresponding *K x 1* parameter vector β. The *N x N* symmetric spatial weights matrix is represented by *W,* and *u* is an independent and identically distributed *N x 1* error term for all i. There are three types of spatial interaction effects included in the Manski model: first, WY is the endogenous interaction effect, i.e. the effect of the neighbouring area's dependent variable. Second, WX is the exogenous interaction effect between the independent variables, and third, Wu is the spatial interaction among the error terms of different spatial entities. The corresponding coefficients of the spatial interactions are denoted as λ, θ, and ρ, respectively. λ denotes the spatial autoregressive coefficient. It can be interpreted as the average influence of the neighbouring dependent variables on the location of interest. θ represents a *K x 1* vector of parameters, similar to β. The single coefficients of vector θ represent the mean spill-over-effects of neighbouring independent variables on the dependent variable at the location of interest. While ρ is the error autocorrelation coefficient. The Manski model is not particularly useful in empirical applications because the estimated coefficients cannot be clearly separated from each other. As a result, estimated coefficients from the Manski model will be biased. Nonetheless, the model is helpful for illustrating the hierarchy of spatial models, because all other cross-section models are nested in it. For instance, setting any of the three spatial interaction coefficients equal to zero, i.e. λ, θ, and ρ, will result in another model. For instance, setting θ and ρ to zero results in the spatial lag model (SLM), which is defined as

$$Y = \lambda WY + X\beta + u$$

If λ and θ equal zero, the outcome will be the spatial error model (SEM), formally

$$Y = X\beta + u$$

$$u = \rho Wu + \varepsilon$$

If only one interaction is set to zero, say ρ, the result will be the Spatial Durbin model (SDM)

$$Y = \lambda WY + X\beta + WX\theta + u$$

Setting all spatial coefficients, λ, θ, and ρ, to zero will result in the standard OLS model. Table 1.2 summarises all combinations of spatial interactions and the resulting model for cross-section data. The Manski model is the most general model and nests six spatial models, i.e. SLM, SEM, SDM, KPM, SDE, SLX, plus the OLS model. The spatial lag model and spatial error model are considered the two most important "classical" spatial models and were established by Anselin (1988). Both models include one type of spatial interaction effect: The spatial lag model contains a spatially lagged dependant variable, while the spatial error model establishes a spatial autoregressive process in the error term. Both models were – and still are – considered workhorse models in spatial econometrics.

Table 1.2: Spatial interaction models for cross-section data

Model	**Independent interaction** (WY)	**Dependent interaction** (WX)	**Error term interaction** (Wu)
Manski model	√	√	√
Spatial Durbin model (SDM)	√	√	-
Kelejian-Prucha model (KPM)	√	-	√
Spatial Durbin error model (SDE)	-	√	√
Spatial lag model (SLM)	√	-	-
Spatial error model (SEM)	-	-	√
Spatial lag in X (SLX)	-	√	-
OLS	-	-	-

During the last ten years, interest has grown in models which contain more than one spatial interaction. Based on Anselin's work, the Kelejian-Prucha model (KPM) combines two spatial interaction effects, i.e. spatial interaction in the dependent variable and the error term, in model. The foundation for KPM was laid by Kelejian, Prucha (1998) and later emphasised by Harry Kelejian at the 2007 Annual Spatial

Econometric conference. In contrast to the classical models with one type of spatial dependence, the Kelejian-Prucha model allows for more efficient estimates and outperforms the SLM and SEM model. In the same year, James LeSage highlighted the advantages of models with spatially lagged independent variables. He introduced the Spatial Durbin model (SDM), which incorporates a spatial structure in the dependent variable and the independent variables at the North American Meeting of the Regional Science Association in 2007. The mathematical foundation for the Spatial Durbin model can be found in the work by LeSage and Pace (2009). The authors argue that the Spatial Durbin model is the most versatile spatial model for cross-section data because it fits to a larger share of empirical applications due its robustness. To be precise, the Spatial Durbin model produces unbiased coefficient estimates, even if the true data generating process is the SLM or SEM model. In order to shed some light on the model selection, the next session will discuss some basic decision rules for spatial econometric models.

1.5 Model Selection

Ultimately, a spatial econometric model should serve two purposes. First, a model should fit the scope of a research question, i.e. it should be consistent with the research motivation. Second, the model should be specified correctly, i.e. minimise coefficient bias and increase coefficient efficiency. The spatial interactions, which were introduced previously, serve both purposes. Especially spatial interactions in the endogenous and exogenous variables are often the key interest because they can help understanding spatial spill-over-effects and externalities. On the other hand, spatial interactions in the error term serve primarily to increase estimation efficiency (Anselin, 1988). Model selection in cross section spatial econometrics is mostly concerned with two questions. First, is a spatial model necessary or is a non-spatial model sufficient? Second, which spatial model is the most useful choice in terms of answering the research motivation? Whether a model is a spatial model or a non-spatial model ultimately depends on the existence of spatial dependence. If a variable does not depend on the spatial distribution of the observations, a spatial model will not improve the estimates over a non-spatial model like OLS. Hence, tests for spatial dependence are necessary to determine whether a spatial model or a non-spatial model is used.

LeSage and Pace (2009) argue that the costs of misspecification are highest, if the spatial dependence in the dependent variable is ignored. Therefore, most tests for spatial dependence focus on the parameter λ which is the corresponding coefficient to the spatial lag WY. A great share of the literature focuses on the Lagrange Multiplier (LM) test statistics, which requires two components: (1) the least squares residuals e, and (2) a spatial weights matrix W (Burridge, 1980). The LM statistics is defined as

$$LM = \left[\frac{e'We}{\frac{e'e}{n}}\right]^2 / \, tr(W^2 + W'W)$$

The LM test is based on Moran's I test statistics and has a similar purpose. It can be used in a t-test setting where the null hypothesis assumes no existence of spatial dependence. For example if we are interested whether a model is a spatial lag model, the null hypothesis will be $H_0: \lambda = 0$. If the null hypothesis is rejected, there is spatial structure in the dependent variable and the SLM model is a likely choice. This test can also be used to test independent variables and the error term for spatial dependence. Since the test depends purely on OLS residuals and a spatial weights matrix, the test is considered rather simple. However, the major drawback of the LM test is its inability to distinguish several spatial interactions at once. The test is restricted to one type of spatial dependence at a time. For example, the test can be also used to study spatial interactions in the error term. However, it will then ignore potential presence of spatial dependence in Y. In other words, the test is useful to detect spatial dependence, but not particularly useful to determine which model describes the true data generating process best.

Therefore, LeSage and Pace (2009) propose that model selection should not be purely based on test statistics, but rather on rational considerations about the costs of misspecification. They argue, that the cost of ignoring spatial dependence in Y is relatively high, because it results in biased coefficients. On the other hand, ignoring spatial dependence in the error term will bear fewer costs because it results in lower efficiency, with which it is easier to deal. One major reason why the authors consider inefficiency as not too big of a concern are improvements in data availability over the last years. The average sample size for spatial data has increased and with larger spatial datasets, efficiency may be less concerning compared to bias. They argue that bias also sustains with larger samples. Hence, ignoring spatial dependence in Y can be considered more "costly" in comparison to spatial dependence in the error term. Based on that reasoning, LeSage and Pace (2009) propose four model selection criteria.

1. If the true DGP is a SEM model, i.e. there is spatial dependence in the error term, using SLM, KPM and SDM models will produce unbiased but inefficient coefficients.
2. If the true DGP is a SLM model, i.e. with spatial dependence in the dependent variable, using SEM will produce biased coefficients due to the omitted variable bias. KPM, SDM will produce unbiased coefficients.

3. If the true DGP is a SDM model, i.e. with spatial structure in the dependent and the independent variables, then SLM, SEM and KPM will suffer from omitted variable bias, since all models exclude the spatial interaction of the independent variables, WX. However, SEM will be even more biased because it also excludes the spatial interaction in the dependent variable WY.
4. If the true DGP is a KPM model, i.e. with spatial structure in the dependent variable and the error term, SLM and SDM will compute unbiased coefficients but will lack some efficiency. On the other hand, SEM will produce biased coefficients.

Unfortunately, the authors do not explicitly consider the SLX and SDE model in their comparison. However, the SDM model produced unbiased coefficients even if the true DGP is the SLX model. If the true DGP is a SDE model, then the SDM model produces unbiased but inefficient coefficients. In conclusion, out of the six spatial models, the SDM model is the only model which is able to produce unbiased coefficients. Yet in some cases, it remains inefficient. Another advantage of the SDM model is its simultaneous consideration of spatial structure in the dependent and the independent variable. One major motivation for applying spatial models are unobserved spatial interactions, such as regional spill-overs. Modelling spatial interactions in the dependent and the independent variable at the same time, is the most extensive approach to account for unobserved effects.

2 An Ecological Approach to Electronic Gambling Machines and Socioeconomic Deprivation in Germany [3]

Keywords

EGM, ecological analysis, Germany, risk factors

Author

Stergios Xouridas, Johannes Jasny, Tilman Becker

Abstract

In Germany, gambling research has primarily focused on the broader population in prevalence studies, neglecting the importance and influence of the local socioeconomic context in the development and maintenance of gambling disorders. To analyze the interplay between contextual and compositional factors in the market for electronic gambling machines (EGMs) in Germany, we assessed the EGM densities and socioeconomic deprivation in 244 local communities within Baden-Wuerttemberg. Our results suggest that EGM density is statistically associated with 3 socioeconomic determinants: The shares of migrants, unemployed, and high-school-educated people in the communities are statistically significant variables in our linear regression model, whereas younger age, male gender, and marital status exhibit no statistical associations with EGM density. The share of unemployed people is the only variable of statistical and practical significance. Our analysis advocates area-based policy measures to minimize gambling-related harm. By decreasing EGM densities in communities with high levels of unemployment, we expect to protect at-risk population strata that are most vulnerable to gambling exposure.

[3] This article has been originally published as: Xouridas, S., Jasny, J., and Becker, T. (2016). „An ecological approach to electronic gambling machines and socioeconomic deprivation in Germany". *Journal of Gambling Issues*, 33: 82-102.

2.1 Introduction

In recent years, the German market for electronic gambling machines (EGMs) has increased by a respectable amount: From 2009 to 2013, the absolute number of EGMs increased from 233,000 to 263,000, while the total turnover of the gaming machine industry increased from 4,965 to 5,550 billion euros (Vieweg, 2013). In Germany, gambling is generally regulated at the federal state level. Authorities in the federal states govern the sectors of lotteries, sports betting, and casinos. EGM gambling in gaming halls, restaurants, and bars is subject to the federal laws of the German Industrial Code (Gewerbeordnung) and the Gambling Ordinance (Spielverordnung). In 2006, the fifth amendment of the Gambling Ordinance considerably improved the framework conditions of gaming operators and increased the incentives for individuals to gamble at EGMs. This was achieved primarily by introducing an accelerated game with higher stakes, profits, and losses. The negative effects of problematic gambling have been recognized as a public health concern by public authorities, and legislators have enacted a new regulatory framework to prevent the development of gambling disorders, the 2008 Interstate Treaty on Gambling (Glücksspielstaatsvertrag). The federal states revised the Interstate Treaty in 2012 (Glücksspieländerungsstaatsvertrag) to implement a more consistent gambling law to include legislative measures for the EGM sector.

Erbas and Buchner (2012) have shown that the quantity of inpatient and outpatient treatments for pathological gambling in Germany has increased significantly during the last decade. However, increased treatment seeking for pathological gambling may not be exclusively attributable to increased gambling availability. The rise in the number of treatment-seeking individuals can also be explained by increased awareness of the dangers associated with gambling. The public has been made aware of the dangers by mass media campaigns and by labels and warnings on gambling tickets. The Federal Centre of Health Education (BZgA) provides a nationwide telephone helpline and Internet self-help programs to counsel on problematic gambling behavior. To advertise lotteries, providers have been legally obligated since 2008 to provide gambling addiction warnings and telephone helplines for those who want confidential advice. Furthermore, legal regulations require operators to display information leaflets on problematic gambling in gaming halls, and EGMs are required to display a telephone helpline linked to the BZgA. All of the above-mentioned measures have increased public awareness of gambling and produced a profound effect on treatment-seeking rates. In response, some federal states have started to finance specialized counseling centers for pathological gamblers (Buchner et al., 2015).

Multiple epidemiological studies in Germany have provided estimates of the proportion of pathological and problematic gamblers. Large community-based epidemiological studies that reported gambling prevalence rates for Germany were conducted by Bühringer, Kraus, Sonntag, Pfeiffer-Gerschel, and Steiner (2007) and Buth and Stöver (2008); other notable studies in this context include periodic studies from the BZgA (Haß, Orth, & Lang 2012), a study by Sassen et al. (2011), and the study of Pathological Gambling and Epidemiology (PAGE) conducted by Meyer et al. (2011).

To give an overview of the German gambling market, we will briefly discuss the BZgA studies. They were the only repeated cross-sectional analyses in Germany (BZgA, 2008, 2010, 2012, 2014). The four population-wide representative studies were conducted in the years 2007, 2009, 2011, and 2013 and assessed gambling behavior and gambling-associated attitudes and problems in the population aged 16 to 65 years. By using random digit-dial sampling, the study investigators collected data by means of computer-assisted telephone interviews. Sample sizes for the years 2007, 2009, and 2011 were 10,000 with only landline telephone numbers. The latest BZgA study in 2013 used a dual-frame approach, in which 10,001 landline telephone numbers and 1,500 mobile phone numbers were included in the analysis. To ensure comparability between the studies, we considered only the results of the sample with landline telephone numbers. Response rates for the BZgA studies were 63.3% in 2007, 61.6% in 2009, 59.9% in 2011, and 56.8% in 2013. Gambling participation rates in the 12 months preceding the BZgA surveys decreased significantly over time, from 55.0% in 2007 to 53.8% in 2009 to 50.7% in 2011 to 44.9% in 2013. The decline in participation rates could be observed for both genders and across all age groups. The 12-month prevalence of the most popular gambling activity, “lotto 6 out of 49,” decreased significantly from 35.5% in 2007 to 28.7% in 2013. In contrast, the 12-month prevalence of gambling on EGMs has increased steadily from 2.2% in 2007 to 2.7% in 2009 to 2.9% in 2011 to 3.0% in 2013. The greatest increases were evident among young males. For example, the 12-month prevalence of EGM gambling increased for 18-to 20-year-old males from 5.8% in 2007 to 19.3% in 2013 and for 21-to 25-year-old males from 5.1% in 2007 to 14.4% in 2013.

For 2009, 2011, and 2013, the BZgA surveys assessed the severity of gambling problems with the South Oaks Gambling Screen (Lesieur & Blume, 1987). The 12-month prevalence of pathological gambling decreased from 0.45% in 2009 and 0.49% in 2011 to 0.38% in 2013. The 12-month prevalence of problematical gambling decreased from 0.64% in 2009 to 0.51% in 2011 to 0.45% in 2013. In describing the socioeconomic profiles of problematic and pathological gamblers, a multivariable analysis showed that male gender, immigration background, and being unemployed were significant predictor variables at the 5% significance level.

The German literature has primarily focused on prevalence studies to delineate the distribution of gambling across population segments. However, the population is not the only relevant focal point in an epidemiological study. For a complete epidemiological description of gambling, we need to include spatial or temporal dimensions in our analysis (Suzuki, 2012). Our analysis focuses on the viewpoint of place-based health research, which perceives gambling as a complex phenomenon with interconnected relationships between gambling activities and the social environment, consisting of not only the gambling venues, but also the socioeconomic, cultural, and political contexts (Korn & Shaffer, 1999).

Contextual factors that characterize the social and physical environment of gamblers play an increasingly important role in the study of gambling by demonstrating how the local context of the neighborhood influences the gambling behavior of individuals (Marshall, 2009). One possible contextual effect that helps explain the development and maintenance of gambling disorders concerns the availability and accessibility of EGMs. Various authors have shown that employees of casino and gaming venues exhibit higher rates of gambling disorders than the general population does, providing evidence that increased exposure to gambling at an individual level is associated with gambling-related harm (Hing & Gainsbury, 2011; Shaffer & Hall, 2002; Shaffer, Vander Bilt, & Hall, 1999; Wu & Wong, 2008). Analyzing gambling at an area level, Marshall (2005) noted that gambling providers could induce demand for gambling by placing EGMs at favorable geographical locations. Increased gambling prevalence rates and higher gambling expenditures would then be supply-driven, or at least demand for gambling would be encouraged by the supply side. The location and number of EGMs are situational determinants of gambling and may explain why some people start gambling and why certain socioeconomic classes are particularly vulnerable to specific forms of gambling (Griffiths, 1999).

The proposition that the supply side encourages or induces gambling activity remains controversial. For example, Govoni, Frisch, Rupcich, and Getty (1998) found no difference in the prevalence rates of pathological and problem gambling or the per capita gambling expenditure after a new casino opened, whereas Ladouceur, Jacques, Ferland, and Giroux (1999) established an association between increased opportunities for gambling and the prevalence of pathological gambling. Other studies have established significant associations between gambling problems and residential proximity to casinos (Welte, Barnes, Wieczorek, Tidwell, & Hoffman, 2007; Welte, Wieczorek, Barnes, Tidwell, & Hoffman, 2004), as well as between gambling prevalence and per capita density of EGMs (Storer, Abbott, & Stubbs, 2009). Vasiliadis, Jackson, Christensen, and Francis (2013) reviewed the existing empirical literature on the interrelation between physical accessibility and gambling involvement in Australia, Canada, New Zealand, Norway, and the United States and reiterated the substantiated evidence of a positive relationship between higher EGM density and higher gambling participation and expenditure.

In addition to the inclusion of contextual effects, gambling behavior at an area-unit level can be described by compositional factors such as socioeconomic characteristics or lifestyle-related variables of the local population (Welte, Wieczorek, Barnes, & Tidwell, 2006). For instance, various international studies have shown that EGM density is higher in socioeconomically deprived areas (Pearce, Mason, Hiscock, & Day, 2008; Wardle, Keily, Astbury, & Reith, 2014; Wheeler, Rigby, & Huriwai, 2006). Socioeconomic characteristics of the area, such as the unemployment rate, may also be considered risk factors. These variables predict or are closely associated with the development and maintenance of gambling disorders, and there are many ways to select variables in order to measure relative deprivation. For example, to assess ecological associations between small-area population characteristics and the location of gaming machines, Wheeler et al. (2006) used New Zealand's Deprivation Index, which is composed of census-based measures of income, unemployment, amenity access, and education. In addition to socioeconomic risk factors, we can also include other risk factors in our analysis, such as genetic vulnerabilities (Lobo & Kennedy, 2009), neurobiological factors (Iancu, Lowengrub, Dembinsky, Kotler, & Dannon, 2008), and personality traits and disorders (MacLaren, Fugelsang, Harrigan, & Dixon, 2011), as well as mental and substance use disorders (Dowling et al., 2015). Not all of the above-mentioned factors are available at the aggregate level. Hence, our analysis focused on socioeconomic factors, where group-level data were collected and made available by the Federal Statistics Office of Germany and the Statistical Offices of the federal states. To the best of our knowledge, no German study has examined the spatial variation of EGMs. Therefore, we set out to analyze the statistical associations between EGM densities and socioeconomic characteristics in small-area units.

2.2 Method

2.2.1 Regional Data for the State of Baden-Wuerttemberg

For our analysis, we focused on the state of Baden-Wuerttemberg (BW), one of 16 federal states in Germany. The state of BW lies in southwestern Germany and is bordered by the states of Rhineland-Palatine to the northwest, Hessen to the north, and Bavaria to the east; it shares borders with two countries, France to the west and Switzerland to the south (see Figure 2.1). BW is the third largest of the federal states, with an area of 35,751 square kilometers and a population of approximately 10.8 million people.

