Network Effects on Domestic Migration Flows Across Germany – A Spatial Autoregressive Perspective with Spatially Structured Origin and Destination Effects and Heteroskedastic Innovations –

Markus Zimmer*

*University of Munich and Ifo Institute for Economic Research

Abstract:

Domestic migration accounts by far for the major proportion of total migration in Germany. While cross-country migration has been intensively investigated in the literature we address the less prominent issue of regional migration in this paper. In our model we study the spatial interaction between German districts in order to explain the domestic flows of migration. Recently empirical research emphasises that beside an increase in expected earnings, network effects seem to have the strongest influence on the decision to migrate. Social networks rise expected income by increasing the employment probability. They also lower migration costs by reducing the fear of loosing social integration. They decrease the search costs and the amount of uncertainty by increasing circulation of information. These networks can either be established formally established like in the case of relatives or informally by linking the migrants through a common regional background. We analyze the migration between the 439 administrative districts in Germany. We show that networks have a strong significant impact on migration flows. We explore the differences in the migration-behaviour depending on the gender or nationality of the migrant. We identify strong influences of amenities and of traditional origin and destination characteristics. Our expansion of the log-linear specification of the gravity migration model by spatial lag structures requires adequate econometric methods to account for spatial autocorrelation. We extend the model of Kelejian and Prucha (2008) to a second order process in both, the spatial endogenous lag and the autoregressive disturbance. Thus we are able to separately analyze origin and destination effects on bilateral flows of migration.

Network Effects on Domestic Migration Flows Across Germany

– A Spatial Autoregressive Perspective with Spatially Structured Origin and Destination Effects and Heteroskedastic Innovations –

Markus Zimmer*

zimmer@lmu.de

*Center for Economic Studies, University of Munich
Schackstrasse 4, 80539 Muenchen (Munich), Germany
Phone: +49-89-2180-2089, FAX: +49-89-397303

*Ifo Institute for Economic Research
Poschingerstr. 5, 81679 Muenchen (Munich), Germany
Phone: +49-89-9224-1260, FAX: +49-89-985369

Introduction

In contrast to the demographic processes of fertility and mortality, domestic migration has a much bigger impact to the spatial distribution of the population. Obviously the migration out of one region within a multiregional domestic system has an impact on several other regions while the birth of a new citizen affects only one region. Observing this multiregional system as a whole will generate a detailed matrix of migration flows between each pair of regions. Since each region can act as an emitter and a receiver of migrants we will distinguish between these two effects in examining each region as origin and as destination. Our main objective is to identify these effects and to determine the determinants of migration.

From the perspective of an individual migrant, the decision to relocate is a question of the net-benefits of migrating, meaning the difference of the expected income gains and the economic and psychological movement costs. Expectations are largely influenced by uncertainty which decreases the willingness to migrate for risk-averse individuals. Social networks can help to increase the expected income, reduce uncertainty and lower migration costs through granting access to housing or integration into a familiar social community. As migrants with a common background exert positive network externalities within their community this gives a rational for herd behaviour. Consequently coordination migration flows in timing and destination increases the efficiency[1]. The means to realize this coordination is the positive signal that migrants exert on each other by their location choice.

It can be expected, that phenomena including a spatial component are also likely to exhibit a spatial structure explaining them. It is plausible to assume units of observation which are spatially close show interdependencies in their development. The formulation of such models in the works of Cliff and Ord (1973, 1981) led to the terminology as Cliff-Ord type spatial model. In the model presented here we will consider several channels of spatial structure. Beside simple spatially weighted exogenous variables these structures typically feature endogeneity and therefore demand adequate econometric procedures. In our model we will consider spatial spillovers in the endogenous variable, spatial autoregression in the disturbance term and will allow for heteroskedasticity in the remaining disturbance which is

[1] Economic theory would suggest that in the presence of a positive externality the amount of migration to the same location is inefficiently small if it is based on the decision of the individual migrants. Yet theoretical and empirical work on how much the efficiency is increased through signalling and herd behaviour is missing but would be an interesting field for further research.

typically referred to as innovation. Since we will consider lags of second order in both of the processes of spatial autocorrelation we will refer to this model as SARAR(2,2) as suggested by Kelejian and Prucha who adopted the original terminology from Anselin and Florax (1995). The procedure developed by Kelejian and Prucha is the preferred choice for such a model since other available approaches are inconsistent or yet fail to show their asymptotic properties. Even though they explicitly prove consistency only for the SARAR(1,1) model with first order autoregressive lags, it is apparent that the asymptotic properties will also hold for lags of higher order. However, a generalization of the proof has yet to be formulated but will not be part of this work.

The Background of the Economic Model

The main motivation to control for spatial structures in our work is the existence of inter-district networks. Therefore by employing appropriate empirical methods we will not only improve the estimation of the effects of the classical factors influencing the migration flows but we will also be able to determine the size of the network effects. These networks have been widely discussed in migration literature but yet had not been observed in their spatial structure[2]. While this structure might still be negligible on the country level it certainly becomes effective as you scale down the size of the observed unit. As we observe migration on a very small scale, the typical migrants to two neighbouring districts will probably only live a few kilometres away from each other or are likely to daily commute to the same district for working. From the literature there are two main channels through which these spatial phenomena can affect migration behaviour: The traditional effect of networks of common family-, friend-, ethnical-, social- or regional background on the one hand and the more recent argument of herd behaviour being motivated from the idea of imperfectly informed migrants, signalling effects and the impact of the own migration behaviour on the location choice of the succeeding migrants.

The previous literature on migrant networks had its main focus on international migration. A prominent example is the autoregressive dynamic of Mexican immigrants in the United States

[2] Notable exceptions are the migration studies of LeSage and Pace (2007) and Goetz and Rupasingha (2004) who consider a spatial autoregressive structure in a non-network oriented analysis.

of America examined in many empirical works as for example in Munshi (2003). Other empirical works confirm the existence of migrant networks for domestic migration within the United States (Bartel, 1989, Frey, 1995), for international migration to the United Kingdom, (Nigel and Pain, 2003), to Australia (Chiswick Lee and Miller, 2001) and to Canada (McDonald, 2002) or for regional clustering of Ethnic German immigrants to Germany (Bauer and Zimmermann, 1997)[3].

The channels through which these networks operate are manifold: Firstly, they can increase the expected income gains by increasing the hiring probability (Cocoran, Datcher and Ducan, 1980), reducing the uncertainty about the job-market conditions (Massey, 1987) or lowering the search costs for a job (Mortensen and Vishwanath, 1994). Secondly, they can reduce moving costs, irrespectively from being of financial or psychological nature, like the loss of ethnical integration or separation from family or friends (Schwartz, 1973, Mincer, 1978, Church and King, 1993, Chiswick and Miller, 1996). For studies reviewing the literature about the effects of networks on migration see Greenwood (1985) and Cohen, Reed, Montgomery and Stren (2003).

Thirdly, due to the large amount of uncertainty about the conditions at a possible destination and because of the increase of the positive network externality with the size of the network it is rational for migrants to show a certain amount of herd behaviour. Thus under imperfect information the destination choice of other migrants is interpreted as positive signal about the quality of their location choice which has positive impact on the own perception of this destination (Epstein and Hillman, 1998). In that respect this signal might not only influence the spatial structure of migration, but also the time dimension (Burda, 1995). Viewing migration from a community perspective it is sensible to deliberately or unconsciously coordinate the timing and the location of migration flows. In that way the network externalities at the destination are maximized. Therefore the signal through the own migration act increases in value if my choice to follow the earlier migrants encourages the subsequent ones to act alike. This results in regional concentration of ethnical groups which is very

[3] The Term of Ethnic Germans refers to the German expression “Aussiedler”. These are persons who themselves or whose ancestors had the citizenship of the Second German Reich as of the boarders of 1937 and who lived in territories which after the Second World War felt to the Allied powers. These Ethic Germans could up to 1990 freely immigrate to the Federal Republic of Germany and apply for citizenship. After 1990 qualification requirements for citizenship have been slightly tightened but still grant a preferential status to this group.

similar to the clustering that would follow from pure network effects but which moreover gives a justification for the dynamics of migration flows which frequently contradict the expectations from traditional explanations (Epstein and Gang, 2004). Empirical work on the relative importance of network versus herd effects can be found in Bauer, Epstein and Gang (2007). A summary of the functioning of network and herd effects is displayed in table 1.

Table 1: differentiation of network versus herd effects

	Network effects	Herd effects
Contribution of the migrants	Circulation of information and provision of ethnical and economic resources	Informal group-dynamical process to internalize the positive externality of migrant networks which results in the coordination of migration flows
Benefit for the migrants	Reduction of economic and psychological migration costs and increase of the expected benefits from migration	Signalling reduces subjective amount of uncertainty and positive externalities of migrant networks increase
Result for the economy	Local clustering of migrants with common background	Efficient amount of migration

Source: own illustration

Commonly, though rarely explicitly stated, the driving forces for migration flows are separated into microeconomic and macroeconomic effects: Firstly, from the microeconomic perspective of the individual the probability of migrating is monotonously increasing function of the net expected utility gain through relocating. The expected utility gains depend on the relative characteristics of the district of origin versus the destination. The most commonly used variable in that manner would be the expected income gains as the relation of the destinations GDP per capita versus the origins GDP per capita[4]. We will call these effects the relative size effects. Secondly, from a macroeconomic perspective the total number of individuals that will migrate (driven by their individual utility maximization) will also depend on the pure size of the two districts. This scaling effect is usually covered by employing a gravity model which for example would predict the migration flows to vary with the product of the populations of both districts. We will label this effect the joint size effect. In a log-

[4] For uniformity we will define all relative effects the other way around as origin versus destination for our estimation.

linear specification we will model for each location specific characteristic its joint size effect as the sum of the logs of the corresponding origin and destination attribute. Likewise the relative size effect will be the difference of the same two values so that our basic specification of the gravity model is the following:

$$\ln(migration_{OD}) = b_R(\ln(x_O) - \ln(x_D)) + b_J(\ln(x_O) + \ln(x_D))$$

For convenience let us define for our further proceeding that $\ln(migration_{OD}) = y$ and $\ln(x) = X$ which we will use in our specification of the econometric model later.

