Explaining spatial homogamy: compositional, spatial and regional cultural determinants of regional patterns of spatial homogamy in the Netherlands

Haandrikman, K.; van Wissen, L.J.G.; Harmsen, C.N.

Link to publication in KNAW Research Portal

Explaining Spatial Homogamy. Compositional, Spatial and Regional Cultural Determinants of Regional Patterns of Spatial Homogamy in the Netherlands

Karen Haandrikman · Leo J. G. van Wissen · Carel N. Harmsen

Abstract Spatial homogamy, or sharing a similarity in geographical origin, is an under-researched dimension in homogamy studies. In the Netherlands, people tend to choose spatially homogamous partners. Moreover, there is considerable regional variation in spatial homogamy, even when residential location and population density are controlled for. This study aims to explain the regional 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. The data used consists of a unique geo-coded micro dataset on all new cohabiters in the Netherlands in 2004 (*N*=289,248), combined with other data from varying sources. In the spatial regression, the dependent variable is the standardized distance coefficient, based on the distance between partners before cohabitation, standardised for the average distance to other inhabitants. We find that especially educational, income and cultural differences contribute to the regional variation in spatial homogamy.

Keywords Partner choice · Homogamy · Spatial patterns · Standardized distance coefficient · Spatial econometrics · The Netherlands

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

K. Haandrikman (✉) · L. J. G. van Wissen
Population Research Centre, Faculty of Spatial Sciences, University of Groningen, Groningen, the Netherlands
e-mail: karenhaandrikman@gmail.com

C. N. Harmsen
Statistics Netherlands, The Hague, the Netherlands

married or cohabitating 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 shared similarity in geographical background, is the topic of the present study.

In a recent study, new cohabiters in the Netherlands were found to choose spatially homogamous partners (Haandrikman et al. 2008a). The explorative study found considerable regional variation in spatial homogamy. This article aims to explain the regional variation in spatial homogamy by means of a spatial regression. Three sets of explanations are taken into account. First, based on the literature on marital distances, compositional factors that have been found to affect spatial homogamy—most importantly demographic and socio-economic status attributes— are considered. Second, specific spatial determinants are examined so as to account for the variation in spatial homogamy. Third, regional cultural differences, particularly variation in religion, language and value orientations, may be related to regional differences in marital distances.

Recent developments in the compilation and linkage of large micro-level datasets have enabled us to conduct a large-scale study on spatial homogamy in the Netherlands. As we are interested in the spatial dimension, we aggregated micro data of all new cohabiters in 2004, taken from the population register, and linked these to geographic coordinates for each separate household address, in order to make a regional comparison. The dependent variable that was used is a so-called standardized distance coefficient, which corrects the average distance to partners to the average distance to all other Dutchmen. Subsequently, explanatory variables were derived from different sources. Besides using annual regional statistics from Statistics Netherlands and regional cultural indicators, micro-level data on educational enrolment was linked to all cohabiters, out of which regional indicators were constructed. Exploratory spatial data analysis was used to analyse the dependent and independent variables using GeoDa, and spatial regression techniques were applied to explain regional patterns of spatial homogamy.

The Spatial Dimension of Partner Choice: Background and Expectations

The spatial dimension is a relatively unexplored dimension of homogamy. 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 was examined (e.g. Bossard 1932; Davie and Reeves 1939; Ellsworth 1948; Koller 1948). 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 blocks from each other before marriage. Van Poppel and Ekamper (2005) provide an overview of different historical studies that prove the existence of geographical endogamy in the Netherlands. However, most studies are outdated, based on historical data, and usually restricted to cities or regions. A recent study (Haandrikman et al. 2008a) showed that Dutch people choose spatially homogamous partners: half of all new cohabiters find their partner within a 6-km distance.

Geographical distance influences partner choice in four ways, as described by Haandrikman et al. (2008a). Proximity increases the likelihood of spontaneous encounters, and therefore distance decay is highly pertinent in partner choice. Second, notwithstanding increases in mobility, educational enrolment and leisure time, bridging distance (still) involves time, energy and costs, and therefore partner choice still occurs at a local scale. Thirdly, physical barriers, population density and degree of urbanisation influence the access to potential partners and therefore impact meeting opportunities. Living in peripheral areas leads to average longer travel distances to partners given the accessibility to potential partners, which is further limited by spatial barriers such as water masses and mountain ranges. Fourth, the spatial pattern of potential candidates with certain characteristics influences partner choice. Geographical clustering of religion, dialect or other cultural assets, but also of socio-economic attributes may imply cultural proximity, leading to the preference of a spatially homogamous partner. 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 share the same language and are assumed to share the same ideas concerning partnerships, family, and religion (Van Poppel and Ekamper 2005).

Regional variation in spatial homogamy results from different processes. The following paragraphs describe these explanatory processes and discuss the expectations for the current study.

