

Agent Based Modelling of Inter-Racial Partnership Patterns in New Zealand

Helen Broome

New Zealand is a multi-ethnic society with increasing diversity both across and within the ethnic groups. The two largest ethnic groups are European and Maori which accounted for 67.6 percent and 14.6 percent of the population in 2006 respectively [17]. The major immigrant groups come from the Pacific and Asia. Between 2001 and 2006 the numbers of Pacific peoples and Asians continued to increase by 14.7 percent and 48.9 percent respectively with the majority settling in Auckland.

To understand the integration of different ethnic groups and immigrant generations we look at the pattern of inter-ethnic partnership; this is typically understood as an indicator of group boundaries although it is unclear whether inter-ethnic partnerships are a cause of integration or because of integration.

We will explore the use of agent based modelling (ABM) techniques to find the conditions that best explain the pattern of inter-ethnic partnership. The trend in the formation of same-ethnic partnership, or homogamy, is of particular interest. By partnership we refer to a cohabiting couple.

The agents in the model are assigned four individual characteristics; age, gender, education level and ethnicity. The model is assortative meaning individuals are assumed to prefer someone who is similar to them in terms of age and education. How agents consider ethnicity, and whether there is a preference for homogamy, is what we are investigating.

The focus of the project is on the emergence of ethnic partnership patterns understood as “stable macroscopic patterns arising from the local interaction of agents” [4]. In particular we focus on *second-order emergence* where the micro level interactions of reflexive agents involves an awareness of the macro level patterns they create [16]. The agents meet partners from within their network based on two constraints; there is a structural constraint on who they can partner with determined by the distribution of characteristics in their network and secondly a cultural constraint based on their recognition of the societal norm, or macro pattern, that current exists around particular inter-ethnic partnerships.

To evaluate which matching algorithm performs the best we use census data and compare the simulated couples against the observed ones. The level of error reduces noticeably as we extend the model from a random matching process to include an assortative matching process and more sophisticated measures of homogamy.

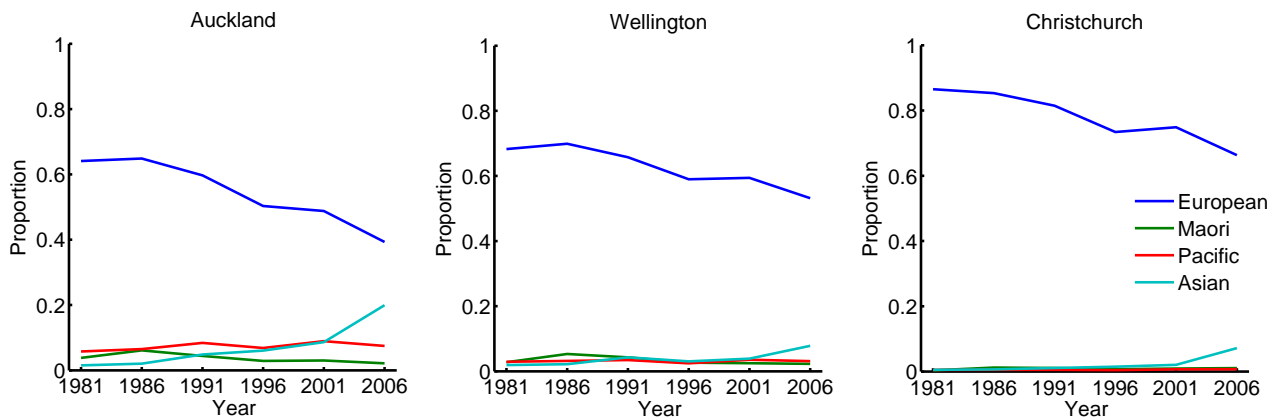
In the first section we discuss ethnicity in the context of New Zealand focusing on the census data we have and exploring ways to quantify the pattern of inter-racial partnerships. The second section describes the model and discusses the choice of minor parameters. The final section evaluates the different models based on ethnicity outcomes when the main parameters used in the matching algorithm are optimised with an evolutionary algorithm.

1 Ethnicity in New Zealand

The majority of couples are homogamous European couples; in 2006 there were nearly ten times as many homogamous European couples as the next highest group of homogamous Asian couples. The number of couples is partly due to the fact that Europeans are the majority ethnic group but as group size changes over time it is important to disentangle the effect of demographics from changing social norms when we look at the pattern of partnerships.

The following images give an indication of the proportion of partnerships of a given ethnic combination over time. While there is definitely a change in the proportion of homogamous European couples this is partly driven

by the changing proportion of Europeans.



Proportion of Homogamous Couples Aged 18 - 30 from the Major Ethnicities

Our simulations are done for the three major cities; Auckland, Wellington and Christchurch. Each city is treated in isolation and there is no finer grained geographical data at the city level. Of these cities Auckland is the largest with 1.37 million people in 2006 while Christchurch and Wellington had 540,000 and 470,000 respectively. Auckland has the highest levels of ethnic diversity as seen by the lower proportion of homogamous European couples.

We are interested in the incidence, rather than the prevalence, of homogamy as this tells us about the current norms surrounding partnership formation. As the census data does not record relationship duration we will use the younger 18 - 30 year old cohort as a proxy for new relationships.

In this section we will discuss the ethnic data we used from the New Zealand census, briefly looking at the relationship between ethnicity and other characteristics such as language and education. Then we focus on finding a measure for the pattern of inter-racial partnerships and look at how this measure has changed over time in the census data.

1.1 Recording of Ethnicity in Census Data

In the census data ethnicity has been recorded in terms of affiliation rather than blood lines since the 1970s and multiple ethnicities were allowed to be specified. Subsequent investigation showed that the affiliation with multiple ethnicities was independent of gender [3]. By 2006 nearly one in eight people had multiple ethnic affiliations and nearly a half of those who affiliated with Maori also identified another ethnicity [6].