The Data

The data was provided by the Forschungsdatenzentrum der statistischen Landesämter (Research-Data-Centre of the Statistical State Offices) which is a recent research cooperation of the 16 statistical state offices of Germany to make regional data available for academic research. All variables are used in the econometric estimation in the logs of their absolute values except for the cases in which the variables could possibly take on the value of zero. In this case for example for the number of students in a district we took the log of the number of students plus one. We estimate the effects for the distance of two districts, the population, the GDP (as in the sense of Gross District Product), the number of unemployed and the number of employed differentiated into the sectors of production of primary industrial goods, manufacturing, services and residual employment which covers inter alia agriculture and fishery. Additionally to characterize the structure of the landscape as well as rurality versus urbanity we use the area of the district subdivided into the agricultural, the urban, the recreational, the forest and the bodies of water area. To cover further effects and amenities we employ the number of students, the tourist overnight stays as measure of general non-economic attractiveness, the number of holiday homes as measure of landscape attractiveness[5]

[5] A closer look at tourist overnight stays and holiday homes reveals that these two measures are not redundant, still a relatively high correlation of 0.64 might lead to the conclusion only to include one of the measures due to collinearity. Nevertheless the exclusion of either or contributes only marginally to changes in the coefficients, which is why we left both measures in the regression.

and the number of medics as measure of health spending[6]. As health care in Germany is practically free for the citizens and the tariffs for treatments and medication that are paid by the public health insurance are largely fixed on a federal level, health expenditures will be for the most part proportional to the un-health of the district population. As such high health care spending is an indicator for disease rather than for health. Concerning its true nature disease-care spending would be the more appropriate term for these expenditures anyway. The additional data necessary to construct the distance matrices of the districts was derived from shape files provided by the Federal Office for Cartography and Geodesy. These shape files contain the district borders and are commonly used as interchange format for computer based geographic information systems. There main purpose is the generation of maps but they serve our needs in locating the centroids of the districts as well.

The Spatial Structure of Origin-Destination Flows

The type of spatial interaction used in this paper was originally developed in the context of bilateral international trade flows and has recently been adapted in a spatial econometrics context by LeSage and Pace (2007). A main characteristic of these models is that the number of observations rises quadratically with the number of regions observed. This is due to the fact that each region is a possible origin and destination as well. The migration flows span up a full matrix instead of a symmetric one because the flows of migration differ depending on the direction. Comparable to LeSage and Pace we will use the distance as common characteristic to an origin and destination pair in our gravity model. But unlike them we will - as motivated above - use the origin and destination specific characteristics as differences (in the relative size effects) and as sums (in the respective joint size effects). The implementation of a spatial weights structure in our origin-destination setting might appear challenging at first. But a simple solution is to split the effects. Therefore we will employ separate weights matrices for

---

[6] The connection between health risks and migration has also been examined in a recent empirical work by Goetz and Rupasingha (2004). While they use differences in the cancer risk due to environmental pollution we use a more direct measure. In Germany the amount of health spending is largely proportional to the number of medics which is not at least due to the fact the health care is mainly publicly financed with common more or less binding fixed budgets per medic. Further research relating especially retirement migration with health care can be found in Graves and Knapp (1988) and Gale and Heath (2000).

the origin and destination. In this fashion we are able to stick to the standard spatial autoregressive models by just considering second order lags.

Let Y be a square matrix of size $l \times m$ with l being the number of origins assigned to the rows and m being the equal number of destinations assigned to the columns. The elements $y_{l,n}$ of matrix are the logarithmized migration flows from a origin to a destination

$$Y = \begin{matrix} y_{1,1} & \cdots & y_{1,m} \\ \vdots & \ddots & \vdots \\ y_{l,1} & \cdots & y_{l,m} \end{matrix}$$

To use these flows in our model we need to rearrange this matrix into a vector which will be of size $1 \times n$ with $n = n \bullet m$. We will employ an origin-centric ordering by sorting the flows by the origin first and by the destination subordinated. Since this origin-centric ordering will be used to calculate the spatially weighted destination characteristics and destination oriented autoregressive effects we refer to this vector as y_n^{des}. For the calculation of the spatial lags it proofs helpful to first employ the whole procedure origin-centric for the destination effects and then to reshuffle the vector destination-centric to calculate the origin effects. That way the spatial structure of the weights matrices can be exploited to employ the simplifications indispensable to make the problem feasible given the computational hardware constraints. Since the procedure is identical we will only state it once for the destination effects and will later only differentiate in the nomenclature between the two types of centring if necessary.

$$\begin{array}{lccc} & & \textit{origin} & \textit{destination} \\ & y_{1,1} & 1 & 1 \\ & \vdots & \vdots & \vdots \\ & y_{1,m} & 1 & m \\ y_n^{des} = & \vdots & \vdots & \vdots \\ & \vdots & \vdots & \vdots \\ & y_{l,1} & l & 1 \\ & \vdots & \vdots & \vdots \\ & y_{l,m} & l & m \end{array}$$

Let X be the $m \times g$ matrix of g different district characteristics and $i(m)$ a $m \times 1$ vector of ones. Given the vector of migration flows we can easily construct the corresponding destination characteristics by taking the Kronecker-product $X_{D,n} = i(m) \otimes X$, likewise the origin characteristics can be constructed by $X_{O,n} = X \otimes i(m)$.

As weights matrices we will use row-normalized neighbourhood matrices while being a neighbour is defined by the distance of the centroids of two districts being below a critical threshold. By varying this threshold we will be able to analyse the sensitivity of our results to the assumptions about the spatial structures. If we look on the map of the German districts in figure 1 this approach seems to be the most reasonable because of the large variances in district size depending on the state. A first visual inspection already reveals the districts in the north-east to have much larger area than the ones in the south-west. In that respect a more sophisticated weighting scheme based on a function of the distance wouldn't make sense, since it would be biased by the district size whereas our simple row-normalized contiguity definition at least gives us a consistent average characteristic for a fixed circular area around the observed district.

Figure 1: District borders in Germany (source: based on maps provided by the Federal office for Building Regulations and Regional Planning)

The following core W^{des} of our weights matrix is constructed such that the diagonal elements are zero, thus preventing a district from being neighbour to itself. The values $w_{l,m}$ are positive and row-identical if the distance of the respective origin and destination is equal or below the predefined threshold and zero if the distance is above it. The positive row elements will be normalized such that they add up to unity.

$$W^{des} = \begin{matrix} 0 & w_{1,2} & \cdots & \cdots & w_{1,m} \\ w_{2,1} & 0 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 0 & w_{l-1,m} \\ w_{l,1} & \cdots & \cdots & w_{l,m-1} & 0 \end{matrix} \qquad \begin{matrix} \sum_{j=1}^{m} w_{1,j} = 1 \\ \sum_{j=1}^{m} w_{2,j} = 1 \\ \vdots \\ \sum_{j=1}^{m} w_{l-1,j} = 1 \\ \sum_{j=1}^{m} w_{l,j} = 1 \end{matrix}$$

Note that for the comparable analysis in the destination-centred ordering this core matrix would have to be transposed, then row-equalized and finally row-normalized. To construct the weights matrix for the destination effects we can employ the following Kronecker product:

$$W_{D,n} = I(m) \otimes W^{des} = \begin{matrix} W & 0_{1,2} & \cdots & \cdots & 0_{1,m} \\ 0_{2,1} & W & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & W & 0_{m-1,m} \\ 0_{m,1} & \cdots & \cdots & 0_{m,m-1} & W \end{matrix}$$

With $I(m)$ being an identity matrix of size m x m and $0_{m,l}$ being a matrix of zeros of size m x m. The weights matrix therefore will be of size n x n. Note that if you are not constrained by calculation power or system memory you would just construct the respective origin-effects weights matrix in the origin-centric ordering as:

$$W_{O,n} = W^{des} \otimes I(m)$$

The Econometric Model

Following Kelejian and Prucha we will extend the model from their forthcoming article in the Journal of Econometrics. Specifically we will extend both first order autoregressive processes to a second order autoregressive spatial model with second order autoregressive disturbances. For ordinary least squares the spatial dependence leads to inconsistent estimates and maximum likelihood approaches are first computationally challenging for large problems as the one observed here and second, as mentioned by Kelejian and Prucha, are inconsistent under heteroskedasticity for the existing (quasi) maximum likelihood estimators. The procedure used by Kelejian and Prucha is based on the nonlinear two-steps least-squares method developed by Amemiya (1974). Their spatial regression model is performed in three separate steps.
Firstly: the spatial regression model in (I) is estimated using a two-steps least-squares method to instrument for the spatially lagged endogenous variable. Secondly: using the residuals from the estimation in the first step the autoregressive parameters ρ_O and ρ_D in the disturbance term are estimated using a generalized method of moments approach. Thirdly: the Cochrane-Orcutt type transformation of the regression model in (I) is again estimated via the two-steps least-squares method, instrumenting for the spatial autoregression in the endogenous variable. Following Kelejian and Prucha their formulation of the basic equations changes to:

$$
\begin{aligned}
&y_n = X_{C,n}\beta_C + (X_{O,n} - X_{D,n})\beta_R + (X_{O,n} + X_{D,n})\beta_J + X_{E,n}\beta_E + \lambda_O W_{O,n} y_n + \lambda_D W_{D,n} y_n + u_n \\
&u_n = \rho_O M_{O,n} u_n + \rho_D M_{D,n} u_n + \varepsilon_n \qquad \text{(I)} \\
&|\lambda_O| + |\lambda_D| < 1 \ , \quad |\rho_O| + |\rho_D| < 1
\end{aligned}
$$

y_n is the n -dimensional vector observations the dependent variable of migration flows from the l origins to the m destinations. As we only account domestic migration which crosses at least the district border and the number of originating districts l equals the number of Destinations m, the number of observations is equal to $l \bullet m = m^2 = n$ with the intra-district-flows being set to zero. $X_{E,n}$ is a $n \times m$ matrix of dummy variables in which the j^{th} column has once the value one if the j^{th} origin is equal to the destination and otherwise zero (with

$j = 1, \ldots, m)$[7]. The remaining exogenous variables split in $X_{O,n}$, the n x g matrix of origin characteristics and $X_{D,n}$ the respective n x g matrix of the destination districts characteristics and $X_{C,n}$ the n x c matrix of common characteristics like the distance. With $k = c + g + g + m$ let $X_n = (X_{C,n}, X_{O,n} - X_{D,n}, X_{O,n} + X_{D,n}, X_{E,n})$ denote the n x k matrix of exogenous regressors. Note that $X_{O,n}$ and $X_{D,n}$ only consist of a total of m different observations, one for each district. $W_{O,n}$, $W_{D,n}$, $M_{O,n}$ and $M_{D,n}$ are the constant n x n spatial weight matrices assumed to be known. As typically done, we will assume for the later estimation that $W_{O,n} = M_{O,n}$ and $W_{D,n} = M_{D,n}$ but abstract from that simplification for the moment. Among the resulting parameters of interest from the estimation are the k-dimensional parameter vectors β_R for the relative size effects and β_J for the joint size effects. But our special curiosity belongs to the scalar parameters λ_O and λ_D which measure the autoregressive effects for spatial autoregressive process and ρ_O and ρ_D which do likewise for the spatial autoregressive disturbance process. Finally u_n is a n-dimensional vector of regression disturbances and ε_n the innovations, form a n-dimensional vector of residuals from the autoregressive disturbance which we allow to be heterogeneous. Since the units we observe will differ strongly in their characteristics like area or population such heterogeneous innovations are highly recommendable. The n x p matrix of instruments used to estimate the first stage in the first step will be denoted by H_n[8].

