Two-sex Models – Time for a Rethink?

M Murphy
London School of Economics

Introduction

The issue of two-sex models has been a topic of interest in demography and related disciplines for over half a century, although the development was particularly rapid in the 1960s and 1970s decades. The related topic of analysing marriage patterns has also been a longstanding area of interest. In a symposium in 1972 (Greville 1972) including such distinguished figures in demography as Sam Preston and Jan Hoem (both IUSSP laureates) two of the papers were devoted to the topic (as was a paper in the synthesis volume Convergent Issues in Demography and Genetics (Adams et al 1990)). The opening paragraph of the first paper in the 1972 Volume by David McFarlane, refers to work by, among others, Ansley Coale, Louis Henry, Nathan Keyfitz and Griffith Feeney (three of whom are IUSSP laureates). The second by Partlett considers marriage functions. More recently, work has continued by Robert Schoen (1981), John Pollard (1993-4, 1997), Nico Keilman (1985), Noreen Goldman (1984), Maire Ni Bhrolchain (2001) and Iannelli et al (2005).

McFarlane states that interest grew out of attempts to apply the one-sex standard projection simultaneously to both sexes, and the related but logically distinct issue of constructing meaningful nuptiality rates specific for age and sex that 'might plausibly be supposed to remain constant despite changes in number of people in the various age-sex categories' (McFarlane p 89 in Greville 1972).He set out a series of propositions that he regarded as key: the first is that 'a marriage model is the most problematic (and possibly even the ONLY problematic) component of a full-fledged two-sex population projection model' and he goes on to justify this substantial claim. Marriage is special when compared, with, for example, living arrangements (or household composition), since the fact that there must be one married man for each married women imposes a strong constraint on the system under consideration – this is a logical rather than an actual constraint in many cases such as Western societies (depending on the unit of analysis – one spouse may be abroad and less well-defined status such as cohabitation may be closer to 'marriage' than a separated but legally-married couple). This is clearly much less so in polygynous societies (which lead to inequalities rather than equalities). Even in Western societies, strict equality is becoming relaxed in that people with different practices regarding monogamy need to be accommodated. In practice, partnership, including cohabitation as well as formal marriage, but excluding the non-co-residential married couples might seem more appropriate, but there are considerable logical and operational problems in establishing a one-to-one exclusive relationship between the partners concerned. A key issue to be developed in this paper is how far the topic of marriage models is concerned with the exploration of formal structures among haplotype species (for example, in the number of maternities based on offspring conceived in a suitably short time interval will receive half their genetic material from a male and half from a female who may be regarded in a (possibly) short-term but exclusive partnership (maternity accounts for possible multiple births).

The second proposition underpins the enduring role of the topic that empirical tests tend to be inconclusive.

McFarlane set out a number of axioms that marriage function should behave, although these have later been challenged and modified by Pollard (1997) but the core of these remain and even in a recent major study (Iannelli et al, 2005) this remains a key aspect of work in this area.

McFarlane identified a number of marriage functions in which implicitly a value k (later called a 'forcing function' of 'force of attraction' by Qian and Preston (1990) (Q&P) was hoped to remain relatively constant with changing values of the underling population values. These include divining the number of marriages of men aged m and women aged f by the harmonic mean of the number of non-married man aged m and women aged f favoured by Schoen and Qian and Preston, and (a slightly modified version of) the geometric mean championed by Choo & Siow (2006) (C&S). Both have also been identified as having useful interpretations in terms of random interactions (Q&P) or gains to marriage (C&S). The limitations of these interpretations will be considered later.

The main difference between the McFarlane and Pollard axiomatic formulations is that Pollard requires homogeneity (i.e. if the numbers of men and women are increased by the same fraction, then the number of marriage will do also).

At the same time, there was considerable work looking at joint distributions of spouses, especially in relation to assortative marriage associated with issues of changing social stratification and looking in particular at educational level, and at inter-racial marriage also, almost exclusively in the US even thought rates of inter-racial marriage there are very low. This area of work was particularly associated with sociologists such as Leo Goodman, and typically involved fitting log linear models to contingency tables of numbers of marriages. Such work had a rather different focus and motivation from the more mainstream and core demographic work noted earlier, and, in particular, it has been pointed out that such analyses, which are concerned with the patterns observed among those who actually marry, include no insight of information on those who do not do so. In demographic terminology, it does not take account of the populations at risk of experiencing the event of interest. A consequence of this is that as reiterated recently by Choo and Siow that such models cannot be used for purposes such as forecasting, and neither do they allow for possible influences of different populations at risk and the role of competitors among the populations at risk, i.e. what will be the likely effect on the number of marriages, say, of men aged m with women aged w of changes in the number of men aged m^* ($m^* \neq m$) or women aged w^* ($w^* \neq w$). Such issues have been developed, for example, by the use of 'availability ratios', e.g. Goldman (1984), Ni Bhrolchain (2001) etc. – these are referred to as 'spillover effects' (C&S)).