In the data we have access to the ethnic labels have been grouped together so we do not know country of origin. There are a maximum of thirteen possible ethnic groupings although some years merge one or two of these groups. There are six single ethnicity affiliation groups;

European	Maori	MELAA (Middle East, Latin America and Africa)
Asian	Pacific	One other ethnicity not already specified

The MELAA group is a very broad category that mainly serves to distinguish who is in the 'other' category. For the multiple ethnicities there are five groups of paired ethnic affiliations and two other groups;

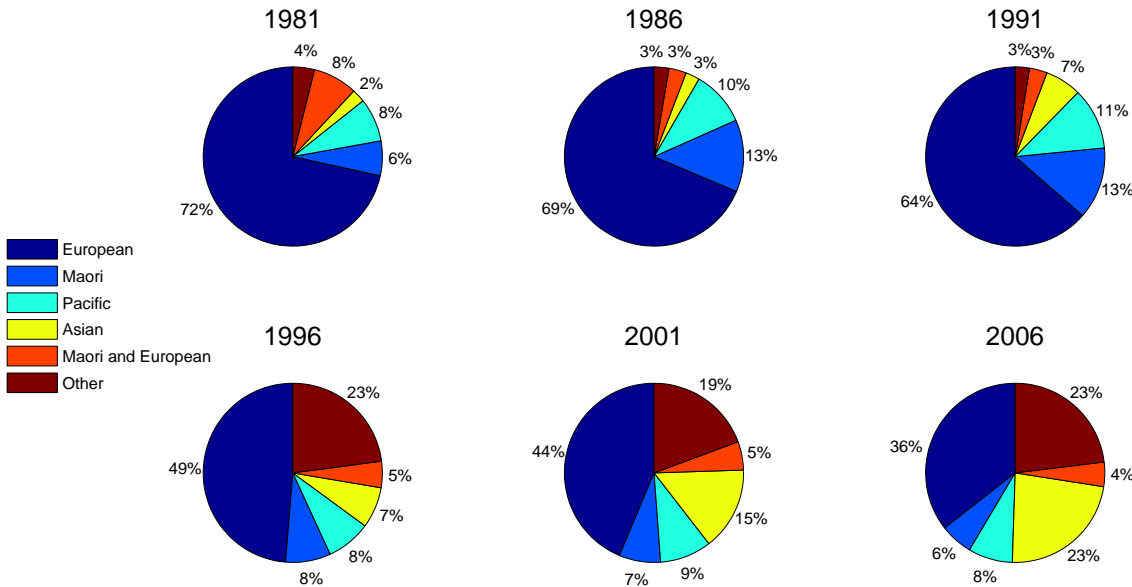
With blended ethnic identification the classification of what counts as homogamy becomes more difficult. We treat these thirteen categories as discrete and so someone from the European group partnering with someone

Maori and European
 Maori and Pacific
 Asian and European
 Pacific and European

Maori and Pacific and European
 Two other ethnicities not already specified
 Other not already specified

from the Maori and European group is not treated as homogamy. However we utilise a measure of inter-racial partnership degree that allows us to recognize more common pairings which is discussed in section 1.3.

While individuals can specify multiple ethnicities it is not clear that the recording of this data is consistent. Across all cities there was a large change from 1991 to 1996.



Ethnic Affiliation of Single People Aged 18 - 30 in Auckland Census Data from 1981 - 2006

It is not clear that this is just due to a trend or a sudden jump in immigration. For instance in 1986 and 1991 there is a dip in the number of single Maori and European people in our data set while over the same time period there is a bump in the number of Maori suggesting that how multiple ethnicities are recorded in our data may not be consistent.

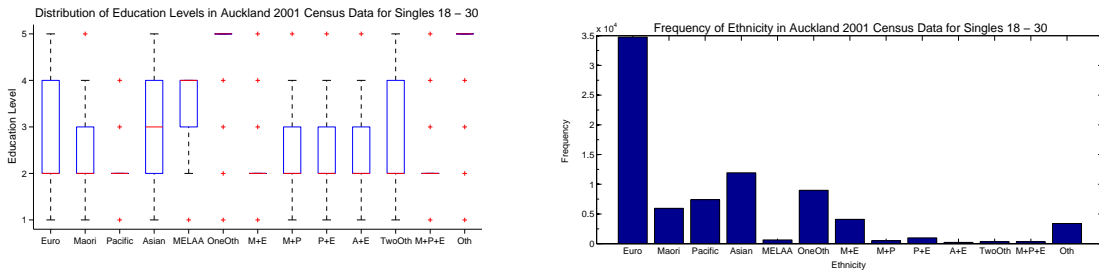
Due to the unexplained jump in ethnic make up from 1991 to 1996 and the use of less educational categories prior to 1996 we base our investigation on the 1996 - 2006 data set where ethnic and education categories are comparable. Although in 1996 there are only 12 categories as 'Maori and Pacific and European' is merged with the 'other not already specified'.

1.2 Ethnicity in relation to other factors

The outcome we are measuring is ethnicity but this is not independent of other characteristics such as education level, language or religion.

Levels of education have been improving, particular in Auckland which partly reflects the larger number of

immigrants whose higher education level is a consequence of the immigration selection process. From 1996 to 2006 the percentage of adults with no qualification dropped from 38.1 to 25.0 across New Zealand while the number with a tertiary qualification rose from 9.5 to 15.8 percent in the same period [19].