A discussion of the limit of the sum of the absolute values of the autoregressive parameters in (I) can be found in Lee and Liu (2006)[9]. Given the expansion to the second order spatial processes we will now state the essential changes to the assumptions by Kelejian and Prucha:

Assumption 1: (a) All diagonal elements of the spatial weighting matrices $W_{O,n}$, $W_{D,n}$, $M_{O,n}$ and $M_{D,n}$ are zero.(c) The matrices $(I - \lambda_O W_{O,n} - \lambda_D W_{D,n})$ and $(I - \rho_O M_{O,n} - \rho_D M_{D,n})$ are non-singular with $|\lambda_O| + |\lambda_O| < 1$ and $|\rho_O| + |\rho_O| < 1$.

---

[7] This is necessary since on the one hand including the variation of this migration flows for the origin equalling the destination (which has been set to zero for all this pairs) would bias the estimates. On the other hand we need this vector to be of the same size as the weights matrix.

[8] As noted in the forthcoming Kelejian and Prucha paper the instruments for the first and the third step of the procedure do not need to be the same. For notational convenience we will stick to the same notation in the third step nevertheless.

[9] A summary of the result is included in the appendix.

Assumption 3: The row and column sums of the matrices $W_{O,n}$ *,* $W_{D,n}$ *,* $M_{O,n}$ *,* $M_{D,n}$ *,* $(I - \lambda_O W_{O,n} - \lambda_D W_{D,n})^{-1}$ *and* $(I - \rho_O M_{O,n} - \rho_D M_{D,n})^{-1}$ *are bounded uniformly in absolute value.*

It follows that given (3) and assuming that an innovation ε in the vector of innovations has an expected value of zero and a finite variance of δ^2 the variance –covariance matrix is:

$$E[u_n u_n'] = \left(I_n - \rho_O M_{O,n} - \rho_D M_{D,n}\right)^{-1} diag\left[\delta_{i,n}^2\right]\left(I_n - \rho_O M_{O,n} - \rho_D M_{D,n}\right)^{-1} \qquad \text{(II)}$$

Where $\delta^2_{i,n}$ is the n x 1 vector of $E\left(u_{i,n} u_{i,n}\right)$ from which we form the diagonal matrix over the single elements i of this vector.

The GM Estimator for the Autoregressive Disturbance Parameters

Sharing Kelejians and Pruchas assumptions for the GM-estimator, Badinger and Egger (2008) derive the following moment conditions for the second order heteroskedastic autoregressive disturbance process[10]:

$$n^{-1}E\left[\bar{\varepsilon}_{O,n}'\bar{\varepsilon}_{O,n} - Tr\left[M_{O,n} diag\left[E\left[\varepsilon_{i,n}\varepsilon_{i,n}\right]\right]M_{O,n}'\right]\right]$$

$$n^{-1}E\left[\bar{\varepsilon}_{O,n}'\varepsilon\right] = 0, \quad \text{with } \bar{\varepsilon}_{O,n} = M_{O,n}\varepsilon_n$$

$$n^{-1}E\left[\bar{\varepsilon}_{D,n}'\bar{\varepsilon}_{D,n} - Tr\left[M_{D,n} diag\left[E\left[\varepsilon_{i,n}\varepsilon_{i,n}\right]\right]M_{D,n}'\right]\right]$$

$$n^{-1}E\left[\bar{\varepsilon}_{D,n}'\varepsilon\right] = 0, \quad \text{with } \bar{\varepsilon}_{D,n} = M_{D,n}\varepsilon_n \qquad \text{(III)}$$

Let us additionally define $\bar{\bar{\varepsilon}}_{O,n} = M_{O,n}\bar{\varepsilon}_{D,n}$, $\bar{\bar{\varepsilon}}_{D,n} = M_{D,n}\bar{\varepsilon}_{D,n}$, $\bar{\bar{\varepsilon}}_{OD,n} = M_{O,n}\bar{\varepsilon}_{D,n}$ and $\bar{\bar{\varepsilon}}_{DO,n} = M_{D,n}\bar{\varepsilon}_{O,n}$. Using the expression for the error process u from (I), solving it for the innovations ε in substituting it in the moment conditions in (III) we get the equation system:

---

[10] Note that the calculation of the terms in the trace expression is due to the size of the weights matrix infeasible with the commonly available hardware. Fortunately it is possible to use the special properties due to the construction of the matrix to reduce the problem to a laptop friendly size. The necessary transformations require only basic matrix algebra and are found in the appendix.

$$\gamma_n - \Gamma_n \alpha_n = 0$$

with $\alpha_n = \left(\rho_{O,n}, \rho_{D,n}, \rho_{O,n}\rho_{D,n}, \rho_{O,n}^{\;2}, \rho_{D,n}^{\;2}\right)'$, $\gamma_n = \left(\gamma_{1,n}, \gamma_{2,n}, \gamma_{3,n}, \gamma_{4,n}\right)'$

and $\Gamma_n = \left(\gamma_{rs,n}, \gamma_{2,n}, \gamma_{3,n}, \gamma_{4,n}\right)_{r=1,\dots 4, s=1,\dots 5}$ (IV)

The single elements of this equation system are:

$$\gamma_1 = n^{-1}E\left[\bar\varepsilon_{O,n}'\bar\varepsilon_{O,n} - Tr\left[M_{O,n}diag\left[E\left[\varepsilon_{i,n}\varepsilon_{i,n}\right]\right]M_{O,n}'\right]\right]$$

$$\gamma_2 = n^{-1}E\left[\bar\varepsilon_{O,n}'\varepsilon_n\right]$$

$$\gamma_3 = n^{-1}E\left[\bar\varepsilon_{D,n}'\bar\varepsilon_{D,n} - Tr\left[M_{D,n}diag\left[E\left[\varepsilon_{i,n}\varepsilon_{i,n}\right]\right]M_{D,n}'\right]\right]$$

$$\gamma_4 = n^{-1}E\left[\bar\varepsilon_{D,n}'\varepsilon_n\right]$$

$$\gamma_{1,1} = 2n^{-1}E\left[\bar{\bar\varepsilon}_{O,n}'\bar\varepsilon_{O,n} - Tr\left[M_{O,n}diag\left[E\left[\bar\varepsilon_{i,O,n}\varepsilon_{i,n}\right]\right]M_{O,n}'\right]\right]$$

$$\gamma_{2,1} = n^{-1}E\left[\bar{\bar\varepsilon}_{O,n}'\varepsilon_n + \bar\varepsilon_{O,n}'\bar\varepsilon_{O,n}\right]$$

$$\gamma_{3,1} = 2n^{-1}E\left[\bar{\bar\varepsilon}_{DO,n}'\bar\varepsilon_{D,n} - Tr\left[M_{D,n}diag\left[E\left[\bar\varepsilon_{i,O,n}\varepsilon_{i,n}\right]\right]M_{D,n}'\right]\right]$$

$$\gamma_{4,1} = n^{-1}E\left[\bar{\bar\varepsilon}_{DO,n}'\varepsilon_n + \bar\varepsilon_{D,n}'\bar\varepsilon_{O,n}\right]$$

$$\gamma_{1,2} = 2n^{-1}E\left[\bar{\bar\varepsilon}_{OD,n}'\bar\varepsilon_{O,n} - Tr\left[M_{O,n}diag\left[E\left[\bar\varepsilon_{i,D,n}\varepsilon_{i,n}\right]\right]M_{O,n}'\right]\right]$$

$$\gamma_{2,2} = n^{-1}E\left[\bar\varepsilon_{O,n}'\bar\varepsilon_{D,n} + \bar{\bar\varepsilon}_{OD,n}'\varepsilon_n\right]$$

$$\gamma_{3,2} = 2n^{-1}E\left[\bar{\bar\varepsilon}_{D,n}'\bar\varepsilon_{D,n} - Tr\left[M_{D,n}diag\left[E\left[\bar\varepsilon_{i,D,n}\varepsilon_{i,n}\right]\right]M_{D,n}'\right]\right]$$

$$\gamma_{4,2} = n^{-1}E\left[\bar{\bar\varepsilon}_{D,n}'\varepsilon_n + \bar\varepsilon_{D,n}'\bar\varepsilon_{D,n}\right]$$

$$\gamma_{1,3} = -2n^{-1}E\left[\bar{\bar\varepsilon}_{OD,n}'\bar\varepsilon_{O,n} - Tr\left[M_{O,n}diag\left[E\left[\bar\varepsilon_{i,D,n}\bar\varepsilon_{i,O,n}\right]\right]M_{O,n}'\right]\right]$$

$$\gamma_{2,3} = -n^{-1}E\left[\bar{\bar\varepsilon}_{OD,n}'\bar\varepsilon_{O,n} + \bar{\bar\varepsilon}_{O,n}'\bar\varepsilon_{D,n}\right]$$

$$\gamma_{3,3} = -2n^{-1}E\left[\bar{\bar\varepsilon}_{D,n}'\bar{\bar\varepsilon}_{DO,n} - Tr\left[M_{D,n}diag\left[E\left[\bar\varepsilon_{i,D,n}\bar\varepsilon_{i,O,n}\right]\right]M_{D,n}'\right]\right]$$

$$\gamma_{4,3} = -n^{-1}E\left[\bar{\bar\varepsilon}_{D,n}'\bar\varepsilon_{O,n} + \bar{\bar\varepsilon}_{DO,n}'\bar\varepsilon_{D,n}\right]$$