The idea that individuals are in competition with others for a pool of potential spouses lead naturally to the idea of a 'marriage market' – a phrase which is widely used and became popular from a painting by Edwin Long (1875), and the discussion of 'Modern marriage markets' by Marie Corelli in the late 1890s. Analogy of a 'marriage market' to a true market as the Babylonian one with contemporary mate selection is widely used (if one assumes that the buyers and sellers are involved in conventional market transactions, rather than the commodities being bought and sold). Neither image is particularly positive, but its original negative connotations has been lost in the mists of time, although one consequence is that the 'marriage market' is not defined and therefore lacks formal validity, even if the phrase is both intuitive and catchy.

In fact, the second article in the Population Dynamics book by Partlett is based on the idea of '<u>the</u> marriage market' (Greville 1972 p 107 – my emphasis – in fact the phrase not used by McFarlane) opens with this as the basis and the author states that 'one of the pleasures of scientific work is that it is not necessary to come back to first principles all the time' – although as noted above, there is a deficit of first principles regarding the scientific basis of what constitutes a marriage market, which serves as possibly an ideal form against which to assess empirical realisations of the ideal form.

The key points to be considered here are
Is there a need for well-defined 'populations at risk' (a.k.a. 'marriage market') for useful further analysis (as is the case for fertility and mortality and even for migration)
If so, what are these?

If not, why not and why do we spend so much time estimating 'marriage markets'?
What approaches are possible?

Marriage market reconsidered

In practice, the 'populations at risk' is synonymous with the marriage market (for example, models that claim to have controlled for the marriage market usually include some indicator such as populations at risk, or even sex ratios, sometime referring to the ratio of unmarried people, but some to overall populations by sex). The pork belly market consists of sellers with a pile of pork bellies and buyers with a pile of cash, and requires some environment whereby transactions involving transfers take place, often requiring additional conditions such as the market must clear and/or buyers and sellers have perfect information (although these are often relaxed to make them more realistic). It is clear that irrespective of the fact that the marriage market, as defined above, cannot clear except in the effectively zero probability case where the number of unmarried men and women are equal (even if this requires an 18 year old to marry an 85 year if these were the last two left). There are questions of what is the population at risk of marrying – cohabiting people presumably are, although it can be argued that their inclusion is balanced by their partner's inclusion or exclusion. However, another and increasingly important issue is what is the theoretical and effective populations at risk of marrying a particular individual – apart from the married population for at least the time interval proscribed by the particular legal system. In theory, the whole world of the opposite sex forms the population at risk, and over time, the impermeability of national borders for marriage has reduced substantially – between 1975 and 2005, the proportion of marriages of Norwegian residents that involved a non-resident increased from 2.7% to 21.6% (Table 1) and, in addition, the proportion of marriages taking place abroad increased from 4.6% in 1986 (first year available to 17.2% in 2005 (Table 2). The assumption that the nation, or a smaller sub-unit forms a reasonably well-defined pool for marriage, which may have been reasonable even in the 1970s, seems much less defensible now. Some of these marriages will include those who were born abroad, but to confine analysis to particular groups such as native born people with other native-borns (and possibly also requiring native-born parents also) seems artificial and fails to relate the increasing variability of marriage, a fact that will become increasingly less plausible in years to come (in Britain, at present, one quarter of births involve at least one overseas born parent) – if migrants in childhood and early adulthood are included, the marriage market among adults cannot ignore such numbers (few studies which discuss issues such as the role of sex ratios consider this issue).