The census data of single 18 to 30 year olds in Auckland from 2001 shows that the distribution of education levels differs across ethnic groups where education levels are recorded from 1 (low) to 5 (high). Among the larger groups we see the Asian group is highly qualified while the Pacific group has a low level of qualification. Furthermore there is a contrast between European, Maori or Maori and European showing the importance of multiple ethnic affiliation for distinguishing group characteristics. Similar patterns are seen across the other cities.

Education has a bearing on homogamy; Maori and Pacific peoples with higher education have lower rates of homogamy while European/Maori partnerships are more common among people with higher education levels. In particular 76 percent of Maori men and 67 percent of Maori women with a degree or higher qualification have a partner who identifies European among their ethnic affiliations compared to only 51 percent of Maori men and 47 percent of Maori women with no formal qualifications [3]. This is in line with findings from the US that an individual with an education level that is not typical for their ethnic group is more likely to partner with someone of a different ethnicity [9].

The model we use includes education level but leaves out religion and language. Both of these factors are relevant to ethnic identification although the links with partnership formation are harder to identify and the information was not in our data set. We discuss these issues briefly but do not investigate them further.

While New Zealanders tend to be quite moderate about religious views there are strong ties between immigration, ethnicity and religion. In 2006 nearly 80 percent of those who identified with Islam, Hindism or Buddhism were immigrants, mainly from Asian, while just over 80 percent of Pacific peoples who answered the question on religion identified as Christian [18]. What is less clear is how this translates to partnership norms as partners may change their religion once beginning a new relationship.

Both English and Te Reo Maori are official languages in New Zealand although the number of Te Reo Maori speakers is quite small. Shared languages do coincide with ethnic partnership patterns with higher rates of homogamy among Te Reo Maori speakers. In 2001 36 percent of Maori females in a homogamous partnership spoke Te Reo Maori compared to only 15 percent of those with a European partner [7]. However among Te Reo Maori speakers 20 percent of females and 22 percent of males were not Maori. Similar to religion it is not clear whether the prevalence of shared language in partnerships is due to individuals who speak the same language being more likely to form a relationship or whether individuals are more likely to learn new languages once in a relationship.

1.3 Measuring patterns of ethnic partnership

Not only are the sizes of ethnic groups quite different but they continue to change over time. Both of these factors impact on how we quantify the pattern we see so that it is a comparable measure between groups and

across years.

In the literature on racial and educational homogamy various measures are used; the proportion [21], the generalized odds ratio [1], and the degree of assortativity that we will adapt [10]. Both the proportion and the odds ratio focus on the binary case of whether there is a trend for homogamy or not. However rates of homogamy differ between ethnicities and in heterogenous relationships trends are not symmetrical when gender is exchanged; for instance there are more couples with an Asian female and European male than an Asian male with a European female.

Following Liu and Lu [10] we will measure the degree of assortative mating as the relative distance between the observed pattern and a case where individuals were randomly matched. This controls for the changing demographics and group size in terms of ethnicity and gender. We will refer to this as the inter-racial degree. As this measure is designed for a dichotomous trait we consider each of the possible 169 ethnic combinations separately and each time group the remaining ethnicities into an ‘other’ category.

	Female Ethnicity j	Other Female Ethnicity o	Total
Male Ethnicity i	$N_{i,j}$	$N_{i,o}$	M_i
Other Male Ethnicity o	$N_{o,j}$	$N_{o,o}$	M_o
Total	F_j	F_o	T

If the matching was perfectly random then $N_{i,j}$, the number of couples with a male of ethnicity i and a female of ethnicity j , would be

$$N_{i,j}^{\text{random}} = R_{i,j} = \frac{M_i \times F_j}{T}$$

where M_i is the total number of males of ethnicity i , F_j is the total number of females of ethnicity j and T is the total number of couples.

Given the distribution of ethnicities $d = \{M_i, F_j, T\}$ we can construct three benchmark measures that tell us the maximum and minimum levels of assortativity at the two extremes and the level that would occur at random. For the maximum and minimum levels once the value for $N_{i,j}$ is known the values for the other three ethnic partnership combinations are fully determined. Maximum partnership occurs when there is the highest number of males with ethnicity i partnering with females of ethnicity j , namely when

$$N_{i,j}^{\text{max}} = \min(M_i, F_j).$$

Similarly the minimum partnership level occurs when there is the highest number of males with ethnicity i partnering with females of any ethnicity *except* i , namely when $N_{i,o} = \min(M_i, F_o)$. This is equivalent to $N_{i,j}$ being the number of surplus males of ethnicity i , namely

$$N_{i,j}^{\text{min}} = \max(0, M_i - F_o).$$

As $N_{i,j}$ is enough to determine the maximum or minimum partnership levels we can restrict the measure to looking at how far $N_{i,j}$ differs from a random matching case, $R_{i,j}$, compared to how far it could differ. When $N_{i,j} = N_{i,j}^{\text{max}}$ this is the maximum partnership case and the inter-racial degree is 1 indicating a high preference for this combination. When $N_{i,j} = N_{i,j}^{\text{min}}$ the inter-racial degree is -1 indicating a tendency to avoid this combination. While for $N_{i,j} = R_{i,j}$ the inter-racial degree is 0 indicating no preference.