$$\gamma_{1,4} = -n^{-1}E\left[\bar{\bar\varepsilon}_{O,n}'\bar{\bar\varepsilon}_{O,n} - Tr\left[M_{O,n}diag\left[E\left[\bar\varepsilon_{i,O,n}\bar\varepsilon_{i,O,n}\right]\right]M_{O,n}'\right]\right]$$

$$\gamma_{2,4} = -n^{-1}E\left[\bar{\bar\varepsilon}_{O,n}'\bar\varepsilon_{O,n}\right]$$

$$\gamma_{3,4} = -n^{-1}E\left[\bar{\bar{\varepsilon}}_{DO,n}'\bar{\varepsilon}_{DO,n} - Tr\left[M_{D,n}diag\left[E\left[\bar{\varepsilon}_{i,O,n}\bar{\varepsilon}_{i,O,n}\right]\right]M_{D,n}'\right]\right]$$

$$\gamma_{4,4} = -n^{-1}E\left[\bar{\bar{\varepsilon}}_{DO,n}'\bar{\varepsilon}_{O,n}\right]$$

$$\gamma_{1,5} = -n^{-1}E\left[\bar{\bar{\varepsilon}}_{OD,n}'\bar{\varepsilon}_{OD,n} - Tr\left[M_{O,n}diag\left[E\left[\bar{\varepsilon}_{i,D,n}\bar{\varepsilon}_{i,D,n}\right]\right]M_{O,n}'\right]\right]$$

$$\gamma_{2,5} = -n^{-1}E\left[\bar{\bar{\varepsilon}}_{OD,n}'\bar{\varepsilon}_{D,n}\right]$$

$$\gamma_{1,5} = -n^{-1}E\left[\bar{\bar{\varepsilon}}_{D,n}'\bar{\varepsilon}_{D,n} - Tr\left[M_{D,n}diag\left[E\left[\bar{\varepsilon}_{i,D,n}\bar{\varepsilon}_{i,D,n}\right]\right]M_{D,n}'\right]\right]$$

$$\gamma_{2,5} = -n^{-1}E\left[\bar{\bar{\varepsilon}}_{D,n}'\bar{\varepsilon}_{D,n}\right] \tag{V}$$

Let $\tilde{\gamma}_n$ be the estimate of γ_n and $\tilde{\Gamma}_n$ be the estimate of Γ_n, then by substituting (V) in (IV) we can get $\rho_{O,n}$ and $\rho_{D,n}$ through solving the resulting nonlinear optimization problem:

$$\left(\tilde{\rho}_{O,n}, \tilde{\rho}_{D,n}\right) = \underset{\rho_{O,n},\rho_{D,n}}{argmin}\left[\left(\tilde{\gamma}_n - \tilde{\Gamma}_n\alpha_n\right)'\Omega\left(\tilde{\gamma}_n - \tilde{\Gamma}_n\alpha_n\right)\right] \tag{VI}$$

Besides deriving the GM-estimator for the second order spatial autoregressive disturbance Badinger and Egger also provide a Mote Carlo study showing a good performance of the estimator even in small samples.

The Choice of Instruments

For the choice of the instruments the changes to the Kelejian and Prucha approach are minor than one would expect. The only essential change is to their equation (30). For our problem the expected values of the lagged endogenous are[11]:

$$E\left(W_{O,n}y_n\right) = W_{O,n}\left(I - \lambda_O W_{O,n} - \lambda_D W_{D,n}\right)^{-1}X_n\beta_n = W_{O,n}\sum_{i=0}^{\infty}\left(\lambda_O W_{O,n} + \lambda_D W_{D,n}\right)^i X_n\beta_n$$

$$E\left(W_{D,n}y_n\right) = W_{D,n}\left(I - \lambda_O W_{O,n} - \lambda_D W_{D,n}\right)^{-1}X_n\beta_n = W_{D,n}\sum_{i=0}^{\infty}\left(\lambda_O W_{O,n} + \lambda_D W_{D,n}\right)^i X_n\beta_n \tag{VII}$$

[11] For the necessary conditions for the expansion see also the explanation of the limiting properties of the autoregressive parameters in the appendix.

We will use the instrument matrix H_n to instrument $Z_{O,n} = (X_n , W_{O,n}y_n)$, $Z_{D,n} = (X_n , W_{D,n}y_n)$, $M_{O,n}Z_{O,n} = (M_{O,n}X_n , M_{O,n}W_{O,n}y_n , M_{D,n}W_{D,n}y_n)$, $(M_{O,n} , M_{D,n})Z_{O,n} = ((M_{O,n} , M_{D,n})X_n , (M_{O,n} , M_{D,n})W_{O,n}y)$ and $(M_{O,n} , M_{D,n})Z_{D,n} = ((M_{O,n} , M_{D,n})X_n , (M_{O,n} , M_{D,n})W_{D,n}y_n)$ by estimating their predicted value with a least squares regression of H_n[12]: If we look exemplarily at $Z_{O,n}$ then with $P_H = H_n(H_n'H_n)^{-1}H_n'$ it follows that $\hat{Z}_{O,n} = P_H Z_{O,n}$. As the ideal instruments would be $E(Z_{O,n}) = (X_n, W_{O,n}E(y_n), W_{D,n}E(y_n))$ and accounting for (VII) it seems reasonable to approximate these instruments by using as H_n a subset of the linear independent columns of $(X_n, W_{O,n}X_n, W^2_{O,n}X_n, \ldots, W_{D,n}X_n, W^2_{D,n}X_n, \ldots, M_{D,n}X_n, M^2_{D,n}X_n, \ldots, M_{D,n}W_{D,n}X_n, M^2_{D,n}W_{D,n}X_n, \ldots, M_{D,n}W_{O,n}X_n, M^2_{D,n}W_{O,n}X_n, \ldots, W_{D,n}W_{O,n}X_n, W^2_{D,n}W_{O,n}X_n, \ldots)$. In the light of having row-normalized weights matrices and setting $W_{O,n} = M_{O,n}$ and $W_{D,n} = M_{D,n}$ the choice of available instruments is drastically reduced due to linear dependency[13]. Therefore we propose to use as excluded instruments the subset of the spatially cross-lagged relative size effects since these include already all spatial lags of the district characteristics[14].

The Estimation Results

In discussion of the estimation results we will first compare the different empirical specifications on the basis of the total population. Then we will present the differences in migration behaviour between female and male migrants and non-German and German nationals and will find striking results for migrants of non-German nationality. For simplicity all models in the tables will be labelled SARAR(*,**) and the term in bracket describes the order of the lag that is included with * being the order of the spatial autoregressive lag in the

---

[12] The latter terms follow from the Cochrane-Orcutt type transformation of (I) into $y_n = \rho_O M_{O,n}y_n + \rho_D M_{D,n}y_n + (I_n – \rho_O M_{O,n} – \rho_D M_{D,n})X_n\beta_n + (I_n – \rho_O M_{O,n} – \rho_D M_{D,n})W_{O,n}y_n + (I_n – \rho_O M_{O,n} – \rho_D M_{D,n})W_{D,n}y_n$ which is estimated by a two-steps least-squares estimator in the last step of the three stages of our procedure.

[13] Due to the construction of the dependent variable and the characteristics of the origins only differing for each origin interaction term but not for each origin destination interaction, the interaction terms with the destination weights matrix become linear dependent for row normalized weighting matrices, e.g.: $W_{D,n}X_{R,n} = W_{D,n}(X_{O,n} – X_{D,n}) = W_{D,n}X_{O,n} – W_{D,n}X_{D,n} = X_{O,n} – W_{D,n}X_{D,n}$ since $W_{D,n}X_{O,n} = W^2_{D,n}X_{O,n} = \ldots = X_{O,n}$ and $W_{O,n}X_{O,n} = W_{O,n}W_{D,n}X_{O,n} = W_{O,n}W^2_{D,n}X_{O,n} = \ldots$. Likewise the argument holds for the destination characteristics. For the interaction of both spatial weights matrices it obviously follows: $W_{O,n}W_{D,n}X_{R,n} = W_{O,n}X_{O,n} – W_{D,n}X_{D,n}$

[14] The cross lag of the relative size effect would be $W_{O,n}W_{D,n}(X_{O,n} – X_{D,n}) = W_{O,n}X_{O,n} – W_{D,n}X_{D,n}$.

endogenous variable and ** being the order of the lag of the spatial autoregressive disturbance. Table 2 shows the estimation results for the different specifications of the econometric model. The first column displays the results of a simple OLS regression including the spatial endogenous lag but disregarding any steps to correct for spatial autocorrelation. The second column only includes the lags in the endogenous variable whereas the third column only accounts for heteroskedastic autoregressive disturbance. The fourth column displays the results for the full model specification and the final column gives information the instruments used in the estimation.