First, from the literature on marital distances, several *compositional factors* have been found to affect spatial homogamy, most importantly demographic and socio-economic status attributes. Spatial clustering of people with the same characteristics, which is very common (e.g. Winch 1971; Goode 1982), may lead to patterns of regionally differentiated behaviour. As age homogamy is more common than age heterogamy (for instance Van Poppel et al. 2001; De Graaf et al. 2003), the availability of potential partners in certain age groups affects meeting and mating opportunities. The most extreme case is the so-called 'marriage squeeze', where men or women are confronted by a shortage of partners their age because of variations in birth numbers (Ni Bhrolcháin 2001). Another compositional effect found to affect spatial homogamy patterns is socio-economic status. Higher social classes are generally associated with longer distances between marriage partners (e.g. Küchemann et al. 1974; Coleman and Haskey 1986; Clegg et al. 1998; Van Poppel and Ekamper 2005; Haandrikman et al. 2008b). A combination of preferences, strong norms to marry within the class, and geographically extensive opportunities to meet partners might lead to greater distances. For the UK, Fielding (1992) found that higher education led to widening horizons of the middle class: these groups tend to find partners in other regions. Perhaps this is also related to the fact that education is a strong proxy for cultural lifestyle (Hendrickx 1998). Especially in the past, the lower social classes were more often locally oriented, partly due to limited (travelling) means. People of similar socio-economic status tend to cluster in space (Winch 1971), which probably leads to regional differences in marital distances between different socio-economic groups.

Second, regional variation in spatial homogamy may also be explained by specific *spatial determinants*. In urban areas, partners may be found at shorter distances, since high concentrations of people, jobs and educational opportunities increase

meeting opportunities. In peripheral areas on the other hand, a fewer number of potential candidates in near proximity might lead to greater mean distances.

Third, regional differences in nuptiality have in the past been found to be related to cultural factors: regions with similar cultural characteristics showed similar patterns of marriage, even after controlling for the level of modernisation (Coale and Watkins 1986). *Regional cultural differences* in religion, language and value orientations are among the most studied and important variables in this regard. The geography of religion in the Netherlands has been surprisingly stable over the centuries. The south is predominantly Catholic, while the northern part is a mixed zone of liberal Protestants and non-denominationalists. Buffered between the two zones is a strip of towns and villages stretching from the southwest to the north known as the Bible belt (e.g. Knippenberg 2005). A large proportion of the people living in the Bible belt are orthodox Calvinists, who are characterised by rather traditional demographic behaviour as compared to the rest of the country. This group holds more traditional views on marriage and their fertility is relatively high. Dutch people tend to marry within their religious group; 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 has declined since the 1930s, an upheaval was experienced in the 1980s (Hendrickx 1998). Religion was found to have a strong influence on marital distances for the first half of the twentieth century in the Netherlands (Polman 1951). 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 more orthodox denominations still have a marked influence on demographic behaviour, through the shaping of attitudes concerning family matters. A recent study by Haandrikman et al. (2008a) revealed particularly high spatial homogamy in the Bible belt.

Language is another key component of culture and therefore a major element of regional cultural differences. Linguistic differences are broad cultural borders, which may create linguistic groups in society (e.g. Van Langevelde 1999). Speaking a dialect or regional language may lead people to prefer partners from the same language group, as was found in the US (Stevens and Schoen 1988). Language then acts as a factor increasing cultural proximity. In the Netherlands, there are three officially recognised regional languages (as proclaimed by the European Charter for Regional or Minority Languages) besides standard Dutch, namely Frisian, Low Saxon, and Limburgish. Regional language speakers are geographically clustered; the dialect map of Daan and Blok (1969) that identified 28 geographically clustered dialect groups on the basis of the perception of dialect speakers is well known. Heeringa (2004) has shown that the three languages are spoken in areas with significant borders around them, as measured by dialect distances.

Differences in value orientations may also lead to different patterns of spatial homogamy, as demographic behaviour has been found to be influenced by value changes (e.g. Van de Kaa 2001). With increasing urbanisation, the probability of wider social circles increases (Blau 1977). As a consequence of the 'urban culture', living in urban areas may nurture new value orientations and open-mindedness, leading to larger networks of friends and acquaintances and increasing opportunities to meet partners in a greater range of meeting places, distributed in a greater area,

thus widening the distance to partners. For the Netherlands, Brons (2006) studied dimensions of regional culture and found considerable regional variation in value orientations. His measurement of value orientations is derived from indirect measures of demographic behaviour, religious adherence, and voting behaviour, and it is based on Hofstede's (1980, 1991) measurement of national cultures. For our study, three dimensions identified by Brons (2006) are pertinent: post-materialism, classic individualism, and Protestant conservatism (see Table 1). These dimensions of regional culture are expected to have an impact on spatial homogamy. High scores on *post-materialism* and *classic individualism* are related to modernisation, as they indicate an increased focus on self-development, little religious influence, and decreased focus on traditional households and families. With increasing modernisation, the geographical horizon of individuals has been found to increase (Beekink et al. 1998), as well as contacts between different groups in society (Hendrickx 1994). As changing geographical horizons are related to changing value orientations, Brons's (2006) indices seem to be useful indicators for regional cultural differences that might account for part of the regional variation in spatial homogamy. The dimension *Protestant conservatism* represents conservative cultures, with high levels of male dominance and uncertainty avoidance. Given the resemblance to characteristics of Bible belt inhabitants, high scores on Protestant conservatism may be related to shorter distances to partners.

Table 2 summarizes the hypotheses.