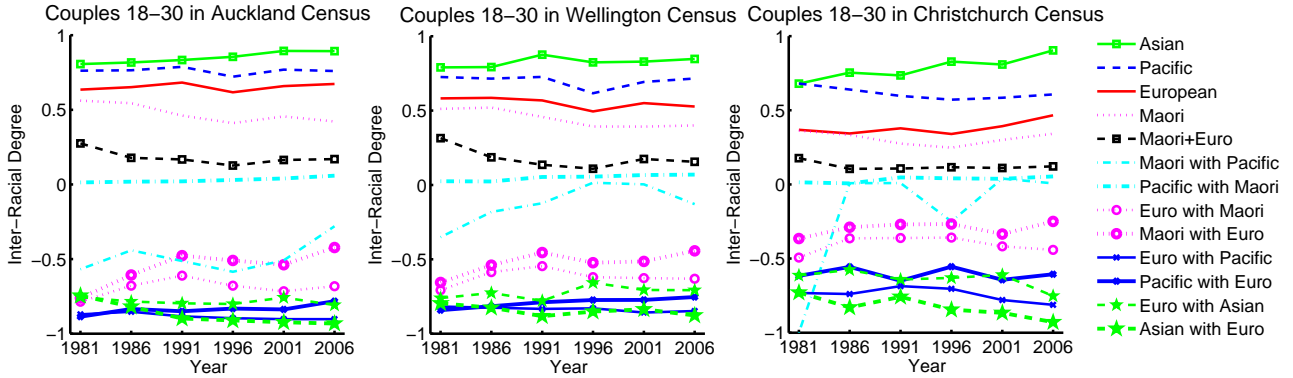
The inter-racial degree D is defined in terms of the distribution $d = \{M_i, F_j, T\}$ and the observed, or simulated, value for $N_{i,j}$.

$$D(N_{i,j}, d) = \begin{cases} \frac{N_{i,j} - \lfloor R_{i,j} \rfloor}{N_{i,j}^{\max} - \lfloor R_{i,j} \rfloor} & \text{if } N_{i,j} \geq R_{i,j} \\ \frac{N_{i,j} - \lceil R_{i,j} \rceil}{\lceil R_{i,j} \rceil - N_{i,j}^{\min}} & \text{if } N_{i,j} < R_{i,j} \end{cases}$$

[Note: $\lfloor x \rfloor$ is the floor function which returns the largest integer n such that $n \leq x$, similarly $\lceil x \rceil = m$ where m is the smallest integer such that $m \geq x$.]

1.4 Inter-racial degree in census data

Calculating the inter-racial degree for various combinations in the three cities from 1981-2006 we can see that the stratification does not change over time or between cities although the relative degree does. The combinations shown are the homogamous partnerships for the five major ethnicities - European, Maori, Asian, Pacific or Maori and European - along with the more common non-homogamous combinations with male ethnicity listed first. The non-homogamous combinations show that there is a gender effect with a noticeable difference in the inter-racial degree when gender is reversed.



The homogamous combinations all occur more frequently than they would in the random matching scenario which corresponds to an inter-racial degree of 0. The Asian and Pacific groups have the highest level of homogamy relative to their group size which may relate to the continued high levels of immigration from these groups. By contrast the non-homogamous combinations occur less frequently than under random matching except for the Pacific male with Maori female combination which occurs at the same frequency as it would under random matching. The strange behaviour of the male Maori with Pacific female in Christchurch is due to the small group size with 0 couples in this combination in 1981.

Over time the trends in inter-racial degree are quite static. These descriptive statistics suggest that changing demographics have a large role to play in explaining changing numbers of couples. However this measure only considers the distribution among people who are already in a couple which may not be representative of the demographics in society.

We investigated a modified inter-racial degree that took account of the distribution of single females along with the couples. While it showed similar trends to the inter-racial degree for couples the error in the model was much higher so we use the original version described above that only uses the distribution of ethnicities within the people who are in a relationship.

1.5 Macro-Micro Feedback

We will use the inter-racial degree as a means of quantifying the current ‘norm’. The inter-racial degree acts as a feedback between the macro level pattern and the micro level partnering process.

Having said that the inter-racial degree we use is calculated and fixed at the start of the simulation from census data so the feedback loop is not within the model but only found in the data we are using. The reason for this is that the inter-racial degree does not change much over a five yearly period and when it dynamically updates based on the couples that were formed in the simulation it has slightly higher error than just using the value from census data. In the Auckland 2001 to 2006 simulation the error was 845 ± 2 for the static inter-racial degree while it was 866 ± 2 when the dynamic inter-racial degree is used. Despite the slightly higher error the dynamic inter-racial degree does have the advantage of utilising the simulation process to explain the developing pattern.

The model we investigate uses the static inter-racial degree to investigate the general model structure although future work could develop on the dynamic inter-racial degree.

2 Simulation

This section outlines the basic model before discussing the parameters. The major parameters we focus on are the weights in the score function and these are discussed in the next section. The minor parameters are discussed here and the age and educational outcomes are checked to make sure they are not unreasonable.

2.1 Basic Form of the Model

Each male agent is assigned a growing network of females. At each iteration they give each female a score based on how ‘similar’ they are. If the best score exceeds their current threshold they ‘propose’ to the woman. Each woman has a set of proposals which typically contains 0 - 4 men and from this set they randomly choose a partner. At the end of the five years simulated a table of partnerships is recorded in terms of male and female ethnicity. The table is used to calculate the error in comparison to the table of observed partnerships obtained from the census data.

Initially we considered having the women look through the set of men that proposed to them and give them a score to pick the best. However at each iteration roughly half the women receive a proposal and less than half of those receive more than one. The number of men to evaluate is quite small and the men tend to have quite similar characteristics so when the women choose someone at random the level of error is no different to when they also evaluate potential partners. Hence random choice is used as it is computationally simpler.

Other significant details are that the initial threshold level is chosen from a uniform distribution whose range is dependent on age. Males that are single reduce the level of their threshold over time while males in a relationship raise their threshold level. When males start a relationship their threshold level is set to the score they gave their new partner and we allow agents that are in a relationship to continue looking and change partners until their relationship exceeds their randomly chosen ‘minimum partnership time’ at which point they are removed from the pool.