Figure 2.1: The Federal state of Baden-Wuerttemberg, one of 16 Federal states in Germany

Because our ecological study design used aggregate-level data, as our unit of spatial analysis, we chose local communities (Gemeinden), which correspond to the lowest administrative tier in BW. In total, there are 1,101 local communities in BW, but we considered only communities with populations over 10,000, leaving us with 244 local communities for our analysis. Our 244 communities comprised approximately 7.3 million inhabitants, which amounts to approximately 68% of the total population of BW. The data originated from three sources. The first data set concerned the EGM locations in BW and was compiled by Trümper and Heimann (2012), who restricted their analysis to communities with a population of at least 10,000. For each of these 244 communities, they gathered information regarding the granted gambling concessions and the number of gambling establishments and EGMs from the respective public order offices (Ordnungsamt; effective date: January 1, 2012). A total of approximately 31,000 EGMs were included in our analysis. The community data sets for our socioeconomic variables originated from the publicly available 2011 European Union Census and from data files for 2012 that were made available to us by the Statistical Office of BW upon our request.

2.2.2 Socioeconomic Risk Factors for Germany

We used the PAGE study to derive valid socioeconomic risk factors for our regression models (Meyer et al., 2011). With 15,023 participants, the PAGE study included the largest community-based sample in Germany. It used computer-assisted telephone interviews with a stratified and clustered sampling design conducted from January 2010 to March 2011, and it included individuals between 14 and 64 years of age.

Of the interviews with the 15,023 participants, 14,022 were based on landline telephone numbers and 1,001 were based on mobile phone numbers, with the aim of including disadvantaged populations that did not have landline telephones. For the landline telephone survey, the response rate was calculated at 44.5% and the cooperation rate at 54.6%. For the mobile phone survey, the response rate stood at 36.8% and the cooperation rate at 57.3%. Using the guidelines of the American Psychiatric Association, the PAGE study categorized gambling severity according to the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association, 1994). The researchers categorized the severity of gambling problems into two subthreshold categories of gambling problems: one with one or two symptoms and one with three or four symptoms from the DSM-IV. Pathological gambling was defined by the presence of five to 10 symptoms from the DSM-IV criteria. In total, the 12-month prevalence rates for pathological gambling stood at 0.3% in comparison to subthreshold gambling problems with one or two criteria at 1.4% and three to four criteria at 0.3% (Meyer et al., 2011). As expected, the lifetime prevalence rates were substantially higher, standing at 1.0% for pathological gambling, 5.5% for subthreshold gambling with one to two criteria, and 1.4% for subthreshold gambling with three to four criteria. Meyer et al. (2011) used a multinomial logistic regression to quantify the relative risk for pathological gambling in relation to given sociodemographic risk factors; Table 2.1 summarizes their results with respect to the diagnosis of lifetime pathological gambling. For the PAGE study, table 2.1 shows that the odds ratios associated with younger age, male gender, less school education, migration background, and unemployment status are statistically significant at the 5% level ($p < .05$).

Table 2.1: Multinomial Logistic Regression of Lifetime Diagnosis of Pathological Gambling on Socioeconomic Risk Factors

Socioeconomic risk factor	*OR*	95% CI
Age	0.96	[0.95, 0.98]
Gender		
Male	10.71	[5.42, 21.17]
Female	Reference value	
School education		
<10 years education	2.43	[1.46, 4.05]
10 years education	1.97	[1.21, 3.20]
>10 years education	Reference value	
Marital status		
Married	1.06	[0.52, 2.14]
Unmarried/divorced	Reference value	
Migration background		
Yes	2.32	[1.39, 3.86]
No	Reference value	
Unemployment status		
Yes	3.26	[1.74, 6.08]
No	Reference value	

Note. The lifetime diagnosis of pathological gambling was the dependent variable in this model, with 1 = lifetime diagnosis and 0 = no lifetime diagnosis. *OR* = odds ratio; CI = confidence interval. Adapted from Meyer et al. (2011).

Consistency of the socioeconomic risk factors. Socioeconomic risk factors may vary according to the types of gambling, implying that not all forms of gambling are homogeneous (Moragas et al., 2015). We needed to make sure that the socioeconomic risk factors from the PAGE study accurately described the profile of EGM players. In the PAGE study, the participants were interviewed about different forms of gambling such as poker, lotteries, sports and horse betting, and illegal gambling. Meyer et al. (2011) used refined models and analyses to determine which gambling type was associated most significantly with gambling disorders. First, in a multivariate logistic regression for the diagnosis of lifetime pathological gambling, the participants who self-reported more than 10 days of gambling activity demonstrated that gambling on EGMs was the most significant variable associated with pathological gambling, exhibiting the highest odds ratio (odds ratio = 6.3, 95% confidence interval [CI] [4.1 , 9.8], $p < .01$), followed by poker (odds ratio = 5.0, 95% CI [2.8 , 8.9], $p < .01$). Second, pathological gamblers were asked to self-report which of the gambling forms had the strongest impact in developing their gambling problems. A total of 108 of 116 pathological gamblers offered an answer, with the majority (n = 54 pathological gamblers) designating EGMs as the most significant contributor in developing gambling problems, followed by poker in second place (n = 16 pathological gamblers).

2.2.3 Ecological Analysis

Selecting variables from the PAGE study and using them in an ecological analysis makes sense because a community-based analysis offers a good starting point to assess the significance of risk factors at the ecological level. In contrast, the indiscriminate choice of explanatory variables may lead to spurious correlations and unfounded statistical associations at the ecological level.

We needed to translate the variables from table 2.1 into operable aggregate-level variables. For our model, we defined younger age as the share of 15- to 29-year-olds in each community population. "Male gender" referred to the share of males in each community, and the share of migrants and unemployed individuals in the communities represented the risk factors "migration background" and "unemployment status" from the PAGE study. "School education" was captured by the share of high school graduates in the community, with high school graduation referring to finishing the 12 to 13 required years of secondary education in Germany (Abitur). At the individual level, marital status was not a predictor for a lifetime diagnosis of pathological gambling in the PAGE study (p = .87, cf. table 2.1). Nevertheless, we used the share of married persons in our ecological study to assess whether this variable was associated with EGM density at an aggregate level. All of our socioeconomic risk factors were expressed per 1,000 individuals to remove variations arising from differences in population size across communities. Table 2.2 summarizes information regarding the variables that we used in our ecological analysis.

Table 2.2: Variable description

Dependent variable	
EGM density	Number of EGMs per 1,000 population for each community
Independent variables	
Younger age	Number of 15- to 29-year-olds per 1,000 population for each community
Male gender	Number of males per 1,000 population for each community
High school graduation	Number of high school graduates (*Abitur*) per 1,000 population for each community
Marital status	Number of married persons per 1,000 population for each community
Migration background	Number of persons with migration background per 1,000 population for each community
Unemployment status	Number of unemployed persons per 1,000 population for each community

Note. EGM = electronic gambling machine.

For the 244 communities in BW, Table 2.3 displays the mean, standard deviation, minimum, and maximum of our variables.

Table 2.3: Description of the data set

Socioeconomic risk factor	*M*	*SD*	Minimum	Maximum
EGM density	4.0	2.6	0.2	16.0
Younger age	172.1	17.3	137.0	301.0
Male gender	490.6	7.4	463.4	510.0
High school graduation	160.9	59.0	68.1	515.0
Marital status	479.1	26.0	347.0	527.6
Migration background	98.3	33.3	31.6	209.0
Unemployment status	20.4	5.1	8.5	40.9

Note. EGM = electronic gambling machine.

2.2.4 Statistical Methods

To analyze our data, we used the ordinary least squares method in a multivariate linear regression model. Because all of our data were continuous, we assessed the statistical significance using Student's two-tailed t test, with the significance level set to .05. We did not adjust alpha levels to compensate for the familywise error rate, as this would have compromised the statistical power of our analysis. Effect sizes were quantified by the unstandardized coefficients of the regression model, as well as by the unadjusted and adjusted coefficients of determination. We confirmed the validity of the regression model, but those results are not shown here; that is, we tested for the absence of heteroscedasticity and multicollinearity, the independence and normality of the error terms, and the proper model specification. The statistical program STATA 13 was used for the analyses.

2.3 Results

A multiple regression analysis for 244 communities in BW was conducted to evaluate the relationship between EGM density and the socioeconomic risk factors of younger age, male gender, high school graduation, marital status, migration background, and unemployment status. The regression equation was found to be statistically significant, $R^2 = .29$, adjusted $R^2 = .27$, $F(6, 237) = 15.8$, $p < .001$. The results of the regression model are depicted in Table 2.4.

Table 2.4: Linear regression of EGM density on socioeconomic risk factors for 244 communities in Baden-Wuerttemberg

Socioeconomic risk factor	*b*	95% CI	*SE*	*p*
Younger age	0.003	[-0.022, 0.029]	0.013	.806
Male gender	-0.027	[-0.072, 0.017]	0.023	.226
High school graduation	-0.014	[-0.021, -0.008]	0.003	< .001
Marital status	-0.013	[-0.032, 0.007]	0.010	.203
Migration background	0.020	[0.011, 0.030]	0.005	< .001
Unemployment status	0.140	[0.076, 0.205]	0.033	< .001

Note. EGM = electronic gambling machine; *b* = unstandardized regression coefficient; CI = confidence interval.

Of six socioeconomic risk factors, only three were statistically significant at the 5% level. Ceteris paribus, communities with larger shares of unemployed individuals or migrants were significantly associated with higher EGM densities ($p < .001$), whereas the share of high school graduates in communities was negatively associated with EGM densities ($p < .001$). The remaining three socioeconomic risk factors—i.e., the share of males, married, and younger people in the communities—failed to reach statistical significance.

As every socioeconomic risk factor was measured in the same units, we were able to compare the unstandardized regression coefficients for each factor. The variable of unemployment status, as indexed by its unstandardized coefficient of 0.140, was shown to have the strongest relationship to EGM density. This indicated that for every additional unemployed person per 1,000 individuals, there was a predicted increase in the quantity of EGMs by 0.140 per 1,000 individuals. In comparison, the unstandardized coefficients of the statistically significant variables of migration background ($b = 0.020$, $p < .001$) and high school education ($b = -0.014$, $p < .001$) were considerably smaller.

The CIs in Table 2.4 provided a range of plausible values for the regression parameters. The 95% CI suggests that if we were to repeat the analysis over and over, the CI would contain the true parameter 95% of the time. The 95% CI for unemployment status was [0.076, 0.205], suggesting that the quantity of EGMs in the communities was raised by at least 0.076, and perhaps by as much as 0.205 per 1,000 individuals. For the statistically significant variables of migration background and high school education, and the nonsignificant variables of younger age, male gender, and marital status, the smallest lower limit of the CIs stood at -0.072 and the highest upper limit at 0.030. The highest upper limit of 0.030 belonged to the variable of migration background. It was considerably smaller than the upper limit of the unemployment status and hence practically nonsignificant. The smallest lower limit of -0.072 for the variable of male gender might have been of practical significance, yet the CI contained the value of 0, nullifying the statistical and practical significance of the effect size. Thus, unemployment status was a statistically and practically significant variable in our analysis, whereas the remaining socioeconomic risk factors failed to reach practical significance, statistical significance, or both.

Table 2.5 illustrates paradigmatically the communities with the five top and bottom numbers of EMGs per 1,000 population, alongside the share of unemployed people.

Table 2.5: Communities with top and bottom five EGM densities, alongside unemployment statistics

Community	EGMs per 1,000 population	Unemployed per 1,000 population
Top Five		
Kehl	16.0	26.3
Oehringen	15.4	22.8
Riedlingen	13.7	25.0
Geislingen	11.9	27.7
Waldshut-Tiengen	10.1	23.3
Bottom Five		
Leingarten	0.4	15.6
Heddesheim	0.3	20.0
Gundelfingen	0.3	18.2
Kraichtal	0.3	15.5
Ubstadt-Weiher	0.2	13.6

Note. EGM = electronic gambling machine.

The descriptive statistics confirm the regression results. The communities with the top five EGM densities consistently had greater unemployment shares than did the communities with the bottom five EGM densities.

2.4 Discussion

2.4.1 Practical Conclusions

Starting from an evaluation of socioeconomic risk factors in the PAGE study, we used those individually based variables at the group level and assessed their usefulness in an exploratory analysis. For the first time in Germany, we used an ecological study design to investigate the associations between EGM density and socioeconomic deprivation at the community level.

Our analysis demonstrates a statistically significant association between EGM density and the three socioeconomic risk factors of unemployment status, migration background, and high school graduation. Judging from the effect size indices, only unemployment status is also of practical interest. With an average density of four EGMs per 1,000 individuals in the 244 communities of BW (cf. Table 2.3), the increase in the quantity of EGMs per additional unemployed individual ($b = 0.140$) is of considerable practical significance. The importance of unemployment status at the aggregate level is in line with findings from other ecological correlation studies (Pearce et al., 2008; Wardle et al., 2014; Wheeler et al., 2006). The results of our study provide preliminary indications that gambling providers place EGMs at geographical locations with relatively high unemployment rates. Further research should investigate whether this relationship also holds true for other geographical areas in Germany.

We cannot establish statistically or practically significant ecological correlations between EGM density and the share of married, male, or younger people. This is in contrast to other community-based epidemiological studies that have consistently associated gambling disorders at the individual level with younger age and male gender (e.g., Blanco, Hasin, Petry, Stinson, & Grant, 2006; Kessler et al., 2008). However, these individually based risk factors do not necessarily hold at an ecological level. The atomistic fallacy in epidemiological research describes situations in which associations between variables at the individual level may differ from associations between variables measured at the ecological level (Diez Roux, 2002). Therefore, in some cases, we may not be able to predict the behavior of groups from the behavior of individuals.

One possible solution is to reexamine statistical associations at a finer scale of ecological analysis. In future studies, we can investigate whether statistical associations exist at the city level, similar to the study of Gilliland and Ross (2005), which assessed the spatial distribution of EGMs and the socioeconomic conditions for the municipalities of Montreal and Laval. For example, Stuttgart, the state capital of BW, would provide an ideal field for carrying out an ecological gambling study. If the socioeconomic risk factors of male gender, marital status, and younger age remain nonsignificant at the city level, we can approve their nonsignificance at the ecological level with greater confidence.

Our analysis is only a starting point for investigating the complex set of interactions between socioeconomic area characteristics and EGM density. Caution is advised when moving from statistical associations at an ecological level to causal interpretations and relationships. In order to strengthen the evidence base for a relationship between socioeconomic deprivation and environmental exposure to EGMs, further longitudinal or repeated cross-sectional studies need to assess how changes in the socioeconomic environment affect the distribution of EGMs. An extension of our ecological model includes time as an important constituent of a time-geographical model. Spatial interdependencies likely vary over space and time, and it is important to identify the most crucial places and environmental contexts over different times to help estimate gambling exposure most effectively (Cummins, 2007).

Studies that examine how participation rates and gambling problems evolve over time provide another valuable research direction. International studies showed that an expansion of EGMs and increased exposure to gambling lead to rapid increases in gambling participation and problem gambling at first, followed by an adaptation period with gradual reductions in gambling participation and problem gambling rates (LaPlante & Shaffer, 2007; Shaffer, Labrie, & LaPlante, 2004; Storer et al., 2009). Factors that scholars believe influence the gradual adaptation of society to gambling include increased public awareness of problem gambling, informal social controls, expansion of treatment and self-help organizations, and regulatory and public health measures (Abbott, Stone, Billi, & Yeung, 2015). In a meta-analysis of 34 surveys in Australia and New Zealand, Storer et al. (2009) demonstrated that increased problem gambling prevalence rates are linked to higher EGM density. However, in a multivariable regression analysis, they also showed how problem gambling prevalence decreased over time at a rate of 0.09% per annum. In a meta-analysis of 202 studies from 1975 to 2012, Williams, Volberg, and Stevens (2012) affirmed the decrease in gambling participation and problem gambling rates. The downward trend began in the late 1990s in North America and in the early 2000s in Australia and Europe.

The authors of two recent studies in Sweden and Australia only partially affirmed the adaptation hypothesis. In comparing the results of the Swedish Gambling Study in 1997–1998 and the Swedish Longitudinal Gambling Study in 2008–2009, Abbott, Romild, and Volberg (2014) showed that past-year gambling participation declined from 88.0% to 71.6%, whereas past-30 days participation in EGMs increased from 3.0% to 4.0%. The authors found, contrary to the prediction of the adaptation hypothesis, a statistically significant increase in the prevalence of lifetime probable pathological gambling from 1997–1998 to 2008–2009, with no significant changes in the prevalence of lifetime problem gambling and of 12-month probable pathological or problem gambling.

For Australia, Abbott et al. (2015) compared the results of the 2003 Victorian Longitudinal Community Attitudes Survey and the 2008 Victorian Gambling Study and found significant reductions in gambling participation rates for the previous 12 months, as well as for the monthly and weekly participation frequencies. Gambling participation on EGMs for the previous 12 months decreased from 33.4% in 2003 to 21.5% in 2008, affirming the adaptation hypothesis. Here again, in contrast to the predictions of the adaptation hypothesis, there is no evidence of prevalence rate reductions in problem and moderate-risk gambling.

For Germany, the BZgA studies offer insight into the processes of exposure and adaptation (BZgA, 2014). The repeated cross-sectional analyses allow us to monitor patterns of change over time. Gambling participation rates in the 12 months preceding the surveys decreased from 55.0% in 2007 to 44.9% in 2013, affirming the adaptation hypothesis. Contrary to the general trend of decreasing participation rates in nearly all gambling activities, the 12-month prevalence of EGM gambling steadily increased from 2.2% in

2007 to 3.0% in 2013. The increases in 12-month prevalence rates of EGM gambling were most evident in 18- to 20-year-old males and 21- to 25-year old males. In the first group, the prevalence rate increased fourfold from 5.8% in 2007 to 23.5% in 2013, and in the second group, it more than doubled from 5.1% in 2007 to 12.8% in 2013. It is interesting to note that the prevalence rate for the male age group of 26- to 35-year-olds increased only slightly from 5.8% in 2007 to 7.7% in 2013. How differential exposures and vulnerabilities to EGMs lead to gender and age disparities merits a systematic analysis. Does increased EGM availability explain the rise in prevalence rates among males aged 18 to 20 and 21 to 25 years? Is the male age group of 26- to 35-year-olds not vulnerable to increased EGM exposure, or has the process of adaptation decreased prevalence rates? It is also possible that the lower prevalence rate for the 26- to 35-year-old males is biased downward because the age bracket covers a 10-year range, whereas the 18- to 20-year bracket covers only a 3-year range and the 21- to 25-year bracket covers only a 5-year range. In epidemiological terms, we need to exhibit the changes in prevalence rates over time by age-period-cohort modeling strategies. In decomposing age, period, and cohort effects, we might be able to determine why certain age groups are more susceptible to increased EGM availability and which age groups are less affected or more able to adapt to the negative effects of gambling by an accumulation of exposure experience.

As far as problematic and pathological gambling behavior is concerned, the adaptation hypothesis seems not to hold from a viewpoint of statistical significance, although a trend is observed toward lower rates. The 12-month prevalence of pathological gambling in the BZgA studies decreased from 0.45% in 2009 and 0.49% in 2011 to 0.38% in 2013. The 12-month prevalence of problematical gambling decreased from 0.64% in 2009 to 0.51% in 2011 to 0.45% in 2013. The decreases in the proportion of problem and pathological gamblers in both male and female respondents were statistically nonsignificant at the 5% level.

With the 2012 Interstate Treaty on Gambling, legislators plan to implement harm minimization policies based on geographical criteria. The minimum distance regulation in BW obliges gaming halls to keep a straight-line minimum distance of 500 m from each other, as well as from child and youth facilities. These provisions will come into effect after a transitional period that ends in July 2017, with some experts expecting a 55% reduction in the number of EGMs and a 68% reduction in the number of gaming halls (Becker & Heinze, 2015; Vieweg, 2013). Area-based policies and locally targeted interventions offer promising possibilities to combat gambling problems; on the basis of our results, public authorities should focus on decreasing EGM density in communities with relatively high unemployment rates.