Table 1 Characteristics of dimensions of regional culture[a]

Dimension of regional culture	High score	Low score
Post-materialism	focus on self-development / self-expression	focus on material well-being
	co-operative and egalitarian	competitive and authoritarian
	(very) small households	large households
	many votes for progressive parties	many votes for conservative parties
	environmentally conscious	
Classic individualism	individual is more important	national or collective interests more important
	postponement of marriage and childbearing	relatively early marriage and childbearing
	many votes for liberal parties	
Protestant conservatism	predominantly Protestant	predominantly Catholic
	early marriage and childbearing	little early marriage and childbearing
	traditional / large households and families	
	male dominance	

[a] Adapted from Brons (2006), p. 562

Table 2 Hypotheses

Compositional effects

| 1 | A shortage of partners in the desired age group leads to increased distances between partners. |
| 2 | In areas with a higher level of socio-economic status, the distance to partners is greater. |

Spatial determinants

| 3 | The more urban the area, the shorter the distance to partners in that area. |

Regional cultural differences

4	In the Bible belt area, distances between partners are shorter.
5	In areas where regional languages are spoken, distances between partners are shorter.
6	With increasing levels of modernisation, the distance between partners increases.

Data and Method

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

Data Sources

Spatial homogamy is often examined by analysing distances between partners before marriage. In the current Dutch context, most couples either cohabit 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, irrespective of whether they are married or not. Geographic similarity is measured before cohabitation. For that reason, a geocoded micro-level database on cohabitation was constructed, based on register data. The Dutch population register, the so-called 'Gemeentelijke Basisadministratie' (GBA), is a decentralised automated population registration system, managed by the different municipalities. The register stores information on each registered inhabitant of the country, such as information on the person, parents, marriage, registered partnership, offspring, and address. As moving house or change in address is reported in the GBA, migration histories can be constructed. Individuals can be linked, through using personal identification numbers, to spouses, children, and parents. The municipal population registers are assessed to be of outstanding quality (Prins 2000). As we are interested in new cohabiters, those individuals who started living together with a partner in the year 2004 were selected. Since marriages and registered partnerships are recorded by the local registrar, these events are directly documented in the GBA. Unmarried cohabiters were identified by using household statistics which are annual statistics 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 to the reference person, marital status, children if any, 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 live 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 of forming a two-person household. To locate new cohabiters, i.e. couples who start living together at the same address, those who experienced a transition in household position, from any other position on January 1, 2004 to being a partner in a couple (with or without children) on January 1, 2005, were selected.[1] The partners were matched to each other based on current address. The resulting dataset for 2004 contains 326,000 individuals (or 163,000 couples).

Subsequently, the (former) addresses of cohabiters were linked to a digital file containing x- and y-coordinates for each known address in the Netherlands, as measured in the national coordinate system. This so-called ACN file (Adrescoördi-naten 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% of all addresses. Spatial homogamy was operationalised by measuring the distance between former addresses of new cohabiters, and it was calculated by computing the Euclidian distance between the geographic coordinates of these addresses, in metres. For each municipality the average distance between partners was then calculated.

Data on explanatory variables were derived from several sources. First, regional statistics were derived from Regional Core Statistics and the Regional Income Distribution 2004, both from Statistics Netherlands. Second, recent developments in the compilation and linkage of large micro-level datasets have provided us with the ability to match our dataset with another micro-level dataset. Our geo-coded micro data on cohabiters were linked 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 data include educational information for each year a person was registered at an institute of higher education, and information pertaining to degrees, majors taken, and so on. By matching the CRIHO files with the cohabiters file, we could establish for each cohabiter whether that person ever studied at an institute of higher education. Finally, regional cultural differences were measured by the earlier-mentioned dimensions of core value orientations proposed by Brons (2006).

Dependent Variable: the Standardized Distance Coefficient

In a previous study, considerable regional variation in spatial homogamy was found in the Netherlands (Haandrikman et al. 2008a). Moreover, partners living in low density areas and in the periphery were found to have lived further apart. One important factor for this result is that the average distance to any other person in the Netherlands is also greater than in the core and densely populated regions. Therefore, the distance between partners should be standardised for *the average distance to all other inhabitants in the Netherlands*. This is done as follows. Firstly, 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

---

[1] 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.

municipality level. Let d_{ij} be the distance between the geometric centres of municipality i and j. Then the average distance for any person living in i to another person in the Netherlands is approximated by:

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

where P_j is the population size of municipality j and N is the population of the Netherlands. As distances to partners within the same municipality are not zero, these distances are approximated by:

$$d_{ii} = \frac{2}{3} \sqrt{\frac{area_i}{\pi}} \qquad (2)$$

where $area_i$ is the area of municipality i in square metres on January 1, 2004.[2] The underlying assumption of this formula is that the population is uniformly distributed within the municipality in the form of a circle.

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

$$SDC_i = \frac{s_i}{d_i} \qquad (3)$$

and is the dependent variable in the analysis. The coefficient of an area can be interpreted as anywhere in the range from a very short distance to a very long distance to partners, given the location of the area within the country. A municipality with a high coefficient might be situated in the Utrecht area (thus centrally located), with partners found at relatively long distances, whereas a low coefficient might be found in an area in the upper north, with relatively short distances to partners. A value of the coefficient of 0.5 implies that the average distance to partners in a specific area is half that of the average distance to all other Dutch people.