Table 1 Marriages by residential status of spouses, 1975-2005 Norway

Year	Total	Bridegroom resident abroad	Bride resident abroad	Either resident abroad %
1975	26316	418	289	2.69%
1985	20764	543	463	4.84%
1995	21677	696	1397	9.66%
2005	24211	1819	3659	21.62%

Source Statistics Norway accessed 3rd April 2007

Table 2 Marriages involving Norwegian Residents by place of solemnisation, 1986 & 2005

Year	Marriages	Abroad	As % of total
1986	19873	913	4.59%

2005	22392	3851	17.20%

Source Statistics Norway accessed 3rd April 2007

Over time, the proportion of marriages which are first marriages has declined (Table 3). Even here, the idea of a well-defined pool of eligibles is fuzzy: it is likely that many of those who re-marry found their new partner when they were married to someone else and may therefore be considered to be in the pool of potential marriage partners.

Table 3 Marriages involving Norwegian Residents by order of marriage, 1979 & 2005

	1979		2005	
	Woman		Woman	
Man	First	Subsequent	First	Subsequent
First	18582	1395	13251	1637
Subsequent	1667	1411	1743	2341
Total		23055		18972
B-S (%)		80.60%		69.85%

Source Statistics Norway accessed 3rd April 2007

Even within a particular country, spouses are disproportionately drawn from groups that are close in terms of age, socio-economic characteristics and geographical proximity as well as a number of other factors. A number of studies have argued that local marriage markets are a better indicator of the availability of partners, but the evidence for this is lacking and the arguments based on plausibility (apart from historical studies where restrictions on mobility and social structures made the restriction of the possible pool of potential partners to local regions plausible – for example of marriages contracted between 1600 and 1850 in the parishes Vang and Slidre in the mountain valley of Valdres in Norway only 54 of 4334 marriages were with residents outside the parishes. The increased mobility and non-uniform mobility of different groups mean that the idea of a meaningful marriage market is well away from reality, but there are few studies that have attempted to justify their choice (or in many cases, there is no choice, and only national-level data may be available).

As marriage markets become more prominent in the development of marriage models, the need for a defensible operationalisation of the concept becomes more pressing, or failing that, there is a need to assess the sensitivity of results to alternative plausible or widely-used definitions to assess the robustness of results to alternative specifications.

Research Questions

The definition of a ‘marriage market’ – attempts to find clear definitions will often lead to blind alleys, e.g. Becker is referred to as a source, but he never defines it.

The boundaries of the marriage market are not clear-cut, but there is a local aspect to it – ‘local’ refers to closeness in social as well as geographical space: people are much more likely to marry persons of similar age, ethnic and socio-economic characteristics as well as geographic proximity. What criteria can be advanced to define and to identify ‘local’ marriage markets? However, I will concentrate on geographical proximity empirically defined as municipalities to consider how disproportionately marriages are likely to occur with others within the group.

Identifying Local marriage markets

In the past, the idea of a pool of eligibles in small and well-defined areas was reasonable. Henry's (1972) derivation of the panmictic model is based on a series on concentric groups. Most work has used national data presumably since this was the only information available, but a number of authors with access to sub-national data have advocated the use of geographically disaggregated data (although, of course, the unit used, such as US state level will often be much larger than the national data for a country such as Norway). The superiority of local markets is often advocated, but none of these studies shoes that the results obtained are superior, and no criteria for assessing such superiority appear to exist).

Given as argued earlier, that the use of national boundaries is increasingly irrelevant for marriage market analysis (either the populations studied have to be artificially restricted, such as Native-born to native-born parents), but even so, some of the more interesting formal results from two-sex analysis such as numbers on men matching the number of women who marry within the population studies cannot be imposed, nor the more recent discussion of sorting or market clearing approaches.

The Norwegian register data provide the possibility of trying to do since municipality of residence every year since birth is available (given cohabitation, the use of current place of residence is unsatisfactory) – cohabitation causes other problems as well but these will be ignored at present.

The municipalities are randomly allocated a number 1 to 434 (for some reason there is no number 90, so there are 433 municipalities. In the period 1974-2003, the number of people marrying (so that the matrix is symmetric) was as follows:

Table 4 Summary of marriage and municipalities

Total	1126198
Mean	2601
s.d.	6255

The data are highly skewed, which is a complication.

In the analysis, I used total number of marriages as a proxy for population size (up to a constant value), since (a) it's not clear what population should be used and (b) total marriages should be proportional to this value in that case in the absence of major sex-age difference between areas.