To support our policy recommendation, we refer to a principle that originated from German national environmental law during the 1970s (Myers, 2002): The "principle of precaution," translated from the German word Vorsorgeprinzip, found its way into international law in the fields of environmental policy, natural resource management, and biodiversity conservation and more recently emerged as a principle of international law in international treaties and national policy statements (Andorno, 2004). The precautionary principle calls for preventive measures in public health affairs whenever potential harm to human health or the environment may be anticipated. This principle of practical decision making provides policymakers with the guidance to forestall potential adverse effects in the face of uncertainty of impact and causality.

We have reasonable grounds for concern that increased availability of EGMs raises the threat of gambling-related harm to human health. With the precautionary principle, we err on the side of caution, shifting the burden of proof about absence of harm from the public to the gambling industry. Policy makers should not wait until evidence of gambling-related harm from increased EGM availability is established beyond all reasonable doubt before taking preventive measures. Instead, they are permitted to act on the basis of evidence that is not conclusive.

2.4.2 Directions for Future Research

Future research should focus on two extensions to our model: multilevel and spatial regression models (Anselin, 1988; Diez Roux, 2002). Multilevel methods consider the concomitant inclusion and analysis of individual and ecological variables within a single model. In a multilevel analysis of gambling, the effect and influences of the neighborhood are a separate contextual level that acts on the individual gambling activity among the local population. For instance, living in a privileged socioeconomic environment might offer a protective effect with regard to at-risk gambling because migrant populations are less likely to reside in these neighborhoods and therefore the local population is less likely to adopt the behaviors of these vulnerable populations. In addition to capturing the simultaneous effects of individual and group-level variables on individual-level outcomes, multilevel models can incorporate interactions across levels. For example, the interaction between EGM density and socioeconomic risk factors can be reciprocal in nature: Neighborhoods with a high density of EGMs might be unattractive to populations of higher socioeconomic status, and when those populations avoid living in those neighborhoods, rental costs and property prices can decrease, which in turn attracts individuals with low socioeconomic status. By focusing on the nested structure of the data, multilevel analysis helps explain the ways that higher level environments affect the decisions of individuals.

Whereas multilevel models consider only the correlation within neighborhoods, spatial regression models focus on the inter-neighborhood correlation structure. We expect an interaction between neighboring units that depends on geographical distance or shared borders. Future research should focus on the use of regression models with spatially varying parameters to capture the local variations over different geographical areas and to improve the understanding of local relationships and spatial variations in gambling behavior.

3 Is Gambling Contagious? An Analysis of Electronic Gambling Machine Clustering in Germany [4]

Keywords

EGM supply, gambling hot spots, spatial analysis, socioeconomic deprivation

Author

Johannes Jasny

ABSTRACT

There are sizeable differences in the Electronic Gambling Machine (EGM) supply among German regions. Furthermore, the EGM supply concentrates in certain regions which results in gambling hot spots. Interestingly the spatial clustering of EGM supply is still observed when we control for agglomeration effects caused by population. This leads to the question why the EGM supply concentrates in some regions and remains low in others. We argue that the concentration of supply can be mostly explained by the socioeconomic characteristics of these regions. This paper makes three central contributions to the location based gambling research. First, it visualizes the absolute and relative supply of EGMs in German communities and highlights the spatial clustering of high and low EGM density regions. Second, it implements socioeconomic and geographical control variables for a more distinct description of regional differences. Third, it employs spatial econometric modelling to quantify and explain the occurrence of EGM hot spots. For our analysis we use census and EGM market data. The main finding implies, that there is a clear clustering of the EGM supply across regions at first, but when considering the socioeconomic characteristics / deprivation of the regions, most of the clustering effect is erased. The model explains most of the clustering effect which appears to exist only when there is no slender consideration of the socioeconomic differences across regions. This result supports the hypothesis that high gambling activity in one region does not affect the gambling activity in neighbouring regions.

[4] This article has been originally published as: Jasny, J. (2016). „Is Gambling Contagious? An Analysis of Electronic Gambling Machine Clustering in Germany“. The Journal of Gambling Business and Economics, 10 (3): 54-70.

3.1 Introduction

Pathological and problem gambling can have severe medical, psychological, and social ramifications for susceptible individuals (Fong, 2005). From an aggregated view, the prevalence rates for pathological gambling in Germany range from 0.2 to 0.6 % (Sassen et al. 2011, BZgA 2014, Meyer et al. 2011). In this regard, Germany exhibits the lowest proportion of past-year problem and pathological gambling among general population samples in Europe (Kun et al. 2012). At the same time, Germany was facing a continuous growth trend in gambling activity during the last decade. For example, the number of electronic gambling machines (EGMs) has increased from 200.000 in 2007 to 263.000 in 2013 (Vieweg, 2014). The Federal Gambling Authority of Germany estimates that revenues of 5.3 billion dollar were generated solely by EGMs in 2014 (FGA, 2015). This amounts to approximately half of the total gambling revenues making EGM gambling the largest gambling branch in Germany.

Besides their placement in casinos, EGMs can be found in gambling arcades, pubs, and bars which makes them widely and easily accessible for the public. The regulation of these EGM venues is implemented in the federal laws of the German Industrial Code (*Gewerbeordnung*) and the German Gambling Ordinance (*Spielverordnung*). Even though gambling legislation has curtailed gambling activity in certain aspects, such as the improved protection of minors, it has also liberalized game characteristics and improved the framework conditions of gambling operators considerably. With the liberalizations in the EGM sector, turnover increased by 42% from 2005 to 2009 (Meyer, 2011). In light of increasing concerns over the development of gambling disorders and to limit the provision of gambling services, the legislators have overhauled the Interstate Treaty on Gambling in 2012 (*Glücksspieländerungsstaatsvertrag*). The Interstate Treaty is transformed into federal state gambling acts for each of Germanys 16 federal states. One major aspect of the casino acts is the requirement for minimum distances between existing gaming venues, and between gaming venues and youth facilities. For example in the German state of Baden-Württemberg gambling arcade providers have to maintain a minimum distance of 500 meters to other arcade providers and vice versa starting 2017. They also have to sustain a minimal distance of at least 500 meters to youth facilities such as kindergarten. The rationale of imposing minimum distances is included in the Interstate Treaty: Gamblers that are leaving their gambling spot should not be tempted to re-enter another gambling venue immediately after finishing their current game. Simply put the gambler should “cool down” from excessive gambling activity and not being tempted into the next arcade.

From an aggregated perspective a vast literature examines the interrelation between availability and accessibility of EGMs and gambling prevalence, expenditure and gambling related problems. In a meta-analysis of 34 studies in Australia and New Zealand, Storer et al. (2009) affirm the positive relationship between higher EGM density and problem gambling prevalence rates. A systematic review conducted by Vasiliadis et al. (2013) reiterates the positive relationship between EGM density, EGM proximity and increased gambling involvement. Ceteris paribus it can be expected that high gambling presence in one region could affect the gambling activity of another region and vice versa. However, there are other studies that disprove these views: For South Australia, Delfabbro (2008) found that the removal of a modest amount of EGMs does not affect EGM expenditure or gambling behavior. Nonetheless the majority of literature proposes an association between exposure to gambling opportunities and gambling involvement. However none of the studies analyzed whether gambling activity in one region affects the gambling activity of neighboring regions and vice versa. This relationship can be a possible explanation, why some regions and their neighbors face high gambling activity and others low gambling activity.

Unfortunately, little research focuses on the interrelation between EGM gambling demand and EGM gambling supply which makes it difficult to tackle the causality question. Whether the demand for gambling induces the supply of EGMs or the presence of EGMs tempts players to play more has not been tackled in the literature on an aggregated level. Following the logic of German legislation, implies that a problematic gambler is tempted by proximate EGM arcades, which in conclusion results in a higher demand for EGMs. Hence the initial impulse which induces the demand is the availability of EGM venues. Since German gambling arcades are limited to a maximum number of 12 EGMs per gambling arcade concession and pubs can setup a maximum of three EGMs, we can assume that the rising demand leads to increasing numbers of EGM venues in the neighborhood. In other words, there could be clustering of gambling activity in proximate areas where the demand for gambling is high.

One major reason for varying EGM supply are regional socioeconomic differences. This topic was addressed by several authors in different research setups. Marshall (2005) argues that gambling providers could induce demand by placing EGMs in favorable regions. Furthermore, Marshall (2009) highlighted the importance of the local environment as a driving factor for the gambling intensity. He stresses to distinguish between compositional and contextual factors which can affect gambling intensity. Welte et al. (2006) and Pearce et al. (2008) used national survey data and census data to study the relationship between local area disadvantages and gambling involvement. Both studies confirm that the odds for gambling involvement are higher when regions face higher unemployment and lower income. Several papers study the relationship between EGM density and socioeconomic deprivation. Wheeler et al. (2006)

find a significant positive correlation between EGM density and socioeconomically deprived regions in New Zealand. Wardle et al. (2014) confirm this finding for Great-Britain. Furthermore, some studies identify individual-level risk factors such as family status, younger age and lower education which can also affect gambling intensity. Johansson et al. (2009) provide a broad overview of gambling related risk factors.

Our aim is to study the spatial clustering of EGM supply in Germany's most populated state, North Rhine-Westphalia (NRW). More precisely we want to test the hypothesis, whether the EGM supply of neighboring regions is self-reinforcing which we interpret as a contagion effect. We focus on the contagion analysis while we control for socioeconomic differences across regions. Using Germany as a sample has two major advantages compared to countries with larger areas like the US or Australia. First, it has many densely populated communities with detailed socioeconomic data and second, those communities are in close proximity to each other which makes them a natural choice for this kind of analysis. The motivation for our research objective is illustrated in figure 3.1. The left side shows the number of EGM per community in Germany's largest federal state NRW. We observe that regions with high EGM numbers are clustered in nearby areas, especially those areas neighboring large cities like Cologne, Dortmund, Düsseldorf and Essen. One of the main reasons for this congestion is the population effect, i.e. regions with a higher population have higher numbers of EGMs. We controlled for this effect by computing the EGM density (number of EGMs per thousand residents) for each community on the right side of the figure. The clustering structure of EGM supply in the right figure is still present but less severe.

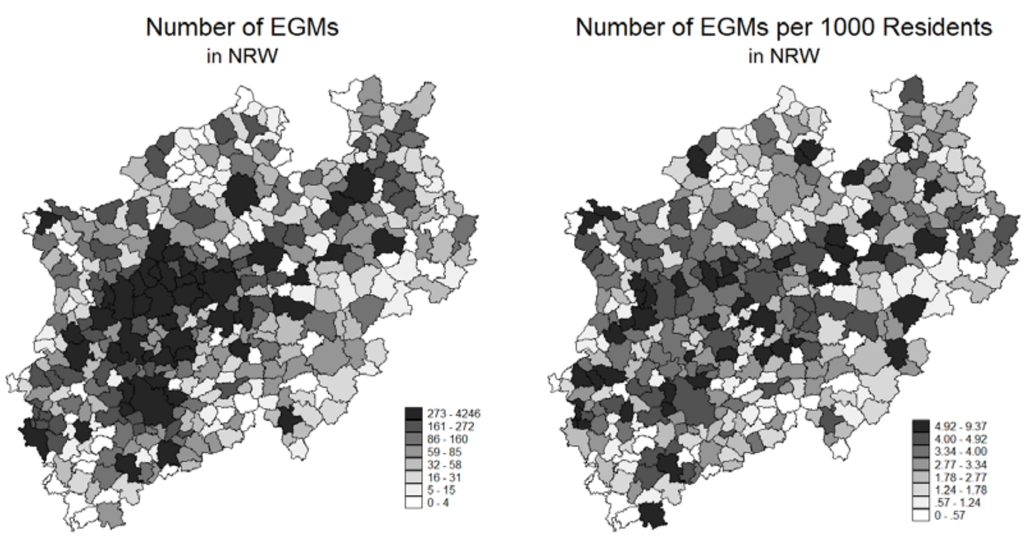

Figure 3.1: The number of EGMs by region (left) and the number of EGMs per 1000 residents by region (right).

The structure of the paper is the following. Part two will present the methodology, define neighborhood in terms of spatial econometrics and introduce tools to analyze the spatial interrelation among neighboring regions. Part three gives an overview of the data. Part four contains the major results. Part five has some concluding remarks and suggestions for further research.

3.2 Methodology

The spatial concentration of EGM supply can be best described by the first law of geography which states that near things are stronger related than distant things (Tobler, 1970). Spatial statistics and spatial econometrics provide techniques which help to visualize data and detect pattern of similarity or dissimilarity such as clustering or hot spots for geographical observations like communities or states. One measure to illustrate the similarity among observed values is the spatial autocorrelation. It describes whether observed values in one region are similar or dissimilar to the values of neighboring regions. When the autocorrelation is positive there is spatial clustering of the observed values among different locations, if the autocorrelation is negative we have no clustering but rather diffused values. Spatial autocorrelation can be tested by Moran's I test statistics which helps to assess spatial similarity or dissimilarity of the gambling supply. To calculate Moran's I statistics it is necessary to define more precisely what is meant by neighborhood.

3.2.1 Measuring Geographical Proximity

The suggested tool to describe proximity or neighborhood is the spatial weight matrix (Drukker et al. 2013a). The spatial weight matrix formalizes which geographical unit is the neighbor of another geographical unit, this can be done in different ways. Assume we wanted to define the nearest neighbors of a single entity in space like a house, whose position can be expressed by one longitude and one latitude coordinate. The neighborhood of this particular house could be defined by its *k*-nearest neighbors, e.g. the six closest houses. Alternatively we can use an arbitrary Euclidian distance radius around that house (e.g. 500 meters) where all other houses within that radius can be considered neighbors. Since our data consists of administrative areas it is useful to think of common boundaries between the communities as a neighborhood definition. In other words, if community i has a common border with community j we consider them as neighbors. This type of spatial weight matrix is called a contiguity spatial weight matrix which is defined as follows:

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix}$$

The contiguity spatial weight matrix contains all neighborhood information about the common borders of different communities. The row and columns are symmetric, i.e. observation 1 represents one community, observation 2 another community, etc. The diagonal values of the matrix are 0 since we rule out intra-community distance (LeSage and Pace, 2009). The single elements w_{ij} of the matrix can take two values, either 0 or 1, defining:

$$w_{ij} = \begin{cases} 1 & if\ i\ is\ neighbor\ to\ j \\ 0 & otherwise \end{cases}$$

When the value is zero, then community *i* and *j* are not sharing a common boundary whereas when the value is one, community *i* and community *j* share a common boundary and can be considered neighbors. There are several ways to define the spatial weight matrix, for example we could use an inverse distance weight matrix where each element of the matrix denotes the inverse distance between two objects. However as we have administrative regions we find the contiguity matrix is a natural choice for our research question.

3.2.2 Testing for Spatial Autocorrelation

The contiguity definition of neighborhood from the previous section can now be used to detect and test for spatial autocorrelation. The most common measure for spatial autocorrelation is Moran's I test statistics which was introduced by Moran (1950) and further developed by Cliff and Ord (1973, 1981). Moran's I statistics can be calculated as follows (Getis, 2010).

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad ; \quad i \neq j \qquad (1)$$

The variable y_i represents the variable of interest for a geographical observation *i*, in our study y_i is the stock of gambling supply in community *i*. The Moran statistics is a test for spatial randomness which means that we reject the null hypothesis if spatial autocorrelation is present and accept it when there is no spatial autocorrelation. Similar to Pearson's correlation the value for Moran's test statistics ranges from minus one to one. Positive autocorrelation indicates clustering values and negative autocorrelation indicates diffusion. Moran's I statistics is useful to determine the spatial autocorrelation for single variables but does not allow a more distinct analysis with several control variables. For example the spatial autocorrelation of EGMs could be also affected by other covariates, like the socioeconomic environment of a region. Since Moran's I does not consider the partial effects of other covariates on the spatial autocorrelation of the variable of interest we will introduce spatial econometric models in the following section.

3.2.3 Spatial Econometric Modelling

The next logical step is to model spatial autocorrelation in a way that other covariates are considered in the analysis. Spatial econometric models allow to disentangle heterogeneities across observations and assess spatial autocorrelation in a controlled setup. In the spatial econometric framework, spatial dependence reflects a situation where values observed for one region *i* depend on the values of neighboring observations at nearby regions *j* and vice versa (LeSage and Pace, 2009). This is a stronger concept in contrast to spatial autocorrelation where we simply observe an association between proximate observations. Spatial dependence can be implemented in the classical linear regression model in two major ways (Elhorst, 2010).[5] First, spatial dependence is present in the dependent variable, therefore one can add the neighbor values of the variable of interest as an additional explanatory variable (spatial lag model) and second, there is spatial dependence in the error term (spatial error model). The combination of both models is called the Kelejian-Prucha Model (KPM). Spatial econometric models are similar to traditional linear regression, with the addition that they allow modeling spatial interaction, help to uncover spatial patterns, reduce estimation bias and increase estimation efficiency (LeSage and Pace, 2009). These models are widely used in economics and social science if the research question has a geographical context or requires to assess neighborhood effects, externalities and detect hot spots. For example Anselin et al. (2000) use spatial econometric methods to analyze the concentration of crime rates and the spill-over effects of crime on neighboring regions, Breustedt and Habermann (2011) apply spatial techniques in agricultural economics and explain differences in farmland prices and Schündeln (2014) models the spatial mobility of migrants across communities in Germany. Although there is some location-based gambling literature which tackles the gambling market from a geographical perspective like Economopoulos (2015), Pearce et al. (2008) and Wardle et al. (2014), to the best of our knowledge no one used spatial techniques to analyze the clustering of EGM supply. We will briefly introduce the major spatial econometric models in the next session.

[5] There can be also spatial dependence in the independent variables. LeSage and Pace (2009) suggest this approach to model externalities which is not the focus of our research.

3.2.4 Spatial Lag Model (SLM)

When there is an autocorrelation in the dependent variable *Y*, the spatial lag model (SLM) is applied which includes average lagged values of the neighbors to the regression (Anselin, 1988). In the context of our research setup this means that the number of EGMs in one region is influenced by the number of EGM of the neighboring regions and vice versa. If a target community is surrounded by communities which have a high EGM supply, the EGM supply of the target community will be positively enhanced by its neighbors which leads to a similarly high EGM supply in the target community. On the opposite a target community with low EGM supply neighbors will be positively affected by the low supply which then dampens the EGM supply in the target community. For the notation we follow LeSage and Pace (2009), equation (2) illustrates a simple SLM model.

$$Y = \lambda WY + X\beta + u \qquad (2)$$

Y is an *N x 1* vector for our depended variable where *N* is the total number of communities in our sample (*N* = 396). The variable *W* is an *N x N* contiguity spatial weight matrix. Furthermore, *X* is an *N x K* vector of explanatory variables similar to regular regression analysis and u is an independent and identically distributed *N x 1* error term. The model has two coefficients, the first coefficient λ is the spatial lag coefficient. This is the core coefficient for our analysis, it determines the strength and direction of EGM clustering comparable to spatial autocorrelation. This parameter can be interpreted as the contagion effect since it implies that neighboring supply affects the supply of the target community. When a clustering of EGM supply is expected, λ takes a positive value. A negative value indicates no clustering but rather a dispersed EGM supply while an insignificant coefficient states that there is no structure in the EGM supply, i.e. there is neither a clustering nor a dispersion. The second model parameter β is a K x 1 coefficient vector for the corresponding exogenous variables. The term WY is defined as the spatial lag term, it applies the neighborhood structure from *W* onto the dependent variable *Y.*

3.2.5 Spatial Error Model (SEM)

The spatial error model is applied when there is spatial autocorrelation in the residuals of the OLS model. The regression function now includes a spatial lagged structure for the errors which is referred to as *spatial heterogeneity*.

$$Y = X\beta + u \qquad (3)$$

$$u = \rho Wu + \varepsilon \qquad (4)$$

The coefficient ρ represents the strength and direction of the spatial structure in the error and ε is normal and i.i.d. white noise. The model assumes an autocorrelation in the error term and not in the dependent or any of the independent variables. It can be interpreted as a weighted average of the individual residuals of neighboring areas.