Operationalisation of Variables

The explanatory variables are listed in Table 3 and they are clarified in the subsequent paragraphs. The spatial units of analysis are the 483 municipalities of the Netherlands in the year 2004.

---

[2] A justification for the use of 2/3 radius for the intra-zonal distance estimate:
We assume a circular shape of the municipality, and a population density function $F(r) = d_0 r^{-a}$, where r is the distance from the centre, d_0 equal to the density in the centre, and a the density decay with increasing distance to the centre. We approximate the average distance to another person in the municipality by the average distance to the centre. The total distance that the population covers to the centre is equal to: $D = \int_0^R 2\pi r^2 F(r) dr$, and the total population in the municipality is equal to: $P = \int_0^R 2\pi r F(r) dr$. The average distance, D/P (total distance/population) is equal to: $D/P = \overline{D} = \frac{2\pi d_0 R^{2-a}}{2\pi d_0 R^{1-a}} \frac{2-a}{3-a} = \left(\frac{2-a}{3-a}\right) R$, with R being the radius of the municipality. If we assume a homogeneous distribution of the population, $a=0$, and the average distance reduces to: $\overline{D} = \frac{2}{3} R$. Finally, we have to make an estimate of R, based on the size of the municipality A. In a circular municipality $A = \pi R^2$, and thus $R = \sqrt{\frac{A}{\pi}}$. Therefore, the intra-zonal distance is estimated as $d_{ii} = \frac{2}{3} \sqrt{\frac{A_i}{\pi}}$.

Table 3 Explanatory variables

Explanatory factors		Variables	Source
Compositional effects			
Demographic characteristics	Age group	Percentage of population aged 25–45 years	Statistics Netherlands[a]
Socio-economic characteristics	Educational level	Percentage of cohabiters that studied at an institute of higher education	CRIHO
	Income	Percentage of population in lowest income group (<€ 6,700)	Statistics Netherlands[b]
		Percentage of population in highest income group (>€ 24,300)	Statistics Netherlands[b]
Spatial determinants			
Degree of urbanisation		Average surrounding address density (/1,000)	Statistics Netherlands
Regional cultural differences			
Religion		Index for Protestant conservatism	Brons (2006)
Language		Living in Frisian-speaking area	
		Living in Low Saxon-speaking area	
		Living in Limburgish-speaking area	
Value orientations		Index for post-materialism	Brons (2006)
		Index for classic individualism	Brons (2006)

[a] From Regional Core Statistics, Statistics Netherlands

[b] From Regional Income Distribution 2004, Statistics Netherlands. The data are based on registers from the Ministry of Finance and the population register (GBA), combined with a sample of 1.9 million households

Compositional effects are measured by demographic and socio-economic variables. Firstly, the percentage of the population that is aged between 25 and 45 years on January 1, 2004 was determined for each municipality. Among the new cohabiters, 65 percent of them fall within this age group (Haandrikman et al. 2008a), making this the most appropriate target group for persons looking for a partner. Socio-economic status was operationalised by educational level and income. The average educational level of cohabiters in each municipality was constructed from the CRIHO file. For each cohabiter, it was determined whether that person was recorded in the CRIHO file, thereby offering an approximation of the educational level 'higher educated'. Then, the percentage of cohabiters that ever studied was calculated for each municipality. Income was operationalised as the total financial income from all jobs and other resources, such as real estate revenues and other assets. The income data is based on persons with 52 weeks of income, including the self-employed. Income units were distributed across ten percent classes, which are of equal size. The data include persons living in one of the 467 municipalities on January 1, 2005. Because of municipal redistributions since 2004, several adaptations were made.[3] For

---

[3] For 20 municipalities that ceased to exist as per January 1, 2005, mostly in the province of Gelderland, income data from the Regional Income Distribution 2003 is used instead. Moreover, two municipalities have missing data on the lowest income percentile, namely Rozendaal and Thorn. For these municipalities, the average of the adjacent municipalities is taken instead.

the regional analysis, the percentage of inhabitants in the lowest and the highest income group in each municipality was included.

Spatial determinants of regional variation in spatial homogamy were 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. The surrounding address density is the number of addresses within a circular area around an address with a radius of 1 km, divided by the square of the circle, and it is calculated for each 500 by 500 m^2 containing at least one address. The resulting variable is expressed in the number of addresses per square kilometre. For the regional analysis, the average surrounding address density per municipality, calculated for each 500 by 500 m^2 per January 1, 2004, was used, divided by 1,000.[4]

Regional cultural differences were operationalised through a set of value orientations and language variables. As religion is not documented on a large scale, it was operationalised through Brons's value orientation 'Protestant conservatism', as it is most strongly related to religion, especially Protestantism (Brons 2006). Language was operationalised by distinguishing three core areas in which Frisian, Low Saxon, and Limburgish are spoken. Frisian is widely spoken in the province of Friesland, whereas Limburgish is the regional language of Limburg. Low Saxon is spoken in a larger area, namely in Groningen, Drenthe, Overijssel and parts of Gelderland, which were classified as Low Saxon-speaking areas. Local value orientations were operationalised by two dimensions of core value orientations, namely post-materialism and classic individualism, as they approximate measures of modernisation.[5] These regional variables were measured at the municipal level, and were based on demographic, religious and voting behaviour in the period from 1997 to 2003. As we expect these value orientations not to have changed within 1 year, the data has been applied to the municipalities of the year 2004.[6] They were matched to the municipality where cohabiters lived before they started living together with their partner on January 1, 2004.

