Explaining Spatial Homogamy

Socio-economic, Spatial and Cultural Aspects of Spatial Homogamy in the Netherlands

Karen Haandrikman[1], Leo van Wissen[1], Carel Harmsen[2] and Inge Hutter[1]

[1] Population Research Centre, Faculty of Spatial Sciences, University of Groningen
[2] Statistics Netherlands, Voorburg

Draft version 23 March 2007

Please do not quote

RuG

Abstract

Spatial homogamy, or the similarity concerning geographical background, is a dimension which has been underexposed in homogamy studies. This paper aims to explain spatial variation in spatial homogamy by means of a spatial regression. Three sets of explanations are taken into account: compositional effects, spatial determinants, and regional cultural differences. A unique geo-coded micro data set on all new cohabiters in the Netherlands in 2004 is linked to educational and income data, and is subsequently combined with indices for local differences in value orientations and regional languages, in order to explain spatial variation in spatial homogamy. Explorative Spatial Data Analysis is used to analyse the dependent and independent variables, and spatial regression techniques are applied to model spatial homogamy, thereby correcting for spatial autocorrelation. Local partner markets are determined by socio-economic and cultural factors, as spatial homogamy is found to be influenced by income, value orientations, and regional languages.

Key words

● Partner Choice ● Spatial Patterns ● Spatial homogamy coefficient ● Spatial econometrics

1 Introduction

Studies on assortative mating have found that around the world, individuals tend to look for a partner with similar characteristics. Homogamy, or the similarity between marriage or cohabitation partners, has mostly been studied from a sociological perspective; similarity in these studies is defined in terms of social class, education, religion, or ethnic background. Implicit in many of these studies is the notion that potential partners are also co-located in space: they tend to live close by. Spatial homogamy, or the similarity concerning geographical background, is the topic of this paper.

In a recent study, new cohabiters in the Netherlands were found to choose spatially homogamous partners (Haandrikman et al. 2006). The explorative study found considerable regional variation in spatial homogamy. This paper aims to explain the spatial variation in spatial homogamy by means of a spatial regression. In the paper, three sets of explanations are taken into account. First, based on literature on marital distances, several compositional effects that have been found to affect spatial homogamy, most importantly socio-economic status attributes, are considered. Specifically, the current study examines whether spatial homogamy in the Netherlands is affected by education, income and socio-economic position. Second, specific spatial determinants are considered to account for the variation in spatial homogamy. Third, the regional variation in spatial homogamy found in Haandrikman et al. (2006) seems to be related to regional cultural differences, such as religion and dialect. Therefore, regional cultural differences are taken into account in the spatial regression as well.

Recent developments in the compilation and linkage of large micro level datasets have provided us with the ability to do a large-scale study on spatial homogamy in the Netherlands. Unique micro data on all cohabiters in the Netherlands from the population register are linked to a geographic register that provides geographic coordinates on household level, to create a unique geo-coded micro level data set on all cohabiters in the Netherlands in 2004. Subsequently, micro level data from the so-called Social Statistical File containing data on income and socio-economic category is linked to all individual cohabiters, while educational data is obtained from registers of higher education. Regional cultural indicators are added to enable a regional analysis of cultural differences.

Exploratory spatial data analysis is used to analyse the dependent and independent variables using the software package GeoDa, and spatial regression techniques are applied in the explanation of spatial patterns of spatial homogamy.

2 The spatial dimension of partner choice

The spatial dimension is a relatively unexplored dimension of homogamy. In a number of international studies, spatial homogamy is mentioned (Mayfield 1972, Küchemann et al. 1974, Coleman 1979, Fisher 1980, Coleman and Haskey 1986, Clegg et al. 1998, Duncan and Smith 2002). Research on the spatial component of marriage markets has predominantly been done in the United States and the United Kingdom. In the United States in the 1940s and 50s, so-called propinquity studies were conducted, in which the proximity of bride and groom before marriage is examined. Examples of these studies are Bossard (1932) in Philadelphia, Davie and Reeves (1939) in New Haven, Koller (1948) in Columbus, Ohio, and Ellsworth (1948) in Connecticut. Most studies found that the number of marriages declines as the distance between potential spouses increases. For example, Bossard (1932) found that one third of all married couples lived within five or less blocks from each other before marriage.

For the Netherlands, the existing studies are mostly based on historical data. An overview of different historical studies that prove the existence of geographical endogamy in the Netherlands is given by Van Poppel and Ekamper (2005). Most studies examine marital horizons of specific cities or provinces, such as the cities of Delft, Arnhem and Gouda (as discussed in Van Poppel and Ekamper 2005) and the province of Zeeland (Kok 1998, cf Van Poppel and Ekamper 2005). A recent study (Haandrikman et al. 2006) showed that Dutch people choose spatially homogamous partners: half of all new cohabiters find their partner within 6 kilometres distance. Spatial homogamy was found to vary by age and former household position, by degree of urbanisation, and by region.

In most studies, spatial homogamy is examined by analysing distances between partners before marriage. In the current Dutch context, most couples either start cohabitation as a prelude to marriage, or cohabit as a substitute to marriage (Manting 1994). Therefore, the geographical similarity of partners in unions is examined for couples that start living together, whether they are married or not. Geographic similarity is then measured before cohabitation.

Geographical distance influences the distance between partners in several ways; figure 1 presents the accompanying conceptual model. First, there is strong distance decay in the probability of partner choice, since proximity increases the likelihood of unplanned social encounters between people that offer opportunities for interaction.

Second, the probability of meeting a partner further away from the home base is lower because bridging distance involves time, energy and costs. In pre-industrial times, the geographical horizon of the activity pattern of most people did not exceed a few kilometres. In the course of the nineteenth century, mobility started to increase. Not only does an increasing portion of the population live outside their birthplace, an ever larger share of the labour force works outside the place they live (Knippenberg and De Pater 1988). Moreover, the increase of participation in higher education has also contributed to the increase in mobility (e.g. Liefbroer 1999). A large share of young people leave the parental home to pursue an education, particularly those who enrol in vocational training institutes and universities. In addition, the increase in leisure time has also contributed to the increase in mobility (Van Poppel and Ekamper 2005). These changes have almost certainly had an influence on the usual pattern of choosing a marriage partner from one's region. More recently, the rise of internet has increased the probability for a geographically distant partner. However, a preliminary analysis of the 2003 Family and Fertility Survey in the Netherlands shows that less than one percent of respondents met their partner through internet[1]. Thus, although the chances for meeting a geographically distant partner have increased, the actual number of people meeting their partner through internet is rather small. Despite increases in mobility, distance is still assumed to play a role in partner choice.

Third, the distribution, size and density of the population determine the number of people who live in close proximity, and therefore influence the opportunity to meet potential partners. High population density may lead to shorter distances between partners, since high concentrations of people imply a large enough supply of potential marriage partners for its inhabitants. Living in peripheral areas decreases the accessibility to partners as the longer average distance to other people in the population implies longer travel distances to partners. Moreover, the geography of a

[1] Preliminary analysis of the 2003 Family and Fertility Survey (Onderzoek Gezinsvorming) by the author showed that 0.7 percent of the 6,728 respondents who answered the question 'How or where did you meet your current partner?' reported that they had met their partner through internet.

region may impose constraints to meeting partners by means of physical barriers such as the presence of borders, water masses or mountain ranges.

Fourth, the spatial pattern of people with desired certain characteristics influences distances to partners. Social and cultural groups tend to cluster together in space, since people with similar characteristics tend to live in same kinds of neighbourhoods, go to the same schools, and so on (Winch 1971/1958). This differential association implies geographical clustering of educational level, occupational class, income, stage in the life course, religion, dialect, and ethnic background. These clusters of similar people influence the type of people in direct proximity, and therefore affect partner choice. Besides, people generally prefer to search for partners in areas where the preferred socio-economic or cultural characteristics are expected to be dominant. Preferred cultural characteristics pertain to views concerning religion, family values, shared dialect, and so forth. This preference is related to cultural or emotional proximity, implying mutual confirmation of each other's behaviour and world views, leading to social confirmation and affection (Kalmijn 1998; Van Poppel and Ekamper 2005). The preference for a partner with the same cultural qualities stimulates the choice of a partner from the same or a culturally related region, since people in the same or related regions are assumed to share the same ideas concerning partnerships, family, and religion, and share the same language (Van Poppel and Ekamper 2005).

Fifth, the spatial pattern of settings that facilitate meeting partners influences partner choice. Bozon and Héran (1989) found that people meet their partners in a much wider range of meeting places than before. The neighbourhood as a meeting place has declined; meetings at work or study have remained stable, whereas meetings at nightclubs, parties and holiday places have increased. Schools, the workplace, the neighbourhood, voluntary associations and family networks account for about 40 percent of the meeting places of Dutch couples (Kalmijn and Flap 2001). The spatial pattern of institutional contexts that increase meeting probabilities, such as bars, schools and churches, influences the distance at which partners are found.

Sixth, differences in value orientations may lead to different patterns of partner choice. As demographic behaviour is influenced by value changes (e.g. Van de Kaa 2001), partner choice may also be affected by changes in value orientations. For instance, living in urban areas may, as a consequence of 'urban culture', facilitate the opportunity to develop new value orientations and open mindedness, leading to a larger network of friends and acquaintances, increasing the opportunity to meet more partners in a larger range of meeting places that are distributed in a larger 'space', thus increasing the distance to partners. Indeed, Blau (1977) found that with increasing urbanisation, the probability that people have wide social circles increases.

3 Regional differences in spatial homogamy

The spatial variation in spatial homogamy is a relatively unexplored dimension of homogamy. However, the findings from marital distance studies in combination with the findings from the explorative study on spatial homogamy in the Netherlands (Haandrikman et al. 2006) lead to a number of expectations. The factors that influence the level of spatial homogamy have consequences on an aggregate level. If groups of people exhibiting the same behaviour are clustered in space, a geographic analysis will yield patterns of regionally differentiated behaviour. These regional differences in spatial homogamy may be explained by compositional effects of the population, spatial determinants, and by regional cultural differences. In the next subsections, these three factors are described, and the corresponding expectations for this study are depicted.