Model	OLS		SARAR(2,0)		SARAR(0,2)		SARAR(2,2)		used as instru-ment
Innovation errors	-		-		Heteroskedastic		Heteroskedastic		
R2	0.7874		0.7877		0.3989		0.4913		
Observations	192721		192721		192721		192721		
Sargan overid			0.3607				0.8228		
Anderson Identification			0.0000				0.0000		
Moran's I Ori.			0.1071 ***		0.2884 ***		0.1071 ***		
Moran's I Des.			0.0992 ***		0.2849 ***		0.0992 ***		
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	
Ori. rho					0.4527		0.3680		
Des. rho					0.4737		0.4351		
Ori. gamma	0.4632 ***	(0.0037)	0.4399 ***	(0.0142)			0.4313 ***	(0.0375)	
Des. gamma	0.4521 ***	(0.0035)	0.4147 ***	(0.0114)			0.4124 ***	(0.0285)	
Distance	-0.3358 ***	(0.0047)	-0.3998 ***	(0.0252)	-0.2851 ***	(0.0023)	-0.1972 ***	(0.0161)	-
Rel. population	-0.0038	(0.0102)	0.0007	(0.0102)	-0.0219	(0.0196)	-0.0159	(0.0182)	++
Rel. employed rest	-0.0155 ***	(0.0033)	-0.0183 ***	(0.0035)	-0.0411 ***	(0.0064)	-0.0217 ***	(0.0059)	++
Rel. employed prim. Industry	0.0185 ***	(0.0062)	0.0168 ***	(0.0062)	0.0307 **	(0.0119)	0.0182 *	(0.0107)	++
Rel. employed manufacturing	-0.0024	(0.0057)	-0.0134 **	(0.0064)	-0.0388 ***	(0.0122)	-0.0192 *	(0.0112)	++
Rel. employed services	-0.0275	(0.0174)	-0.0624 ***	(0.0201)	-0.1701 ***	(0.0371)	-0.0771 **	(0.0347)	++
Rel. unemployed	0.0556 ***	(0.0199)	0.1240 ***	(0.0288)	0.3216 ***	(0.0503)	0.1389 ***	(0.0498)	++
Rel. GDP (districts)	-0.0477 ***	(0.0099)	-0.0651 ***	(0.0123)	-0.1368 ***	(0.0209)	-0.0597 ***	(0.0203)	+
Rel. medics	0.0149	(0.0095)	0.0145	(0.0099)	0.0221	(0.0176)	0.0272 *	(0.0162)	++
Rel. students	-0.0013 ***	(0.0004)	-0.0011 ***	(0.0004)	-0.0020 ***	(0.0007)	-0.0012 **	(0.0006)	++
Rel. tourist overnight stays	-0.0169 ***	(0.0023)	-0.0153 ***	(0.0024)	-0.0303 ***	(0.0042)	-0.0184 ***	(0.0042)	-
Rel. welfare recipients	0.0018	(0.0032)	0.0025	(0.0033)	0.0056	(0.0068)	-0.0002	(0.0060)	++
Rel. holiday homes	0.0026 *	(0.0014)	0.0009	(0.0015)	0.0017	(0.0027)	0.0032	(0.0024)	++
Rel. recreational area	0.0000	(0.0028)	-0.0001	(0.0028)	0.0092	(0.0061)	0.0026	(0.0052)	-
Rel. agricultural area	0.0051 *	(0.0029)	0.0082 ***	(0.0030)	0.0188 ***	(0.0061)	0.0092 *	(0.0054)	++
Rel. forest area	0.0080 ***	(0.0014)	0.0092 ***	(0.0015)	0.0190 ***	(0.0030)	0.0103 ***	(0.0026)	++
Rel. bodies of water	-0.0028 *	(0.0016)	-0.0030 *	(0.0016)	-0.0061 *	(0.0034)	-0.0026	(0.0029)	++
Rel. urban area	0.0078	(0.0078)	0.0033	(0.0079)	0.0044	(0.0171)	0.0074	(0.0147)	-
Joint population	0.1128 ***	(0.0103)	0.1340 ***	(0.0132)	0.6208 ***	(0.0197)	0.3127 ***	(0.0278)	o
Joint employed rest	-0.0168 ***	(0.0033)	-0.0158 ***	(0.0033)	0.0110 *	(0.0064)	-0.0121 **	(0.0057)	o
Joint employed prim. Industry	-0.0783 ***	(0.0062)	-0.0820 ***	(0.0064)	-0.2052 ***	(0.0119)	-0.1113 ***	(0.0124)	o
Joint employed manufacturing	-0.2719 ***	(0.0056)	-0.2822 ***	(0.0069)	-0.2163 ***	(0.0122)	-0.1897 ***	(0.0127)	o
Joint employed services	-0.4264 ***	(0.0171)	-0.4092 ***	(0.0183)	0.2126 ***	(0.0372)	-0.0727 **	(0.0325)	o
Joint unemployed	0.7481 ***	(0.0191)	0.7392 ***	(0.0194)	0.1299 **	(0.0504)	0.3599 ***	(0.0400)	o
Joint GDP (districts)	0.1056 ***	(0.0099)	0.1227 ***	(0.0120)	0.1683 ***	(0.0209)	0.0628 ***	(0.0203)	o
Joint medics	0.2930 ***	(0.0094)	0.3016 ***	(0.0099)	0.2199 ***	(0.0177)	0.1844 ***	(0.0162)	o
Joint students	0.0079 ***	(0.0004)	0.0081 ***	(0.0004)	0.0099 ***	(0.0007)	0.0057 ***	(0.0007)	o
Joint tourist overnight stays	0.0819 ***	(0.0022)	0.0789 ***	(0.0024)	0.0728 ***	(0.0042)	0.0617 ***	(0.0038)	o
Joint welfare recipients	-0.0733 ***	(0.0032)	-0.0742 ***	(0.0032)	-0.0281 ***	(0.0068)	-0.0326 ***	(0.0058)	o
Joint holiday homes	-0.0126 ***	(0.0014)	-0.0084 ***	(0.0021)	-0.0016	(0.0028)	-0.0039	(0.0026)	o
Joint recreational area	-0.0120 ***	(0.0028)	-0.0106 ***	(0.0029)	0.0180 ***	(0.0061)	0.0033	(0.0052)	o
Joint agricultural area	0.0717 ***	(0.0029)	0.0756 ***	(0.0033)	0.0321 ***	(0.0061)	0.0326 ***	(0.0054)	o
Joint forest area	-0.0125 ***	(0.0014)	-0.0133 ***	(0.0014)	-0.0149 ***	(0.0030)	-0.0159 ***	(0.0027)	o
Joint bodies of water	0.0212 ***	(0.0017)	0.0250 ***	(0.0022)	0.0200 ***	(0.0034)	0.0225 ***	(0.0032)	o
Joint urban area	-0.0982 ***	(0.0079)	-0.1147 ***	(0.0102)	-0.1151 ***	(0.0172)	-0.0697 ***	(0.0162)	o
Constant	5.7145 ***	(0.1741)	6.2472 ***	(0.2696)	2.0957 ***	(0.4727)	1.2424 ***	(0.3868)	

Table 2: estimation results for the total population

Comparing the Specification-Specific Results

Foremost it should be recognized that the network effects proof to be important, highly significant and within their theoretical limits regardless of the model specification[15]. The elasticities vary between 0.38 and 0.48 for all observed model specifications. In general the origin based network effects tend to be slightly higher than the destination based ones. The origin based effects cover the influence of migrants from the districts surrounding an origin and choosing the same destination while the destination based effects include the . Yet it is unclear how much the formal or informal networks contribute in relation to the herd effects to each of these effects. The most influential characteristics besides the network effects are the distance and the joint values for the population, the GDP and the prevalence of medical conditions. As motivated above we use the number of medics to measure the health conditions. Comparing these characteristics in the last three columns we can observe how strictly the estimates will be biased if spatial autoregression is neglected. The elasticity of the distance increases by 50% in the SARAR(0,2) model and doubles in the SARAR(2,2) compared to the full SARAR(2,2) specification. Even stronger are the differences in the effect of the population ranging between 0.13 and 0.62 and of the unemployment having a lower bound of 0.13 in the SARAR(0,2) model and an upper level of 0.74 in the SARAR(2,0) specification. Health threats encourage migration in narrower, but yet significantly different limits ranging from 0.18 to 0.30. In general the SARAR(0,2) model strongly overestimates the elasticities. For example all the employment related elasticities are more than doubled compared to the full specification. This again is perfectly in line with theory. Since networks increase the employment probability the actual labour market conditions become less important.

Interpreting the remaining results of the SARAR(2,2) model we will examine the effects of relative differences between origins and destination separately from the general impact of the joint size of the characteristics. Note for the interpretation that the relative characteristics are for the origin versus the destination. These differential effects are barely surprising. Migrants choose destinations with relatively higher employment. An exception is the employment in primary industrial production which discourages migration. Unsurprisingly the strongest

[15] It is indeed possible to produce estimation results that are contrary to this statement but for reasonable model specifications the results vary within narrow limits.

attraction exhibits a higher employment in the service sector. Other attracting factors are higher local income (measured by the districts GDP), the existence and size of universities (measured by the number of students) and the quantity of amenities (measured by the number of tourists). The main factors suppressing migration are relatively higher unemployment and health threats at the destination. Also we can observe a tendency to leave rural regions indicated by the agricultural and forest area. Studying the general effects of the characteristics we can distinguish the supportive joint size effects from those suppressing migration. Obviously combined population of the regions will determine the scale of migration that is possible. Thus anything but a strong significant positive elasticity would cast doubt on the whole estimation. Among the factors that generally encourage migration are the income, the amount of unemployment, health threats, the amount of amenities, the number of students and the rurality. It is not difficult to find plausible explanations for any of these effects. Income is likely to provide resources that increase mobility while unemployment might create the necessity to migrate in order improve the personal job conditions. The existence of regional health risks encourages migration as it would generally decrease the standard of living. The opposing effect holds true for the availability and quantity of amenities. Since we measure the amenities by the overnight stays we capture the influence of cultural venues, sights and other factors attracting tourists. Additionally part of the economic attractiveness is included as business travel also contributes to the accommodation figures. More students will foster migration since education increases mobility and because finishing a university degree is likely to induce a job-related relocation. Taking the agricultural area as measure for rurality and the urban area as respective measure for urbanity suggests that rural population is in general more mobile than townspeople. But it should be mentioned that these two types of migration are very different. If for example you would assume the equivalent of moving from one quarter to another would be to migrate from one village to the next than the rural version of this migration is much more likely to cross a district boarder. Joint size effects repressing the migration tendencies include higher employment which reduces the need to relocate for a job improvement and the number of welfare recipients as they are less mobile due to budget constraints and the localized welfare payments. Unclear is the effect of natural amenities since the bodies of water generally encourage migration while the forest area has a repressive effect. But as mentioned above the interpretation of the forest area is not straight forward since it will also capture other effects as since those areas are also likely to be rural or hilly. The selection of the ideal instruments has been discussed in an earlier passage. The final column in the table describes the deviation from that selection. All instruments not marked by a circle should be included principally. Nevertheless for all the estimations presented here only the instruments marked by a double positive sign were included. The instruments marked by a negative sign were excluded because they showed high correlations with the error term. The instrument with the single positive sign was excluded because it was identified by testing as being redundant. Alternative estimates including it didn't show any significant differences. An impression on how much spatial autocorrelation is captured by the endogenous autoregressive process can be gained by looking at the Moran's I measure for spatial autocorrelation. The values displayed here are the correlation statistics for the given weights matrix. In the presence of beneficial network effects we would expect the correlation to lie between zero and one. Looking at the SARAR(0,2) model we observe a positive correlation in the error terms between 0.28 and 0.29 for the origin as well as for the destination. If we now employ the endogenous spatial autoregressive process this correlation drops to almost one third of its original amount but still being highly significant so that it reasonable to further correct for spatial autoregressive disturbances.

Does Gender and Nationality Matter for the Migration Behaviour?