Methodology of the Spatial Data Analysis

Exploratory spatial data analysis was used to analyse the dependent and independent variables, using ArcGIS and GeoDa (Anselin et al. 2006). Spatial regression techniques were applied to explain spatial patterns of spatial homogamy.

In spatial analyses, spatial autocorrelation may cause problems. In our study, there is a mismatch between the spatial unit of analysis, i.e. municipalities, and the spatial extent of local partner markets. We know that in 2004, the average distance to a cohabitation

---

[4] The Pearson correlation between the degree of urbanisation and the average distance to all other inhabitants (the numerator of the standardized distance coefficient) is −0.26 ($p<0.01$).

[5] We use the index of post-materialism corrected for degree of urbanisation, education and income (Brons 2006). The resulting correlation between post-materialism and degree of urbanisation is only 0.10 ($p<0.05$). Likewise, we used the index of classic individualism corrected for education and income (Brons 2006). The correlation with degree of urbanisation is −0.11 ($p<0.05$).

[6] In 2004, 11 municipalities ceased to exist and were merged into five new municipalities. The indices for the new municipalities were recalculated by weighing the indices with the population of the old municipalities in 2003.

partner before cohabitation was 23 km (Haandrikman et al. 2008a), while the average diameter of a municipality is about 5 km. In other words, in explaining spatial homogamy, neighbouring municipalities should be taken into account as well. 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. In our study, two types of spatial weights matrices are used to test which matrix corrects the problem of spatial autocorrelation in the best way.[7] The first weights matrix is the so-called first-order Queen's contiguity-based matrix, in which municipalities with adjoining borders or corners are neighbours. A second-order Queen's contiguity matrix also takes neighbours of neighbours into account. All spatial regression models are estimated using maximum likelihood methods in GeoDa.

   Although the basis for our study is microdata, the analysis is based on aggregated spatial data, which obviously entails disadvantages of ecological fallacies, spurious relations, and the modifiable areal unit problem (Anselin 2002). However, since our interest is in regional differences in partner choice behaviour and its potential explanatory factors on a regional level, we believe that our methodology is justified, although care is needed in the interpretation of results.

Results

In this section the exploratory spatial data analysis of the dependent and independent variables is described, followed by the specification of the multivariate regression model and the spatial regression model.

Exploratory Spatial Data Analysis

Figure 1 shows the map of the standardized distance coefficients for all 483 municipalities in the Netherlands in 2004. The average standardized distance coefficient for the whole of the Netherlands is 0.23; with the coefficient ranging from 0.09 to 0.48. Areas with a high standardized distance coefficient are municipalities which have a longer distance between partners compared to the expectation on the basis of their geographic location and number of inhabitants, and vice versa. 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 4). Moran's I is significant using both matrices, meaning that the null hypothesis of spatial randomness can be rejected. The positive values of Moran's I indicate positive spatial autocorrelation, or municipalities with low or high standardized distance coefficients are clustered in space.

   Table 5 displays some descriptive statistics of the independent variables taken into account in the regression model.

---

[7] Spatial weights matrices constructed on Rook-based contiguity and distance-based contiguity using the average distance between partners in 2004 were also conducted, but yielded very similar results.
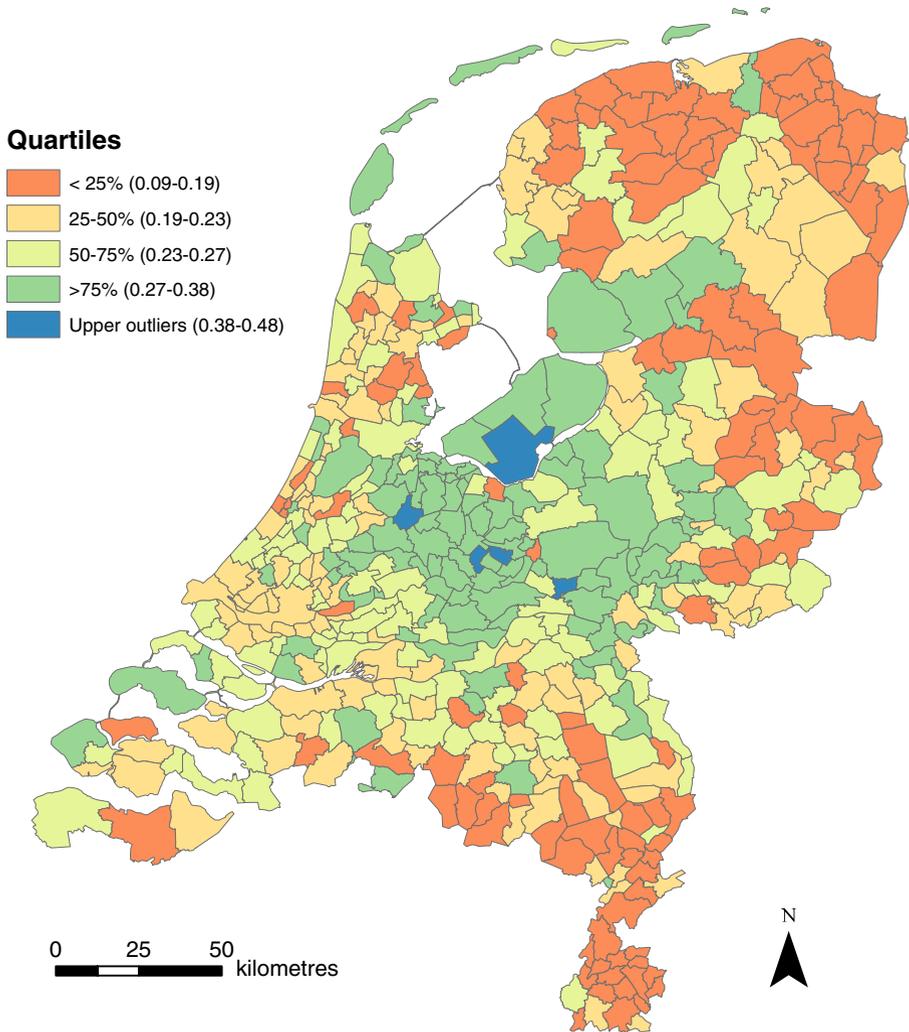
Fig. 1 Map of standardized distance coefficients. Source: © 2005, Statistics Netherlands / Topografische Dienst Kadaster