For confidentiality reasons, the municipalities cannot be identified (to protect anonymity of the persons concerned), but the allocation seems to be non-random since the numbers of marriages occurring within adjacent municipalities in this file is larger than would be expected by chance. However, it does open the possibility to identify 'pure' marriage markets, i.e. not using geographical proximity as a proxy for higher contact municipalities, but simply defining them in terms of the higher propensity of people within a group of municipalities to marry within that group. To see whether behavior in sub-national marriage markets was similar, I attempted to identify groups of municipalities which showed a greater propensity for people to marry within (residence at age 15 was used), about 5 or 6 groups ('clusters') were identified, and the objective was to find groups of municipalities of approximately equal population size (proxied by equal number of marriages) which maximized the number of case where people married within the cluster

In order to identify clusters, an approach using clustering analysis programs to try to identify such groups appears appropriate, but the problems is how to define a suitable dissimilarity matrix, because 'closeness' between two areas is not given simply by the number of marriages involving those in the two areas since the population sizes can be so different for the various pairs of areas in the 433x433 matrix giving the distribution of marriage partners. It is necessary to define a distance function that controls for such difference in size, so I therefore use a number of alternatives:

- an IPF model to this matrix with the matrix marginals set to the same value (e.g. every municipality had 1,000 marriages in total (using 'marriage' as shorthand for 'persons marrying') – McFarlane (1977) advocated a similar approach in a paper but not relating to areas. The resulting matrix gives the expected number of marriages between any two areas with the initial propensities to find a partner in the other municipality, i.e. by keeping the cross-product ratios equal, but with the numbers amended so that more meaningful comparisons can be made between areas of different sizes. I then fit a cluster analysis with {1000-fitted no. of marriages} as the non-Euclidian metric between each pair of areas (since smaller is 'closer' - the method is insensitive to the fixed value as long as there are no non-negative values), and estimated the full cluster tree. Since the idea was to find a number N of 'marriage market' areas similar size – typically 5 or 6, I then worked up the branches combining areas until the number of marriages in these areas in the original matrix was close to 1126198/6.

- the second distance was defined by a Schoen (1981) marriage function , whereby the matrix used for clustering had elements $2*M_{ij}\ M_{.i}\ M_{.j}\ /(M_{.i} + M_{.j}\ )$ where M_{ij} is the number of marriages between a member of municipality i and municipality j and $M_{.i}$ is the total number of marriages involving a person if municipality i. Since 'closeness' is associated with larger function values, whereas near values have smaller distance, the reciprocal was used (the results are insensitive to monotonic transformations, so other choices such the canonical transformation of Mardia, Kent and Bibby would produce similar results). The harmonic mean was if the total number of marriages of those in Municipalities i and j, rather than populations, but a close direct relationship between these would be expected. The clustering procedure was as above.

- the third distance was defined by a Choo-Siow (2006) type marriage function, dividing but the geometric mean of the number of marriages whereby the matrix used for clustering had elements $M_{ij}/\mathrm{sqrt}(M_{.i}*M_{.j}\ )$ where M_{ij} is the number of marriages between a member of municipality i and municipality j and $M_{.i}$ is the total number of marriages involving a person if municipality i.

- the fourth used a chi-square measure, dividing the expected number of marriage between 2 municipalities under independence by the observed number (since small values indicate closeness)

- the fifth and sixth used metrics derived from community ecology studies on number of species in locations (arising in part for information-theoretic approaches). I'll only discuss the Horn-Morisita index where the distance between sites j & k, d[jk], is defined as
  d[jk] = 2*sum(x[ij]*x[ik])/((λ[j]+ λ[k]) * sum(x[ij])*sum(x[ik]))
  where λ[j] = sum(x[ij]^2)/(sum(x[ij])^2) and x[ij] is the ijth element of the marriage matrix

- finally I used the original data as a control set.

Since 'closeness' is associated with larger function values, whereas near values have smaller distance, the

Since the results of such analyses, are known to be sensitive to the order of the data (largely related to the treatment of ties in distance), I then randomly permuted the initial matrix a number of times to assess sensitivity (the matrix is large and sparse and there were many equal distances so results were not identical) and calculated the number of 'endogamous' cases (i.e. marriages taking place with both partners in the same cluster or 'marriage market'). Since cases with multiple ties are sensitive to the clustering method used, I repeated the analysis with single linkage (which is less sensitive) and complete linkages (which takes account of 'friend of friends')