3.2.6 Kelejian-Prucha Model (KPM)

When there is a lag in the dependent variable as well as in the error term we have to use the Kelejian-Prucha Model which combines the SLM and SEM to one model.

$$Y = \lambda WY + X\beta + u \qquad (4)$$

$$u = \rho u + \varepsilon \qquad (5)$$

This model is particularly useful to disentangle the spatial autocorrelation in the dependent variable while controlling for spatial heterogeneity in the error term. The parameter estimation of SLM, SEM and KPM is based on the maximum likelihood estimation introduced by Ord (1970) for spatial econometric models. There are also other estimation methods like the general method of moments and instrument variable estimators but for simplicity we restrict our estimation to the maximum likelihood approach and apply the model selection criteria proposed by LeSage and Pace (2009). For the estimation we follow Drukker et al. (2013a, 2013b) in STATA.

3.3 Data

The analysis aims to analyze spatial clustering of EGM gambling activity across communities in the German state of North Rhine-Westphalia (NRW). It is important to highlight that observations can differ with regard to their aggregation level: a single house or gambling arcade in a certain street can be considered as the lowest aggregation level, the number of EGMs in communities or municipalities is a medium aggregation level and the stock of EGMs in states or countries is a high level of aggregation, this issue is known as the Modified Area Unit Problem (MAUP). For the data selection it is essential to choose the right balance between low aggregation and data availability. For a clean analysis of spatial clustering of EGMs it is necessary to have as many neighboring observations of EGM arcades as possible, on the other hand low aggregation comes along with a lack of data availability. For example we could employ the analysis for single EGM arcades with precise location coordinates, however doing so we won’t be able to include the socioeconomic environment of a single arcade because of lacking or incomplete data. While we have good data coverage for socioeconomic variables on the community and state level, socioeconomic data for single locations is hardly available. Therefore, we focus for our analysis on German communities within the state of NRW since they offer a variety of observable socioeconomic and geographic characteristics as well as a good regional variation in the stock of EGMs among the communities.

The state of NRW is the largest of sixteen German states and lies in western Germany. We analyze this particular state for a number of reasons. First, it is the most populated state in Germany with 17.5 million people which accommodates nearly a quarter of Germany's total population. Second, it is a densely populated state with the largest metropolitan area in Germany covering major cities like Cologne and Düsseldorf. The state consists of 396 communities which are populated by an average of 44.000 residents each. Third, for the analysis of spatial clustering in the EGM market we require a contiguous dataset i.e. it is necessary to gather EGM data for each community within that state so that we can construct a complete map of gambling activity without missing spots.

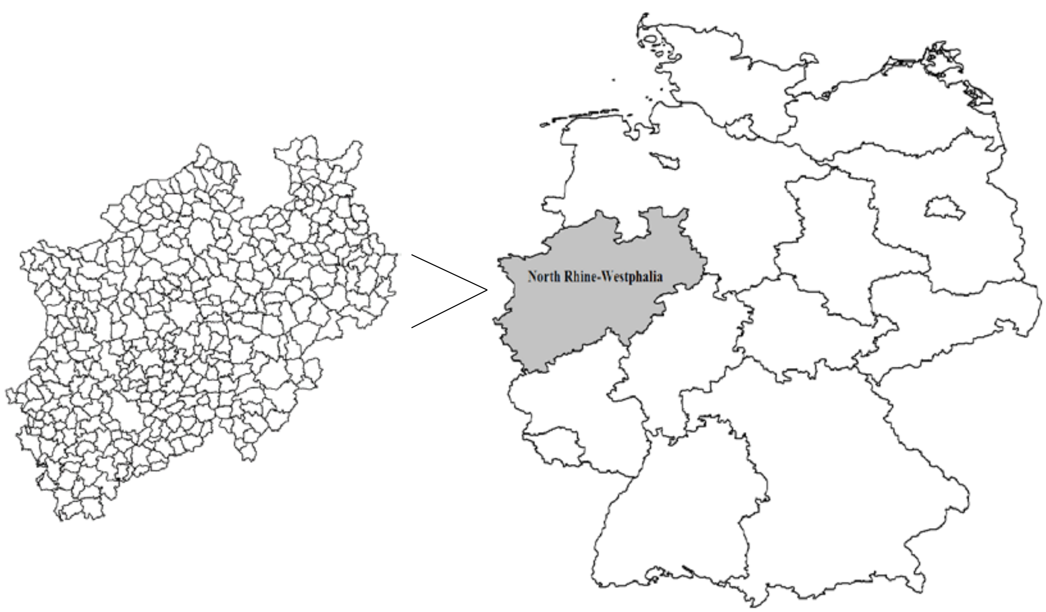

Figure 3.2: Germany, the state of NRW and communities in NRW

Only very few states have complete coverage of the stock of EGMs for each community, NRW is an exception and has nearly complete EGM data. Germany and in particular NRW are ideal to study spatial clustering of EGMs because it is densely populated, has short travel distances from one community to the neighboring communities as well as a high number of communities.

We construct a cross-section dataset which uses three data sources from the years 2011 and 2012. For the stock of EGMs per community we employ data by Trümper and Heimann (2014). They conduct a survey among local regulatory agencies which maintains records on nearly all legal EGMs in the communities. From Trümper and Heimann we draw our gambling supply variables for the year 2012. We correlate the total number of EGMs with the population of a community and find a coefficient of 0.98. We conclude that an analysis in absolute terms is misleading and therefore standardize the EGMs supply per community to one thousand residents, so that we obtain an EGM supply density.

To sharpen our analysis of EGM clustering previous studies suggest to include regional control variables which allow a more distinct description of the socioeconomic environment in the regions of interest and therefore increase the validity of the study. The socioeconomic environment plays a crucial role with regard to the gambling activity within a region (Wheeler et al. (2006), Pearce et al. (2008), Wardle et al. (2012)). The source for the socioeconomic variables is the EU population and housing census for Germany in the year 2011. We standardize each socioeconomic variable to one thousand residents for the same reason as we standardized the EGM supply. Furthermore, we distinguish between rural and urban areas by implementing a dummy variable which takes the value 1 when the region has less than 10.000 residents.

As a third data source, we include the effect of motorway service station (MSS) gambling activity into our analysis by counting the number of motorway service stations within each community. This idea is rather new and has not been used in the gambling literature so far. Truck drivers in Germany are obliged to take a compulsory break after eight hours of driving for the rest of the day. These breaks are usually done at motorway service stations. Besides snacks and beverages service stations also offer the opportunity to play EGMs. We assume that truck drivers are mostly not resident in the community in which they make their break, but their gambling activity will be still addressed to the hosting community. The motorway gambling has to be included into the analysis because bypassing truckers can affect the number of gamblers in a region, but are not included in the number of residents per community. We expect that the number of motorway service stations should positively affect the number of EGM within a community. We collected the location of each motorway service station in NRW and assign them to the corresponding community. A detailed description of our variables can be found in Table 3.1.

Table 3.1: Variable overview

Variable	Measure	Mean	Std. Error
(1) EGM variables			
EGM absolute	Number of EGMs per community	162.44	18.47
EGM relative	Number of EGMs per 1000 residents per community	2.73	0.09
(2) Socioeconomic and geographic variables			
Population	Number of residents	44288.52	4338.03
Male	Density of males	490.14	0.39
Unemployed	Density of unemployed	30.00	0.58
Migrant	Density of immigrants	47.84	1.26
No school degree	Density of residents without education	68.15	1.26
Secondary degree I	Density of residents with a secondary degree I (*10 years of education)*	555.96	2.30
Secondary degree II	Density of residents with a secondary degree II *(12-13 years of education)*	139.21	2.23
Young age	Density of residents between 18 – 24 years	81.09	0.40
	Density of residents between 25 – 29 years	50.52	0.39
Married	Density of married residents	489.50	1.11
Single	Density of singles	378.49	1.21
Divorced	Density of divorced	60.14	0.60
Rural area dummy	= 1 if less than 10.000 residents in community = 0 if more than 10.000 residents in community	0.14	0.02
MSS	Number of motorway service stations in a community	0.17	0.03

Note: All densities are standardized per 1000 residents. Example: For the variable 'Unemployed', 30 residents are unemployed per 1000 residents.

3.4 Results

The main objective of this study is to understand the spatial clustering of EGM supply. Previously Fig. 1 illustrated potential EGM clustering across communities, the next step is to test this initial result analytically by using Moran's I test statistics in Table 3.2. First, we analyze the spatial autocorrelation for the absolute number of EGMs per community. We find moderate and significant clustering of the variable *EGM absolute*. Moran's I statistics shows an autocorrelation of .271 (0.028). For the variable *EGM relative*, i.e. the EGMs per 1000 residents, we still find a highly significant clustering but the autocorrelation is now less severe and takes a values of .101 (0.030). Although the clustering of *EGM relative* is weaker we find that the absolute and the relative EGM supply show significant positive spatial autocorrelation at the 1% level. The result is straightforward: there is significant clustering of EGM activity across regions, even when we look at the population adjusted relative supply.

Table 3.2: Spatial autocorrelation for the number of EGMs per community (absolute) and the number of EGMs per thousand residents per community (relative)

Variable	**Coefficient (Std. Error)**	**Significance**
EGM absolute	0.271 (0.028) ***	0.000
EGM relative	0.101 (0.030) ***	0.001

****1% significant ; ** 5% significant; * 10% significant*

This result supports the contagion argument, however from the literature we expect significant variation in the EGM supply depending on the socioeconomic and geographical composition of the communities. Table 3.3 represents our major results from the regression analysis. For all regressions we used the relative EGM supply as the dependent variable. First we estimated the OLS residuals and included several control variables. Our estimation was overall significant and had a moderate R^2 of 0.41. The second column shows the results from the SLM model. With the inclusion of the control variables we do not find a significant spatial clustering anymore. The spatial lag coefficient λ has a very weak value of 0.003 which is not significantly different from zero. The third column illustrates the SEM results, we find a highly significant autocorrelation in the error term indicating some spatial heterogeneity across regions.

The last column shows the findings for the KPM model. We can see that the combined model is the most likely model. Furthermore, in the KPM setup the spatial lag coefficient λ changed magnitude but remains insignificant. The spatial error is still present and significant.

The coefficient of the control variables present some insight on the gambling supply in different regions. All four models show a lower gambling supply in regions with higher education levels (secondary degree II), we also find significantly higher gambling activity in regions where many young adults in the age from 25 to 29 live. The highest coefficient in our analysis has the rural dummy variable, which ranges from -0.94 to -0.98 depending on the underlying model. This result indicates that rural regions with less than 10.000 residents have on average approximately one gambling machine less per 1000 residents than urban areas, this result is highly significant and consistent across all models. Furthermore, we find weakly significant evidence for higher EGM supply in regions with higher unemployment rates which is in line with other literature. There is also weak evidence for the male variable indicating that gender does not play a role in our aggregated framework. The marital status composition and the density of migrants in a region are insignificant in our analysis. We also find a positive effect of the motorway service stations on the relative EGM supply in the OLS and the KPM models. Although the coefficient is moderately high (0.2), the standard errors are quite high as well. A larger sample size could further improve the significance of the motorway service station effect and should be further analyzed in other works.

Table 3.3: Regression results

Dependent variable EGM supply per 1000 residents	**OLS**	**SLM**	**SEM**	**KPM**
(1) Spatial lag variables				
Spatial lag λ	-	0.003 (0.010)	-	-0.023 (0.014)
Spatial error ρ	-	-	0.034** (0.014)	0.052*** (0.017)
(2) Control Variables				
Male	- 0.029* (0.016)	- 0.029** (0.015)	- 0.029** (0.015)	- 0.030** (0.015)
Unemployed	0.020** (0.010)	0.019* (0.011)	0.019* (0.012)	0.023* (0.012)
Migrant	0.001 (0.005)	0.001 (0.004)	0.000 (0.004)	- 0.0001 (0.004)
No school degree	- 0.009 (0.008)	- 0.009 (0.008)	- 0.010 (0.009)	0.012 (0.009)
Secondary degree I	- 0.006* (0.003)	- 0.006 (0.004)	- 0.006 (0.004)	- 0.006 (0.004)
Secondary degree II	- 0.014*** (0.004)	- 0.014*** (0.005)	- 0.014*** (0.005)	- 0.013*** (0.005)
Young age 18-24	- 0.002 (0.013)	- 0.002 (0.015)	- 0.003 (0.015)	- 0.006 (0.015)
Young age 25-29	0.037*** (0.015)	0.036** (0.017)	0.039*** (0.017)	0.043** (0.018)
Married	- 0.016 (0.011)	- 0.016 (0.012)	- 0.015 (0.012)	- 0.014 (0.013)
Single	- 0.013 (0.011)	- 0.013 (0.012)	- 0.012 (0.012)	- 0.012 (0.012)
Divorced	0.015 (0.015)	0.015 (0.017)	0.018 (0.017)	0.020 (0.017)
Rural Area Dummy	- 0.964*** (0.299)	- 0.962*** (0.260)	- 0.982*** (0.260)	- 0.983*** (0.259)
MSS	0.211* (0.112)	0.208 (0.127)	0.200 (0.125)	0.215* (0.114)
R^2	0.412	-	-	-
Log-likelihood	-	- 686.94	- 684.24	- 682.88
N = 396				

****1%, ** 5%, * 10% significant*

3.5 Conclusion

This paper analyzed the clustering of gambling machines in the German state of NRW. We distinguished between the absolute EGM supply and the relative EGM supply. A first visual inspection showed that EGM supply, absolute and relative, is spatially correlated which indicates a highly clustered EGM supply. This result was confirmed by Moran's I test. A spatial econometric framework was used and included several control variables to disentangle the initial clustering effect. We found that the relative EGM supply is significantly lower in regions with higher education status and significantly higher in regions with higher unemployment rates and larger shares of young adults. This result is consistent with previous studies. We find that gambling activity is significantly lower in rural regions compared to urban regions. We also introduced the effect of motorway service station gambling activity which weakly affects the gambling supply of a region. Finally our spatial econometric analysis shows that the clustering and contagion of EGM supply is mostly explained by the demographic and geographic composition of a region. After controlling for those factors we find no significant clustering of EGM supply in all of our models. In other words the initial clustering effect is completely erased in a distinct framework which provides evidence that potential contagion of EGM supply across communities does not exist.

Our study can be extended in several ways. First, we used a cross section data set to explain the EGM supply by the characteristics of a region. In order to clarify the direction of causality and to delete heterogeneity which we observed in our SEM and KPM model a panel data framework should be considered. Second, we focused on the demographic composition of the region, however other supply and demand factors should be included into the analysis. For example we expect that rental costs or the venue price are substantial factors for EGM arcade location choice as well as regional tax differences. These variables should be incorporated into this framework. Third, we analyzed the clustering effect on the community level, this analysis should be extended to other aggregation levels (e.g. nationwide) to obtain further insight and more consistent results. Overall it can be said that spatial econometric modelling is very useful to analyze EGM supply and should be extended to other markets as well.

4 Refugees Welcome, but not in my Backyard? The Impact of Immigration on Right-Wing Voting - Evidence from Germany[6]

Keywords

Immigration, refugee crisis, right-wing voting, voting behaviour, spatial econometrics

Authors

Johannes Jasny, Tilman Becker

ABSTRACT

During the last few years, Europe has faced an unprecedented influx of immigrants, commonly known as the refugee crisis. The refugee crisis was the major topic in the 2016 German state elections. This paper studies whether immigration in voter's neighbourhoods is a driving factor of the rise of Germany's major right-wing party Alternative fuer Deutschland (AFD) and the decline of Angela Merkel's centre ruling party the Christian Democratic Union (CDU). We use the 2015 refugee crisis as a natural experiment to study the short-run impact of refugee presence on the voting behaviour in German municipalities. This is the first study to use a spatial econometric framework combining small-scale immigration data, election data, and a set of socio-economic factors. Our main finding states that the local immigration boosted AFD votes but did not affect CDU votes directly. Instead, the CDU gained fewer votes in regions which perceived immigration indirectly, i.e. had higher immigration in neighbouring municipalities.

---

[6] This article has been submitted to the *IZA Journal of Development and Migration* and is currently under revision.

4.1 Introduction

In recent years, the world has faced the second largest refugee movement since World War II. The increase in immigration is largely due to the high numbers of refugees who escape from the Syrian civil war and an increasingly instable Middle East.[7] Immigration to Europe, and in particular to Germany, was further fostered in autumn 2015 when German chancellor Angela Merkel gave Syrian refugees permission to travel from the EU border to Germany.[8] As a result, the total number of registered asylum seekers in Germany reached 1.1 million by the end of 2015 (BAMF, 2016). The refugee influx came suddenly and left little room for an elaborate public debate, which could have illuminated chances and potential risks of the arriving immigrants. Throughout the critical year 2015, the majority of politicians as well as the mainstream media followed Merkel's refugee-friendly stance and ignored potential negative aspects of immigration which fuelled popularity of the right spectre (Haller, 2017). Merkel's ruling centre party, the *Christian Democratic Union* (CDU), unilaterally promoted the advantages of immigration and downplayed potential risks and economic burdens caused by the refugee influx. Shortly after the immigration influx, a federal election took place in spring 2016. The election acted as an evaluation of the immigration policy from the previous year and the refugee crisis was the major election topic in 2016. Angela Merkel was the tragic protagonist of the election. Her party lost one quarter of votes while Germany's major right-wing party, the *Alternative fuer Deutschland* (AFD), tripled the votes and became the third largest party. The election outcome evokes the research question for this paper. Did the local presence of refugees in the municipalities lead to the rise of the German right-wing and the downfall of Angela Merkel and the CDU party? In other words, did micro-level exposure to refugees increase local right-wing voting and decrease CDU voting, or was the development a trend which is independent of the refugee allocation? Our research question contributes to the rising interest in the influence of immigration on election outcomes. The refugee influx in 2015 acts as an ideal natural experiment to study the effect of a sudden immigration shock on short-run voting behaviour. The event is similar to other famous immigration shocks, such as the Marial Boatlift from 1980 in the sense that a large number of immigrants from one predominant ethnic group migrate in a short period of time to a certain location. The predominant ethnic group, which accounts for 35.9 per cent of the total immigration in 2015, were Syrians (BAMF, 2016).

---

[7] For the sake of simplicity, we will not distinguish whether an asylum seeker is already a registered refugee, is still in the registration process, or has not been registered yet but claims to receive refugee status at some time. Since the registration of asylum seekers is an ongoing process, we will refer to all three groups as refugees.

[8] Merkel allowed refugees to travel from Hungary to Germany by train over increasing tensions in Hungary. According to the then German Minister of the Interior, Thomas de Maizière, Merkel gave order to the border police on 4 September 2015 to let refugees pass. This decision is a violation of the Dublin regulation which obliges asylum seekers to register at the borders of the EU.