We use the threshold system which allows agents to start and break up numerous relationships instead of the competitive system where all potential relationships are ranked by score and then the highest scoring ones are formed and fixed. This means we are placing less restrictions on the type of relationships that form and allowing the pattern to be determined by the agents local interactions rather than a global constraint.

However we do fix the total number of couples that form in the five yearly period to match the census data. In a

model where the number of couples is fixed the error is about 10 percent lower than when the number is unfixed and the error is calculated by rescaling the observed tables to have the same sum as the simulated table. The problem is that an excess of couples forms, up to 25 percent too many in Christchurch and Wellington and 10 percent too many in Auckland. When the maximum number of couples is reached more couples can still form as long as it involves breaking up a pre-existing relationship.

Using the mechanism of homophily to explain partnering behaviour may not be as relevant to homosexual couples who are less likely to have the same education, age or ethnicity [14]. Furthermore sexual orientation is not recorded for single individuals in the census so it would be difficult to infer the appropriate distribution of age, education and ethnicity that corresponds to homosexual agents. For simplicity we make the erroneous assumption that all agents are heterosexual.

2.2 Parameters; Cultural and Structural

Cultural explanations for the pattern of ethnic partnerships focus on norms, values and preferences around interaction and partnership with someone from the same or different ethnic group [9]. In the model the cultural parameters control how a male scores the females in his network and reflect the relative preferences for certain characteristics. The parameters are listed below;

Scoring Weights for the four variables in the scoring function; age, education, stochastic attraction factor and the inter-ethnic partnership degree. These are optimized using an evolutionary algorithm and are the main parameters investigated here.

Disapproval Factors scales the raw value of age or education level difference to maintain the age and education scores in the range $[0, 2]$ while discouraging large differences in either age or education. There is no data to compare age or educational outcomes against and so these parameters are chosen to give plausible results based on expected behaviour from the literature.

Satisficing and Satisfaction Rate determine the level the threshold drops for single males and increases for ones in a relationship. The intention is to prevent males dropping one partner for someone with the same characteristics.

Initial Threshold is selected from a uniform random distribution where the range of values is age dependent being lower for older males making them more willing to ‘settle down’.

To deal with the variety of parameters we set most of them at ‘reasonable values’ based on some testing. Then we used an evolutionary algorithm to find the optimal weights in the scoring function as these were the main parameters we were interested in. With the optimal weights we then went back and tested the other parameters to see if the error was sensitive to these parameters. None of the other cultural parameters had much effect and various other parameter values produced similar error showing that it was not sensitive to these values.

Structural constraints are concerned with the factors that shape meeting and mating opportunities to partner with someone of the same or different ethnic group [9]. As we did not have any data on the social networks across the city we assigned a network of friends to each agent randomly. Despite the use of random networks many important structural factors are present within our model; for instance the impact of gender ratio, typical education level and group size across different ethnicities is captured by the likelihood of interacting with an agent with those characteristics. Any restrictions on the type of agents that interact is purely determined by the likelihood of their interaction; what we do restrict is the number of agents that interact and the frequency with which this occurs using the following structural parameters.

Number of Female Friends assigned to each male initially and the number added at each subsequent iteration.

Number of Iterations over the five year period.

Probability of Switching Partners for a female who is in a relationship and has a set of new proposals to choose from.

Total Number of Couples which is determined by the census data. Once the maximum is reached the men continue to evaluate their female friends as usual but only the women who are already in a relationship will randomly choose whether to change relationships as this will either maintain or decrease the total number of couples. Without this restriction the level of error is much higher even after the outcomes and simulated results are rescaled to have the same sum.

The structural parameters are partly determined by computational time; while adding extra friends or using more iterations gives the agents more opportunities to interact, find better partners and reduce the overall error it takes more time. For the sake of comparable models we have fixed the number of initial female friends at 50 with 20 extra at each iteration and used one iteration to represent one year. Typically the maximum number of couples is reached in two iterations and the other three iterations allow for some movement with the probability of a female switching partners set at 0.05. The level of error was not sensitive to the probability of females switching partners so we just fixed one value and maintained it for all models.

With more data the structural constraints that impact the type of people an agent regularly interacts with could be included; for instance type of workplace, local suburb or educational and religious institutions. However there is a lot of debate in the literature about the comparative significance of these ‘local marriage markets’ on partnership outcomes [8] [2] and so we were wary over introducing unjustified constraints that would just complicate the model. Furthermore in New Zealand there are low levels of neighbourhood segregation in terms of ethnicity and increasing levels of mixing are occurring [20].

2.3 Scoring Function

The scoring function, s , uses the age difference, educational difference, inter-racial degree for the given ethnic pairing and a stochastic ‘attraction’ factor to find the best match by minimising differences.

The scoring function is modified from previous work by Walker [21] and his development of the DYNASIM model described in [12] which makes our work more comparable with past models. The form is,

$$s = e^{-\sqrt{w_a A + w_e E + w_x X + w_d D}}$$

where A is an appropriately scaled, asymmetric age difference, E is the scaled difference in education level, X is a uniform random number used as the ‘attraction factor’ and D is related to the inter-racial partnership degree. With each of the four variables there is a corresponding weight which are the main cultural parameters we will investigate. By constraining each of the variables to the same range of $[0, 2]$ we can interpret the corresponding weights as a qualitative guide to the significance of the variables. X is a uniform random number from the range $[0, 2]$ while the variables A , E and D are defined below and discussed further in section 2.5.