Table 5 Proportion of persons marrying within cluster under alternative metrics

	Complete		Single	
	Original	Randomised	Original	Randomised
IPF	69.94%	63.47%	67.55%	67.72%
schoen	69.39%	59.63%	70.57%	70.57%
cs	70.41%	62.94%	68.58%	68.58%
chisq	47.95%	47.82%	65.58%	48.34%
horn-morisita	71.46%	71.46%	68.28%	68.28%
raup	64.97%	62.66%	53.18%	48.86%
none	69.60%	47.96%	69.60%	47.96%

The size of the clusters is not fixed, but is constrained to lie within bounds close to 1/N of the total (with N clusters). Different approaches will produce different results. Other methods of trying to estimate these clusters included simulation by selecting random subsets (I'm grateful to David Freedman for this suggestion), but the number of trials needed turned out to be prohibitive. I also tried David's suggestion of simply randomly permuting the matrix and here is the code and results. Given that the initial distribution is likely to be close to the actual maximum, the total proportion of 'endogamous' marriages within the 6 areas ('marriage markets') identified is 67.1%, but 37.4% were within the actual municipality, so 29.7% took place within other areas in the marriage market (on average, there were 71 such areas inside the marriage market, and 361 outside in which the remaining 32.9% took place), so just under one half of those marrying outside the municipality did so in their marriage market.

The results of 999 simulations (excluding the non-random initial one) for the overall proportion endogamous are shown below (Table 6)

Table 6 Distribution of proportion of endogamous marriages under random permutation of areas

N	Mean	.05	.10	.25	.50	.75	.90	.95
999	0.4735	0.4658	0.4672	0.4701	0.4731	0.4765	0.4803	0.4829

Lowest: 0.4596
Highest: 0.5008

The values don't get close to the initial value of 67.1% and lie around 47.4% - since 62.6% take place outside the municipality, these figures give 10.2% in the rest of the market but outside the municipality, which is the expected proportion of 1/6 of all such marriage if marriage was random (i.e. correct, but the number of simulations would be astronomic to get values close to the starting value).

Formally, using definitions of 'distance' defined by these metrics could be applied equally well to different age or any other groups such as educational level, although there is additional information available, here it's a 'pure' question of partitioning a matrix.

The problem can be reformulated as we are told that the true maximum value is above a given value (that using the unpermuted matrix), and the challenge is to find a permutation which produces a larger value, subject to constraints on the marginal totals within that sub-matrices (i.e. close to 1126198/6). It seems likely that this simply has no solution in reasonable time like the 'packing boxes' or 'traveling salesman' problems.

Conclusions

The traditional model of a well-defined marriage market that provides the pool of people that marry other members of the group is becoming increasingly irrelevant as increasingly people find marriage partners from a much wider group of eligibles, many of whom will simultaneously be in a number of different potential marriage markets (e.g. a person living in a popular international holiday area).

Selected References

Adams J et al. (ed) (1990) *Convergent Issues in Genetics and Demography.* New York: Oxford Univ. Press.
Ní Bhrolcháin, M. (2001) Flexibility in the Marriage Market. *Population: An English Selection* 13(2):9–47.
Choo E. and Siow A. (2006) Estimating a marriage matching model with spillover effects. *Demography*, Volume 43: 463–490.
Goldman N, Westoff CF and Hammerslough C. (1984) Demography of the Marriage Market in the United States. Population Index 50(1):5-25.
Greville, TNE (ed) (1972) *Population Dynamics*. London:,Academic Press.
Henry L. (1972) Nuptiality. *Theoretical population biology*, (3):135-152,
Iannelli M, Martcheva M, and. Milner FA (2005) *Gender-Structured Population Modeling: Mathematical Methods, Numerics, and Simulations. Frontiers in Applied Mathematics 31*. Philadelphia: SIAM.
Keilman N (1985) Nuptiality models and the two-sex problem in national population forecasts. European Journal of Population 1:207-35.
Pollak R A. (1990) Two-Sex Demographic Models. *Journal of Political Economy* 98(2): 399-420.
Pollard JH. (1997) Modelling the Interaction between the Sexes." *Mathematical and Computer Modelling* 26: 11–24.
Pollard JH, and Ho¨hn C. (1993–94) The Interaction between the Sexes. *Zeitschrift fur Bevolkerungswissenschaft* 19 (2): 203–8.
Qian Z and Preston SH. (1993) Changes in American Marriage, 1972 to 1987: Availability and Forces of Attraction by Age and Education. *American Sociological Review*. 58(4):482-495.
Schoen R. (1981) The harmonic mean as the basis of a realistic two-sex marriage model. *Demography*, 18(2):201-216.