The inspection the different population sub-groups in table 3 mostly confirms the results for the entire population. Comparing the gender sub-groups the differences manifest especially in the employment related elasticities and in the effects of health risks. For the male population the effects of sectoral relative employment differences are close to the total population with the relative employment in the production of primary industrial goods being insignificant. For the female population the differences in employment between the origin and the destination are all insignificant but for the residual employment. Even though this effect indicates that women migrate to destinations with relatively higher employment they are still less affected than their male counterparts. This is also especially true for the relative unemployment with the elasticity of men being more than twice than that of women. In general these results display a higher importance of the availability of jobs for the male migrants than for the female ones. Nevertheless the joint effect of income is only significant for women. Thus mobility of male migrants is probably independent of the financial resources will the mobility of the female population strictly increases with income. Unfortunately we cannot further specify the source of the results that are related to the expected earnings. To do so we would need to distinguish between individual migrants and those relocating as a complete household or into an existing household. We are also unable to control for the employment status of the migrants before and after migrating. Another striking observation is that the migration behaviour of men is in general influenced stronger by health threats than that of women. But while women prefer locations with relatively lower health risks the male migrants show no significant reaction to those differences.

Model	SARAR(2,2)		SARAR(2,2)		SARAR(2,2)		SARAR(2,2)	
Population sub-group	Female		Male		Non-German nationality		German nationality	
Innovation errors	Heteroskedastic		Heteroskedastic		Heteroskedastic		Heteroskedastic	
R2	0.4599		0.4654		0.4309		0.4818	
Observations	192721		192721		192721		192721	
Sargan overid	0.9436		0.8682		0.4560		0.7945	
Anderson Identification	0.0000		0.0000		0.0000		0.0000	
Moran's I Ori.	0.0936 ***		0.0946 ***		0.0534 ***		0.0987 ***	
Moran's I Des.	0.1021 ***		0.1050 ***		0.0794 ***		0.1050 ***	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Ori. rho	0.3618		0.3562		0.2361		0.3740	
Des. rho	0.4374		0.4449		0.4559		0.4281	
Ori. gamma	0.4202 ***	(0.0395)	0.4495 ***	(0.0418)	0.4817 ***	(0.0535)	0.4427 ***	(0.0368)
Des. gamma	0.4003 ***	(0.0294)	0.4223 ***	(0.0337)	0.4243 ***	(0.0347)	0.4150 ***	(0.0295)
Distance	-0.1931 ***	(0.0160)	-0.1790 ***	(0.0175)	3.2860 ***	(0.3454)	-0.1926 ***	(0.0160)
Rel. population	-0.0168	(0.0176)	-0.0207	(0.0175)	-0.1440 ***	(0.0208)	-0.0095	(0.0182)
Rel. employed rest	-0.0202 ***	(0.0058)	-0.0165 ***	(0.0057)	-0.0465 **	(0.0167)	-0.0193 ***	(0.0058)
Rel. employed prim. Industry	0.0138	(0.0104)	0.0159	(0.0103)	-0.0108	(0.0054)	0.0174	(0.0106)
Rel. employed manufacturing	-0.0209 *	(0.0109)	-0.0145	(0.0108)	0.0029	(0.0100)	-0.0185 *	(0.0112)
Rel. employed services	-0.0935 ***	(0.0341)	-0.0422	(0.0326)	-0.0068	(0.0099)	-0.0789 **	(0.0348)
Rel. unemployed	0.1602 ***	(0.0490)	0.0835 *	(0.0465)	0.0032	(0.0300)	0.1292 ***	(0.0497)
Rel. GDP (districts)	-0.0614 ***	(0.0197)	-0.0441 **	(0.0194)	0.0342 ***	(0.0373)	-0.0478 **	(0.0198)
Rel. medics	0.0242	(0.0162)	0.0337 **	(0.0153)	-0.0503 ***	(0.0185)	0.0175	(0.0162)
Rel. students	-0.0010 *	(0.0006)	-0.0009	(0.0006)	0.0807	(0.0150)	-0.0013 **	(0.0006)
Rel. tourist overnight stays	-0.0170 ***	(0.0040)	-0.0170 ***	(0.0040)	0.0006 **	(0.0006)	-0.0173 ***	(0.0041)
Rel. welfare recipients	0.0023	(0.0059)	-0.0030	(0.0057)	-0.0093 ***	(0.0039)	0.0036	(0.0060)
Rel. holiday homes	0.0029	(0.0024)	0.0030	(0.0023)	-0.0204 *	(0.0054)	0.0028	(0.0024)
Rel. recreational area	0.0028	(0.0051)	0.0001	(0.0050)	0.0046	(0.0023)	0.0032	(0.0052)
Rel. agricultural area	0.0069	(0.0052)	0.0070	(0.0052)	-0.0032 **	(0.0048)	0.0055	(0.0053)
Rel. forest area	0.0095 ***	(0.0026)	0.0087 ***	(0.0026)	0.0107	(0.0051)	0.0104 ***	(0.0026)
Rel. bodies of water	-0.0018	(0.0028)	-0.0016	(0.0028)	0.0014	(0.0024)	-0.0033	(0.0029)
Rel. urban area	0.0082	(0.0142)	0.0125	(0.0141)	0.0018	(0.0027)	0.0080	(0.0146)
Joint population	0.2818 ***	(0.0262)	0.2639 ***	(0.0280)	0.0204 ***	(0.0133)	0.3089 ***	(0.0275)
Joint employed rest	0.0021	(0.0057)	-0.0111 **	(0.0055)	0.0765	(0.0225)	-0.0101 *	(0.0057)
Joint employed prim. Industry	-0.1169 ***	(0.0123)	-0.1052 ***	(0.0126)	0.0073 ***	(0.0066)	-0.1146 ***	(0.0125)
Joint employed manufacturing	-0.1794 ***	(0.0123)	-0.1785 ***	(0.0133)	-0.0797 ***	(0.0114)	-0.1840 ***	(0.0130)
Joint employed services	-0.0992 ***	(0.0312)	-0.0847 ***	(0.0311)	-0.1599 ***	(0.0094)	-0.0363	(0.0322)
Joint unemployed	0.3641 ***	(0.0388)	0.3758 ***	(0.0387)	-0.3709 ***	(0.0304)	0.3410 ***	(0.0399)
Joint GDP (districts)	0.0309	(0.0189)	0.0619 ***	(0.0202)	0.5509 ***	(0.0414)	0.0298	(0.0190)
Joint medics	0.2254 ***	(0.0167)	0.1388 ***	(0.0153)	0.0970 ***	(0.0275)	0.1851 ***	(0.0162)
Joint students	0.0046 ***	(0.0006)	0.0045 ***	(0.0006)	0.1745 ***	(0.0156)	0.0047 ***	(0.0006)
Joint tourist overnight stays	0.0563 ***	(0.0037)	0.0511 ***	(0.0036)	0.0043 ***	(0.0006)	0.0623 ***	(0.0038)
Joint welfare recipients	-0.0364 ***	(0.0057)	-0.0253 ***	(0.0056)	0.0333	(0.0036)	-0.0411 ***	(0.0058)
Joint holiday homes	-0.0006	(0.0026)	-0.0023	(0.0026)	-0.0008	(0.0056)	-0.0047 *	(0.0026)
Joint recreational area	0.0041	(0.0051)	0.0062	(0.0051)	-0.0018	(0.0027)	0.0051	(0.0052)
Joint agricultural area	0.0240 ***	(0.0051)	0.0208 ***	(0.0051)	0.0051 ***	(0.0051)	0.0358 ***	(0.0055)
Joint forest area	-0.0201 ***	(0.0027)	-0.0158 ***	(0.0027)	-0.0205 ***	(0.0063)	-0.0131 ***	(0.0026)
Joint bodies of water	0.0184 ***	(0.0031)	0.0167 ***	(0.0031)	-0.0270 ***	(0.0030)	0.0216 ***	(0.0032)
Joint urban area	-0.0536 ***	(0.0154)	-0.0401 ***	(0.0154)	0.0074	(0.0027)	-0.0661 ***	(0.0161)
Constant	1.9439 ***	(0.3778)	1.3087 ***	(0.3762)	0.0120 ***	(0.0134)	1.3162 ***	(0.3967)

Table 3: estimation results for male vs. female

As we observe the flows of migrant on a small regional scale we face the problem of an increasing proportion of zero migration events between an origin-destination pair if we continuously reduce the size of the observed sub-groups. To avoid this problem we employed the commonly accepted practice to add one to the total of migrants before taking the logs. For the total population an unproblematic share of less than 7% of the possible one directional flows between two districts showed no migration. For the gender sub-groups this proportion increased to 12% for the male migrants and 13% for the respective female ones. For the sub-groups of differing nationality that we will observe next this fraction is around 7% percent for the German-nationals but reaches almost 53% percent for the migrants of non-German nationality which should be kept in mind in the interpretation of the results. The migration behaviour of the German sub-group does not differ fundamentally from the total population. The main differences are not in the size but in the significance of the elasticities. As a result the effects are insignificant for the differences in the health risk, the agricultural area and the employment in the production of primary industrial goods. The joint employment in the service sector and the joint GDP also no longer show a significant effect but the joint effect of the holiday homes is now significantly negative. This indicates that generally German nationals living in areas with an attractive landscape are less likely to move. The findings for the sub-group of non-German nationals are remarkably different. The insignificance of relative unemployment and employment (except for the redundant sector) might be a sign that for this group networks are more important for finding a job than the job-market differences. Also the negative joint effect of unemployment shows that non-German nationals are unlikely to move as general job-perspectives get worse. This is rational if they dependent on networks and thus rather confront the situation by using their local connections. The importance of networks could also explain the tendency to move to districts with lower relative income. Intense networks give a competitive advantage compared to the German-nationals and allow to access the informal job-market so that the legal wage levels are less important. The importance of the sectoral employment also differs essentially from their counterparts of German nationality. A rise in the joint employment in the service sector has a much stronger repressing tendency for non-German nationals and is beside unemployment and distance the strongest inhibitory factor at all. Employment in manufacturing has a much lower elasticity and joint employment in the production of primary industrial goods even encourages migration. Some differences might also be attached to sub-groups of certain nationality or social status. The migration to districts with relatively high agricultural area could originate from mostly Polish farm labourers and the attraction to regions with relatively many welfare recipients might mirror the fact the share of non-German nationals is relatively high within this group. While the attraction to districts with relatively higher health risks is just puzzling the different reaction to the population size and the distance just indicate the fundamental differences in migration behaviour. If non-German nationals move they tend to move far. They react highly elastic to distance with a positive (!) value of 3.29. This makes sense if networks are important. The numbers of non-German nationals are relatively low and networks need a certain size to exert a positive externality. Thus the clusters of persons with common nationality will be few, far away from each other and probably located within larger agglomerations. Therefore if a migrant wants to profit from this networks, she will have to move far and preferably to a region with a relatively higher population size.