The refugee crisis is a unique natural experiment for two reasons. First, within a year, more than one million immigrants arrived in Germany, which is the highest immigration number to a single European country in recent history. The sheer mass of people required quick allocation and housing solutions. The housing market does not provide the necessary volume of accommodations at a sufficiently low price. Hence, the refugee allocation was managed by the German government. Since vacancies were scarce, new refugee shelters were built and existing houses were converted to refugee homes. This situation creates a quasi-experimental setting, because most of the contemporary refugee accommodation were either non-existent or served another purpose before the immigration influx in 2014. Second, the refugee crisis was the main topic in German media throughout 2015/2016 and the major political issue in the elections in spring 2016. Due to the catenation of events, the ensuing election acts as an evaluation of the refugee influx and the deciding factor at the ballot box during the 2016 election (Haller, 2017). Although immigration and the refugee crisis were still subject of discussion in later elections, the issue was most important during the spring elections of 2016, shortly after immigration peaked. Figure 4.1 illustrates the uneven distribution of refugee accommodations and the AFD vote shares in 2016 for a total number of 1,101 municipalities in the German state of Baden-Wuerttemberg (BW). Sixty-one per cent of the municipalities accommodate refugees, while thirty-nine per cent do not accommodate refugees.[9] Our identification strategy exploits the regional variation of refugee settlement, voting behaviour, and socio-economic differences among the municipalities in the state of BW. We use a spatial econometric framework, which allows modelling spill-over-effects among the regions and reduces estimation bias. The framework is able to model two channels through which refugee presence can influence election outcomes directly and indirectly. We distinguish between direct refugee presence, i.e. the backyard effect, and indirect refugee presence, i.e. the neighbourhood effect. The backyard effect measures how the number of refugees within a municipality affect the election outcome, while the neighbourhood effect describes how the number of refugees in neighbouring municipalities affects the election outcome. Both effects are measures for the degree of refugee presence through physical distance. While the first effect expresses direct proximity, the second effect measures the influence of immigration to neighbouring municipalities, which are further away but still close enough to be perceivable for voters. Although there is a rising number of literature which studies the influence of immigration on voting behaviour, to the best

[9] The distribution among German counties follows a distribution scheme, namely, the *Koenigssteiner Schluessel*, which directs more refugees to higher populated federal states with higher tax incomes. However, the scheme does not apply to municipalities within the states, which results in an increasingly uneven distribution.

of our knowledge, there is no paper that investigates how physical presence of immigrants affects voting behaviour directly and indirectly.

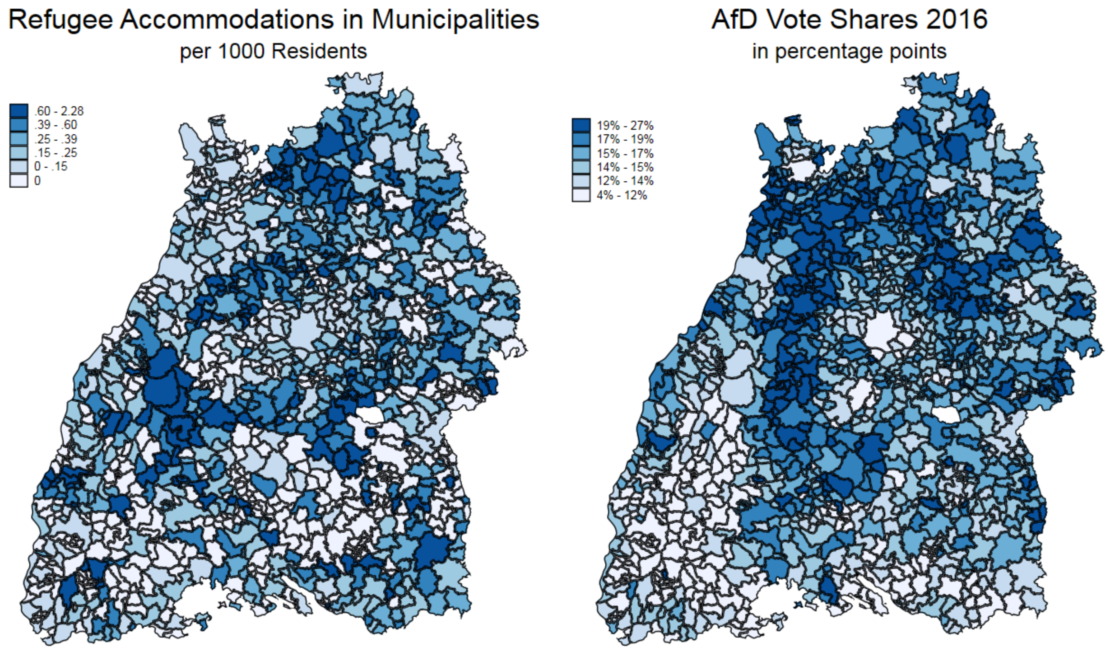

Figure 4.1: (a) The number of refugee accommodations per 1,000 residents, and (b) the AfD vote shares in 2016 in Baden-Wuerttemberg (BW)

4.1.1 Immigration and Voting Behaviour

Recent literature shows mixed evidence for the effects of immigration on voting behaviour. A significant share finds a positive relationship between right-wing voting and immigration (Barone et al., 2016; Halla, Wagner and Zweimüller, 2017; Otto and Steinhardt, 2014). Barone et al. (2016) exploit municipality-level data using an IV estimation strategy to analyse the impact of immigrants on the centre-right and centre-left parties in Italy. They conclude that immigration causes an increase in votes for the centre-right coalition. This result is accompanied by a loss of votes for centre and centre-left parties as well as lower turnout rates. Halla, Wagner and Zweimüller (2017) study the influence of immigration on Austria's major right-wing party (FPOE). They find significantly higher vote shares for FPOE in regions with higher migration inflows. They argue that voters are concerned about the labour market effects as well as a decline in quality of their neighbourhood caused by immigration. There are also some studies which find either no relationship or a negative relationship between right-wing parties and immigration (Gehrsitz and Ungerer, 2017; Steinmayr, 2016; Dustmann, Vasiljeva and Damm, 2016). The studies by Gehrsitz and Ungerer (2017) and Steinmayr (2016) focus on the 2016 refugee crisis as well. Gehrsitz and Ungerer analyse the short-run effects of refugee exposure on AFD vote shares in Germany using county level data. Their analysis suggests that regions with higher refugee presence had no significantly different AFD vote shares than regions with lower refugee presence. Although using county level data is common practice in political geography, it allows vague conclusions about micro-level exposure to refugees due to an aggregation bias.[10] Steinmayr (2016) deals with this bias by using municipality level data from the 2015 federal state elections in Austria. He uses an IV approach and defines pre-existing group accommodations as an instrument for refugee presence. Steinmayr's findings suggest that municipalities, which hosted more refugees, exhibited lower vote shares for Austria's major right-wing party.[11] These results are in contrast with the anti-immigration surge in Austria at the macro level. He argues that micro exposure to refugees increases the sympathy for refugees and therefore dampens local right-wing voting. However, the results are in line with the contact hypothesis which states that the interaction between members from different (racial) groups can reduce prejudice and discrimination (Allport, 1954; Dixon, Durrheim and Tredoux, 2005; Ellison, Shin and Lael, 2011). Although the IV approach is a useful tool to reduce endogeneity, Steinmayr's instrument is potentially flawed, because group accommodations have a limited ability to represent the true refugee presence. One innovation in our paper is the interaction between spatially distributed variables. While most of the literature makes a commendable effort to discuss potential endogeneity between immigration and voting behaviour, the majority ignores the spatial dimension. Figure 4.1 shows how regions with high AFD vote shares are located in a "circle" around the Stuttgart metropolitan area. Hence, it is evident that AFD vote shares are not randomly distributed. Ignoring this fact is another source of bias. We tackle the shortcomings of the literature in three ways: First, we use municipality level data which is finer scaled than county level data. Second, we use factually recorded refugee settlement data instead of proxy variables. Third, we implement a spatial econometric framework which accounts for biased estimates that result from misspecification of the spatial structure.

4.1.2 Socioeconomic Environment and Voting Behaviour

Several studies show that immigration is not the sole cause for the rise of right-wing parties and political polarisation. Demographic differences such as race, gender, and age as well as economic prosperity measures such as income and unemployment crystallise as important factors for the decision at the ballot box. According to the Kramer model (Kramer, 1971; Kiewiet and Udell, 1998), high income and low unemployment increase the chance of re-election of the ruling party. Hersh and Nall (2016) argue that voting behaviour in the US is a function of income as well as ethnic and racial heterogeneity. Furthermore,

---

[10] For further information, see: "Modifiable area unit problem" (MAUP).

[11] The major Austrian right-wing party is the *Freedom of Austria* (FPOE) party.

they conclude that partisanship is stronger in regions with higher cultural diversity. Noelle-Neumann (1980) explains that individuals with socially and economically unprivileged background, i.e. individuals with low income or unemployed, are more likely to elect opposition parties. McCarty, Poole and Rosenthal (2006) argue that political polarisation coincides with income inequality in the US. Garand (2010) argues that both major US parties', i.e. the Democrats' and the Republicans', ideological positions are more polarised during times of higher inequality in the US. Hence, the ideological positions of parties are a function of state-income inequality. Funke, Schularick and Trebesch (2016) study political extremism after financial crises for different European countries during the period between 1870 and 2014. They argue that right-wing parties increased their vote share significantly after financial crises while government majorities shrank. Financial crises induce a reallocation of wealth which can disrupt political ideology. Han (2016) analyses links between income and voter support for radical right-wing parties in Western Europe. He concludes that individuals with lower income are more likely to vote for radical right-wing parties. Furthermore, he finds a positive, yet weakly significant link between right-wing voting and unemployment. Krieckhaus et al. (2014) study the effect of higher income disparity on the support for democracy in general. They describe opposing forces, which can either increase or decrease the support for democratic parties. On the one hand, increasing income inequality could strengthen support for democratic parties as democracies can work as an income redistribution mechanism. On the other hand, they argue that inequality can induce a "political disillusion" which leads to dissatisfaction and ultimately passes voters into the hands of right-wing parties. They inspect forty democracies and find that an increased inequality reduces the acceptance of democracy among all social classes. Dahlberg, Edmark and Lundqvist (2012) study the relationship between increased immigration and preference for wealth redistribution in Sweden. They find a significant negative influence of immigration on the support for redistribution. This effect is particularly strong among high-income earners. Hence, high-income earners are more likely to oppose financial support for refugees as well. Following the literature, we include a set of socioeconomic control variables to account for political preferences among various social milieus.

4.2 Background and Data

For a better understanding, it is useful to shed some light on the political landscape in Germany. In 2016, there were five federal state elections in Germany, namely in Baden-Wuerttemberg (BW), Berlin, Mecklenburg-Western Pomerania, Rhineland-Palatinate, and Saxony-Anhalt. Due to incomplete or lacking data regarding the latter four states, we restrict the analysis to the largest of the five states, namely, Baden-Wuerttemberg. The state of BW has a population of 10.9 million which corresponds to 13 per cent of the total German population. BW is divided in 1,101 municipalities with an average of 9,900 residents per municipality. For the analysis, we chose the election year 2016 due to its temporal closeness to the refugee crisis.[12] Since the election took place during spring of 2016, it is considered an immediate policy evaluation of the preceding refugee crisis (Haller, 2017). The majority of all votes were split among four parties. The *Green Party* (30.3%), the conservative *Christian Democratic Union* (CDU, 27%), the right-wing *Alternative fuer Deutschland* (AFD, 15.1%) and the *Social Democrats* (SPD, 12.7%).[13] These four parties gathered eighty-five per cent of all votes which makes them the key parties in the election. During the refugee crisis, Angela Merkel of CDU stood out politically as the main proponent of immigration and the leader of the ruling coalition. The CDU party was the main loser in the 2016 election since they lost twelve percentage points. The Green Party is the ruling party in the state of BW but not a ruling party in parliament. Although the Green Party won the 2016 election in BW, it is not accountable for the refugee crisis at the federal level where it acts as an opposition party. The right-wing AFD party is a successful newcomer that was founded in 2013. Within three years, the party tripled its votes and became the third largest party in the federal election 2016. In comparison to the other parties, the AFD openly opposed the long-term settlement of refugees in Germany. The Greens and SPD were either supporting Angela Merkel's immigration-friendly policy or remained vague in their political positions throughout 2015 and spring of 2016. As a conclusion, we focus on two parties of interest: Angela Merkel's CDU and the right-wing AFD party. Both parties were in the spotlight throughout the crisis with the AFD acting as Germany's major anti-immigration party and the CDU being the ruling party and holding responsibility for the crisis. Recent literature shows that the AFD and CDU party were the key players in the immigration debate in the 2016 elections (Haller, 2017).[14]

For our analysis, we require three categories of data: election data, refugee settlement data, and socioeconomic data at the municipality level. Voting data is made available by the Federal Statistical Office of Baden-Wuerttemberg (FSOBW, 2016). We measure the election outcome for each party in percentage points per municipality. Furthermore, we include the vote share of the 2013 election to control for previous election outcomes, where refugees had not been an issue yet. We also account for the election turnout rate from previous elections since the refugee crisis might have mobilised non-voters to vote. Table 4.1 gives a variable overview. Second, the Ministry of Social Policy and Integration provided refugee settlement data. We measure the immigration shock by the number of refugee accommodations per

[12] The federal state election in Baden-Wuerttemberg was held on 13 March 2016.

[13] The Green Party is the ruling party in Baden-Wuerttemberg. It was an opposition party in the German parliament at the date of the

[14] We have tested this hypothesis and found no significant impact of refugee presence on the green party as well as the social democrat party election outcomes.

municipality. The record date is March 2016, shortly before the state election took place. We do not distinguish between newly-built refugee shelters and existing buildings that were converted into refugee accommodations. At the record date, there were 2,452 refugee accommodations in BW hosting a total number of 185,000 refugees. This makes an average of 2.23 refugee locations per municipality and an average refugee exposure of 168 refugees per municipality. As the population and the number of refugee locations are correlated (r = .70), we standardise refugee accommodations per one thousand residents.

Table 4.1: Variable overview and descriptive statistics

Variables	**Measure**	**Mean**	**Std. Dev.**
Ref home	Refugee accommodations per 1,000 residents	0.26	0.35
policy variables			
AFD 16	Vote share of the AFD party in state election 2016 in percentage points	15.35	3.69
CDU 16	Vote share of the CDU party in state election 2016 in percentage points	31.10	7.15
AFD 13	Vote share of the AFD party in parliamentary election 2013 in percentage points	5.21	1.35
CDU 13	Vote share of the CDU party in parliamentary election 2013 in percentage points	50.18	6.87
Turnout rate	Turnout rate in the parliamentary election in 2013 in percentage points	71.93	4.93
socioeconomic controls			
GDP	Gross domestic product per 1,000 residents (in €)	35.57	7.16
Unemployment	Unemployment rate in percentage points	3.74	0.53
Mean age	Mean age of the population	43.53	1.64
Males	Male population per 1,000 residents	495.10	12.01
Married	Married population per 1,000 residents	486.09	23.30
Population density	The number of residents per hectare	3.24	3.45
Migrant background	German citizens with migrant background per 1,000 residents	160.45	42.07
Low education	Highest degree = Basic education (second step) per 1,000 residents	337.87	35.71

N = 1,101

Third, the Federal Statistical Office of Baden-Wuerttemberg and the Federal Institute for Employment Research provided socioeconomic data. We include a set of socioeconomic control variables to account for voting differences among social groups. The main variables are GDP per municipality, the unemployment rate, mean age, gender, marital status, the share of residents with immigration background, and the share of residents with basic education. We also include the population density per hectare to distinguish between rural and urban voting preferences. Although panel data is undoubtedly an integral part to reduce endogeneity, it fails to address our research question. There are two reasons for which we chose a cross section approach. First, the AFD party became an anti-immigration party in recent years. It evolved from a Euro-sceptic party, which focused on European economic policy, to a right-wing party with a restrictive stance towards immigration. In other words, voters who oppose immigration did not necessarily vote for the AFD party in 2013. Against this backdrop, it is unreasonable to compare election outcomes from 2013 and 2016. Second, the party was established in 2013 and participated in only two federal elections, i.e. 2013 and 2017. As for the state elections, the party participated in only one election cycle. Due to the recent transition in AFD's political agenda and the lack of election periods, we chose a cross section approach.

4.3 Model and Empirical Implementation

Using municipality level data, we build a simple model where the vote share for a political party depends on the number of refugee accommodations and a set of policy and socioeconomic control variable. The policy and socioeconomic control variables are pooled in variable X.

$$vote\ share = \alpha_1 refhome + \beta\, X + u \qquad (1)$$

We exploit the variation among i municipalities. The parameter α_1 is the marginal effect of refugee presence (backyard effect), the parameter vector β represents the influence of socioeconomic, political and regional control variables. Since all the data has a spatial context, i.e. all variables are attributes of spatial units, we must account for spatial spill-overs among the variables. Ignoring a spatial structure by assuming randomness will lead to biased and inefficient coefficients (Anselin, 1988). This makes spatial econometric techniques necessary for estimation. Non-randomness in space can occur when, for example, right-wing voting concentrates in a certain neighbourhood resulting in right-wing voting clusters. Another advantage of spatial models is the analysis of the neighbourhood effect itself. One motivation for the neighbourhood effect stems from the literature review. Immigration to certain regions can lead to better election outcomes for right-wing parties. However, none of the literature has measured how immigration to neighbouring regions affects the voting behaviour. Spatial econometric models allow to distinguish between the number of refugees in a municipality, i.e. the backyard effect, and the number of refugees in neighbouring municipalities, i.e. the neighbourhood effect. We are not only interested in the backyard effect on the voting outcome but also in the neighbourhood effect. Before we define the latter effect, we require a plausible definition of neighbourhood. The most common way to implement a neighbourhood structure in the sense of spatial econometrics is the spatial weight matrix (Drukker et al. 2013a). The spatial weight matrix defines which spatial entities are neighbours. A common way to define a neighbourhood is a common border between two geographical units. If two municipalities *i* and *j* share a common border, the single element of the matrix w_{ij} takes a value of one or zero otherwise. The binary neighbourhood relations among all spatial entities form a neighbourhood structure which is defined in a contiguous spatial weight matrix W.

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{ij} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{pmatrix} \quad (2)$$

With

$$w_{ij} = \begin{cases} 1 & if\ i\ is\ neighbour\ to\ j \\ 0 & otherwise \end{cases} \quad (3)$$

The contiguous spatial weight matrix W is an N x N matrix where the number of rows and columns corresponds to the total number of municipalities in our sample (N = 1,101). We define the diagonal elements of the matrix as zero. There are many possible definitions of neighbourhood, however, we use the common border contiguity definition.[15] Wang, Kockelmann and Wang (2013) argue that variations in the spatial weight matrix setting do not significantly change parameter estimation. For the sake of simplicity, we follow this argument and restrict the model to one type of neighbourhood structure. The spatial weight matrix is the starting point for most spatial econometric models. An in-depth overview on spatial econometric models is provided by Anselin (1988), Elhorst (2010), and LeSage and Page (2009). There are three common ways to implement a spatial structure, i.e. the spatial weight matrix, into an econometric model, namely, in the dependent variable, in the independent variable, and in the error term. Implementing all three types of structure is unfeasible and leads to misspecification (Elhorst, 2010). LeSage and Page (2009) argue that implementing spatial structure in the dependent and independent variables results in unbiased coefficient estimates in a majority of cases. Potential spatial structure in the error term does not affect consistency, however, it improves estimation efficiency. LeSage and Page suggest

[15] For example, w_{ij} can take the value of one, if two locations lie within a certain distance to another. Any other boundary measure can be taken as well. Another method is the implementation of an inverse distance, so that the neighbourhood is measured in terms of distance rather than connectivity.

estimating the Spatial Durbin Model (SDM) with spatial structures in the dependent and independent variables. The SDM is defined as follows.

$$vote\ share = \alpha_1 refhome + \alpha_2 W\ refhome + \lambda W\ vote\ share + \beta\ X + \theta WX + u \qquad (4)$$

λ is the spatial autoregressive parameter for the dependent variable and can be interpreted as the average effect of neighbouring vote shares. Note that λ is a scalar. When λ is positive, a high vote share for a party in municipality i reinforces the vote share for the same party in neighbouring municipality $-i$. λ is also positive, when a low vote share further weakens the vote share for the same party in neighbouring municipalities. In other words, a positive λ represents the degree of spatial clustering of similar values in the dependent variable at nearby locations.[16] The parameters, α_2 and θ constitute the spatial structure in the independent variables. Parameter α_2 represents the average effect of neighbouring refugee accommodations on the initial community, i.e. the neighbourhood effect. The second parameter θ shows the average influence of neighbouring socioeconomic and policy variables. u represents the error term. If there is no spatial structure in the independent variables, the parameters α_2 and θ are insignificant. In that case, the Kelejian-Prucha Model (KPM) can improve estimation efficiency. The model omits the spatial structure in the independent variables but adds a spatial structure to the dependent variable and the error term. The KPM model is defined as follows.

$$vote\ share = \alpha_1\ ref\ home + \lambda W\ vote\ share + \beta X + u$$

With

$$u = \rho W u + \varepsilon \qquad (5)$$

ρ is the coefficient which shows the average effect of neighbouring disturbances that improves the efficiency. Therefore, ρ is considered a nuisance parameter. ε is an i.i.d. noise term. If there is no structure in the independent variable, but a spatial structure prevails in the dependent variables as well as the error term, the most efficient and consistent model is the Spatial Durbin Error Model (SDE).

$$vote\ share = \alpha_1\ ref\ home + \alpha_2 W\ ref\ home + \beta X + \theta WX + u$$

With

$$u = \rho W u + \varepsilon \qquad (6)$$

[16] The case of a negative λ is a rather exotic and less common case.