Compositional effects

From the existing studies on spatial homogamy, a number of factors have been found to influence spatial homogamy. Marital distances have been found to vary by age (Clegg et al. 1998; Fisher 1980; Coleman and Haskey 1986), social class (Coleman and Haskey 1986; Van Poppel and Ekamper 2005), occupational class (Clegg et al. 1998), and religion (Polman 1951). Affluence or high social status has often been associated with increased marital distance. In general, higher social classes are associated with higher distances between marriage partners (e.g. Küchemann et al. 1974; Coleman and Haskey 1986; Clegg et al. 1998). In a historical study on the Dutch city of Gouda, Van Poppel and Ekamper (2005) found greater marital distances among higher social classes in the Netherlands. People in the lower social classes were much more often locally or regionally oriented, which was associated with low educational level, dialect speaking, few leisure time, and too little resources to travel outside the region. According to Knippenberg and De Pater (1988), in the beginning of the nineteenth century, locally oriented people were mainly Catholics of the lower classes. Beekink et al. (1998) studied homogamy in the Dutch town of Woerden, and found that social class homogamy declined in the period of 1830 to 1930. However, Hendrickx (1998) found no significant increases in educational homogamy, whereas Uunk (1996) found a decline in occupational homogamy in the last decades in the Netherlands. Worldwide, increases as well as decreases in educational homogamy are found (Kalmijn 1998).

Mulder and Kalmijn (2005) studied the geography of family networks in the Netherlands, and found that level of education has a significant influence on the geography of family networks, i.e. higher educated individuals were found to live much further from their family members. Moreover, persons with high socio-economic status were living further away from their family members.

Following from the above, the following hypothesis is derived.

Hypothesis 1: With increasing *socio-economic status,* the distance between partners increases.

Spatial determinants

Regional variation in spatial homogamy may also be explained by specific spatial determinants. As discussed in the previous section, living in peripheral areas leads to average longer travel distances to partners given the accessibility to potential partners. The accessibility is further limited by spatial barriers, such as water masses and mountain ranges. In a flat country such as the Netherlands, no barrier effects are expected due to differences in altitude, but the many water masses could possibly function as barriers. Moreover, the population is mainly concentrated in the western part (the so-called 'Randstad'), leaving a number of peripheral areas, mainly in the north and southwest. Indeed, in a previous study, partners living in low density areas and in the periphery were found to live further apart (Haandrikman et al. 2006). Therefore, the distance between partners should be standardized for the average distance to all other inhabitants in the Netherlands. This procedure is described in the data and methods section.

Another spatial determinant of regional patterns of spatial homogamy is the division between urban and rural. The relation between degree of urbanisation and partner choice is ambiguous. As discussed before, high population density may lead to increased distances between partners, because of more open mindedness and different value orientations in urban cultures. On the other hand, a higher population density in urban areas may also lead to shorter distances between partners, since high concentrations of people imply a large enough supply of potential marriage partners for its inhabitants. Moreover, in urbanized areas, the higher availability of jobs and educational opportunities increase meeting opportunities and therefore lower distances to partners might be expected. Summarizing, with increasing degree of urbanisation, we expect lower distances between partners.

Hypothesis 2: With increasing levels of urbanisation, the distance between partners decreases.

Regional cultural differences

Third, regional cultural differences may affect regional patterns of spatial homogamy. In a recent study, spatial homogamy was found to be particularly high in certain urban neighbourhoods, as well as in some fisherman's villages and areas in the so-called Bible Belt, where high proportions of orthodox Calvinists reside (Haandrikman et al. 2006). The reasons for these high levels of spatial homogamy may be related to underlying cultural and sociological phenomena in society that reveal something about social groups and social borders in a society. The spatial dimension of marriage patterns may reflect cultural factors such as religion, dialect and regional differences in value orientations.

Religious factors were found to have a strong influence on marital distances for the first half of the twentieth century in the Netherlands by Polman (1951). Dutch people tend to marry within their religious group, and the level of endogamy differs per denomination (Hendrickx 1994). Especially protestant denominations are more endogamous than the more liberal denominations. While religious endogamy of Catholics and re-reformed Protestants declined since the 1930s and in the post-Second World War-period until the 1980s, an upheaval was experienced in the 1980s (Hendrickx 1998). Nowadays, religion still serves as a strong predictor of spatial demographic differences in the Netherlands (Sobotka and Adigüzel 2002). In spite of the ongoing secularisation, some churches still have a marked influenced on demographic behaviour, through the shaping of attitudes concerning family matters.

Moreover, the geography of religion has been surprisingly stable over centuries. The south is predominantly Catholic, while the northern part is a mixed zone of liberal Protestants and non-denominationalists. In between the two zones, a strip of towns and villages stretching from the southwest to the north is known as the Bible Belt (e.g. Knippenberg 2005). A large share of inhabitants of the Bible Belt are Orthodox Calvinists, who are characterized by somewhat traditional demographic behaviour as compared to the rest of the country, with more traditional views on marriage, and relatively high fertility levels. The combination of Orthodox Calvinists being more endogamous, and the geography of this religious group, leads to the expectation that Orthodox Calvinists find their partners at shorter distances.

Language is another basic component of culture and therefore a basic element of regional cultural differences. Linguistic differences are broad cultural borders, which may create linguistic groups in society (e.g. Van Langevelde 1999). There are three officially recognized regional languages in the Netherlands (as recognized by the European Charter for Regional or Minority Languages) besides standard Dutch, namely Frisian, Low Saxon, and Limburgish. People who speak dialects are geographically clustered: studies have found a strong relation between geography and dialects, from which the most well-known is the dialect map of Daan and Blok (1969) that identifies 28 geographically clustered dialect groups on the basis of the perception of dialect speakers. One of the consequences of speaking a regional language might be that people within those language groups prefer partners within that group, as dialect may act as a factor increasing cultural proximity. The spatial dimension of partner choice may therefore reflect regional language as a cultural asset that is preferred by those speaking the same language. We expect that speakers of the three recognized regional language groups are more inclined to look for a partner who speaks the same dialect, and therefore that the distance between partners for dialect speakers will be shorter.

Besides religion and regional language, other regional cultural differences may be causing the regional variation in spatial homogamy. Brons (2006) studied dimensions of regional culture in the Netherlands, and found considerable regional variation in value orientations. His measurement of basic value orientations is based on indirect measures of demographic behaviour, religious adherence, and voting behaviour. This measurement implies that culture can be derived from data on collective behaviour and is based on Hofstede's measurement of national cultures (1980, 1991). The resulting five dimensions of Dutch regional culture include post materialism, individualism, egalitarian anti-conservatism, dissatisfaction, and protestant conservatism. Table 1 gives the summary characteristics for the five dimensions of regional culture. We expect high scores on *post materialism, classic individualism* and *egalitarian anti-conservatism* to be related to high distances between partners, as they indicate an increased focus on self-development, little religious influence, and less focus on traditional households and families, and therefore these characteristics may be seen as expressions of modernisation. With increasing modernisation, the geographical horizon of individuals has been found to increase (e.g. Beekink et al. 1998), and contacts between different groups in society have increased (Hendrickx 1994). As changing geographical horizons are related to changing value orientations, Brons' (2006) indices seem useful indicators for regional cultural differences that might account for part of the spatial variation in spatial homogamy. The dimension *protestant conservatism* represents conservative cultures, with high levels of male dominance and uncertainty avoidance. Given the resemblance with the previous description of the Bible Belt, we expect lower distances between partners for areas with high scores on protestant conservatism. The fifth dimension, *dissatisfaction*, represents dissatisfaction with life and society in general, and may imply a focus on the own region, and therefore, may be related to choosing a partner at shorter distances.

Hypothesis 3: With increasing levels of *protestant conservatism*, the distance between partners decreases.
Hypothesis 4: In areas where the *Frisian, Low Saxon or Limburgish regional language* is spoken, distances between partners are shorter.
Hypothesis 5: With increasing levels of *post materialism, individualism and egalitarian anti-conservatism*, the distance between partners increases.
Hypothesis 6: With increasing levels of *dissatisfaction*, the distance between partners decreases.

4 Data and Method

In this section the several data sources that are used in this study are discussed, followed by a description of the dependent variable, the operationalisation of variables, and the methodology of the spatial data analysis.

Data sources

In this paper, spatial homogamy is operationalised by measuring the distance between the addresses of new cohabiters before cohabitation. In order to construct a geocoded micro level database on cohabitation, the following data sources are used: the Dutch population register (GBA), Address Coordinates Netherlands (ACN), the Social Statistical File (SSB), the CRIHO file on educational enrolment, and a dataset on regional cultural indices (Brons 2006).

The Dutch population register, the so-called 'Gemeentelijke Basisadministratie' (GBA), is a decentralised automated population registration system, managed by the different municipalities. In the register, information on each registered inhabitant of the country is stored, such as information on the person, the parents, marriage, registered partnership, widowhood and divorce, the offspring, and the address. As house moves are reported in the GBA, migration histories can be constructed. Moreover, individuals can be linked, through using personal identification numbers, to spouses, children, and parents. The municipal population registers are of outstanding quality (Prins 2000).

As we are interested in new cohabiters, those individuals who start living together in the year 2004 are selected. These new cohabiters include both married and unmarried couples. Since marriages and registered partnerships are recorded by the local registrar, these events are directly documented in the GBA. Unmarried cohabiters are identified by using household statistics. These annual statistics are constructed by linking the personal lists of persons living at the same address. Statistics Netherlands use a set of rules to derive household positions, based on the relationships of persons to the reference person, marital status, possible children, and an imputation model to determine the remaining group. If two people moved to the same address at the same date, they are classified as a single household. The imputation model is used to determine whether the remaining persons who are living at the same address, form a single household. This logistic regression model, described in Israëls and Harmsen (1999) and Harmsen and Israëls (2003), is based on findings from the Labour Force Survey about relations between background variables and the probability to form a two-person household.