Sensitivity Analysis

In the weights matrix contiguity was defined through a threshold distance. If the distance of the centroids of a district pair was below that limit they were considered neighbours. Since the specification of the threshold is not based on theory but out of reasonable considerations and practical reasons. Considering what is reasonable anecdotal evidence and intuition would tell that the effects of networks will strongly diminish with distance. Commuting 100 kilometres or more is not uncommon and still seem to be a sensible maximum distance to make use of network benefits. Out of practical considerations 75 kilometres is the minimum distance that insures that every district does have a neighbour at all thus we choose a distance of 100 kilometres to ensure that a district has several neighbours. Nevertheless the validity of this assumption should be tested.

Model	SARAR(2,2)		SARAR(2,2)		SARAR(2,2)		SARAR(2,2)	
Max. distance neighbours	50 kilometre		75 kilometre		150 kilometre		200 kilometre	
Innovation errors	Heteroskedastic		Heteroskedastic		Heteroskedastic		Heteroskedastic	
R2	0.7025		0.5632		0.4466		0.4386	
Observations	192721		192721		192721		192721	
Sargan overid	0.0000		0.4946		0.4054		0.7588	
Anderson Identification	0.0000		0.0000		0.0000		0.0000	
Moran's I Ori.	0.0472 ***		0.0932 ***		0.0938 ***		0.0825 ***	
Moran's I Des.	0.0621 ***		0.1017 ***		0.0989 ***		0.0865 ***	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Ori. rho	0.1126		0.3035		0.4031		0.4024	
Des. rho	0.1627		0.3536		0.4693		0.4650	
Ori. gamma	0.4590 ***	(0.0171)	0.4363 ***	(0.0263)	0.4307 ***	(0.0539)	0.4570 ***	(0.0512)
Des. gamma	0.4229 ***	(0.0139)	0.4082 ***	(0.0207)	0.3818 ***	(0.0371)	0.3757 ***	(0.0363)
Distance	-0.2360 ***	(0.0247)	-0.2059 ***	(0.0176)	-0.2738 ***	(0.0186)	-0.3889 ***	(0.0196)
Rel. population	-0.0028	(0.0120)	-0.0147	(0.0154)	0.0036	(0.0222)	0.0201	(0.0220)
Rel. employed rest	-0.0093 **	(0.0042)	-0.0184 ***	(0.0052)	-0.0227 ***	(0.0066)	-0.0228 ***	(0.0071)
Rel. employed prim. Industry	0.0042	(0.0071)	0.0150	(0.0093)	0.0150	(0.0127)	0.0150	(0.0130)
Rel. employed manufacturing	-0.0133 *	(0.0076)	-0.0149	(0.0099)	-0.0319 **	(0.0129)	-0.0435 ***	(0.0134)
Rel. employed services	-0.0428 *	(0.0233)	-0.0656 **	(0.0303)	-0.0992 **	(0.0395)	-0.1293 ***	(0.0410)
Rel. unemployed	0.1315 ***	(0.0313)	0.1329 ***	(0.0445)	0.1962 ***	(0.0562)	0.2515 ***	(0.0582)
Rel. GDP (districts)	-0.0692 ***	(0.0131)	-0.0653 ***	(0.0179)	-0.0703 ***	(0.0237)	-0.0809 ***	(0.0252)
Rel. medics	0.0128	(0.0113)	0.0288 **	(0.0142)	0.0184	(0.0188)	0.0275	(0.0200)
Rel. students	-0.0012 ***	(0.0004)	-0.0013 **	(0.0005)	-0.0012	(0.0007)	-0.0009	(0.0008)
Rel. tourist overnight stays	-0.0143 ***	(0.0026)	-0.0171 ***	(0.0035)	-0.0170 ***	(0.0050)	-0.0114 **	(0.0050)
Rel. welfare recipients	-0.0056	(0.0040)	-0.0020	(0.0052)	0.0041	(0.0068)	0.0018	(0.0070)
Rel. holiday homes	0.0022	(0.0017)	0.0021	(0.0021)	0.0028	(0.0028)	0.0017	(0.0030)
Rel. recreational area	-0.0004	(0.0034)	0.0033	(0.0045)	0.0051	(0.0057)	0.0046	(0.0058)
Rel. agricultural area	0.0094 **	(0.0037)	0.0102 **	(0.0047)	0.0152 **	(0.0059)	0.0221 ***	(0.0062)
Rel. forest area	0.0060 ***	(0.0018)	0.0092 ***	(0.0023)	0.0092 ***	(0.0030)	0.0081 **	(0.0031)
Rel. bodies of water	-0.0016	(0.0020)	-0.0037	(0.0025)	0.0002	(0.0033)	0.0005	(0.0034)
Rel. urban area	0.0007	(0.0096)	0.0066	(0.0129)	-0.0131	(0.0158)	-0.0347 **	(0.0162)
Joint population	0.0265	(0.0197)	0.2140 ***	(0.0249)	0.3953 ***	(0.0319)	0.3556 ***	(0.0312)
Joint employed rest	-0.0553 ***	(0.0038)	0.0014	(0.0049)	-0.0036	(0.0064)	0.0008	(0.0067)
Joint employed prim. Industry	-0.0002	(0.0076)	-0.1177 ***	(0.0099)	-0.1365 ***	(0.0147)	-0.0973 ***	(0.0153)
Joint employed manufacturing	-0.2333 ***	(0.0087)	-0.2403 ***	(0.0097)	-0.2263 ***	(0.0156)	-0.2678 ***	(0.0165)
Joint employed services	-0.2616 ***	(0.0217)	-0.2811 ***	(0.0290)	-0.0883 **	(0.0373)	-0.1345 ***	(0.0380)
Joint unemployed	0.4046 ***	(0.0230)	0.6494 ***	(0.0337)	0.3571 ***	(0.0456)	0.4617 ***	(0.0453)
Joint GDP (districts)	0.2149 ***	(0.0152)	0.0459 **	(0.0179)	0.1326 ***	(0.0221)	0.1530 ***	(0.0241)
Joint medics	0.2646 ***	(0.0113)	0.2494 ***	(0.0135)	0.1691 ***	(0.0196)	0.1818 ***	(0.0206)
Joint students	0.0108 ***	(0.0004)	0.0066 ***	(0.0006)	0.0059 ***	(0.0008)	0.0056 ***	(0.0008)
Joint tourist overnight stays	0.0649 ***	(0.0026)	0.0653 ***	(0.0033)	0.0488 ***	(0.0045)	0.0342 ***	(0.0048)
Joint welfare recipients	-0.0622 ***	(0.0038)	-0.0488 ***	(0.0050)	-0.0346 ***	(0.0066)	-0.0433 ***	(0.0069)
Joint holiday homes	0.0152 ***	(0.0019)	-0.0030	(0.0023)	0.0084 ***	(0.0028)	0.0237 ***	(0.0030)
Joint recreational area	0.0023	(0.0035)	0.0177 ***	(0.0045)	0.0040	(0.0058)	0.0033	(0.0060)
Joint agricultural area	0.0555 ***	(0.0044)	0.0376 ***	(0.0050)	0.0401 ***	(0.0059)	0.0636 ***	(0.0063)
Joint forest area	-0.0246 ***	(0.0017)	-0.0161 ***	(0.0023)	-0.0189 ***	(0.0032)	-0.0268 ***	(0.0034)
Joint bodies of water	0.0063 **	(0.0029)	0.0174 ***	(0.0031)	0.0264 ***	(0.0043)	0.0492 ***	(0.0040)
Joint urban area	0.0298 **	(0.0142)	-0.0690 ***	(0.0157)	-0.1229 ***	(0.0172)	-0.1997 ***	(0.0172)
Constant	0.3230	(0.2239)	3.6902 ***	(0.3088)	2.6253 ***	(0.4459)	5.5379 ***	(0.4185)

Table 4: estimation results for different weights matrices

In table 4 we explore the effect of varying the assumptions about the spatial structure of the autoregressive processes by varying the threshold distance. We can observe that the estimates for the autoregressive disturbance deplete as we reduce the threshold but the elasticities for the endogenous autoregressive process vary only little. They lie between 0.43 and 0.46 for the origin and slowly declining from 0.42 to 0.38 for the destination. However, the elasticities for the other effects can vary essentially, as for relative or joint employment in the service sector or the joint GDP. Other effects again are influenced less by the choice of the spatial structure as elasticity of relative GDP or the joint employment in manufacturing. The partial sensitivity of the results to the choice of the weights matrix shows that more research towards a better theoretical foundation or econometric determination of the weights matrix is needed. Because of the persistent significance of spatial network effects we have to conclude that neglecting the spatial structure at all would be the worse choice.

Conclusions

We found strong empirical evidence for network effects being a major determinant of domestic migration.

Appendix

(Ap. I) Explanation of the limiting properties of the autoregressive parameters:

Let $S_n = I_n - \lambda_O W_{O,n} - \lambda_D W_{D,n}$ be a term which we want to invert. Then with $\|\bullet\|$ being an arbitrary matrix norm it holds true that $\left\|\sum_{j=O,D} \lambda_j W_{j,n}\right\| \leq \sum_{j=O,D} |\lambda_j| \cdot \|W_{j,n}\| \leq \left(\sum_{j=O,D} |\lambda_j|\right) \cdot \max_{j=O,D} \|W_{j,n}\|$ such that for a row normalized weights matrix the last term of the expression will be equal to one e.g. Now as we now that the inverse of S_n is equal to the expansion $S_n^{-1} = \sum_{k=0}^{\infty} \left(\sum_{j=O,D} \lambda_j W_{j,n}\right)^k$ and observing the above statement it is clear that this term will certainly converge if $\sum_{j=O,D} |\lambda_j| < 1$ and thus the added values of the higher order terms in the expansion of S_n^{-1} converge to zero. Hence S_n will be invertible. For a deeper discussion see also Lee and Liu (2006).