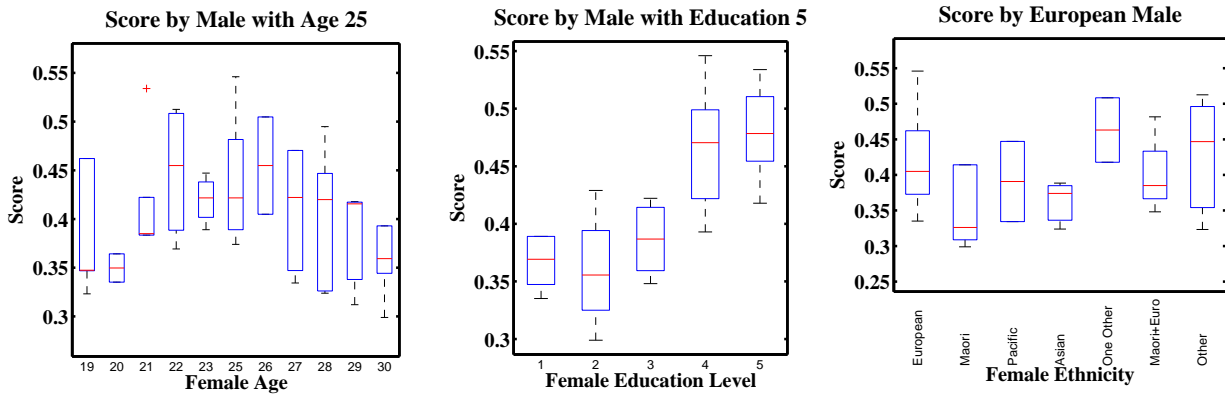
$$\begin{aligned} A &= \left(0.85 - \frac{\text{age}_{\text{male}}}{40}\right) \times |(\text{age}_{\text{male}} - 1) - \text{age}_{\text{female}}| \\ E &= 0.5 \times |\text{education}_{\text{male}} - \text{education}_{\text{female}}| \\ D &= 1 - \text{inter-racial degree} \end{aligned}$$

If a male is giving a score to the females in his network then the females who are most similar to him will receive a higher score. We interpret the variables as an indication of how different two agents are, with the preference for homophily meaning the less different the more desirable. This means the highest score comes from

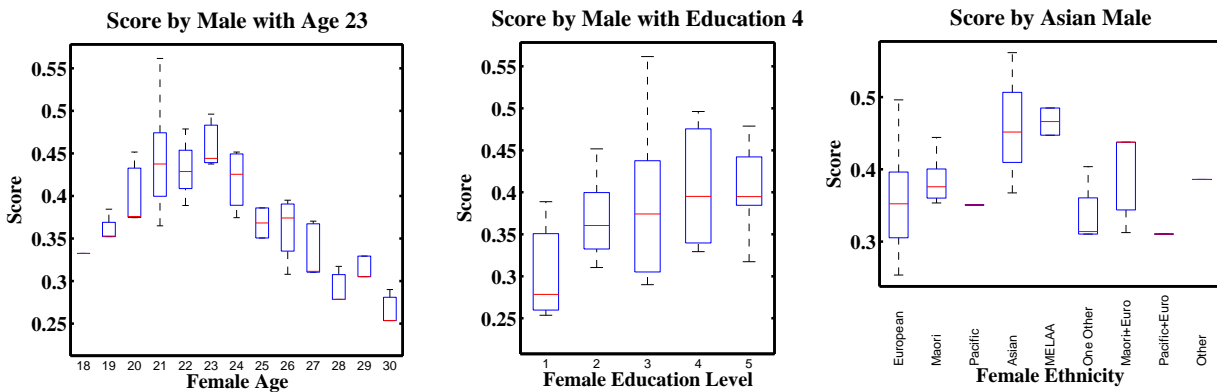
minimising age and educational difference, minimising the ‘attraction’ factor and minimising the inter-racial degree which occurs when the particular ethnic combination is more in line with the societal norm or current pattern i.e. an inter-racial degree of 1 represents the highest preference for that combination and corresponds to a score of 0 for the variable D .

2.4 Range of Scores Across the Network of Female Friends

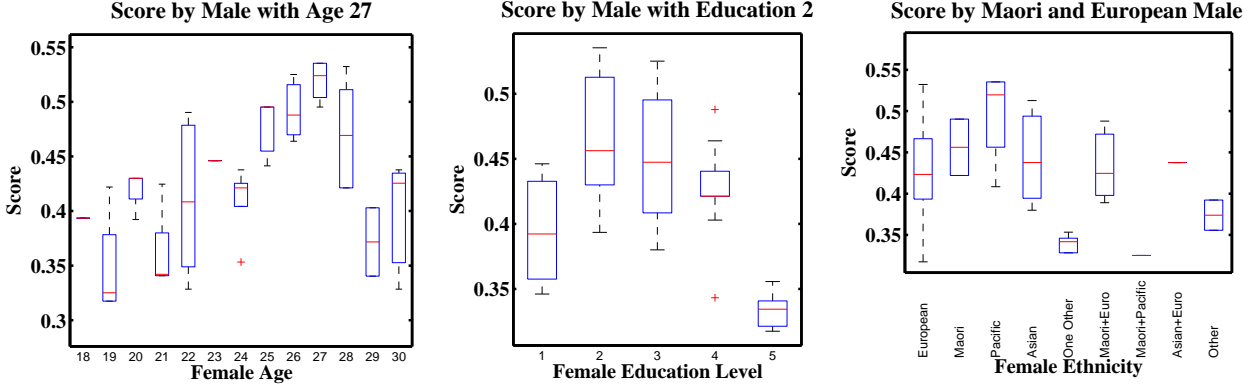
To demonstrate how the scoring function responds to the different characteristics the following graphs show the range of scores given to the network of females for a 25 year old European male with high education, a 23 year old Asian male with high education and a 27 year old Maori and European male with low education. Their ages are given as at 2001 when they were single in Wellington and the weights used in the scoring function were 0.45 for age, 0.35 for education and 0.2 for inter-racial degree. These are the optimal weights for the Wellington 2001 simulation as found from the evolutionary algorithm results in section 3.3.