Multivariate Regression Model

An OLS 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. Regression results are presented in Table 6. Coefficients and t-statistics are summarised in the first two columns of the table, and model fit statistics are provided below the coefficients. As spatial autocorrelation was found in the data, the OLS coefficients are likely to be biased in the absence of a spatial lag. After correcting for the problem of spatial autocorrelation, the model gives more reliable coefficients.

Table 4 Moran's I of the standardized distance coefficient for different weights matrices

Type of weights matrix	Queen's first-order contiguity	Queen's second-order contiguity
Moran's I	0.539 ***	0.475 ***

***: significant at 0.001. This is a pseudo significance calculated with a randomisation process with 999 permutations

The spatial diagnostics based on the two different spatial weights matrices show that both types of models could be appropriate (Anselin 2005). However, based on theoretical considerations—we expect spatial autocorrelation in the residuals—a spatial error model was chosen.

Spatial Regression

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

$$y = X\beta + \varepsilon, \tag{4}$$

$$\varepsilon = \lambda W \varepsilon + \xi \tag{5}$$

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 $W\varepsilon$.

Table 5 Descriptive statistics of independent variables

Variable	Mean	Standard deviation	Range	N^c
Population aged 25–45 years (%)	28.2	2.3	20.2–38.4	478
Cohabiters that studied at an institute of higher education (%)	25.6	7.1	10.2–61.5	478
Population in lowest income group (<€ 6,700) (%)	10.8	1.6	7.0–19.0	478
Population in highest income group (>€ 24,300) (%)	21.0	5.1	9.0–44.0	478
Average surrounding address density/ 1,000	0.89	0.7	0.1–6.0	478
Index for Protestant conservatism	0.00	1.0	−1.4–6.2	478
Living in Frisian-speaking area	0.06	0.2	0–1	478
Living in Low Saxon-speaking area	0.28	0.5	0–1	478
Living in Limburgish-speaking area	0.10	0.3	0–1	478
Index for post-materialism[a]	0.01	0.5	−1.83–1.53	478
Index for classic individualism[b]	0.00	0.4	−1.72–1.48	478

[a] Corrected for degree of urbanisation, education and income

[b] Corrected for education and income

[c] The 5 Wadden islands are not considered

Table 6 OLS and spatial regression results[a]

Dependent variable: standardized distance coefficient

	OLS	t-statistic	Spatial error model based on			
			Queen's first-order contiguity		Queen's second-order contiguity	
	coefficients		coefficients	z-value	coefficients	z-value
Constant	−0.043	−0.977	0.130	3.113 ****	0.127	3.037 ***
Compositional effects						
% Population aged 25–45 years	0.006	4.570 ****	0.001	1.121	0.001	0.896
% High educated	0.001	3.250 ***	0.001	3.463 ****	0.001	3.670 ****
% In lowest income group (< € 6,700)	−0.002	−1.123	−0.003	−1.750 *	−0.002	−1.500
% In highest income group (> € 24,300)	0.005	8.428 ****	0.003	4.768 ****	0.003	4.875 ****
Spatial determinants						
Address density / 1,000	−0.005	−0.954	0.006	1.314	0.007	1.571
Regional cultural differences						
Index for protestant conservatism	0.012	4.778 ****	0.002	0.548	0.002	0.695
Living in Frisian-speaking area	0.003	0.317	−0.011	−0.658	−0.001	−0.078
Living in non-Frisian-speaking area (ref)	0.000		0.000		0.000	
Living in Low Saxon-speaking area	0.014	2.281 **	0.004	0.446	0.004	0.439
Living in non-Low Saxon-speaking area (ref)	0.000		0.000		0.000	
Living in Limburgish-speaking area	−0.029	−3.363 ****	−0.033	−2.579 ****	−0.025	−1.619
Living in non-Limburgish-speaking area (ref)	0.000		0.000		0.000	
Index for post-materialism	0.035	7.400 ****	0.022	5.036 ****	0.023	5.364 ****
Index for classic individualism	0.019	3.330 ****	0.020	3.918 ****	0.022	4.483 ****
Lambda			0.592	13.002 ****	0.788	17.707 ****
Test statistics						
Log likelihood		805.33		855.89		876.17
R^2		0.40		0.55		0.58
Akaike info criterion		−1,586.67		−1,687.77		−1,728.35
Schwarz criterion		−1,536.63		−1,637.74		−1,678.31