To locate new cohabiters, i.e. couples who start living together at the same address, those persons who experienced a transition in household position, from any other position on 1 January 2004 to being a partner in a couple (with or without children) on 1 January 2005, are selected[2]. The partners are matched to each other based on current address. The resulting dataset for 2004 contains 326,000 individuals (or 163,000 couples), and consists of those people who started living together in the year 2004. Of these individuals, their birth year, and former and current household position and marital status are known.

Subsequently, the addresses of cohabiters are linked to a digital file containing an x- and y-coordinate for each known address in the Netherlands, as measured in the national coordinate system. This so-called ACN-file (Adrescoördinaten Nederland) uniquely identifies each individual address through the 6-digit postal code and the house number. There are about 7 million addresses identified through ACN-coordinates, covering 95 percent of all addresses[3]. In most cases, the location of the coordinates is in the building itself.

Spatial homogamy is operationalised by measuring the distance between former addresses of new cohabiters, and are calculated by computing the Euclidian distance between the ACN-coordinates of the former addresses of partners. The resulting distance is in meters. Then, the distance between partners is corrected for population density and location, which is described in the next subsection.

Recent developments in the compilation and linkage of large micro level datasets have provided us with the ability to link up our cohabitation data to other micro level datasets. Our unique geo-coded micro level data set on all cohabiters in the Netherlands in 2004 is linked to data from the so-called Social Statistical File (SSB). The SSB consists of several linked datasets based on registrations from official sources such as tax offices, as well as survey data. On the basis of a unique identification number (based on the social security number), cohabiters are linked to data on socio-economic category and income[4].

The second linkage is to data from the so-called CRIHO-files, in which all persons who studied at an institute of higher education in the Netherlands in the period 1986-2004 are included. The CRIHO data, based on data from the Informatie Beheer Groep (IBG), include educational information for each year a person was

[2] Since the imputation model may lead to overestimation of the number of cohabiting same-sex couples (Steenhof and Harmsen, 2003), only heterosexual couples were selected for analysis.
[3] Only addresses in the Netherlands are considered.
[4] Persons registered in the GBA are identified through the so-called 'A-number', while the same persons in the SSB and the CRIHO file are documented on the basis of the so-called 'RIN-number'. Statistics Netherlands replaces A-numbers with RIN-numbers on request; however, it is not allowed to combine the two numbers.

registered at an institute for higher education, and contains degrees, majors taken, and so on. By matching the CRIHO files with the cohabiters file, for each cohabiter it is known if that person ever studied at an institute of higher education.

Lastly, regional cultural differences are measured by using Brons' (2006) dimensions of core value orientations. As discussed in the previous section, these dimensions are combinations of aspects of meta-behaviour, and represent post materialism, individualism, egalitarian anti-conservatism, dissatisfaction, and protestant conservatism. As these indices of culture are measured at municipal level, they are added to each municipality where cohabiters lived before they started living together with their partner in 2004.

Definition of the spatial homogamy coefficient

In a previous study, we found that there is considerable regional variation in spatial homogamy. Moreover, partners living in low density areas and in the periphery were found to live further apart. One important factor for this result is that the average distance to any other person in the Netherlands is also larger than in the core and densely populated regions. Therefore, the distance between partners should be standardized for *the average distance to all other inhabitants in the Netherlands*. This is done as follows. First, for a person living in municipality i we calculate the distance to all other persons in the Netherlands. For practical purposes this is approximated by aggregating to the municipality level. Let d_{ij} be the distance between the geometric centroids of municipality i and j. Then the average distance for any person living in i to another person in the Netherlands is approximated by:

$$\bar{d}_i = \frac{1}{N}\sum_j d_{ij}P_j$$

where P_j is the population size of municipality j. Distances within the same municipality are approximated by:

$$d_{ii} = \tfrac{2}{3}\sqrt{\frac{surf_i}{\pi}}$$

where $surf_i$ is the area of a municipality in square metres on 1 January 2004. The underlying assumption of this formula is that the population is uniformly distributed within the municipality and that the form is a circle.

Next, let $\bar{s}_i$ be the average distance to cohabitation partners of all persons who started cohabiting in 2004 and who were living in municipality *i* on January 1, 2004. The *spatial homogamy coefficient* for municipality *i* is calculated as:

$$SHC_i = \frac{\bar{s}_i}{\bar{d}_i}$$

A value of 0.5 of the coefficient means that for a person in municipality *i* the average distance to his or her partner before cohabitation is half that of the average distance to the average person in the Netherlands. Municipalities with a high spatial homogamy coefficient are municipalities which have a longer distance to partners compared to the expectation on the basis of their geographic location and number of inhabitants; municipalities with a low score on the spatial homogamy coefficient are municipalities which have a shorter distance than one would expect if geographic location and population were the only determinants of spatial homogamy. The application of the spatial homogamy coefficient shows that the high average distances found in the northern provinces and in the southwest are due to their peripheral position and low population density, and that conditional on these geographical factors, partner choice in these regions is not different from other regions. The spatial homogamy coefficient is a methodological novelty in analyzing geographical differences in spatial homogamy.

Operationalisation of variables

The dependent variable in the analysis is the spatial homogamy coefficient, which was discussed in the previous section. The regional analysis is based on the 483 municipalities of the Netherlands in the year 2004. The independent factors are listed in table 2 and are clarified below.

Compositional effects on spatial homogamy are measured by examining the socio-economic status of cohabiters. This is done by dividing socio-economic status into three subgroups: socio-economic category, income, and education, based on data from the Social Statistical File (SSB) and CRIHO. *Socio-economic category* is a variable dividing persons into one of the following categories, based on the category in which the highest income is earned (based on tax registration): employed, self-employed, student, retired, inactive, and recipient of benefits (social benefits, sickness benefits, benefits on grounds of disablement, or unemployment benefits). As we expect socio-economic status of new cohabiters to affect the distance to their partner before cohabitation, we have taken the socio-economic category on 1 September 2003. September was chosen as it is normally a stable month regarding changes in socio-economic category and income. For each municipality the percentage of cohabiters in each of these categories is calculated. In the regional analysis, only the category 'students' is taken into account, which reflects the proportion of new cohabiters in a municipality that is currently studying or attending some type of school.

Income is measured as the total financial income from all jobs and other resources, such as real estate revenues and other assets, for each inhabitant (SSB file). Income percentiles represent a relative distribution of income, in which all Dutch inhabitants are divided over 1-percent income groups. For the regional analysis, for each municipality the percentage of cohabiters in the first, second, third and fourth quartile of these income percentiles is calculated. Moreover, a category is added for cohabiters with missing values on income. Again, we have taken the income data on 1 September 2003.

The *education* variable is constructed from the CRIHO file[5], in which all persons that ever studied at an institute of higher education in the Netherlands in the period of 1986 to 2004 are documented. For every year that a person ever studied at one or several of these institutes, data is available on courses and degrees taken, including whether the person graduated or not, and the specific institutes of higher education. By linking the cohabiters file with the CRIHO file, it is determined whether a new cohabiter ever studied at an institute of higher education. The education variable is a dummy variable, indicating whether or not a person ever studied. For the regional analysis, the percentage of cohabiters that ever studied is calculated for each municipality.

Spatial determinants of regional variation in spatial homogamy are operationalised by examining the *degree of urbanisation* of the municipalities. Statistics Netherlands annually measures the extent of concentration of human activities (houses, jobs, schools, shops, pubs, and so forth) by calculating the average surrounding address density[6]. The surrounding address density is the number of addresses within a circular area around an address with a radius of one kilometre,

[5] The specific file used is the '1cyferHO_2005_v.0' file.
[6] 'Omgevingsadressendichtheid' in Dutch.

divided by the square of the circle, and is calculated for each 500 by 500 metre square according to the national coordinate system. The resulting variable is expressed in number of addresses per square kilometre. For the regional analysis, we use the average surrounding address density per municipality, calculated for each 500 by 500 metre square, per 1 January 2004.

Regional cultural differences are operationalised through value orientations and dialect. We measure local *value orientations* by dimensions of core value orientations, as developed by Brons (2006). The culture indices post materialism, individualism, egalitarian anti-conservatism, dissatisfaction, and protestant conservatism are measured at municipal level, and are based on demographic, religious and voting behaviour in the period from 1997 to 2003. As we expect that these value orientations have not changed within one year, we apply the data to 2004. Since in between 1 January 2003 and 1 January 2004, some municipalities changed or merged borders, some corrections were made[7].

As *religion* is not documented for individuals in the Netherlands, it is operationalised through Brons' (2006) value orientation 'protestant conservatism', as it is most strongly related to religion, and Protestantism especially (Brons 2006).

As Frisian, Low Saxon, and Limburgish are officially recognized regional languages in the Netherlands, they are chosen as independent variables for the analysis. Moreover, the geography of speakers of *Frisian, Low Saxon and Limburgish dialect* gives more reason to include these languages. Heeringa (2004) measured Dutch dialect distances by creating composite cluster maps of distances between dialects. These maps show significant borders around three language areas. The first distinct region is the province of Friesland, where Frisian is spoken. A second significant border was found south of the area in which Low Saxon is spoken, halfway the province of Gelderland. Lastly, the province of Limburg almost coincides with the area in which Limburgish is spoken, and is a third area which significantly differs with surrounding areas. Therefore, in the current regional analysis, those cohabiters residing in Friesland were classified as Frisian, those living in either Groningen,

[7] In 2004, 11 municipalities that had existed until then, were merged into 5 new municipalities: Kesteren became Neder-Betuwe, Rijssen became Rijssen-Holten, Geldrop and Mierlo merged into the new municipality Geldrop-Mierlo, the 5 municipalities of 's Gravezande, De Lier, Monster, Naaldwijk and Wateringen were absorbed in the new municipality Westland, and the municipalities of Maasland and Schipluiden were merged to the new municipality Midden-Delfland. The indices for the new municipalities were recalculated by weighing the indices with the population of the old municipalities in 2003.