For this European male education is the feature with the greatest polarity between the range of scores with only the high education levels having high scores. As this agent has no 24 year old women in his network the age score is similar across a range of ages close to his age. When it comes to ethnicity the ‘other’ categories have higher median and upper quartile than the European category which is probably driven by higher levels of education in immigrant groups. However the highest score is a female of the same ethnicity, slightly lower education and slightly older age.



For the 23 year old Asian male the effect of age is quite clear with a rapid tailing off of scores for older females. There is also a prominent effect of ethnicity with Asian and MELAA groups having much higher score ranges than other ethnicities. Education has some effect but not as much as in the case above, partly due to having an education level of 4 instead of 5. Again the best match is homogamous with a slightly less than ideal education and age level.



Unlike the previous two examples the best match here is not homogamous. In part this is due to the smaller size of the Maori and European group meaning there are less chances of finding someone with the same age and education within that ethnic group. The best match was with a Pacific woman with the same age and education level.

As the scores are for the 70 females in their respective networks the plot for a single feature only gives some indication of how the scoring focuses on that feature. However the scoring function seems to pick up on a preference for homophily without ruling out the inclusion of some difference which seems intuitive.

2.4.1 Age and Education Outcomes

With no data to compare simulated age and educational outcomes against this section sketches the intuition behind these features and checks that the simulated results are not unreasonable.

As men are giving a score to females and men tend to be older than women in a relationship we treat the 'ideal' age difference as being one year younger. In New Zealand the median age gap decreased from 2.28 years in 1983 to 1.94 years in 2003 however this is considering couples of all ages [13]. We use a one year age difference because it goes some way to recognising a pattern for age asymmetry without penalising the younger men in the model too much given that all the men are looking for partners in the same, fixed pool of 18 - 30 year old women. For the 18 and 19 year old men we drop the condition on being one year younger to not disadvantage them for the lack of 16 and 17 year old women included.

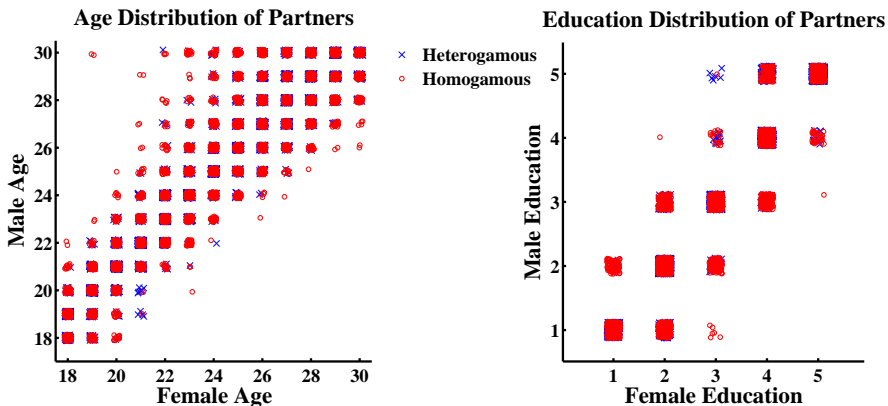
The age difference is scaled by an age dependent constant $(0.85 - \frac{\text{age}_{\text{male}}}{40})$ to capture the intuition that a 30 year old man is more likely to have a 25 year old partner (i.e. four years away from the ideal age difference) than a 20 year old man is to have either a 15 year old or a 23 year old partner.

For a 20 year old male only women in the age range 18 - 24 have a score in the range [0, 2] and only 18 - 21 year old women have a score in the range [0, 1]. An age score that is greater than 2 represents a highly unlikely pairing. For a 30 year old man these ranges extend so that all women in the age range 18 - 30 have a score in the range [0, 2].

For the education difference there is a constant scaling of 0.5. Consequently only males with an education

level of 1 or 5 will use the full education score range of $[0, 2]$ whereas for a male with education level 3 the educational score range will be $[0, 1]$. As educational homogamy is observed at higher levels at the extremes of the educational spectrum [5] it seems reasonable that someone with a middle education level would consider anyone else to be relatively similar where as those at the extreme ends would be more likely to notice differences in education level.

As we did not have access to data on partnership outcomes described in terms of age or education level the best we can do is check that the simulated outcomes seem reasonable. The scatter plots are similar to those from a recent US study [11]. We expect to see a general clustering around the diagonal in the scatter plots which show male age or education compared to female age or education within couples. Whether the relationship was homogamous or not was also recorded to check for difference in patterns although none was observed.



Age and Education Distribution for Simulated Couples in Wellington 2001 - 2006

Changes to the educational or age terms do not alter the level of error. This is because the error is based solely on ethnicity outcomes. The above results for the simulated age and educational distributions do not seem unreasonable and so without further data to test against we have kept these features constant across all the models tested.

3 Evaluation of the Model

3.1 Error Calculation from Census Data

From each census we have a table of couples categorised by male and female ethnicity for the 13 ethnic categories. By subtracting the number of couples aged 18 - 30 at one census from the number aged 23 - 35 at the subsequent census we get a general indication of the number of couples that have formed. This is not an ideal benchmark as it does not account for relationships that have ended in the same period and so there are often a few negative values in our table of 'observed partnership formation'. The negative values involve one or both partners belonging to a smaller ethnic group and so by restricting our focus to the outcome for the major groups we have a more realistic benchmark to use for comparison. Furthermore the creation or dissolution of a small number of partnerships between a male and female of different ethnic groups with one or both belonging to small ethnic groups is not as indicative of partnership trends as when we are more selective about the combinations to focus on.