We follow the model selection procedure proposed by the literature and estimate three models for each party. First, we estimate a simple OLS model without any spatial interactions for both parties. Second, we estimate the SDM model from equation (4) for the parties to check the existence of the neighbourhood effect. As for the third model, we follow the model selection procedure proposed by LeSage and Page. We estimate equation KPM model from equation (5) by the generalised spatial two stage least squares estimator (GS2SLS) for the AFD party. Since we do not find significant spatial autocorrelation for CDU vote shares, we estimate equation (6) for the CDU party for higher efficiency. Ord (1975) and Anselin (1980) provide a detailed description for the GS2SLS estimation procedure. For estimation, we follow the procedure of Drukker et al. (2013b) in STATA 15.

4.4 Refugee Allocation and Endogeneity

There is some potential for endogeneity, primarily because refugee allocation is not a random process and depends on various factors itself. Following the literature as well as the allocation schemes, we identify five major factors that influence the refugee allocation, namely, (1) population density, (2) tax revenues, (3) presence of initial entry centres (IEC), (4) the vacancy rate, and the (5) political affiliation of the decision makers in the municipalities. The allocation of refugees follows different schemes depending on the administrative level. For instance, the allocation regime among states differs from the allocation among municipalities. The asylum seeker management is a top-to-bottom system. This means that the allocation starts at the highest administrative level (federal states), is then followed by the allocation among counties, and ends at the lowest administrative level (municipalities). In practice, every asylum seeker picked up at the German border undergoes a quick check by the border police followed by a registration in a federal asylum seeker database (EASY-System). The EASY-System assigns a number of refugees to each of the sixteen states based on a quota, namely, the "*Koenigssteiner Schluessel*". The quota allocates refugees based on the state's population and tax revenues and ensures a rather balanced distribution of the burden.[17] Refugees assigned to a certain state are send to one of the large-scale initial entry centres (IEC) within the state, where the majority of the paperwork takes place. IECs are rather large housing-facilities and provide short-term housing for several thousand refugees as long as immigrants are waiting for the asylum admission. Municipalities that operate IECs subsequently receive fewer refugees as municipalities without IECs. Hence, the number of immigrants correlates negatively with the number of IECs. After successful registration at the IEC, the refugees are redirected to the counties (*Landkreise*) within

[17] The distribution rests to one third on the population and two thirds on tax revenues.

the same state. The allocation to counties follows the same quota.[18] Finally, the authorities within the counties allocate the refugees to the municipalities, which ultimately provide middle- and long-term refugee accommodations. It is important to note that, in contrast to the allocation among states and counties, the quota is not obligatory for the allocation among municipalities. The allocation among municipalities mostly depends on the presence of IECs, the population density, and, most importantly, on the number of vacant flats and houses. The immigration influx happened in a short time span and suitable accommodations were scarce. House owners received a lucrative rent from the German government for hosting refugees. Consequently, municipalities with high vacancy rates have higher numbers of refugee accommodations.[19] Hence, a high vacancy rate induces a self-selection bias. Finally, the refugee distribution could be affected by the party affiliation of the incumbent mayor. For example, the green party supports immigration and has a rather refugee-friendly stance. Hence, a green party mayor could decide to provide a larger number of refugee accommodations. We believe the last argument is somewhat controversial, because the decision of a mayor is not bound to the party leadership. Furthermore, hosting larger numbers of refugees voluntarily leads to a higher financial burden for the municipalities and could threaten the popularity of a mayor. As a result, it would be politically rational to oblige to the assigned immigration numbers without deviating. In order to deal with endogeneity, we run a regression where the number of refugee accommodations is a function of IECs, GDP, vacancy rate, and the political affiliation. Since tax revenue data is only available for a quarter of municipalities, we use regional variation in GDP as a proxy for tax income. Furthermore, we account for the influence of population by standardising the refugee accommodations per one thousand residents. For the political affiliation, we include a dummy variable which takes the value one if the mayor affiliates to the CDU party. Note that the majority of mayors have either no party membership or affiliate to the CDU, while not a single mayor affiliates to the AFD party and only few mayors are in other parties. The estimates are shown in table 4.2. The regression shows that the refugee allocation can be partly explained by initial entry centres and the vacancy rates. Both coefficients are significantly different from zero and in line with the official allocation procedures. IECs correlate negatively with refugee accommodations and the vacancy rate is beneficial for refugee settlement. Furthermore, we confirm that the refugee allocation is unaffected by variations in GDP. We also find that the political affiliation of the mayor does not affect the refugee allocation procedure. Although two variables are significant, the explanatory power of the regression is weak. Note that the coefficient of determination is quite low with $R^2 = 3.4\%$. In other words, the refugee allocation is not

[18] The protracted registration procedure led to overcrowded IECs during 2015. Consequently, many applicants were redirected to the counties (*Landkreise*) before their application was completed.
[19] Note that existing houses and apartments which host refugees are also considered refugee accommodations.

completely random, but it is also little affected by the main four factors that can induce endogeneity. The variables are considered as the main drivers which could potentially cause a non-random allocation. However, they explain only a small proportion of the refugee allocation mechanisms.

Table 4.2: Determinants of refugee allocation

Dependent variable: Refugee accommodations per 1000 residents	
IEC	- 0.027 *** (- 4.33)
GDP	0.001 (1.25)
Vacancy rate	3.977 *** (4.43)
Mayor dummy (CDU)	- 0.020 (- 0.84)
R^2	0.034
N = 1,101	
*** 1%, ** 5%, * 10% level of significance	

Although the estimates suggest that the bias caused by non-randomness is minimal, we control for it by adding the predicted values of the estimation from table 4.2 as a regressor ($\hat{y}$) to the estimates in the following chapter.

4.5 Results

The aim of this paper is to study how immigration in voter's neighbourhoods affects the rise of the right-wing AFD party and the decline of Angela Merkel's CDU party. We measure the influence of refugee immigration through two effects: the backyard effect (refugee presence) and the neighbourhood effect (indirect refugee presence). The coefficients α_1 and α_2 represent the magnitude and direction of both effects respectively. The estimates for the AFD party are summarised in table 4.3. Regarding the first coefficient α_1, we find strong evidence for the backyard effect on AFD vote shares. All three models indicate a positive influence of refugee presence on AFD vote shares. The magnitude of the coefficients varies slightly between 0.59 and 0.67.

Table 4.3: Estimates for the AFD vote share in the year 2016 in percentage points

Dependent variable: AFD '16 vote share	**Standard (OLS)**	**SDM (GS2SLS)**	**KPM (GS2SLS)**
(1) Refugee variables			
Backyard effect ($\boldsymbol{\alpha_1}$)	0.667 **	0.594 **	0.678 ***
	(2.47)	(2.27)	(2.99)
Neighbourhood effect ($\boldsymbol{\alpha_2}$)	-	0.046	-
		(0.06)	
(2) Socioeconomic variables			
AFD 13	1.087 ***	1.033 ***	0.878 ***
	(12.05)	(15.41)	(13.26)
Turnout rate	- 0.179 ***	- 0.170 ***	- 0.151 ***
	(- 7.14)	(- 8.53)	(- 8.34)
GDP	0.035 ***	0.024	0.024
	(2.85)	(1.66)	(1.55)
Unemployment	1.234 ***	1.094 ***	0.938 ***
	(7.47)	(6.12)	(4.80)
Mean age	- 0.472 ***	- 0.456 ***	- 0.415 ***
	(- 6.32)	(- 7.64)	(- 7.41)
Males	0.005	0.008	- 0.001
	(0.41)	(0.92)	(- 0.13)
Married	0.037 ***	0.039 ***	0.038 ***
	(7.23)	(9.60)	(9.81)
Population density	0.078 ***	0.066 **	0.102 ***
	(2.81)	(2.04)	(3.02)
Migrant background	0.002	0.002	0.003
	(0.96)	(0.96)	(1.39)
Low education	0.012 ***	0.012 ***	0.012 ***
	(4.43)	(4.04)	(3.84)
(3) Spatial parameters			
λ	-	0.010 ***	0.007 ***
		(3.70)	(3.10)
θ	-	✓	-
ρ	-	-	0.094 ***
			(20.77)
$\hat{y}$	5.521 ***	5.170 ***	4.377 ***
	(2.93)	(3.52)	(3.21)
R^2	0.42	-	-

N = 1,101
*** 1%, ** 5%, * 10% level of significance

The direction of the effect is clear: A higher refugee presence significantly boosts AFD vote shares. To be precise, one additional refugee accommodation per one thousand residents increases the AFD vote share by 0.6 percentage points on average.[20] We find a positive and significant spatial autoregressive structure in the dependent variable, which justifies the use of spatial econometric models. Both spatial models, SDM and KPM, indicate that the intensity of AFD vote shares is spatially clustered in certain regions. In other words, voters with AFD preference are likely to live in nearby municipalities. With regard to the second coefficient (α_2) we find no evidence for the neighbourhood effect on AFD vote shares.

When it comes to the CDU, we do not find convincing evidence for the backyard effect (table 4.4). Although α_1 is negative in all models, it is only significant in the least likely model (OLS). Hence, refugee presence does not have a direct negative influence on the CDU vote shares. However, we find some evidence for an indirect influence, i.e. the neighbourhood effect. The SDM and SDE models confirm that refugee presence in neighbouring municipalities negatively affects the CDU election outcome. In addition, we have also conducted several robustness checks by including and excluding various regressors. In all of our variations, the direction and significance of both refugee effects remain unchanged and the coefficients vary slightly. Overall, we conclude that the influence of the immigration shock is rather robust for both parties. The backyard effect was beneficiary for the AFD party in the 2016 elections while the neighbourhood effect was disadvantageous to the CDU party.

Regarding other policy variables, we find a strong positive correlation between the 2013 election outcome and the 2016 election outcome for both parties. Regions that had a high vote share for either party in the previous election have a similarly high vote share for the same party in the recent election. This indicates some regional and temporal stickiness of party preferences. For both parties, the control variables for the previous election, *AFD 13* and *CDU 13*, exhibit the highest t-values among all variables. In fact, the previous election is the driving force for the high coefficient of determination for the CDU party. It explains nearly eighty per cent of the CDU vote share in 2016. Regarding the election turnout, we find that AFD voting is more pronounced in regions with traditionally high voting abstinence. In 2016, the AFD gained more votes in regions that had a low turnout rate in the 2013 election. On the contrary, CDU vote shares are higher in regions with low voting abstinence. This result indicates that right-wing voting is an attractive option in regions with a traditionally higher voting abstinence. The AFD is often considered a "catch all" party for protest voters. One alternative to protest voting is voting abstinence. Hence, it is possible that the party mobilised a significant share of non-voters which changed to the AFD party. Unfortunately, the AFD is a

[20] The true coefficients in the SDM and KPM model will vary slightly since we have to account for a positive autoregressive structure (λ = 0.007 – 0.010) in the dependent variable.

relatively young party and extensive time series data of election results is not available. Hence, further studies should be conducted in the future to prove this hypothesis.

Table 4.4: Estimates for the CDU vote shares in the year 2016 in percentage points

Dependent variable: CDU ’16 vote share	**Standard (OLS)**	**SDM (GS2SLS)**	**SDE (GS2SLS)**
(1) Refugee variables			
Backyard effect ($\boldsymbol{\alpha_1}$)	- 0.604 *	- 0.334	- 0.113
	(- 1.91)	(- 1.14)	(- 0.43)
Neighbourhood effect ($\boldsymbol{\alpha_2}$)	-	- 0.187 **	- 0.184 **
		(- 2.22)	(- 2.08)
(2) Socioeconomic variables			
CDU 13	0.934 ***	0.925 ***	0.910 ***
	(36.90)	(52.75)	(49.63)
Turnout rate	0.094 ***	0.096 ***	0.111 ***
	(3.40)	(4.33)	(5.25)
GDP	- 0.069 ***	- 0.063 ***	- 0.037 **
	(- 5.29)	(- 3.95)	(- 2.01)
Unemployment	- 0.016	0.000	0.122
	(- 0.09)	(0.00)	(0.54)
Mean age	- 0.046	- 0.045	0.034
	(- 0.49)	(- 0.67)	(0.52)
Males	- 0.014	- 0.014	0.003
	(- 1.08)	(- 1.49)	(0.30)
Married	- 0.025 ***	- 0.026 ***	- 0.024 ***
	(- 4.73)	(- 5.71)	(5.24)
Population density	- 0.053 *	- 0.042	0.015
	(- 1.64)	(- 1.15)	(0.37)
Migrant background	0.001	0.001	0.004
	(0.29)	(0.36)	(1.47)
Low education	- 0.002	- 0.000	0.007 *
	(- 0.50)	(- 0.02)	(1.91)
(3) Spatial parameters			
λ	-	- 0.000	-
		(- 0.03)	
θ	-	✓	✓
ρ	-	-	0.088 ***
			(16.37)
$\hat{y}$	3.471 *	3.802 **	0.425
	(1.69)	(2.31)	(0.27)
R^2	0.81	-	-

N = 1,101

*** 1%, ** 5%, * 10% level of significance

We find some interesting links between the socioeconomic environment and the election outcome. Our results show that right-wing voting is closely related to unemployment. This result is supported by the aforementioned literature. Among all models, the unemployment rate is positively correlated with AFD vote shares at the highest significance level. Unemployment has also a relatively high coefficient. On average, one additional percentage point in the unemployment rate increases the AFD vote share by 0.87 to 1.08 percentage points. With regard to the municipality output, we find no significant relationship between regional GDP an AFD votes.

Although GDP has a positive effect on AFD vote shares in the standard model, the OLS estimator is most likely biased, since it does not implement spatial structures. The better fitting models, SDM and KPM, show no significant influence of GDP on AFD vote shares. In contrast, the CDU party is unaffected by regional variation in unemployment. However, it exhibits significantly lower vote shares in wealthier regions with high GDP. Regarding other demographic factors, our estimates show that the AFD party obtained more votes in municipalities with a certain demographic milieu. The AFD party was more successful in municipalities with younger age, higher share of married couples, and a lower level of education. Moreover, the AFD received more votes in regions with higher population density. Surprisingly, the AFD election outcome is unaffected by the share of residents with immigration background. This group consists of people who live in Germany for a longer period of time, hold a German citizenship, and are entitled to vote. Due to the anti-immigration rhetoric of the AFD party, one could expect that Germans with migration background might averse the party's position and vote against it. Hence, a high share of Germans with migration background could drive down the AFD votes. However, since we use aggregated data, the setup does not allow drawing conclusions on individual voting behaviour. Hence, this result could be the outcome of three scenarios. First, preferences for the AFD party of Germans with migration background do not differ significantly from those of native Germans. Second, higher exposure to Germans with migration background does not lead to higher AFD vote shares which would indicate that the voter population tolerates Germans with migration background. Third, Germans with migration background oppose the AFD party and native Germans oppose those with immigration background. However, both effects work in opposite directions and cancel each other out. Our research setup is unable to disentangle these effects which leaves space for further research.

With regard to the CDU, there is no clearly-cut demographic environment, with the exception of married people, which has significantly higher CDU vote shares. Variation in age, gender, migration background, and population density do not affect the CDU election outcome significantly. We find some evidence for a positive correlation between a low education level and CDU votes in the SDE model. Overall, the results for the CDU party are little surprising, since the CDU is Germany's centre party which attracts voters from various social groups. We also do not find a spatial clustering for the CDU vote shares, which indicates that CDU voting is spatially dispersed. Finally, some remarks on the spatial model parameters. The Moran's I test for spatial autocorrelation finds significant spatial autocorrelation in some of the independent variables as well as the error term. In addition, we find a positive autocorrelation of the dependent variable for the AFD party. Our estimates confirm the existence of spatial autocorrelation. Hence, we conclude that the OLS model is the weakest performing model among the estimates and all estimated spatial models are more consistent and efficient.

4.6 Conclusion and Discussion

This paper studies the short-run impact of immigration on right-wing voting behaviour. We use the refugee crisis from 2015/2016 as a natural experiment to study the influence of a large and unexpected immigration influx on the voting behaviour in German municipalities during the 2016 election. We model the influence of immigration, directly and indirectly, through two channels, i.e. the backyard effect and the neighbourhood effect. The former effect measures the impact of refugee presence in municipalities on the election outcome while the latter effect measures the impact of refugee presence in neighbouring municipalities. Our estimates for the right-wing AFD party are straightforward - the party clearly benefits from the immigration influx since the backyard effect is positive and significant across all models. We find no evidence for the neighbourhood effect on AFD vote shares. Our results support the hypothesis that a significant share of the population opposed refugee settlement in their home municipality by voting for the right-wing AFD party. We suppose that voters in municipalities with larger numbers of refugee settlements found the AFD party program more appealing, since it opposes refugee settlement in Germany and criticises Merkel's refugee policy. With regard to the massive loss of the CDU party in the election, it is surprising that local refugee presence did not directly affect CDU vote shares. Hence, other factors have caused the decline of Angela Merkel's CDU. According to Haller (2017) the media coverage about the refugee crisis was one major factor which caused CDU vote shares to plummet. He argues that the subjective perception through media consumption was rather important to form an opinion about the matter and stimulated disagreement with the refugee-welcoming policy of the CDU. We do not prove Haller's argument point in our work, however, it is a potential extension to our work in the future. Although we do not find evidence for the backyard effect on CDU votes, we do find some evidence that the neighbourhood effect negatively affects CDU vote shares. This is an interesting result because it implies that direct exposure to immigration does not decrease CDU votes while indirect exposure does. The contact hypothesis might partly explain this observation. The hypothesis states that social interaction with immigrants can reduce prejudice and hostility. Having few migrants in the “backyard” while having many immigrants in the “neighbourhood” describes a situation where social interaction between natives and immigrants is potentially limited. Note that AFD and CDU voters are two different population segments. Contrary to AFD voters, CDU advocates could be more open to intercultural contact with refugees, which could reflect in an insignificant backyard effect but a significant neighbourhood effect. According to the contact hypothesis, intercultural contact can induce empathy and acceptance among different cultural groups. This would explain the insignificant backyard effect and the significant negative neighbourhood effect on the CDU. In other words, CDU advocates, who experience refugee presence in their hometown, also have a higher chance to become familiar with refugees and develop empathy and tolerance, while refugee presence in neighbouring municipalities physically limits intercultural contact. This would ultimately explain why the CDU received fewer votes in regions which neighbour municipalities with many refugee accommodations. This behaviour is also in line with the CDU party programme of 2016, which has a tolerant and integrative stance towards refugees. Furthermore, we find that CDU obtained more votes in regions which had higher turnout rates in the previous elections. In contrast, the right-wing AFD was more successful in regions, which had low turnout rates in the previous election. This indicates that the AFD party was more attractive in regions which previously had a higher voting abstinence. Another interesting result is the spatial clustering of AFD voters. The SDM and KPM models show that regions with high AFD vote-shares at the 2016 election are likely to be neighbour to other regions which prefer the AFD likewise. This is an interesting observation, because it raises the question why right-wing voting is spatially clustered. There seems to be some sort of “contagion effect” of right-wing voting among neighbouring municipalities. This result leaves some space for further research. However, one potential explanation could be the public perceptibility of the AFD in regions where the AFD is strong. Regions with higher AFD vote shares have more AFD supporters and therefore, more financial means and manpower. This allows more party advertisement and a professional structure. Since the AFD party is a new party, it is all the more important to make a professional impression on voters, which is evidently easier achievable with solid party funding. We assume that the “professional” AFD presence is perceived in the neighbourhood which attracts voters in neighbouring municipalities. It would be interesting to study the role of party funding on election results in a spatial context to check this hypothesis.