Drenthe or Overijssel as Low Saxon, and the inhabitants of Limburg as Limburgish, thereby approximating three language regions.

Methodology of the spatial data analysis

Exploratory spatial data analysis is used to analyse the dependent and independent variables, using ArcGIS 9.1 and the software package GeoDa (Anselin et al. 2006). Spatial regression techniques are applied in the explanation of spatial patterns of spatial homogamy.

In spatial analyses, spatial autocorrelation may cause problems. Misspecifications in models may lead to considering spatial dependence as a side effect (Anselin and Bera 1998). In our study, there is a mismatch between the spatial unit of analysis, i.e. municipalities, and the spatial extent of local partner markets. From a previous study (Haandrikman et al. 2006) we know that in 2004, the average distance to a cohabitation partner before cohabitation was 23 kilometres, while the average diameter of a municipality is about 5 kilometres. In other words, in explaining spatial homogamy, neighbouring municipalities should be taken into account as well.

Ignoring the addition of a spatial lag to the regression equation, and estimating the model using Ordinary Least Squares may lead to an overestimation of the magnitude of the parameters, to the extent that the spatial error parameter lambda is statistically significant.

Spatial autocorrelation can be defined as the coincidence of value similarity with locational similarity (Anselin and Bera 1998). In the case of positive spatial autocorrelation, high or low values of a variable are clustered in space, while a checkerboard pattern of observations indicates negative spatial autocorrelation. To detect any possible spatial autocorrelation in data, the spatial dependence between observations needs to be modelled by means of the definition of a spatial weights matrix. The definition of this spatial weights matrix is crucial in the analysis, as it defines the neighbour structure of the units of analysis. In this matrix, each unit of analysis - in this case municipalities- is connected to other municipalities according to a predefined format. A spatial weight matrix is an N by N symmetric matrix W which states for every location those locations that belong to its neighbourhood set as nonzero elements. Two different types of spatial weight matrices are generally defined: contiguity and distance-based matrices. The contiguity criterion assumes that all municipalities with touching borders or corners are neighbours. In distance-based spatial weights matrices, a predefined distance between polygons determines which areas are defined as neighbours. Since spatial analyses require areas to share borders, islands are excluded from the analysis, which in the case of the Netherlands means that the five 'Wadden-islands' are excluded from analysis.

The choice for the specification of a spatial weights matrix often seems arbitrary. In our study, we will test three types of spatial weights matrices to see which matrix corrects the problem of spatial autocorrelation in the best way. The first weights matrix is a so-called Queen's contiguity based matrix, in which municipalities with touching borders or corners are neighbours. A Queen's 2^{nd} order contiguity matrix also takes neighbours of neighbours into account. In distance-based spatial weights matrices, the definition of a neighbour is based on the distance between points or polygons. In this study, this distance is the average distance between partners of 23 kilometres that was found in Haandrikman et al. (2006). The spatial weights matrices are defined using GeoDa.

Whether spatial autocorrelation is present in the data, can be tested with Moran’s I, using the different spatial weights matrices. If Moran's I is significant, spatial dependence should be taken into account in the specification of the regression model. Subsequently, statistics to test spatial dependence are calculated using different Lagrange Multipliers (LM), that are developed to determine whether misspecification exists in the OLS model due to the existence of spatial dependence in the form of spatial autocorrelation of the residuals (spatial errors) or spatial autocorrelation of the dependent variable (spatial lags) (Anselin 2003). The methodology for the proper diagnosis of either a lag or an error model is fixed (Anselin 2005)[8]. All spatial regression models were estimated using Maximum Likelihood methods in GeoDa.

---

[8] If neither LM-lag nor LM-error is significant, there is no reason to add a spatially lagged variable to the regression equation. However, when the LM-lag is found (more) significant, lag autocorrelation is more likely the correct the error structure. When the LM-error is found (more) significant, error autocorrelation is more likely the correct the error structure. When both LM-lag and LM-error are significant, the Robust LM-lag and LM-error should be considered. When the Robust LM-lag is more significant, a lag model should be chosen, while an error model should be opted for when the Robust LM-error is more significant. When both Robust LM-lag and LM-error are significant, the model with the largest value for the test-statistic should be chosen (based on Anselin 2005).

5 Results

In this section the exploratory spatial data analysis of the dependent and independent variables are described, followed by the specification of the multivariate regression model, and the spatial regression model. The spatial regression model includes socio-economic, spatial and cultural variables.

Exploratory spatial data analysis

Figure 2 shows the map of the spatial homogamy coefficient for all 483 municipalities in the Netherlands in 2004. The average spatial homogamy coefficient for the whole of the Netherlands is 0.23; the coefficient ranges from 0.09 to 0.48. The spatial variation in spatial homogamy is evident, even when corrected for population density and geographic location of municipalities. A cursory visual assessment demonstrates a clustering of high values in centrally located municipalities, and local clusters of low values in the north, east and south of the country.

Using the different spatial weights matrices, Moran's I is calculated to test for spatial autocorrelation (table 3). For all three weights matrices, Moran's I is significant, meaning that the null hypothesis of spatial randomness can be rejected[9]. The positive values of Moran's I indicate positive spatial autocorrelation, or municipalities with low or high spatial homogamy coefficients are clustered in space.

While Moran's I gives an indication of global spatial autocorrelation, or the extent to which clustering is present in the data, local spatial autocorrelation statistics (LISA) test whether clusters appear in the data, and indicates whether significant spatial autocorrelation appears for each location. Figure 3 shows a LISA cluster map, depicting locations with significant local Moran statistics, according to different significance levels[10]. The bright red and bright blue locations are indications of spatial clusters (high surrounded by high, and low surrounded by low, respectively). The pink and light blue municipalities can be defined as spatial outliers, as they reflect high surrounded by low, and low surrounded by high values. The overall pattern depicts a cluster of high values on the spatial homogamy coefficient in centrally located municipalities, while a cluster of low values on spatial homogamy is found in the north, north east, and the south east (Limburg).

---

[9] This is a pseudo significance calculated with a randomisation process with 999 permutations.

[10] The significance of the local Moran statistics is based on a conditional permutation procedure. In replicating the procedure, slightly different results appear. However, the global pattern of local clusters is robust.

Multivariate regression model

An Ordinary Least Squares estimation of a linear regression model is conducted to understand the global relationships between spatial homogamy and compositional effects, spatial determinants, and regional cultural differences. Table 4 displays the list of independent variables taken into account in the regression model. The construction of variables was described in the data and methods section.

Regression results are presented in table 5. Coefficients and z-values are summarized in the first two columns of table, and model fit statistics are provided below the coefficients. The R^2 is 0.43, indicating a relatively good fit. As spatial autocorrelation is found in the data, the OLS coefficients are likely to be biased in the absence of a spatial lag. Once correcting for the problem of spatial autocorrelation, the model gives more reliable coefficients.

The spatial diagnostics, based on the three different spatial weights matrices are shown in table 6. For all three matrices, the LM-lag as well as the LM-error are highly significant. For the matrix based on the Queen's criterion, Robust LM-lag is more significant than Robust LM-error, and therefore a spatial lag model is more likely to correct for spatial bias in the model. However, for the other two matrices, a spatial error model may fit the data better. For comparative purposes, a spatial error model is run for all three weights matrices.

Spatial regression

A spatial regression is conducted to explain geographical variation in spatial homogamy at the municipal level, taking a spatial process for the disturbance terms into account. The spatial error model is specified as follows:

$$y = X\beta + \varepsilon\ ,$$

$$\varepsilon = \lambda W\varepsilon + \xi$$

where y is a vector of observations for the dependent variable, X is a matrix of observations for the explanatory variables, β is a vector of parameters to be estimated, and ε is a vector of spatially correlated residuals. W is the spatial weights matrix, ξ is a vector with residuals, and λ is the spatial autoregressive coefficient for the error lag Wε.

The models are run by means of maximum likelihood, where the spatial regression models include a spatial autoregressive error term. Coefficients, z-values and accompanying significance levels are displayed in table 3. To begin the comparison, it is useful to examine the model fit statistics for both the OLS and the spatial error models. It is not appropriate to use the R^2 as an indicator for model fit, since the R^2 given by the maximum likelihood are so-called pseudo-R^2, which cannot be compared to OLS results. The proper measures are log-likelihood, AIC and SC. The log-likelihood is highest for the model based on the Queen's criterion with second order contiguity, so when neighbouring municipalities and adjacent municipalities are taken into account. Compensating the improved fit for the added variable, the AIC and SC both decrease relative to OLS, again suggesting an improvement of fit.

The spatial autoregressive coefficient is estimated as 0.6447 for the Queen's model, as 0.8578 for the Queen's 2nd order model, and as 0.8547 for the distance-based model, and is highly significant for all three models. The addition of the extra spatial variable in the model leads to some changes in the coefficients of the error model in comparison with the OLS model, which is discussed below.

Our first hypothesis is partly confirmed using the spatial error models. Three of the income variables, cohabiters in the second and third quartile of income percentiles, and those with missing values on income, turn out to contribute significantly to the model. These variables are robust throughout the different error models. Compared to cohabiters in the highest income quartile, cohabiters in the second and third income groups have lower spatial homogamy coefficients, or shorter distances to partners. Missing values on income make a significant positive contribution to the model, indicating that a missing value on income increases the distance to a partner. Although being in the lowest income quartile makes a significant contribution to the OLS model, in the spatial error model it does not, which may be caused by lambda taking the spatial effects into account.