[a] Levels of significance: * 0.1; ** 0.05; *** 0.01; **** 0.001

The models were run by means of maximum likelihood, where the spatial regression models included a spatial autoregressive error term. Coefficients, z-values and accompanying significance levels are displayed in Table 6. 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 maximum likelihood are so-called pseudo-R^2, which cannot be compared to OLS results. The proper measures are log-likelihood, the Akaike information criterion (AIC) and the Schwartz criterion (SC). The log-likelihood is highest for the model based on the Queen's criterion with second-order contiguity, thus 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 error model based on a Queen's second-order spatial weights matrix gives the best results.

Ignoring the addition of a spatial lag to the regression equation, and estimating the model using OLS may lead to an overestimation of the magnitude of the parameters, to the extent that the spatial error parameter lambda is statistically significant (Anselin 2005). The spatial autoregressive coefficient is estimated at 0.592 for the Queen's first-order model and 0.788 for the Queen's second-order model, and the coefficient is highly significant for both 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, as discussed in the following.

The first hypothesis was not confirmed using the spatial error models. The percentage of 25 to 45-year-olds in a region does not have an impact on the standardized distance coefficient of that area. Additional models for different age groups and different stages in the life course were also specified, i.e. models for young singles, cohabiters who were living in the parental home before, and those who were living with children before, but these yielded no evident differences.

Not surprising, socio-economic differences between regions do explain variation in spatial homogamy (hypothesis 2). Higher percentages of higher educated persons lead to an increased standardized distance coefficient. This finding is robust throughout the error models. Income differences also contribute to differences in spatial homogamy. Especially high shares of high income groups have a large impact: areas with higher average income have higher standardized distance coefficients. In addition, in municipalities with a high concentration of low income groups, partners are found at significantly shorter distances.

Contrary to expectation, level of urbanisation does not influence distances between partners; none of the models supports our third hypothesis.

The third set of hypotheses yields some mixed results. We do find an effect of modernisation on spatial homogamy: with increasing levels of post-materialism and individualism, distances between partners increase (hypothesis 6). This finding is robust throughout the models. Religion, measured by the index of Protestant conservatism, was expected to negatively influence the standardized distance coefficient (hypothesis 4), but we find a positive effect in the OLS model and no effect using the spatial error model. The fifth hypothesis, on the effect of speaking a regional language on spatial homogamy cannot be supported either. Using first-order contiguity, only speaking Limburgish is found to decrease the value of the standardized distance coefficient, but this effect disappears when adjacent municipalities are taken into account.

Conclusions and Discussion

This article 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 *standardized distance 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 different types of spatial weights matrices. By including this added spatial variable, spatial bias in the results is avoided, which would otherwise have resulted in distorted findings.

The study is unique since it has been able to relate spatial homogamy to a set of variables, on a very detailed level, for a whole country. The scale of the current analyses is the strength of the study, in which results from previous studies were largely confirmed. Of the three sets of explanations taken into account, compositional effects, and particularly socio-economic characteristics, together with regional cultural indicators are the most important in explaining regional variation in spatial homogamy.

Demographic composition of the population, measured as the percentage of 25 to 45-year-olds, does not affect distance to partners, although it does when spatial autocorrelation is not taken into account. The reasons for this result are not clear. Completely in line with previous studies, the effect of socio-economic characteristics is highly evident. New is that higher income and a higher educational level not only lead to increased distance to partners at the individual level, but also at the regional level.

The impact of spatial determinants on spatial homogamy was partly accounted for in the definition of the standardized distance coefficient, but was also considered by testing degree of urbanisation as an explanatory factor in the model. However, this spatial factor was found to be non-significant throughout the models. There is no effect of population density when compositional and regional cultural indicators are controlled for. It might be that (some) partners are found at shorter distances in highly populated areas, but that others are found at (very) long distances. Indeed, analysis of distances to partners across lower regional units shows that there are many urban low-income neighbourhoods in which partners are found at (very) short distances. Our finding might be resulting from this combined effect.

In fact, this study adds to existing work that it is not urbanisation that is causing wider spatial horizons, but it is the value orientations of people which leads to a more global outlook. The modernisation indicators post-materialism and individualism partly explain regional patterns of spatial homogamy: the more post-materialistic and individualistic the area, the greater the distances to partners. The finding that regional cultural differences do account for part of the regional differences in spatial homogamy is consistent with and adds to studies conducted in the framework of the European Fertility Project (Coale and Watkins 1986). Modernisation theory assumes that boundaries between groups become less strong as modernisation proceeds. 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 enhanced the autonomy of individuals and have decreased the effectiveness of sanctions on social norms. These economic, social and cultural changes have had a major impact on interpersonal relationships: Beekink et al. (1998) found for the Netherlands that geographical horizons tended to widen in the last two centuries, when these changes took place. We state that value

orientations have an impact on spatial homogamy, thereby complementing to studies on increasing openness of societies during modernisation processes (e.g. Smits 1996; Van de Putte 2003).

