EPC 2008 Extended Abstract

Explaining the flight of Cupid's arrow: a spatial micro model of partner choice

Karen Haandrikman and Leo van Wissen
Population Research Centre, Faculty of Spatial Sciences, University of Groningen

Abstract
Spatial homogamy may be defined as: everyone may be attracted to everyone else, but near candidates are more attractive than distant candidates. In this paper we propose a model of partner choice, where homogamy is defined in terms of demographic, socio-economic, cultural and spatial similarity. A spatial choice model using random utility theory is formulated, taking into account a relaxation of the independence from the irrelevant alternatives property. Unique micro data on all new cohabiters in the Netherlands in 2004 is used, including demographic, socio-economic, educational and spatial attributes. The model is estimated for a random sample of all cohabiters, and for a number of subpopulations, segmented by place of origin. The coefficients of the models reflect the degree to which partners take into account (dis)similarity in choosing each other.

Key words
● Partner Choice ● Homogamy ● Spatial Choice ● Random Utility Theory ● Competing Destinations Model

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, examining the socio-economic and cultural dimensions of homogamy. Geographers have added the spatial dimension to this concept (Mayfield 1972, Küchemann et al. 1974, Coleman 1979, Fisher 1980, Coleman and Haskey 1986, Clegg et al. 1998, Duncan and Smith 2002).

The starting point for this research is that everyone may be attracted to anyone else, but similar partners are more attracted to each other than others. Applying partner choice based on Tobler's (1970) general law of geography, who stated that everything is related to everything else, but near things are more related than distant things, spatial homogamy is defined as: everyone may be attracted to everyone else, but near candidates are more attractive than distant candidates.

In this paper we propose a model of partner choice, where homogamy is defined in terms of demographic, socio-economic, cultural and spatial similarity. We formulate a spatial choice model using random utility theory, and taking into account a relaxation of the Independence from Irrelevant Alternatives property (IIA) (Pellegrini and Fotheringham 2002). We use unique micro data on all new cohabiters in the Netherlands in the year 2004. The geo-coded micro data is based on population register data, and is linked to several educational and socio-economic data sources.

BACKGROUND

In a recent study, new cohabiters in the Netherlands were found to choose spatially homogamous partners (Haandrikman et al. 2006). This study showed that partner choice is subject to strong distance decay: geography does matter. In the study, spatial homogamy was operationalized by examining the exact geographic distances between partners before cohabitation. The explorative study found considerable demographic and spatial variation in spatial homogamy. With increasing age, individuals tend to find their partner at shorter distances, with an exception of cohabiters in their twenties, who find their partners at significantly shorter distances than younger and older cohabiters. Moreover, demographic variation was explored through position in the household before cohabitation. Those living in the parental home and those living in single person households found their partners significantly more nearby than singles and other household members. Besides considerable regional variation in spatial homogamy, degree or urbanisation was found to matter. With increasing degree of urbanisation, the distance between partners decreases. In a follow-up paper, the regional variation in spatial homogamy was found to be related to socio-economic, cultural and spatial characteristics of areas (Haandrikman et al. 2007).

Although we do know that demographic, socio-economic, cultural and spatial factors influence partner choice, as yet we cannot separate their effects and determine how they interact. As people tend to live amongst people like themselves, geographical clustering is found for characteristics such as stage in the life course, educational level, income, religion, language. The current paper aims to single out the effects of demographic, socio-economic, cultural and spatial factors on partner choice.

MODEL FORMULATION

The model can be formulated as follows. For a given individual *i* who is actively seeking a partner in period *t*, we define an attractiveness function U_{ij} for potential partner *j* who is also active on the partner market. The attractiveness function of potential partner *j* vis-à-vis person *i* is composed of a deterministic part V_{ij} and a random component ε_{ij}, which accounts for the unobserved part of the attractiveness to the modeller: $U_{ij}=V_{ij}+\varepsilon_{ij}$. Due to this unobservable part, we are at best able to produce probabilistic statements about the outcome of the choice process. A person *i* will choose potential partner *j* over *k* if the following condition is satisfied: $U_{ij}>U_{ik}$, which may be rewritten as: $V_{ij}+\varepsilon_{ij} > V_{ik}+\varepsilon_{ik}$, or: $\varepsilon_{ik} > V_{ij} - V_{ik} + \varepsilon_{ij}$. Since the error terms are unknown, we can at best make probabilistic statements about the likelihood of choosing partner *j* over *k*:

$$P(j \mid j \in C_i) = \Pr(U_{ij} > U_{ik}, \forall k \in C_i, k \neq j) \qquad (1)$$

Or, rearranging:

$$P(j \mid j \in C_i) = \Pr(\varepsilon_{ik} > V_{ij} - V_{ik} + \varepsilon_{ij}, \forall k \in C_i, k \neq j) \qquad (2)$$

Equations (1) or (2) express the probability that partner *j* in the choice set *C* of individual *i* is preferred over *k* by individual *i*. Different assumptions about the joint distribution functions for the error terms lead to different models. If we assume a multivariate normal distribution we arrive at the probit model; the multinomial logit model (MNL) results if we assume a so-called type I independently and identically distributed extreme value distribution (McFadden 1974). This assumption leads to the well known and computationally tractable form:

$$P_{ij} = \frac{\exp V_{ij}}{\sum_{k \in C_i} \exp V_{ik}}, \; j \neq k \qquad (3)$$

The systematic part of the attractiveness function V_{ij} is determined by the degree of similarity or dissimilarity between both candidates in a number of dimensions: demographic (D), socio-economic (*E*), cultural (*K*) and spatial (*S*). We define the *dissimilarity* in the demographic dimension between both potential partners as: $D_{ij}=D_i-D_j$ (and likewise for *E*, *K* and *S*). D_i is operationalized in terms of stage in the life course, and D_{ij} is then the difference in age between both potential partners. Similarly we could define a socio-economic (dis)similarity index E_{ij}, on the basis of differences in individual income and years of schooling of potential partners, a cultural (dis)similarity index K_{ij}, on the basis of differences in individual scores on cultural indicators, and a spatial dissimilarity index S_{ij} on the basis of geographic similarity or distance between both candidates.