Surprisingly, the fact that a person ever studied at an institute of higher education does not contribute to variation in spatial homogamy in the Queen's 2nd order and distance-based error model. The fact that this variable was highly significant in the OLS model indicates that it is largely influenced by spatial effects. Also, currently being a student does not affect spatial homogamy in any of the models.

Level of urbanisation is not found to have a significant influence on the distances between partners; our second hypothesis is not supported by any of the models.

The third set of hypotheses gives some mixed results. We expected that religion, measured by the index of protestant conservatism, would negatively influence the spatial homogamy coefficient. However, in all three models, there is no effect of protestant conservatism.

In areas where regional languages are spoken, we expected shorter distances to partners. The three language variables show the least robust results. In the OLS model and the spatial error model based on the Queen's criterion, all language variables are significant, but in the two other spatial error models different results arise. When adjacent municipalities are taken into account, and when the average distance between partners is taken as a distance cut-off point, only living in the Low-Saxon language area contributes significantly and negatively to spatial homogamy.

Hypothesis 5 stated that with increasing levels of post materialism, individualism and egalitarian anti-conservatism, distance between partners increases. This hypothesis can be confirmed, in all three spatial error models. All three dimensions of basic value orientations have a positive and significant contribution to spatial homogamy. When we correct for spatial effects, the indices individualism and egalitarian anti-conservatism have an even greater effect on spatial homogamy.

The extent to which areas are dissatisfied with life and society was expected to lead to lower distances between cohabitation partners. This hypothesis can also be confirmed: the higher the index of dissatisfaction, the lower the distances between partners.

If the spatial regression is conducted correctly, the residuals of the spatial error model should not contain any remaining spatial autocorrelation. Indeed, the Moran's I for the residuals is found to be essentially zero for all three spatial weights matrices (see table 5), indicating that the addition of the spatial autoregressive coefficient in the spatial error model has eliminated all spatial autocorrelation. Figure 4 maps the residuals of the spatial regression based on the Queen's 2nd order criterion, showing a dispersed rather than a clustered pattern.

6 Summary and discussion

This paper examined spatial variation in spatial homogamy in the Netherlands, by taking three sets of explanations into account: compositional effects, spatial determinants, and regional cultural differences. Spatial homogamy is measured by means of a methodological novelty, a *spatial homogamy coefficient* that measures the distance between partners before cohabitation, and standardises for regional differences in residential location and population density. Since partner markets operate on a local level, neighbouring municipalities were taken into account in the spatial regression. A spatial autoregressive coefficient was estimated and was found highly significant, using three different types of spatial weights matrices. Income, living in the Low Saxon regional language area, post materialism, individualism, egalitarian anti-conservatism, and dissatisfaction were found to account for a large share of the spatial variation in spatial homogamy.

Compositional effects on spatial homogamy, operationalised by socio-economic category, income, and education, showed mixed results. We should note here that the aggregation of individual data to municipal level may affect the results. Aggregation of individual data to municipal level may lead to less variation compared to individual differences. Only income was found to contribute significantly to variations in spatial homogamy, thereby confirming previous studies. Low income was found to be related to lower distances to partners. Although previous studies did not find solid evidence on the effect of education on spatial homogamy, it is quite surprising to find that both the percentage of cohabiters that ever studied at an institute of higher education and the percentage of cohabiters that are currently students had no significant impact on variation in spatial homogamy. In the standard regression model, the percentage ever studied did significantly contribute to the model, but when spatial effects were taken into account, the effect disappeared. This could mean that educational level does not play a role in the distance at which partners are found. However, it could also be a result of measurement. The education variable indicates whether a person ever studied at a university of other institute of higher education. A distinction in different levels of education might yield other results, for instance differences in spatial homogamy between persons with only primary education and persons who have completed vocational training. The analysis could be improved by adding more detailed data on education; however, currently these data are not available.

The impact of spatial determinants on spatial homogamy was accounted for by degree of urbanisation, and was found non significant in the model. The ambiguous relation between urbanisation and partner choice may be a cause of this result. High concentrations of people (and jobs and educational institutes) in urban areas can lead to shorter distances between partners because of the mass of people available, but urban culture may also lead to increased distances between partners by the facilitation of opportunities to develop new value orientations. In this study we found that living in urban areas per se does not seem to influence spatial homogamy, however, certain value orientations that are more common in urban areas may affect results.

Indeed, four out of five basic value orientations that were included in the model had a significant impact on spatial homogamy. With increasing post materialism, individualism and egalitarian anti-conservatism, distance between partners increases. Post materialism is a value orientation related to self-expression, feminine values and anti-conservatism, while individualism is focused on personal freedom and is more materialistic and egoistic (Brons 2006). Egalitarian anti-conservatism indicates little religious influence, low power distance and modern families. These three value orientations are closely related and can be seen as indicators of modernisation. The results of our study are complementary to studies on increasing openness of societies during modernisation processes (e.g. Smits 1996, Van der Putte 2003). Modernisation theory assumes that boundaries between groups become less strong as modernisation proceeds. For instance, Beekink et al. (1998) describe economic, social and cultural changes in the Netherlands in the last two centuries that have had a major impact on interpersonal relationships. The growth in education, the increased importance attached to education, the increase in social and geographical mobility, and the expansion of the welfare state have widened the autonomy of individuals and have decreased the effectiveness of sanctions on social norms. The association of a decrease in spatial homogamy with high values of post materialism, individualism and egalitarian anti-conservatism can be interpreted as a result of modernisation leading to a more open partner choice, in which geography is less of a factor of importance.

The relation education and culture is intriguing. When the model was run without the cultural indicators, there was a positive and significant effect of both educational level and percentage of cohabiters that are students, on the variation in spatial homogamy. This may mean that culture and education are related. However, the indices of post materialism and classic individualism are corrected for education, although the correction was done by standardizing for the average percentage of *higher* educated in the labour force (Brons 2005). Similar to our measurement of education, this only accounts for differences between higher and lower educated persons, and does not distinguish between other educational groups. Again, more detailed educational data would improve the analysis.

The extent to which people are dissatisfied with life and society in general was found to increase spatial homogamy. In areas with higher values on dissatisfaction, distances between cohabitation partners were shorter. This may be seen as a reverse development to modernisation processes, expressed by finding a partner at shorter distances.

Although we expected religion to affect spatial homogamy, no significant results were found in this regard. In the standard regression model, protestant conservatism did explain part of the variation in spatial homogamy, however, after correcting for spatial effects, the effect disappeared. The municipality with the shortest distances to partners, Urk, with a median of 800 metres before cohabitation, also has the highest score on protestant conservatism. However, Urk has one of the highest residual levels, indicating that spatial homogamy in this area is related to other factors not accounted for in this paper. A case study approach on outliers might shed some light on remaining determinants of spatial homogamy.

The finding that living in the Low Saxon language area significantly decreases the distance to partners might indicate that the preference for cultural similarities in a partner leads to finding a partner at close range. Why this is not valid for Frisian and Limburgish language, is not straightforward. Further research should clarify causal relationships between dialect and partner choice.

In this paper, the evident clusters of spatial homogamy (in the north, east and south east) and spatial heterogamy (in the centre of the country) were explained by a combination of socio-economic and cultural determinants. The remaining spatial variation is a stimulation for further research.

By applying spatial interaction methods, the flows of partners of specific localities can be mapped. In this way, the extremely low distances between partners in for instance Urk, and very high distances in centrally located areas such as Almere can be explored. The influence of the spatial pattern of meeting places, such as workplace, schools, and pubs will be explored. As the CRIHO data provides information on where individuals studied, and which specific courses they took, educational homogamy can be combined with spatial homogamy. Moreover, the SSB files also provide information on work place for each individual, so that schools and workplaces as meeting places can be studied.

As the socio-economic and educational data we have is available for each cohabiter, a multilevel approach could be adopted to include both individual level data and area level data.

As a previous study found that on individual level age and household position also affect spatial homogamy (Haandrikman et al. 2006), further studies aim at comparing results for different stages in the life course, such as for those who start cohabitation directly from the parental home, or those who were living alone before.

This paper has provided new insights into spatial assortative mating, by applying methods from spatial econometrics. Cupid's wings are not adapted for long flights, but high income, and post modernist, individualist and anti-conservatist value orientations make Cupid fly further from home.

References

Anselin, L. and A.K. Bera (1998), Spatial dependence in linear regression models with an introduction to spatial econometrics. In: A. Ullah and D.E.A. Giles, *Handbook of applied economic statistics*. Marcel Dekker Inc, New York, Basel, Hong Kong, pp. 237-289.

Anselin, L. (2003), Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26(2): 153-166.

Anselin, L. (2005), Exploring spatial data with GeoDa: A workbook. Spatial Analysis Laboratory, University of Illinois, and Center for Spatially Integrated Social Science.

Anselin, L., I. Syabri, and Y. Kho (2006), GeoDa, an introduction to spatial data analysis. *Geographical Analysis* 38: 5-22.

Beekink, E., A.C. Liefbroer and F.W.A. van Poppel (1998), Changes in choice of spouse as an indicator of a society in a state of transition: Woerden, 1830-1930. *Historical Social Research* 23 (1/2): 231-253.

Blau, P.M. (1977), *Inequality and heterogeneity. A primitive theory of social structure*. The Free Press, New York, and Collier Macmillan Publishers, London.

Bossard, J.H.S. (1932), Residential propinquity as a factor in marriage selection. *American Journal of Sociology* 38 (2): 219-224.

Bozon, M. and F. Héran (1989), Finding a spouse: A survey of how French couples meet. *Population: an English selection* 44 (1): 91-121.

Brons, L.L. (2005), Rethinking the culture-economy dialectic. Dissertation University of Groningen. Available at http://www.lc.dds.nl/brons/CED.htm.

Brons, L.L. (2006), Indirect measurement of regional culture in the Netherlands. *Tijdschrift voor Economische en Sociale Geografie* 97 (5): 547-566.