One exception to the above statement is that we did not find an effect of religion on spatial homogamy, measured as protestant conservatism, although this was expected. Religion can also be seen as part of regional culture or as an indicator of modernisation, which makes it more surprising, that in conservative cultures, partners are not found at shorter distances. The municipality with the shortest distances to partners, Urk, with a median of 800 m between cohabiters before cohabitation, also has the highest score on Protestant conservatism. However, Urk, and other protestant strongholds such as Rijssen en Spakenburg, have some of the highest residual levels, indicating that spatial homogamy in this area is related to religion or related cultural factors, that were not accounted for in this article. Better data on religiosity of the inhabitants of different areas might shed a different light on the matter. Besides, the definition of the dependent variable might lead to overestimation of the distances between partners in the middle of the country, which partly coincides with the Bible belt.

Another component of culture, language, was partly found to account for regional differences in spatial homogamy. Further research using micro data on language might shed more light on the interaction between linguistic and spatial homogamy.

This study is part of a PhD research into the spatial dimension of partner choice. In a subsequent paper, the available microdata is optimally used by applying a random utility model to the probability to find a partner given mutual demographic, socioeconomic, cultural and spatial characteristics, also including information on where people work and study, as a considerable number of people meet at these places (Haandrikman 2010). We would have liked to include data on meeting places in the current analysis, for instance the number of bars, voluntary associations, or schools. Unfortunately, most data is not detailed enough at the municipal level. Second, sometimes too high correlations might result between meeting places and the degree of urbanisation, the percentage higher educated and the percentage young population.

We have provided new insights into spatial assortative mating, by applying methods from spatial econometrics. Cupid's wings are not adapted for long flights, but higher educational level, high income, and post-materialist and individualist value orientations make Cupid fly further from home.

References

Anselin, L. (2002). Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*, 247–267.

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

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

Beekink, E., Liefbroer, A. C., & Van Poppel, F. W. A. (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*. New York: Free; London: Macmillan.

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

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

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

Coale, A. J., & Watkins, S. C. (1986). *The decline of fertility in Europe. The revised proceedings of a conference on the Princeton European Fertility Project*. Princeton: Princeton University Press.

Coleman, D. A., & Haskey, J. C. (1986). Marital distance and its geographical orientation in England and Wales, 1979. *Transactions of the Institute of British Geographers, 11*, 337–355.

Daan, J., & Blok, D. P. (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*. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.

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

De Graaf, N. D., Smeenk, W., Ultee, W., & Timm, A. (2003). The when and whom of first marriage in the Netherlands. In H. P. Blossfeld & A. Timm (Eds.), *Who marries whom? Educational systems as marriage markets in modern societies* (pp. 79–112). European Studies of Population, Vol. 12. Dordrecht: Kluwer Academic.

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.

Fielding, A. J. (1992). Migration and social mobility: South East England as an escalator region. *Regional Studies, 26*(1), 1–15.

Goode, W. J. (1982). *The family*. Englewood Cliffs: Prentice Hall.

Haandrikman, K. (2010). Waar ontmoeten partners elkaar? Sociale differentiatie in ontmoetingsplaatsen. *Mens en Maatschappij* 2, forthcoming.

Haandrikman, K., Harmsen, C., Van Wissen, L. J. G. & Hutter, I. (2008a). Geography matters: Patterns of spatial homogamy in the Netherlands. *Population, Space and Place, 14*(5), 387–405.

Haandrikman, K., Harmsen, C., Van Wissen, L. J. G. & Hutter, I. (2008b). De geografische dimensie van partnerkeuze. *Bevolkingstrends, 56*(3), 19–28.

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

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

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

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*. Beverly Hills: Sage.

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

Israëls, A., & Harmsen, C. (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. Voorburg: Centraal Bureau voor de Statistiek.

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

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 Sociological Review, 13*, 613–616.

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

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

Ni Bhrolcháin, M. (2001). Flexibility in the marriage market. *Population: An English Selection, 13*(2), 9–48.

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

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*. Nijmegen: Katholieke Universiteit Nijmegen.

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

Steenhof, L., & Harmsen, C. (2003). *Same-sex couples in the Netherlands*. Paper presented at the workshop on Comparative Research, 30 June–2 July 2003, Rome.

Stevens, G., & Schoen, R. (1988). Linguistic intermarriage in the United States. *Journal of Marriage and the Family, 50*, 267–279.

Van de Kaa, D. (2001). Postmodern fertility preferences: From changing value orientation to new behaviour. In R. A. Bulatao & J. B. Casterline (Eds.), *Global fertility transition* (pp. 290–331). Population and Development Review, a supplement to Vol. 27. New York: Population Council.

Van de Putte, B. (2003). *Het belang van de toegeschreven positie in een moderniserende wereld. Partnerkeuze in de 19e eeuwse Vlaamse steden (Leuven, Aalst en Gent)*. Leuven: Katholieke Universiteit Leuven.

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

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

Van Poppel, F. W. A., Liefbroer, A. C., Vermunt, J. K., & Smeenk, W. (2001). Love, necessity and opportunity: changing patterns of marital age homogamy in the Netherlands, 1850–1993. *Population Studies, 55*, 1–13.

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