(Ap. II) The simplification of the trace calculation:

The calculation of the traces in the moment conditions involves the multiplication of the matrices of size n^2. As n is roughly 200,000 one such matrix would need about 30GByte of RAM just to load it. To be able to calculate with three matrices of this size it seems recommendable to either find a computer with more than 100GByte of or a method to reduce the memory demand. As we look on one of the moment conditions in (3)

$$n^{-1} E\left[\bar{\varepsilon}_{D,n}' \bar{\varepsilon}_{D,n} - Tr\left[M_{D,n} diag\left[E\left[\varepsilon_{i,n} \varepsilon_{i,n}\right]\right] M_{D,n}'\right]\right]$$

we can focus our interest on the trace $Tr[M_{D,n} diag[E[\varepsilon_{i,n}\varepsilon_{i,n}]] M_{D,n}']$ and use the condition that the innovations from the first step of the three step procedure are an unbiased estimate for the error term such that $E[\varepsilon_{i,n}\varepsilon_{i,n}] = \hat{\varepsilon}_{i,n}\hat{\varepsilon}_{i,n}$.

Additionally remember the construction of the weights matrix as being $M_{D,n} = W^{destination} \otimes I(m)$, then:

$$Tr\left[M_{D,n} diag\left[E\left[\varepsilon_{i,n}\varepsilon_{i,n}\right]\right]M_{D,n}'\right]$$

$$= Tr\left[\left(I(m)\otimes M^{des}\right) diag\left[\hat{\varepsilon}_{i,n}\hat{\varepsilon}_{i,n}\right]\left(I(m)\otimes M^{des}\right)'\right]$$

$$= Tr\left[diag\left[\hat{\varepsilon}_{i,n}\hat{\varepsilon}_{i,n}\right]\left(\left(I(m)I(m)'\right)\otimes\left(M^{des}M^{des}{}'\right)\right)\right]$$

$$= Tr\left[diag\left[\hat{\varepsilon}_{i,n}\hat{\varepsilon}_{i,n}\right]\left(I(m)\otimes\left(M^{des}{}'M^{des}\right)\right)\right]$$

$$= \sum_{k=0}^{m-1} Tr\left[diag\left[\hat{\varepsilon}_{k}\hat{\varepsilon}_{k}\right]\left(M^{des}{}'M^{des}\right)\right]$$

with $\hat{\varepsilon}_k$ being the elements $i = \{m*k+1 , \ldots , m*(k+1)\}$ of the vector $\hat{\varepsilon}_{i,n}$. So the whole calculation just involves the multiplication of the sub-matrices which are only of size m x m with $m = n^{1/2}$.

(Ap. III) A note of the simple calculation of the Moran's I statistics:

As the most common test for spatial autocorrelation we listed the Moran's I statistic. The statistic itself ranges between –1 and 1 and can be interpreted as the spatial correlation of the error terms given the assumed spatial structure (the weights matrix). For our row-normalized weights matrices it is equal to:

$$I = \frac{\hat{\varepsilon}_{i,n}' M_{D,n}\hat{\varepsilon}_{i,n}}{\hat{\varepsilon}_{i,n}'\hat{\varepsilon}_{i,n}}$$

Referring to (Ap. II) we notice that the whole problem breaks down to feasible size because:

$$M_{D,n}\hat{e}_{i,n} = I(m) \otimes M^{des}\hat{e}_{i,n} = \begin{pmatrix} M^{des}\hat{e}_{k=1} \\ M^{des}\hat{e}_{k=2} \\ \vdots \\ M^{des}\hat{e}_{k=m} \end{pmatrix}$$

As the Moran’s I statistic doesn’t tell us anything as long as we don’t know if it is significant, we also observe that:

$$z_I = \frac{\hat{e}_{i,n}{}'M_{D,n}\hat{e}_{i,n}}{n^{-1}\hat{e}_{i,n}{}'\hat{e}_{i,n}\left(Tr\left[\left(M_{D,n}{}'+M_{D,n}\right)M_{D,n}\right]\right)^{1/2}}$$

The critical trace term simplifies to:

$$Tr\left[\left(M_{D,n}{}'+M_{D,n}\right)M_{D,n}\right]$$

$$=Tr\left[\left(\left(I(m)\otimes M^{dis}\right)'+I(m)\otimes M^{dis}\right)\left(I(m)\otimes M^{dis}\right)\right]$$

$$=Tr\left[\left(I(m)'I(m)\right)\otimes\left(M^{dis}{}'M^{dis}\right)\right]+Tr\left[\left(I(m)I(m)\right)\otimes\left(M^{dis}M^{dis}\right)\right]$$

$$=Tr\left[I(m)'I(m)\right]+Tr\left[I(m)I(m)\right]+Tr\left[M^{dis}{}'M^{dis}\right]+Tr\left[M^{dis}M^{dis}\right]$$

$$=2m+Tr\left[M^{dis}{}'M^{dis}\right]+Tr\left[M^{dis}M^{dis}\right]$$

Literature

Amemiya, T. (1974). *The Nonlinear Two-Stages Least-Squares Estimator*. Journal of Econometrics, Volume 2, Number 2, pp. 105-110.

Anselin, L. and R. Florax (1995). *New Directions in Spatial Econometrics*. Springer, London.

Badinger, H. and P. Egger (2008). *Intra- and Inter-Industry Productivity Spillovers in OECD Manufacturing: A Spatial Econometric Perspective*, CESifo-Working-Paper Series, Number 2181.

Bartel, A.P. (1989). *Where do New U.S. Immigrants Live?.* Journal of Labor Economics. Volume 7, Number 4, pp. 371-91.

Bauer, T., G.S. Epstein and I.N. Gang (2007). *Herd Effects or Migration Networks? The Location Choice of Mexican Immigrants in the U.S.*. Research in Labor Economics, Volume 26, pp.199-229.

Bauer, T. and K.F. Zimmermann (1997). *Network Migration of Ethnic Germans.* International Migration Review, Volume 31, Number 1, pp.143–149.

Burda, M.C. (1995). *Migration and the Option Value of Waiting*. The Economic and Social Review, Volume 27, Number 1, pp.1-19.

Chiswick, B.R., Y.L. Lee and P.W. Miller (2001). *The Determinants of the Geographic Concentration Among Immigrants: Application to Australia.* Australasian Journal of Regional Studies. Volume 7, Number 2.

Chiswick, B.R. and P.W. Miller (1996). *Ethnic Networks and Language Proficiency among Immigrants.* Journal of Population Economics, Volume 9, Number 1, pp.19–35.

Church, J. and I. King (1993). *Bilingualism and Network Externalities.* Canadian Journal of Economics, Volume 26, Number 2, pp.337–45.

Cliff, A. and J. Ord (1973). *Spatial Autocorrelation*, Pion, London.

Cliff, A. and J. Ord (1981). *Spatial Processes, Models and Applications*. Pion, London.

Corcoran, M., L. Datcher and G. Duncan (1980). *Most Workers Find Jobs through Word of Mouth*. Monthly Labor Review, Number 103, Number 8, pp. 33–35.

Cushing, B. (2005). *Specification of Functional Form in Models of Population Migration.* Research Paper, Regional Research Institute, West Virginia University

Epstein, G.S. and I.N. Gang (2006). *The Influence of Others on Migration Plans.* Review of Development Economics, Volume 10, Number 4, pp. 652–665.

Epstein, G.S. and A.L. Hillman (1998). *Herd Effects and Migration.* CEPR Discussion Paper No. 1811.

Frey, W.H. (1995). *The New Geography of US Population Shifts: Trends toward Balkanization.* In: Farley, R. (Ed.) The State of the Union: Social Trends, pp. 271-336. Russell Sage, New York.

Gale, L.R. and W.C. Heath (2000). *Elderly Internal Migration in the United States.* Public Finance Review, Volume 28, Number 2, pp. 153-170.

Graves, P.E. and T.A. Knapp (1988). *Mobility Behavior of the Elderly*. Journal of Urban Economics, Volume 24, Number 1, pp.1-8.

Greenwoood, M.J. (1969). *An Analysis of the Determinants of Geographic Labor Mobility in* *the United States.* Review of Economics and Statistics, Volume 51, Number 2, pp.189–194.

Kelejian, H. and I. R. Prucha (1999). *A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model.* International Economic Review, Volume 40 Number 2, pp. 509-533.

Lee, L.F. (2003). *Best Spatial Two-Stage Least Square Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances.* Econometric Reviews, Volume 22, Number 4, pp.307-335.

Lee, L. and X. Liu (2006). *Efficient GMM Estimation of High Order Spatial Autoregressive Models.* Ohio State University, Working Paper.

LeSage, J.P. and R.K. Pace (2007). *Spatial Econometric Modeling of Origin-Destination Flows*, SSRN Working Paper.

Massey, D.S. (1987). *Do Undocumented Migrants Earn Lower Wages than Legal Immigrants? New Evidence from Mexico.* International Migration Review, Volume 21, Number 2, pp. 236–274.

McDonald, J.T. (2002). *The Location Choice of New Immigrants to Canada: The Role of Ethnic Networks*. In: Beach, C., Green, A. and Reitz, J. (Eds.) Canadian Immigration Policy for the 21st Century , John Deutsch Institute for the Study of Economic Policy, Queen's University, Kingston, Canada.

Mincer, J. (1978) *Family Migration Decisions.* The Journal of Political Economy, Volume 86, Number 5, pp.749–773.

Mitchell, J. and N. Pain (2003). *The Determinants of International Migration into the UK: A Panel Based Modelling Approach*. NIESR Discussion Paper, Number 216.

Montgomery, M.R., R. Stren, B. Cohen and H.E. Reed (2003). *Cities Transformed: Demographic Change and its Implications in the Developing World.* Panel on Urban Population Dynamics, Committee on Population, Division of Behavioural and Social Sciences and Education, The National Academies Press, Washington, DC.

Mortensen, D.T. and T. Vishwanath (1994*). Personal Contacts and Earnings: It is Who You Know!*. Labour Economics, Volume 1, Number 1, pp. 187–201.

Munshi, K. (2003). *Identification of Network Effects: Mexican Migrants in the U.S. Labor Market.* Quarterly Journal of Economics, Volume 118, Number 2, pp. 549–599.

Rupasingha, A. and S.J. Goetz (2004). *County Amenities and Net Migration.* Agricultural and Resource Economics Review, Volume 33, Number 2, pp. 245-254.

Schwartz, A. (1973). *Interpreting the Effect of Distance in Migration.* The Journal of Political Economy, Volume 81, Number 5, pp.1153–1169.