CBS (Statistics Netherlands) (2006), Statline, central database of Statistics Netherlands. Accessible at http://statline.cbs.nl/.

Clegg, E.J., T.J. Ringrose and J.F. Cross (1998), Some factors affecting marital distances in the Outer Hebrides. *Journal of Biosocial Science* 30 (1): 43-62.

Coleman, D.A. (1979), A study of the spatial aspects of partner choice from a human biological viewpoint. *Man NS* 14 (3): 414-435.

Coleman, D.A. and J.C. Haskey (1986), Marital distance and its geographical orientation in England and Wales, 1979. *Transactions of the Institute of British Geographers* 11: 33 7-355.

Daan, J. and D.P. Blok (1969), *Van Randstad tot Landrand*; toelichting bij de kaart: Dialecten en Naamkunde, volume XXXVII of Bijdragen en mededelingen der Dialektencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam. Noord-Hollandsche Uitgevers Maatschappij, Amsterdam.

Davie, M.R. and R.J. Reeves (1939), Propinquity of residence before marriage. *American Journal of Sociology* 44 (4): 510-517.

Duncan, S. and S. Smith (2002), Geographies of family formations: Spatial differences and gender cultures in Britain. *Transactions of the Institute of British Geographers* 27: 471-493.

Ellsworth, J.S. (1948), The relationship of population density to residential propinquity as a factor in marriage selection. *American Sociological Review* 13 (4): 444-448.

Fisher, W.A. (1980), *The Soviet marriage market. Mate selection in Russia and the USSR.* Praeger Scientific, New York.

Haandrikman, K., C. Harmsen and L.J.G. van Wissen (2006), Spatial homogamy in the Netherlands. Distances between partners before cohabitation. Unpublished paper.

Harmsen, C. and A. Israëls (2003), Register-based household statistics. Paper presented at the European Population Conference 2003, 26-30 August 2003, Warsaw, Poland.

Heeringa, W. (2004), *Measuring dialect pronunciation differences using Levenshtein distance*. Dissertation University of Groningen, Groningen.

Hendrickx, J. (1994), *The analysis of religious assortative marriage. An application of design techniques for categorical models.* Thela Thesis, Amsterdam.

Hendrickx, J. (1998), Religious and educational assortative marriage patterns in the Netherlands, 1940-1985. *The Netherlands Journal of Social Sciences* 34 (1): 5-22.

Hofstede, G. (1980), *Culture's consequences: International differences in work-related values*. Sage, Beverly Hills.

Hofstede, G. (1991), *Cultures and organizations: Software of the mind*. McGraw-Hill, London.

Israëls, A. and C. Harmsen (1999), Imputatiemodel voor jaarlijkse huishoudensstatistiek; adressen met twee niet-in-gezinsverband-levende personen. Interne notitie Centraal Bureau voor de Statistiek, Divisie Research en Ontwikkeling, Sector Statistische Methoden. Centraal Bureau voor de Statistiek, Voorburg.

Kalmijn, M. (1998), Intermarriage and homogamy: causes, patterns, trends. *Annual Review of Sociology* 24: 395-421.

Kalmijn, M. and H. Flap (2001), Assortative meeting and mating: unintended consequences of organized settings for partner choices. *Social Forces* 79 (4): 1289-1312.

Knippenberg, H. (2005), The Netherlands. Selling churches and building mosques. In: H. Knippenberg (ed.), *The changing religious landscape of Europe*. Het Spinhuis, Amsterdam, pp. 88-106.

Knippenberg, H. and B. de Pater (1988), *De eenwording van Nederland: Schaalvergroting en integratie sinds 1800*. SUN, Nijmegen.

Koller, M.R. (1948), Residential propinquity of white mates at marriage in relation to age and occupation of males, Columbus, Ohio, 1938 and 1946. *American Sociologial Review* 13: 613-616.

Küchemann, C.F., G.A. Harrison, R.W. Hiorns and P.J. Carrivick (1974), Social class and marital distance in Oxford city. *Annals of Human Biology* 1: 13-27.

Liefbroer, A.C. (1999), From youth to adulthood: understanding changing patterns of family formation from a life course perspective. In: L.J.G. van Wissen, P.A. Dykstra (eds.), *Population issues: An interdisciplinary focus*. Kluwer Academic/ Plenum Publishers: New York, pp. 53-85.

Manting, D. (1994), *Dynamics in marriage and cohabitation. An inter-temporal, life course analysis of first union formation and dissolution.* Thesis Publishers, Amsterdam.

Mayfield, R.C. (1972), The spatial structure of a selected interpersonal contact: A regional comparison of marriage distances in India. In: P.W. English and R.C. Mayfield (eds.), *Man, space, and environment. Concepts in contemporary human geography.* Oxford University Press, New York, pp. 385-401.

Mulder, C.H. and M. Kalmijn (2005), The geography of family networks. Paper for the IUSSP XXV International Population Conference, Tours, France, 18-23 July 2005.

Polman, A. (1951), Geografische en confessionele invloeden bij de huwelijkskeuze in Nederland. Stenfert Kroese, Leiden, nr. 8.

Prins, C.J.M. (2000), Dutch population statistics based on population register data. *Maandstatistiek van de Bevolking* 48 (2): 9-15.

Smits, J. (1996), *Trouwpatronen en sociale openheid. Opleidingshomogamie en beroepshomogamie in een zestigtal landen landen.* Dissertation Katholieke Universiteit Nijmegen.

Sobotka, T. and F. Adigüzel (2002), Religiosity and spatial demographic differences in the Netherlands. SOM Research Report 02F65. University of Groningen, Groningen.

Steenhof, L. and C. Harmsen (2003), Same-sex couples in the Netherlands. Paper for workshop on Comparative Research, 30 June-2 July 2003, Rome.

Ultee, W.C. and R. Luijck (1990), Educational heterogamy and father-to-son occupational mobility in 23 industrial nations. *European Sociological Review* 6: 125-149.

Uunk, W. (1996), *Who marries whom? The role of social origin, education and high culture in mate selection of industrial societies during the twentieth century.* Proefschrift Katholieke Universiteit Nijmegen.

Uunk, W. and M. Kalmijn (1996), Wie trouwt met wie binnen de opleidingselite? De invloed van studierichting en onderwijsniveau. *Sociologische Gids* 43 (3): 183-203.

Uunk, W. and W.C. Ultee (1995), Partnerkeuze in Nederland: Trends in herkomst- en opleidingshomogamie tussen 1947 en 1992. *Sociale Wetenschappen* 38 (2): 23-46.

Van de Kaa, D. (2001), Postmodern fertility preferences: From changing value orientation to new behaviour. In; R.A. Bulatao and J.B. Casterline (eds), *Global fertility transition*. Population and Development Review, a supplement to vol 27. Population Council, New York.

Van der Putte, B. (2003), *Het belang van de toegeschreven positie in een moderniserende wereld. Partnerkeuze in de 19e eeuwse Vlaamse steden*

(Leuven, Aalst en Gent). Dissertation, Faculteit Sociale Wetneschappen, Katholieke Universiteit Leuven.

Van Langevelde, A. (1999), *Bilingualism and regional economic development. A Dooyeweerdian case study of Fryslan*. Netherlands Geographic Studies, Utrecht/ Groningen.

Van Poppel, F. and P. Ekamper (2005), De Goudse horizon verruimd. Veranderingen in de herkomst van Goudse bruiden en bruidegoms bruidegoms. In: M. van Leeuwen and J. Kok (eds.), *Genegenheid en gelegenheid. Twee eeuwen partnerkeuze en huwelijk.* Aksant, Amsterdam, pp.181-212.

Van Tulder, J. (1972), Het sociaal gemengd huwen. In: R. Beerling, J.K. Gevers and H. Philipsen (eds.), *Onderzocht en overdacht. Sociologische opstellen voor prof.dr. F van Heek*, speciale aflevering Mens en Maatschappij, Universitaire Pers Rotterdam, Rotterdam, pp. 92-110.

Winch, R.F. (1971/1958), *Mate-selection. A study of complementary needs*. Brown Reprints, Dubuque, Iowa. Originally from Harper and Brothers, New York, 1958.