The analysis of the socioeconomic environment shows that unemployment is a major driving factor for right-wing voting. The AFD clearly benefitted in regions with higher unemployment rates. A large share of the literature on that matter points out that right-wing voting is often reinforced by a high unemployment rate. With regard to the demographic factors, we find strong evidence for AFD support in municipalities with high degrees of married couples, low education, and lower mean age. In light of German history, a recent study shows that political extremism is the second largest concern among Germans in 2015 (RAV, 2016). We presume that the fear of political extremism is pronounced more strongly among older voters. Hence, older generations might be less willing to vote for right-wing parties. Another surprising result is the seemingly indifferent stance towards right-wing voting in regions with many Germans with immigration background. We cannot say whether immigrants in Germany have the same stance towards "new" immigrants as native Germans do or whether native Germans' voting behaviour is not affected by the presence of Germans with immigration background. As a final conclusion, the finely-scaled spatial methodology proofed to be useful to study the influence of local immigration shocks and local socioeconomic conditions on voting behaviour since it outperformed the non-spatial model. It could serve as a useful template for future policy evaluations of various local shocks and spatial clustering of voting behaviour.

5 General Discussion and Conclusion

This work studies spatial data analysis techniques in economics and illustrates these techniques in three applications. In the articles, we employ techniques from exploratory spatial data analysis such as maps of spatial distributions, measures from spatial statistics such as Moran's I, and spatial econometric models for regression analysis.

The first article (chapter two) explores the determinants of electronic gambling machine (EGM) supply in the state of Baden-Württemberg. We use municipality level data to assess the influence of socioeconomic determinants on the EGM density, i.e. the number of EGMs per thousand residents. Our results suggest that the EGM supply is significantly higher in regions with higher rates of unemployment, immigrants, and lower levels of education. Hence, the EGM supply is more likely to occur in a certain socioeconomic milieu rather than being independent from socioeconomic factors. For the model selection we have chosen a simple OLS approach because the test statistics do not indicate the existence of spatial dependence. This is presumably due to a rather incomplete sample. We restrict the sample to 244 municipalities from a total number of 1,101 municipalities, because some data is not available for small municipalities, which constitute more than seventy percent of the population (figure 5.1). Hence, we include twenty two per cent of the available municipalities which results in a holey sample.

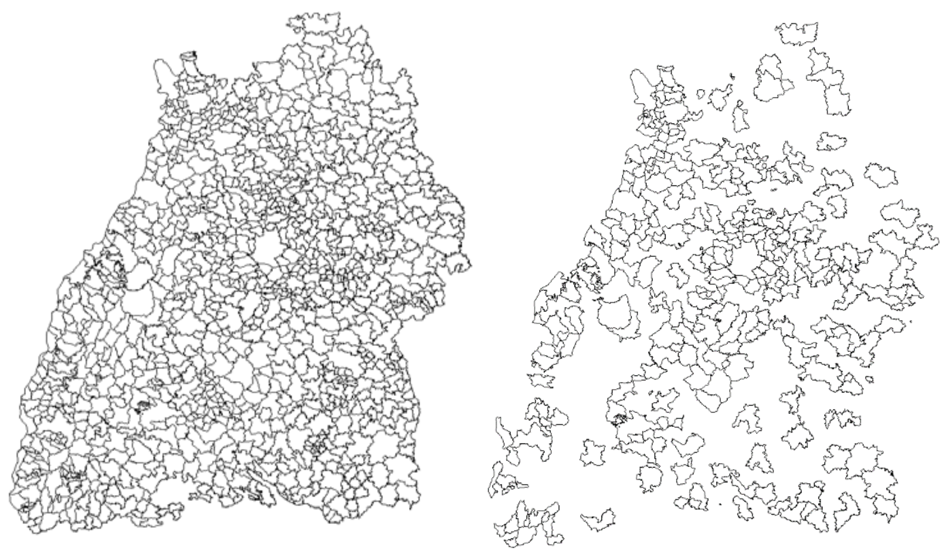

Figure 5.1 : Municipalities in Baden-Württemberg. Complete sample N=1.101 (left side), and incomplete sample N=244 (right side).

Thus, a lot of the contiguous neighbourhood structure is missing resulting in an incomplete spatial weights matrix which does not fully represent the complexity of the neighbourhood structure. However, the spatial weights matrix is required for spatial autocorrelation/dependence testing. Furthermore, the incomplete sample consists of several “islands”, i.e. regions which do not share a common border with other regions, while the complete sample has no islands. Too many islands in the sample lead to a higher number of regions which are excluded from potential spatial interaction. As a result, the tests for spatial dependence are insignificant. From a methodological point of view, the first article highlights the necessity of a complete sample for meaningful spatial econometric analysis. Compared to other econometric techniques, spatial econometric models build upon spatial dependence, which heavily relies on the completeness of the neighbourhood structure among geographical entities.

The article also highlights the trade-off between data availability and completeness in spatial data analysis. For instance, we could obtain a complete sample which covers all neighbourhood relations, if we used data from a higher aggregation level, i.e. counties. Figure 5.2 shows a total of 44 counties (*Landkreise* and *Stadtkreise*) in Baden-Württemberg in comparison to the incomplete sample of 244 municipalities. It illustrates a complete neighbourhood structure of counties, where all areas are linked together – directly and indirectly. On the other hand, the 244 municipalities represent a larger sample in terms of observations, but does not exhibit a contiguous neighbourhood structure. Hence, higher aggregated data can be a solution to the island problem.

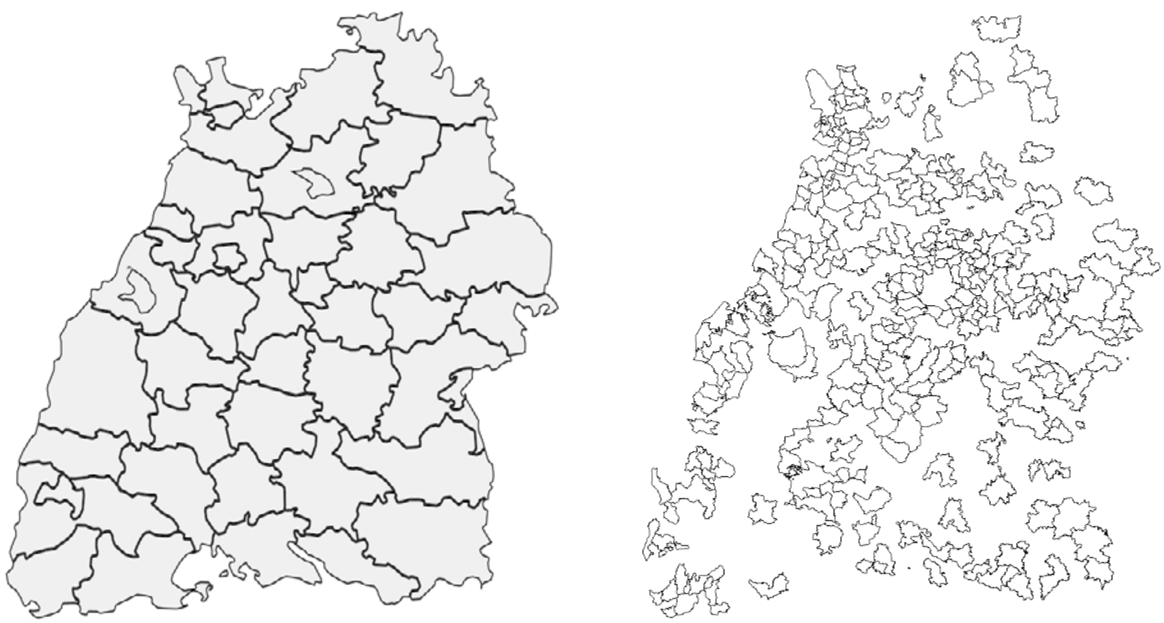

Figure 5.2: A total of forty-four counties (left) and 244 municipalities (right) in Baden-Württemberg

At the same time, higher aggregated data changes the scope of research and reduces the number of observations from 244 observations to 44, which affects estimation efficiency. Furthermore, the coefficients might change completely because the true data generating process at a higher aggregation level might be different compared to lower aggregation levels. This problem is known as the *Modified Area Unit Problem* (MAUP), which indicates that the outcome of a geographical analysis changes, if the aggregation level or scope of perspective changes (Fischer and Wang, 2011). Although data at higher aggregation levels is a convenient way to deal with the island problem and obtain a completely contiguous sample, it often suffers from the MAUP problem. There is no definite solution to this issue, however, alternative definitions of the spatial weights matrix (e.g. k-nearest neighbours) and better data availability for small-scale regions can soothe the issue (Gibbons and Overman, 2012).

The second article (chapter three) studies the spatial distribution of EGM supply while controlling for socioeconomic factors in the German state of North Rhine-Westphalia (NRW). Using data from NRW resolves a majority of the aforementioned data issues which were present in the first article. The NRW sample consists of 396 municipalities and contains records for socioeconomic factors and EGM supply data for nearly all municipalities. Hence, the analysis contains a complete sample, which is also at a low aggregation level. The fine scaled data does not suffer from the island-problem and biased spatial dependence tests. A first view at the relative EGM supply map, i.e. the EGMs per thousand residents, shows that EGM supply is rather high in certain neighbouring regions (hot spots), while it remains low in other neighbouring regions (cold spots). The aim of this work is to analyse the market concentration and link it to the socioeconomic environment within the municipalities.

The article introduces two novelties: First, it visualises the spatial distribution of EGM supply, and illustrates spatial clustering of EGM supply in NRW. Second, since Moran's I statistics and the LM test indicate the presence of spatial dependence, it is feasible to use spatial econometric modelling. The main findings imply that there is a clustering of EGM supply across regions at first, but when controlled for socioeconomic differences across the regions, the clustering effect is erased. In other words, the social and economic factors explain most of the supply concentration among municipalities. The results are partially in line with the first article. For instance, we find a higher EGM supply in regions with lower degree of education and higher unemployment rates. However, the link between unemployment and EGM supply is not as significant as in the first article. We also find that there is a higher gambling activity in urban regions compared to rural areas. Furthermore, we find some evidence for higher EGM supply in regions with motorway service station (MSS), which indicates that gambling activity is partially linked to passing traffic. In conclusion, the state of NRW is a suitable sample for further spatial socioeconomic research, because it has well documented data for nearly all municipalities.

The third article studies the influence of immigration on local election outcomes during the 2016 refugee crisis. To be precise, we analyse how variations in refugee settlement affected the election outcomes in the municipalities during the 2016 BW state election. The analysis focuses on two parties of interest, namely, the right-wing *Alternative für Deutschland* (AFD), and Angela Merkel's conservative ruling party, the *Christian Democratic Union* (CDU). It is the first study which applies spatial econometric techniques and combines small-scale immigration data, election data, and a set of socioeconomic factors for a complete sample of 1,101 municipalities in Baden-Württemberg. In contrast to the first article, our sample covers all municipalities in Baden-Württemberg. Hence, there are no issues regarding the island problem. We thereby obtain a correctly specified spatial weights matrix and meaningful spatial dependence tests. One novelty compared to the previous articles is the use of spatial lags in the independent variables, which are incorporated in the Spatial Durbin Model (SDM). Since we find significant spatial interaction in the dependent as well as some independent variables, the SDM model is the best fitting model for our purpose (LeSage and Pace, 2009).

Regarding the results, we find that right-wing voting is a local phenomenon. This means AFD vote shares are highly clustered in certain regions and the clustering still sustains in the SDM model, which controls for several factors. There are two important conclusions from this study. First, there are strong spill-over-effects in right-wing voting. However, we do not explore the reasons, why right-wing voting is concentrated in certain regions. Hence, the determinants of AFD clustering remain an open question, which should be further studied. Second, we find a significant positive relationship between refugee settlement and AFD vote shares. This result supports the hypothesis that local refugee presence strengthened the AFD party. Furthermore, we find that the AFD party was popular in regions with lower turnout rates, higher GDP, higher unemployment rates, lower mean age, higher share of married people, and lower degrees of education. Regarding the CDU party, the results towards immigration are not as clear. The direct influence of immigrants turns out insignificant. Considering the disastrous election outcome for the CDU, it is surprising that "backyard" refugee presence didn't play a role for the election outcome. The CDU lost roughly a quarter of its votes and local refugee presence didn't contribute to it. However there are negative spill-over-effects from neighbourhood immigration on CDU votes. In other words, immigration to neighbouring municipalities negatively affects the CDU voting outcomes. We do not deepen this observation, however, the causes for this effect should be further explored. When it comes to the socioeconomic milieu, the CDU was preferred in regions with higher turnout rate, lower GDP, and a higher share of unmarried people.

Regarding the interpretation of the results, it is tempting to draw conclusions about individuals. For example in this setup, it is not possible to state that the average AFD voter is either unemployed, young, married, has a low degree of education, or is a combination of these attributes. Since spatial data analysis in this work is based on area data, it is not possible to trace the estimated outcomes back to individuals. Conclusions about individuals will cause the ecological fallacy problem (King, 1997). The problem describes that statements about individuals, which are based on aggregated data, will be either biased or infeasible. There are some solutions to the ecological fallacy problem, such as King's EI approach (King, 1997). However, these methods have not been incorporated in spatial data analysis yet and should be further studied and implemented in statistical software.

Although spatial data analysis has become increasingly popular among economists within the last years, it has some disadvantages compared to non-spatial econometrics and leaves room for improvement. For instance, causal inference is still an issue for applied spatial econometricians. One reason for that stems from the ecological fallacy problem. Since conclusions on individuals are not feasible from aggregated data, the applicability for micro-level analysis is rather limited. Another issue for causal inference is described by Gibbons and Overman (2012). They argue that the spatial lag in the dependent variable and the error term cannot be separated which leads to identification issues. This is due to the fact that spatial autocorrelation in Y and spatial autocorrelation in something unobserved cannot be distinguished. Hence, there should be more theoretical considerations which explains why certain spatial structures occur. Moreover, some of the important spatial standard literature, such as LeSage and Pace (2009), either downplay or ignore the causality issue. This leads to a significant number of publications based on this literature, which avoids serious discussions about causality in spatial models. One reason for that is the traditional understanding of populations and samples. Usually, statistical inference assumes that there is an unobserved population, which can be studied by drawing samples from that population. However, since spatial data analysis often uses samples which correspond to the whole population, as we did in article two and three, one can assume that no inference from the sample is required because the population itself is studied.

In view of this criticism, it is clear that model selection and causality issues in spatial econometrics need more revision. However, these issues have partially been addressed and refuted by recent publications. Bramoulle et al. (2009) present a set of assumptions, which have to be fulfilled in order to identify the

Spatial Durbin Model correctly. Vega and Elhorst (2015) study model specifications of the SLX model, while Paelinck and Mur (2018) review causality issues in spatial data analysis in general. These articles show that methodological issues are addressed and further developed. Despite all the criticism, most economists agree with the basic idea of the discipline, which stems from Tobler's first law of geography: near things *are* indeed more related than distant things. Ignoring these structures in applied work will result in biased estimates. Finally, we should not underestimate the role of spatial data visualisation and pattern recognition. Visualising spatial data makes otherwise complex issues easier to grasp and spatial pattern recognition allows to detect and locate economic activity. Hence, the techniques are valuable for applied research and didactical purposes likewise.

6 Literature

(1) Abbott, M. W., Romild, U., and Volberg, R. A. (2014). "Gambling and problem gambling in Sweden: Changes between 1998 and 2009". *Journal of Gambling Studies,* 30: 985–999.

(2) Abbott, M., Stone, C. A., Billi, R., and Yeung, K. (2015). "Gambling and problem gambling in Victoria, Australia: Changes over 5 years". *Journal of Gambling Studies.* Advance online publication.

(3) Ahrens, A. and Bhattacharjee, A. (2015). "Two-Step Lasso Estimation of the Spatial Weights Matrix". Econometrics, 3: 128-155.

(4) Allport, G.W. (1954). "The nature of prejudice." Reading, MA: Addison-Wesley Publishing Company.

(5) American Psychiatric Association. (1994). "Diagnostic and statistical manual of mental disorders". Vol. 4. Washington, DC: Author.

(6) Andorno, R. (2004). "The precautionary principle: A new legal standard for a technological age". *Journal of International Biotechnology Law*, 1, 11–19.

(7) Anselin, L. (1980). "Estimation Methods for Spatial Autoregressive Structures." Regional Science Dissertation and Monograph Series. Ithaca, NY: Cornell University.

(8) Anselin, L. (1988). "Spatial Econometrics: Methods and Models." Dordrecht: Kluwer Academic Publishers.

(9) Anselin, L. (2010). "Thirty years of spatial econometrics". *Regional Science,* 89: 3-25.

(10) Anselin, L., Cohen, J., Cook, D., Gorr, W. and Tita, G. (2000). "Spatial Analyses of Crime". *Criminal Justice*, Vol. 4: 213-262.

(11) Anselin, L., Rey, S.J., (2014). "Modern Spatial Econometrics in Practice". Chicago, IL: GeoDa Press.

(12) Arbia, G. (2011). “A Lustrum of SEA: Recent Trends Following the Creation of the spatial Econometrics Association (2007-2011)”. *Spatial Economic Analysis*, 6(4): 377-395.

(13) Arbia, G. (2014). “A Primer for Spatial Econometrics”. Basingstoke: Palgrave Macmillan.

(14) Arbia, G. (2016). “Spatial Econometrics: A Rapidly Evolving Discipline”. *Econometrics,* 4(18).

(15) BAMF (2016). “Migration Report 2015.” Nuremberg: Federal Office for Migration and Refugees (BAMF).

(16) Barone, G., D’Ignazio, A., De Blasio, G. and Natichhioni, P. (2016). “Mr. Rossi, Mr. Hu and politics. The role of immigration in shaping natives’ voting behaviour”. *Journal of Public Economics,* 136: 1-13.

(17) Becker, T., and Heinze, K. (2015). „Auswirkungen geplanter Abstandsregelungen und Regelungen zu Konzessionsgrößen auf Spielhallen am Beispiel ausgewählter Kommunen in Baden-Württemberg“ (Research Report No. 3). Stuttgart: University of Hohenheim, Gambling Research Center.

(18) Blanco, C., Hasin, D. S., Petry, N., Stinson, F. S., and Grant, B. F. (2006). “Sex differences in subclinical and DSM-IV pathological gambling: Results from the National Epidemiologic Survey on Alcohol and Related Conditions”. *Psychological Medicine:* 36, 943–953.