However, we could specify more complicated spatial similarity functions that include similarity indices in work-location, school-location, or other relevant meeting places. Our aim is to determine the parameters of the deterministic part of the attractiveness function, i.e. to estimate the weights of the different dimensions of (dis)similarity of homogamy in the attractiveness function. The deterministic part of the attractiveness function can be written as:

$$V_{ij}=\alpha D_{ij}+\beta E_{ij}+\gamma K_{ij}+\delta S_{ij} \quad (4)$$

where the *D, E*, *K* and *S* are vectors of random variables, and the *α, β*, *γ* and *δ* vectors of coefficients.

The coefficients may be estimated by the method of maximum likelihood. However, in theory the choice set C_i can be very large, and this becomes infeasible when determining the likelihood function. However, we may construct a much smaller choice set by creating a subset consisting of the chosen alternative (i.e. the chosen partner *j*, and a random sample $\{k=1,..,n,\ k \in C_i,\ k \neq j\}$ out of the set of feasible alternatives (see McFadden 1978). *n* is usually in the range between 5 and 10.

A strong assumption of the MNL model is the IIA property. This assumption may especially be too strong in a spatial choice context, where there are many alternative choices clustered in space that are more alike each other than alternatives further away. Therefore we will reformulate the model as a Competing Destinations model (CDM), that relaxes the IIA property of the MNL model (Pellegrini and Fotheringham 2002). The Competing Destinations model adds the probability π_{ij} that a potential partner is evaluated to the attractiveness function:

$$P_{ij}=\frac{\exp V_{ij}\cdot\pi_{ij}}{\sum_{k\in C_i}\exp V_{ik}\cdot\pi_{ik}},\ j\neq k \quad (5)$$

In this additional probability, the spatial structure of the choice set may be reflected.

DATA

We use vital statistics from the Dutch population register (GBA). This register is a decentralised automated population registration system, managed by the individual municipalities. In the GBA, information on each registered inhabitant of the country is stored, and each individual can be identified through a personal identification number, which enables linkage to spouses, children, and parents. We include unions of both married and unmarried cohabiters. The first group is recorded by the local registrar and is therefore directly documented in the GBA. Unmarried cohabiters are tracked down using household statistics. Statistics Netherlands assigns household positions to persons based on the relationship of an individual to the reference person, his or her marital status, and possibly, children. If two people moved to the same address at the same date, they are classified as a single two-person household. The remaining cohabiters are tracked down by using an imputation model to determine which persons living at the same address form a household (see Israëls and Harmsen 1999 and Harmsen and Israëls 2003). Those living with a partner on 1 January 2005 but not living with a partner on 1 January 2004 were selected. The resulting dataset contains 326,000 starting cohabiters.

The addresses of cohabiters are geo-coded using the Geographical Base Register, which assigns geographic coordinates to each known address based on 6-digit postal codes and house numbers.

Next, the micro level data set 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. Based on their social security number, cohabiters are linked to data on socio-economic category and income. Moreover, the cohabiters file is matched with the so-called CRIHO-file, in which all persons who studied at any institute of higher education in the Netherlands in the period 1986-2004 are included. For each year that a person is registered at an institute for higher education, information on degrees, majors taken, and so on, is available.

The cultural dimension is operationalized by combining linguistic data with Brons' (2006) dimensions of core value orientations. These dimensions are combinations of aspects of meta-behaviour, and represent post materialism, individualism, egalitarian anti-conservatism, dissatisfaction, and protestant conservatism. The cultural indices are measured at municipal level.

Spatial factors are operationalized by the current and former spatial locations of partners. For both partners, the birth place, the residential location 5 years prior to cohabitation, the address just before cohabitation, and the cohabitation address is known (from the GBA), as well as the workplace before cohabitation (from the SSB file), and the place where persons studied (from the CRIHO files).

ANTICIPATED RESULTS

We will estimate the parameters of this model: (1) for a random sample of all persons involved in partner choice in 2004; (2) for a number of subpopulations, segmented by place of origin. The results for these subpopulations may give additional information if the partner choice process in specific parts of the Netherlands is different from the overall picture.

The estimated coefficients of these models reflect the degree to which partners take into account (dis)similarity in choosing each other. In this model, the joint contribution of similarity in demographic, socio-economic, cultural and spatial dimensions may be evaluated statistically.

References

- 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., 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: 337-355.
- 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.
- 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. Submitted.
- Haandrikman, K., L.J.G. van Wissen, C. Harmsen and I. Hutter (2007), Explaining spatial homogamy. Socio-economic, spatial and cultural aspects of the partner market in the Netherlands. Paper presented at the Annual Conference of the Population Association of America, 29-31 March 2007, New York.
- Harmsen, C. and A. Israëls (2003), Register-based household statistics. Paper presented at the European Population Conference, 26-30 August 2003, Warsaw.
- 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.
- 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.
- 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.
- McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (ed.), Frontiers in econometrics. Academic Press, New York.
- McFadden, D. (1978), Modelling the choice of residential location. In: A. Karlqvist et al. (eds.), Spatial interaction theory and planning models. North Holland, Amsterdam.
- Pellegrini, P.A. and A.S. Fotheringham (2002), Modelling spatial choice: a review and synthesis in a migration context. *Progress in Human Geography* 26 (4): 487-510.
- Tobler, W.R. (1970), A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, Supplement: 234-240.