Tables and figures

Figure 1. Determinants of distance to partners

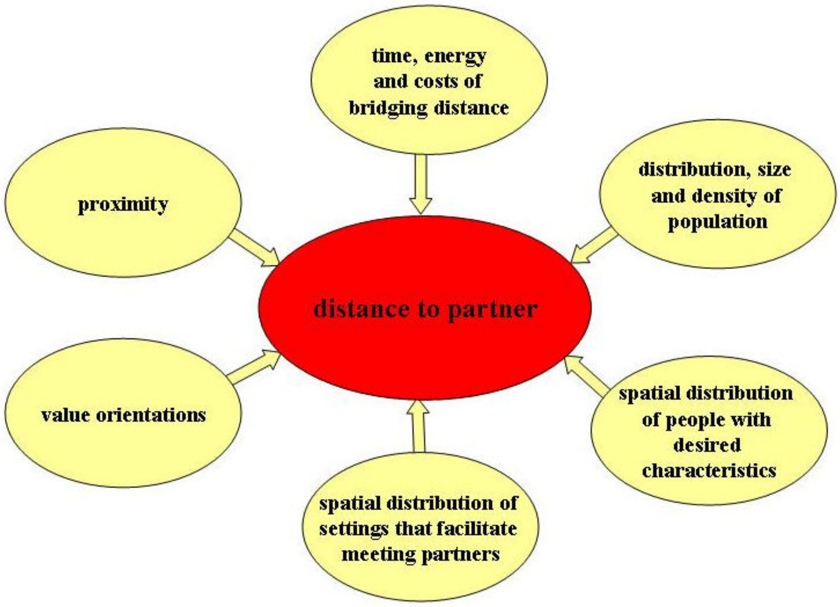


Table 1. Summary characteristics of dimensions of regional culture

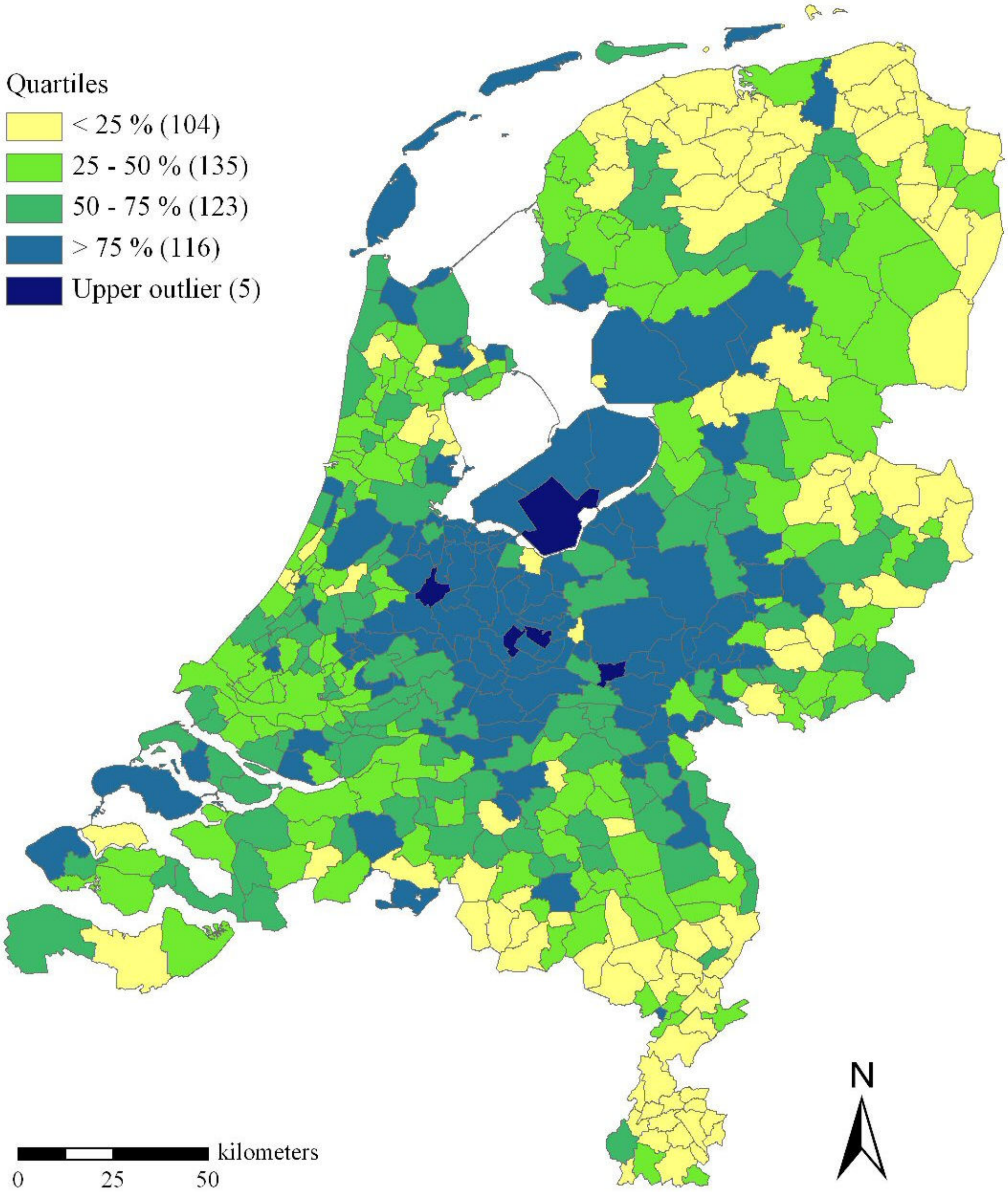
	high score	**low score**
Post Materialism	focus on self-development / self-expression co-operative and egalitarian (very) small households many votes for progressive parties environmentally conscious	focus on material wellbeing competitive and authoritarian large households many votes for conservative parties
Protestant Conservatism	predominantly Protestant early marriage and childbearing traditional / large households and families male dominance	predominantly Catholic little early marriage and childbearing
Classic Individualism	individual is more important postponement of marriage and childbearing many votes for liberal parties	national or collective interests more important relatively early marriage and childbearing
Egalitarian Anti-Conservatism	little religious influence modern households and families many votes for social democratic parties egalitarian and co-operative	predominantly Catholic traditional / married households and families many votes for conservative parties authoritarian and competitive
Dissatisfaction	dissatisfied with life and society many votes for political reform movement	relatively little dissatisfaction

Adapted from Brons (2006), p. 562.

Table 2. Independent variables

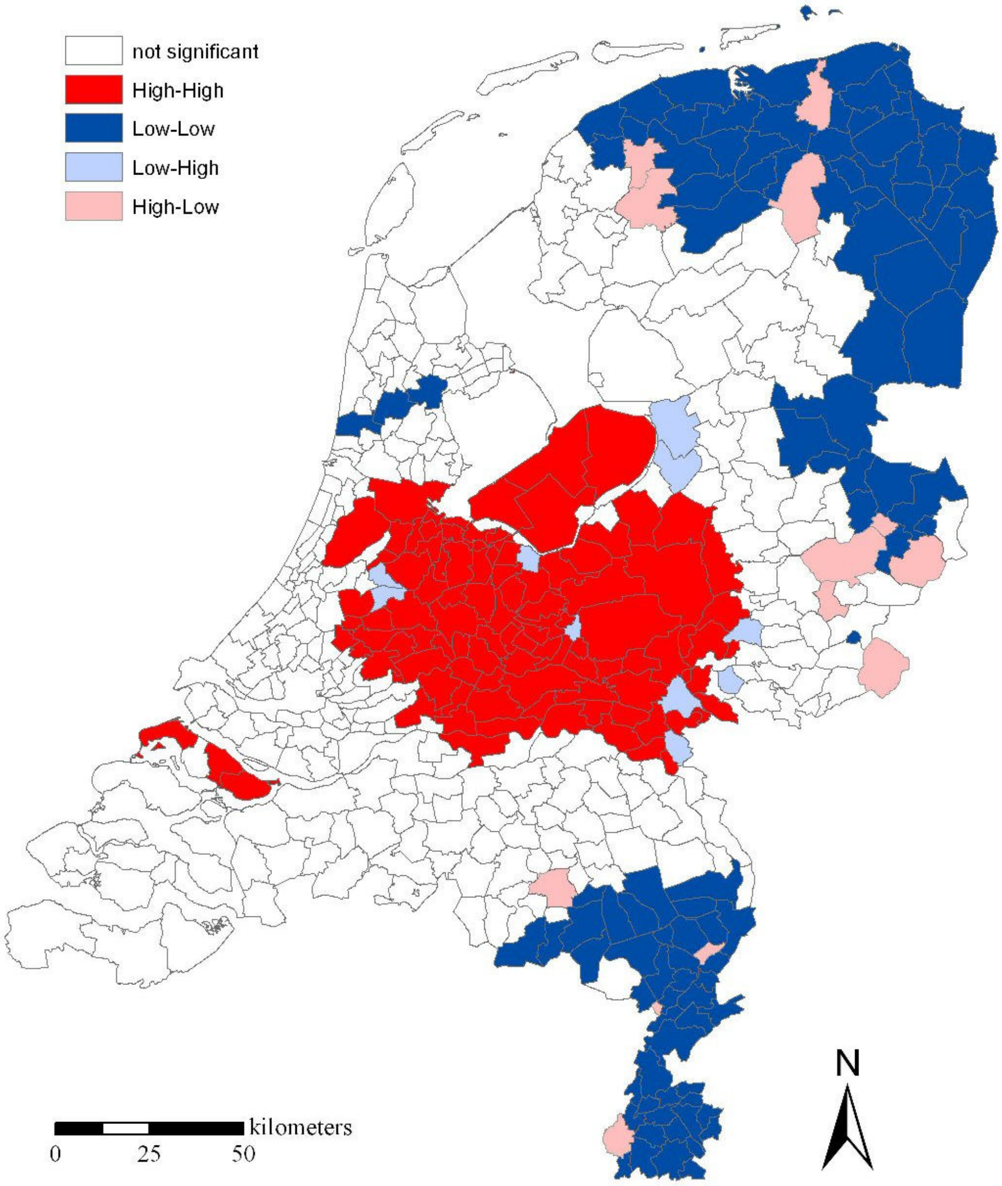
Factors influencing the spatial variation in spatial homogamy	**Operationalisation**	**Variables**	**Source**
Compositional effects: Socio-economic status	Socio-economic category	- Percentage of cohabiters that are students	SSB
	Income	- Percentage cohabiters in first quartile of income percentiles	SSB
		- Percentage cohabiters in second quartile of income percentiles	SSB
		- Percentage cohabiters in third quartile of income percentiles	SSB
		- Percentage cohabiters with missing value on income percentiles	SSB
	Education	- Percentage of cohabiters that ever studied	CRIHO
Spatial determinants	Degree of urbanisation	- Average surrounding address density	CBS (2006)
Regional cultural differences	Religion	- Protestant conservatism score	Brons (2006)
	Value orientations	- Post materialism score, corrected for degree urbanization, education and income - Protestant conservatism score - Classic individualism score, corrected for education and income - Egalitarian anti-conservatism score - Dissatisfaction score	Brons (2006)
	Dialect: Frisian, Low Saxon and Limburgish	- Residing in province of Friesland / Groningen, Drenthe and Overijssel / Limburg	

Figure 2. Map of the spatial homogamy coefficient



Source: © 2005, Statistics Netherlands / Topografische Dienst Kadaster

Figure 3. LISA cluster map of spatial homogamy coefficient



Source: © 2005, Statistics Netherlands / Topografische Dienst Kadaster

Table 3. Moran's I for different weights matrices

	Queen's contiguity	Queen's contiguity 2nd order	Distance between partners
Moran's I	0,5392 ***	0,4752 ***	0,4481 ***