(19) Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). “Identification of Peer Effects through Social Networks,” *Journal of Econometrics*, 150: 41–55.

(20) Breustedt, G. and Habermann, H. (2011). “The Incidence of EU Per-Hectare Payments on Farmland Rental Rates: A Spatial Econometric Analysis of German Farm-Level Data”. *Journal of Agricultural Economics,* 62(1): 225-243.

(21) Buchner, U. G., Erbas, B., Stürmer, M., Arnold, M., Wodarz, N., and Wolstein, J. (2015). „Inpatient treatment for pathological gamblers in Germany: Setting, utilization, and structure”. *Journal of Gambling Studies,* 31: 257–279.

(22) Bühringer, G., Kraus, L., Sonntag, D., Pfeiffer-Gerschel, T., and Steiner, S. (2007). „Pathologisches Glücksspiel in Deutschland: Spiel- und Bevölkerungsrisiken“. *Sucht*, 5: 296–308.

(23) Burridge, P. (1980). “On the Cliff-Ord Test for Spatial Correlation”. *Journal of the Royal Statistical Society*. Series B: Methodological, 42(1).

(24) Buth, S., and Stöver, H. (2008). „Glücksspielteilnahme und Glücksspielprobleme in Deutschland: Ergebnisse einer bundesweiten Repräsentativbefragung“. *Suchttherapie*, 9: 3–11.

(25) BZgA. (2008). „Glücksspielverhalten und problematisches Glücksspielen in Deutschland 2007“. Cologne: Bundeszentrale für gesundheitliche Aufklärung.

(26) BZgA. (2010). „Glücksspielverhalten in Deutschland 2007 und 2009. Ergebnisse aus zwei repräsentativen Bevölkerungsbefragungen“. Cologne: Bundeszentrale für gesundheitliche Aufklärung.

(27) BZgA. (2012). „Glücksspielverhalten und Glücksspielsucht in Deutschland. Ergebnisse aus drei repräsentativen Bevölkerungsbefragungen 2007, 2009 und 2011“. Cologne: Bundeszentrale für gesundheitliche Aufklärung.

(28) BZgA. (2014). „Glücksspielverhalten und Glücksspielsucht in Deutschland 2013“. Cologne: Bundeszentrale für gesundheitliche Aufklärung.

(29) Cliff, A.D., and Ord, J.K. (1973). “Spatial Autocorrelation”. London: Pion.

(30) Cliff, A.D., and Ord, J.K. (1981). “Spatial Processes Models and Applications”. London: Pion.

(31) Cummins, S. (2007). “Commentary: Investigating neighbourhood effects on health-Avoiding the ‘Local Trap’”. *International Journal of Epidemiology,* 36: 355–357.

(32) Dahlberg, M., Edmark, K., and Lundqvist, H. (2012). “Ethnic Diversity and Preference for Redistribution.” *Journal of Political Economy,* 120 (1): 41-76.

(33) Delfabbro, P. (2008). “Evaluating the Effectiveness of a Limited Reduction in Electronic Gaming Machine Availability on Gambling Behaviour and Expenditure”. *International Gambling Studies*, 8: 151-166.

(34) Diez Roux, A. V. (2002). “A glossary for multilevel analysis”. *Journal of Epidemiology & Community Health*, 56: 588–594.

(35) Dixon, J., Durrheim, K., and Tredoux, C. (2005). “Beyond the Optimal Contact Strategy: A Reality Check for the Contact Hypothesis.” *American Psychologist,* 60 (7): 697-711.

(36) Dowling, N. A., Cowlishaw, S., Jackson, A. C., Merkouris, S. S., Francis, K. L., and Christensen, D. R. (2015). “Prevalence of psychiatric co-morbidity in treatment-seeking problem gamblers: A systematic review and meta-analysis”. *The Australian and New Zealand Journal of Psychiatry*, 49: 519–539.

(37) Drukker, D.M., Peng, H., Prucha, I.R. and Raciborski, R. (2013a). “Creating and Managing Spatial-Weighting Matrices with the SPMAT Command”. *The Stata Journal*, 13 (2): 242-286.

(38) Drukker, D.M., Prucha, I.R. and Raciborski, R. (2013b). “Maximum Likelihood and Generalized Spatial Two-Stage Least-Squares Estimators for a Spatial-Autoregressive Model with Spatial-Autoregressive Disturbances”. *The Stata Journal,* 13 (2): 221-241.

(39) Dustmann, C., Vasiljeva, K., and Damm A.P. (2016). “Refugee Migration and Electoral Outcomes.” *The Rockwool Foundation Research.* Study Paper No. 111.

(40) Economopoulos, A.J. (2015). “Examining Impact of Casinos on Economic Development: A Spatial Analysis of the Counties in the Mid-Atlantic Region“. *Journal of Gambling Business and Economics,* 9(1): 77-92.

(41) Elhorst, J.P. (2010). “Applied Spatial Econometrics: Raising the Bar“. *Spatial Economic Analysis,* 5(1): 9-27.

(42) Ellison, C.G., Shin, H., and Leal, D.L. (2011). “The Contact Hypothesis and Attitudes Toward Latinos in the United States.” *Social Science Quarterly,* 92 (4): 938-958.

(43) Erbas, B., and Buchner, U. G. (2012). „Pathological gambling: Prevalence, diagnosis, comorbidity, and intervention in Germany”. *Deutsches Ärzteblatt International*, 109: 173–179.

(44) FGA (2015). „Der Deutsche Glücksspielmarkt 2014 – Eine ökonomische Darstellung“. Jahresreport 2014 der obersten Glücksspielaufsichtsbehörde in Hessen. Hessen.

(45) Fischer, M.M. and Getis, A. (2010). “Handbook of Applied Spatial Analysis“. Heidelberg: Springer.

(46) Fischer, M.M. and Wang, J. (2011). “Spatial Data Analysis“. Heidelberg: Springer.

(47) Fong, T.W. (2005). “The Biopsychosocial Consequences of Pathological Gambling“. *Psychiatry,* 2(3): 22-30.

(48) FSOBW (2016). “Statistisches Monatsheft Baden-Wuerrtemberg.” Stuttgart: Federal Statistical Office of Baden-Wuerttemberg 4.

(49) Funke, M., Schularick, M., and Trebesch, C. (2016). “Going to extremes: Politics after financial crises, 1870-2014.” *European Economic Review,* 88: 227-260.

(50) Garand, J.C. (2010). “Income Inequality, Party Polarization, and Roll-Call Voting in the U.S. Senate.” *The Journal of Politics,* 72 (4): 1109-1128.

(51) Geary, R. (1954). “The Contiguity Ratio and Statistical Mapping”. *The Incorporated Statistician*, 5(3): 115-146.

(52) Gehrsitz, M., and Ungerer, M. (2017). “Jobs, Crime, and Votes: A Short-run Evaluation of the Refugee Crisis in Germany.” *Institute of Labor Economics (IZA) Discussion Paper Series*: 10494.

(53) Gilliland, J. A., and Ross, N. A. (2005). “Opportunities for video lottery terminal gambling in Montréal: An environmental analysis”. *Canadian Journal of Public Health*, 96: 55–59.

(54) Govoni, R., Frisch, G. R., Rupcich, N., and Getty, H. (1998). „First year impacts of casino gambling in a community”. *Journal of Gambling Studies*, 14: 347–358.

(55) Griffiths, M. (1999). “Gambling technologies: Prospects for problem gambling”. *Journal of Gambling Studies*, 15: 265–283.

(56) Halla, M., Wagner, A.F., and Zweimüller, J. (2017). Immigration and voting for the far right. *Journal of the European Economic Association,* 15 (6): 1341-1385.

(57) Haller, M. (2017). “Die „Flüchtlingskrise“ in den Medien”. Frankfurt: Otto Brenner Foundation.

(58) Han, K.J. (2016). “Income Inequality and voting for radical right-wing parties.” *Electoral Studies,* 42: 54-64.

(59) Haß, W., Orth, B., and Lang, P. (2012). „Zusammenhang zwischen verschiedenen Glücksspielformen und glücksspielassoziierten Problemen. Ergebnisse aus drei repräsentativen Bevölkerungs-Surveys der Bundeszentrale für gesundheitliche Aufklärung (BZgA)“. *Sucht*, 58: 333–345.

(60) Hersh, E.D., and Nall, C. (2016). “The Primacy of Race in the Geography of Income-Based Voting: New Evidence from Public Voting Records.” *American Journal of Political Science,* 60 (2): 289–303.

(61) Hing, N., and Gainsbury, S. (2011). “Risky business: Gambling problems amongst gaming venue employees in Queensland, Australia”. *Journal of Gambling Issues*, 25: 4–23.

(62) Iancu, I., Lowengrub, K., Dembinsky, Y., Kotler, M., and Dannon, P. N. (2008). “Pathological gambling: An update on neuropathophysiology and pharmacotherapy”. *CNS Drugs*, 22: 123–138.

(63) Jasny, J. (2016). „Is Gambling Contagious? An Analysis of Electronic Gambling Machine Clustering in Germany“. *The Journal of Gambling Business and Economics,* 10 (3): 54-70.

(64) Johansson, A. Grant, J.E., Kim, S.W., Odlaug, B.L. and Götestam, K.G. (2009). “Risk Factors for Problematic Gambling: A Critical Literature Review“. *Journal of Gambling Studies*, 25(1): 67-92.

(65) Kelejian, H. H. and Prucha, I. R. (1998). “A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances”. *Journal of Real Estate Finance and Economics*, 17: 99-121.

(66) Kessler, R. C., Hwang, I., LaBrie, R., Petukhova, M., Sampson, N. A., Winters, K. C., and Shaffer, H. J. (2008). “DSM-IV pathological gambling in the National Comorbidity Survey Replication”. *Psychological Medicine*, 38: 1351–1360.

(67) Kiewiet, D.R., and Udell, M. (1998). “Twenty-five Years after Kramer: An Assessment of Economic Retrospective Voting based upon Improved Estimates of Income and Unemployment.” *Economics and Politics,* 10: 219-248.

(68) King, G. (1997). “A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data.” Princeton: Princeton University Press.

(69) Korn, D. A., and Shaffer, H. J. (1999). „Gambling and the health of the public: Adopting a public health perspective”. *Journal of Gambling Studies*, 15: 289–365.

(70) Kramer, G.H. (1971). “Short-Term Fluctuations in U.S. Voting Behavior, 1896–1964.” *American Political Science Review,* 65 (1): 131-143.

(71) Krieckhaus, J., Son, B., Bellinger, N.M., and Wells, J.M. (2014). "Economic inequality and democratic support." *Journal of Politics,* 76 (1): 139–151.

(72) Kun, B., Balázs, H., Arnold, P., Paksi, B. and Demetrovics, Z. (2012). "Gambling in Western and Eastern Europe: The Example of Hungary". *Journal of Gambling Studies,* 28(1): 27-46.

(73) Ladouceur, R., Jacques, C., Ferland, F., and Giroux, I. (1999). "Prevalence of problem gambling: A replication study 7 years later". *Canadian Journal of Psychiatry,* 44: 802–804.

(74) LaPlante, D. A., and Shaffer, H. J. (2007). „Understanding the influence of gambling opportunities: Expanding exposure models to include adaptation". *American Journal of Orthopsychiatry*, 77: 616–623.

(75) LeSage, J., and Pace, R.K. (2009). "Introduction to Spatial Econometrics." Boca Raton, LLC: Taylor & Francis Group.

(76) Lesieur, H. R., and Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): "A new instrument for the identification of pathological gamblers". *American Journal of Psychiatry*, 144: 1184–1188.

(77) Lobo, D. S. S., and Kennedy, J. L. (2009). "Genetic aspects of pathological gambling: A complex disorder with shared genetic vulnerabilities". *Addiction*, 104: 1454–1465.

(78) MacLaren, V. V., Fugelsang, J. A., Harrigan, K. A., and Dixon, M. J. (2011). „The personality of pathological gamblers: A meta-analysis". *Clinical Psychology Review*, 31: 1057–1067.

(79) Manski, C. (1993). "Identification of Endogenous Social Effects: The Reflection Problem". *The Review of Economic Studies*, 60(3): 531-542.

(80) Marshall, D. (2005). "The gambling environment and gambler behaviour: Evidence from Richmond-Tweed, Australia". *International Gambling Studies*, 5: 63–83.

(81) Marshall, D. (2009). "Gambling as a public health issue: The critical role of the local environment". J*ournal of Gambling Issues*, 23: 66–80.

(82) McCarty, N., Poole, K.T., and Rosenthal, H. (2006). "Polarized America: The Dance of Ideology and Unequal Riches." Cambridge, MA: MIT Press.

(83) Meyer, C., Rumpf, H.-J., Kreuzer, A., de Brito, S., Glorius, S., Jeske, … John, U. (2011). "Pathologisches Glücksspielen und Epidemiologie (PAGE): Entstehung, Komorbidität, Remission und Behandlung. Endbericht an das Hessische Ministerium des Innern und für Sport". Greifswald, Lübeck: University Greifswald and Lübeck.

(84) Moragas, L., Granero, R., Stinchfield, R., Fernández-Aranda, F., Fröberg, F., Aymamí, N., … Jiménez-Murcia, S. (2015). "Comparative analysis of distinct phenotypes in gambling disorder based on gambling preferences". *BMC Psychiatry*, 15: 173.

(85) Moran, P.A.P. (1950). "Notes on continuous stochastic phenomena". *Biometrika,* 37: 17-23.

(86) Myers, N. (2002). “The precautionary principle puts values first”. *Bulletin of Science, Technology & Society*, 22: 210–219.

(87) Noelle-Neumann, E. (1980). “Die Schweigespiralle. Öffentliche Meinung – unsere soziale Haut.” Munich: Langen Mueller.

(88) Ord, K. (1975). “Estimation Methods for Models of Spatial Interaction.” *Journal of the American Statistical Association,* 70 (349): 120-126.

(89) Otto, A.H. and Steinhardt, M.F. (2014). Immigration and election out-comes-evidence from city districts in Hamburg. *Regional Science and Urban Economics,* 45: 67-79.

(90) Paelinck, J., and L. Klaassen (1979). “Spatial Econometrics”. Farnborough: Saxon House.

(91) Paelinck, J., and Mur, J. (2018). “Some Issues on the Concept of Causality in Spatial Econometric Models”. *Estudios de Economia Aplicada*, 36 (1): 107-118.

(92) Pearce, J., Mason, K., Hiscock, R., and Day, P. (2008). “A national study of neighbourhood access to gambling opportunities and individual gambling behaviour”. *Journal of Epidemiology & Community Health*, 62: 862–868.

(93) RAV (2016). “Die Ängste der Deutschen 2015.” Wiesbaden: R+V Versicherung.

(94) Sassen, M., Kraus, L., Bühringer, G., Pabst, A., Piontek, D., and Taqi, Z. (2011). „Gambling among adults in Germany: Prevalence, disorder and risk factors”. *Sucht*, 57: 249–257.

(95) Schündeln, M. (2014). “Are immigrants more mobile than natives? Evidence from Germany“. *Journal of Regional Science*, 54(1): 70-95.

(96) Shaffer, H. J., and Hall, M. N. (2002). „The natural history of gambling and drinking problems among casino employees”. *The Journal of Social Psychology*, 142: 405–424.

(97) Shaffer, H. J., LaBrie, R. A., and LaPlante, D. (2004). „Laying the foundation for quantifying regional exposure to social phenomena: Considering the case of legalized gambling as a public health toxin”. *Psychology of Addictive Behaviors*, 18: 40–48.

(98) Shaffer, H. J., Vander Bilt, J., and Hall, M. N. (1999). „Gambling, drinking, smoking and other health risk activities among casino employees”. *American Journal of Industrial Medicine*, 36: 365–378.

(99) Steinmayr, A. (2016). “Exposure to Refugees and Voting for the Far-Right: (Unexpected) Results from Austria.” *Institute of Labor Economics (IZA) Discussion Paper Series*: 9790.

(100) Storer, J., Abbott, M., and Stubbs, J. (2009). “Access or adaptation? A meta-analysis of surveys of problem gambling prevalence in Australia and New Zealand with respect to concentration of electronic gaming machines”. *International Gambling Studies*, 9: 225–244.

(101) Suzuki, E. (2012). “Time changes, so do people”. *Social Science & Medicine*, 75: 452–456.

(102) Tobler, W.R. (1970). “A computer movie simulating urban growth in the Detroit region“. Economic Geography, 46: 234-240.

(103) Trümper, J., and Heimann, C. (2012). „Angebotsstruktur der Spielhallen und Geldspielgeräte in Deutschland 2012“. Arbeitskreis gegen Spielsucht e.V.

(104) Trümper, J., and Heimann, C. (2014). „Angebotsstruktur der Spielhallen und Geldspielgeräte in Deutschland 2014“. Arbeitskreis gegen Spielsucht e.V.

(105) Tukey, J.W. (1997). “Exploratory Data Analysis”. London: Pearson.

(106) Vasiliadis, S. D., Jackson, A. C., Christensen, D., and Francis, K. (2013). “Physical accessibility of gaming opportunity and its relationship to gaming involvement and problem gambling: A systematic review”. *Journal of Gambling Issues*, 28: 1–46.

(107) Vega, S. H., and Elhorst, J. P. (2015). “The SLX Model”. *Journal of Regional Science*, 55: 339-363.

(108) Vieweg, H.-G. (2013). „Wirtschaftsentwicklung Unterhaltungsautomaten 2012 und Ausblick 2013“. *Institute for Economic Research* (IFO) at the University of Munich. Retrieved from: http://www.vdai.de/ima2013/ifo-wirtschaftsstudie-dt.pdf

(109) Wang, Y., Kockelmann, K.M., and Wang, X.C. (2013). “The impact of weight matrices on parameter estimation and inference: A case study of binary response using land-use data.” *Journal of Transport and Land Use,* 6 (3): 75-85.

(110) Wardle, H., Keily, R., Astbury, G., and Reith, G. (2014). “‘Risky places?’: Mapping gambling machine density and socio-economic deprivation”. *Journal of Gambling Studies*, 30: 201–212.

(111) Welte, J. W., Barnes, G. M., Wieczorek, W. F., Tidwell, M.-C. O., and Hoffman, J. H. (2007). “Type of gambling and availability as risk factors for problem gambling: A tobit regression analysis by age and gender”. *International Gambling Studies*, 7: 183–198.

(112) Welte, J. W., Wieczorek, W. F., Barnes, G. M., and Tidwell, M.-C. O. (2006). “Multiple risk factors for frequent and problem gambling: Individual, social, and ecological”. *Journal of Applied Social Psychology*, 36: 1548–1568.

(113) Welte, J. W., Wieczorek, W. F., Barnes, G. M., Tidwell, M.-C., and Hoffman, J. H. (2004). “The relationship of ecological and geographic factors to gambling behavior and pathology”. *Journal of Gambling Studies*, 20: 405–423.

(114) Wheeler, B., Rigby, J., and Huriwai, T. (2006). “Pokies and poverty: Problem gambling risk factor geography in New Zealand”. *Health & Place*, 12: 86–96.

(115) Whittle, P. (1954). “On Stationary Processes in the Plane”. *Biometrika*, 41(3/4): 434-449.

(116) Williams, R. J., Volberg, R. A., and Stevens, R. M. G. (2012). “The population prevalence of problem gambling: Methodological influences, standardized rates, jurisdictional differences, and

worldwide trends". Report prepared for the Ontario Problem Gambling Research Centre and the Ontario Ministry of Health and Long Term Care. May 8, 2012.

(117) Wu, A. M. S., and Wong, E. M. W. (2008). "Disordered gambling among Chinese casino employees". *Journal of Gambling Studies*, 24: 207–217.

(118) Xouridas, S., Jasny, J., and Becker, T. (2016). „An ecological approach to electronic gambling machines and socioeconomic deprivation in Germany". *Journal of Gambling Issues*, 33: 82-102.