***: significant at 0,01.

Table 4. Descriptive statistics of independent variables

Variable	Mean	St.dev	Range	N
Percentage cohabiters that are students	0,0942	0,0396	0 - 0,3373	478
Percentage cohabiters in first quartile of income percentiles	0,1664	0,0409	0,0714 - 0,3132	478
Percentage cohabiters in second quartile of income percentiles	0,2285	0,0416	0,0952 - 0,3455	478
Percentage cohabiters in third quartile of income percentiles	0,3649	0,0570	0,0714 - 0,5481	478
Percentage cohabiters with missing value on income percentiles	0,0527	0,0224	0 - 0,1560	478
Percentage of cohabiters that ever studied	0,2559	0,0712	0,1019 - 0,6149	478
Average surrounding address density per municipality	891	683	109 - 5987	478
Post materialism score*	0,0107	0,4985	-1,8300 - 1,5300	478
Classic individualism score**	-0,0003	0,4253	-1,7200 - 1,4800	478
Egalitarian anti-conservatism score	-0,0083	0,9825	-2,7300 - 2,7700	478
Dissatisfaction score	-0,0032	1,0001	-3,1200 - 3,5800	478
Protestant conservatism score	0,0009	1,0072	-1,4200 - 6,1500	478
Residing in Frisian regional language region	0,0565	0,2311	0 - 1	478
Residing in Low Saxon regional language region	0,1318	0,3386	0 - 1	478
Residing in Limburgish regional language region	0,0983	0,2981	0 - 1	478

* corrected for degree urbanization, education and income
** corrected for education and income

Table 5. OLS and spatial regression results*

	OLS		**Spatial error model based on**					
			Queen's contiguity		**Queen's contiguity 2nd order**		**Distance between partners**	
	coefficients	**z-value (sign.)**	**coefficients**	**z-value (sign.)**	**coefficients**	**z-value (sign.)**	**coefficients**	**z-value (sign.)**
Constant	0,3343	7,9389 ***	0,3154	8,6740 ***	0,2870	7,6003 ***	0,2814	7,1594 ***
Socio-economic variables								
Students (nonstudents REF.)	0,0017	0,0174	0,0302	0,3673	0,1154	1,4776	0,1065	1,3326
1st quartile income (4th quartile REF.)	-0,1829	-2,0914 **	-0,0572	-0,7719	-0,0555	-0,7680	-0,0079	-0,1060
2nd quartile income (4th quartile REF.)	-0,2030	-3,2220 ***	-0,1858	-3,2426 ***	-0,1277	-2,2522 **	-0,1616	-2,7627 ***
3rd quartile income (4th quartile REF.)	-0,1819	-3,2141 ***	-0,1683	-3,5377 ***	-0,1565	-3,3538 ***	-0,1332	-2,7719 ***
Missing value income (4th quartile REF.)	0,3513	3,0048 ***	0,2760	2,9325 ***	0,3701	4,0253 ***	0,3428	3,6025 ***
Ever studied	0,1393	2,6096 ***	0,0729	1,6751 *	0,0407	0,9670	0,0519	1,2102
Spatial variables								
Address density	0,0000	-0,5700	0,0000	0,8827	0,0000	0,5379	0,0000	0,2602
Cultural variables								
Protestant conservatism	0,0118	4,7571 ***	0,0019	0,6910	0,0012	0,4515	0,0019	0,6910
Frisian language (Dutch REF.)	-0,0502	-4,2193 ***	-0,0550	-3,1478 ***	-0,0258	-1,3262	-0,0321	-1,3380
Low Saxon language (Dutch REF.)	-0,0555	-6,4616 ***	-0,0484	-4,0015 ***	-0,0233	-1,7780 *	-0,0279	-1,9752 **
Limburgish language (Dutch REF.)	-0,0517	-5,9456 ***	-0,0336	-2,6098 ***	-0,0199	-1,2911	-0,0235	-1,3317
Post materialism	0,0238	5,0314 ***	0,0150	3,7164 ***	0,0148	3,7807 ***	0,0172	4,2758 ***
Individualism	0,0107	1,8604 *	0,0112	2,2621 **	0,0114	2,4775 **	0,0149	3,1341 ***
Egalitarian anti-conservatism	0,0058	1,8589 *	0,0121	3,7211 ***	0,0117	3,8614 ***	0,0136	4,3040 ***
Dissatisfaction	-0,0110	-3,6373 ***	-0,0083	-2,5219 **	-0,0097	-3,0187 ***	-0,0081	-2,5658 **
Lambda			0,6447	15,4108 ***	0,8578	25,7260 ***	0,8547	23,2192 ***
Log-likelihood		815,97		885,64		909,37		900,16
R^2		0,43		0,61		0,65		0,63
Akaike information criterion		-1599,93		-1739,27		-1786,73		-1768,32
Schwarz criterion		-1533,22		-1672,56		-1720,02		-1701,61
Moran's I of residuals				-0,0755		-0,0258		-0,0220

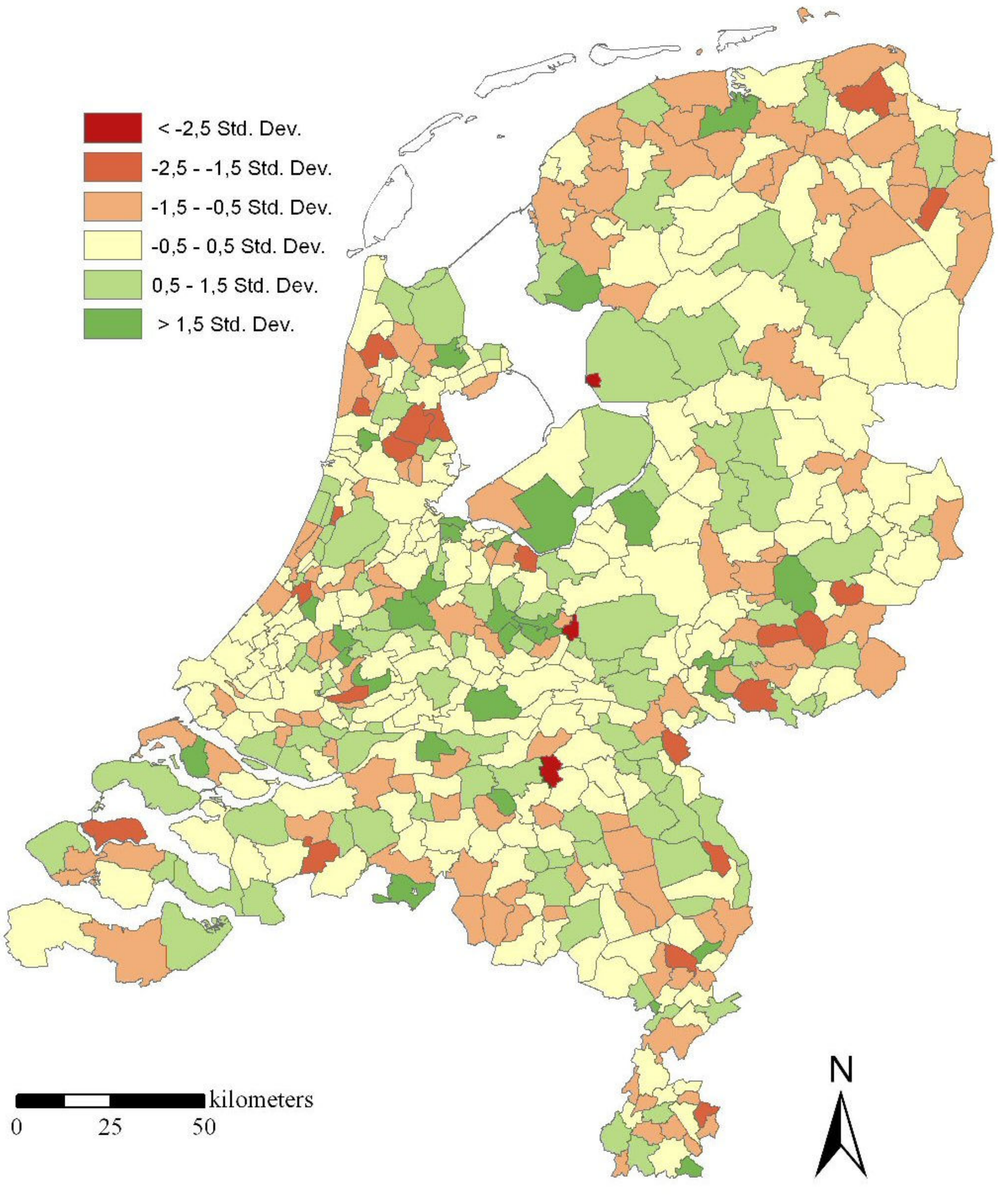
* Significance levels: 0,10 (*), 0,05 (**), 0,01 (***).

Table 6. Spatial diagnostics of OLS regression based on three different spatial weights matrices

	Queen's contiguity		Queen's contiguity 2nd order		Distance between partners	
Test statistic	**Value**	**Probability**	**Value**	**Probability**	**Value**	**Probability**
LM (lag)*	150,61	0,00000	281,86	0,00000	239,66	0,00000
Robust LM (lag)	12,07	0,00051	17,69	0,00003	12,69	0,00037
LM (error)	145,11	0,00000	326,38	0,00000	292,95	0,00000
Robust LM (error)	6,56	0,01042	62,21	0,00000	65,98	0,00000
LM (SARMA)	157,17	0,00000	344,07	0,00000	305,64	0,00000

* LM = Lagrange Multiplier

Figure 4. Residuals of the spatial regression model using Queen's 2nd order contiguity matrix



Source: © 2005, Statistics Netherlands / Topografische Dienst Kadaster