The error calculation is the root mean square of the difference between 33 out of the 169 ethnic combinations in the simulated and observed partnership tables over a five year census period. We look at the 25 combinations

of male and female ethnicity pairings from the five largest ethnic groups - European, Maori, Pacific, Asian or Maori and European - along with the 8 other possibilities for a homogamous partnership.

Occasionally one of the values of observed partnership formations in the homogamous partnerships within the smaller ethnic groups is negative. In this case we set the value to 0 and calculate the error as above.

3.2 Evolutionary Algorithm

Given an initial set of weights 134 different perturbations are tested and the one with minimal average error is used for the next iteration of the evolutionary algorithm. The 134 perturbations involve changes by increments of ± 0.05 with the number of increments being increased or decreased involving the following combinations;

- pairs of ± 0.05 , (12 varieties)
- pairs of ± 0.1 (12 varieties)
- pairs of pairs ± 0.05 , ± 0.05 (6 varieties)
- triples with $+0.1$, -0.05 , -0.05 or its negative (24 varieties)
- triples with $+0.2$, -0.1 , -0.1 or its negative (24 varieties)
- triples with $+0.15$, -0.1 , -0.05 or its negative (48 varieties)
- quadruple with $+0.15$, -0.05 , -0.05 , -0.05 or its negative (8 varieties)

The changes are coarse grained as we are only looking for a qualitative indication of the relative importance of each factor and similar levels of error are found with small perturbations. Various different initial weights are then used to confirm that the algorithm is not just settling on a local minima.

3.3 Comparing the Error from Different Models

Using five independent realisations for each weight combination and model we calculated the minimal error to compare four different models.

Random Matching where males propose to a female in their network at random and females pick at random under their usual constraints.

Age and Education is an assortative model where males score females in their network based on how similar they are in age and education. A stochastic ‘attraction’ factor is also included.

Homogamy Proportion extends the assortative model above to also include a term that represents the proportion of homogamous couples that formed at the last iteration. If there is a higher proportion of homogamous couples then homogamous matches will have a slightly higher score.

Inter-Racial Degree replaces the measure of the proportion of homogamous couples with the inter-racial degree. This is set from the table of 18 to 30 year old couples at the start of the census period being simulated with a value given to each of the 169 possible ethnic combinations to capture differences between groups and gender.

As the decreasing level of error shows the use of the inter-racial degree picks up on the importance of distinguishing trends between ethnicities and gender. However the low or zero weighting on the educational term in the inter-racial degree model suggests our use of educational difference is not picking up on a relevant factor. The corresponding high weighting on the stochastic factor in the 1996 - 2001 model suggests either a reworked educational term or some other important factor is missing.

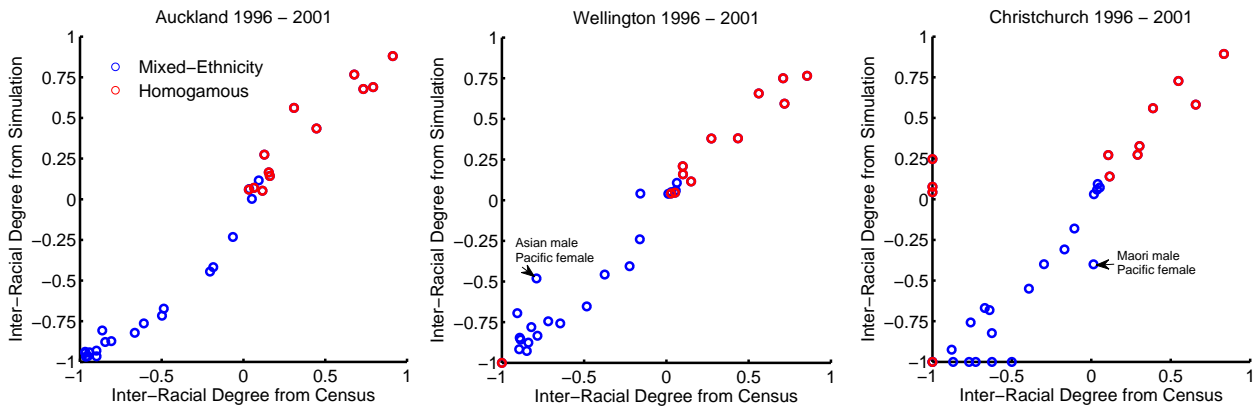
	1996 - 2001					2001 - 2006				
	Weights				Error	Weights				Error
	Age	Ed	Att	Race		Age	Ed	Att	Race	
Auckland										
Random Matching	-	-	-	-	1292 ± 2	-	-	-	-	1845 ± 2
Age and Education	0.30	0.60	0.10	-	1128 ± 2	0.25	0.70	0.05	-	1583 ± 2
Homogamy Proportion	0.10	0.10	0.30	0.50	680 ± 1	0.65	0.05	0.00	0.30	1091 ± 2
Inter-Racial Degree	0.35	0.00	0.30	0.35	525 ± 2	0.75	0.05	0.00	0.20	845 ± 2
Wellington										
Random Matching	-	-	-	-	370 ± 2	-	-	-	-	339 ± 3
Age and Education	0.50	0.50	0.00	-	308 ± 2	0.10	0.65	0.25	-	316 ± 2
Homogamy Proportion	0.45	0.00	0.15	0.40	134 ± 1	0.60	0.10	0.00	0.30	174 ± 1
Inter-Racial Degree	0.55	0.00	0.15	0.30	103 ± 1	0.45	0.35	0.00	0.20	149 ± 1
Christchurch										
Random Matching	-	-	-	-	339 ± 3	-	-	-	-	317 ± 2
Age and Education	0.30	0.50	0.20	-	315 ± 4	0.30	0.40	0.30	-	256 ± 1
Homogamy Proportion	0.20	0.10	0.20	0.50	75 ± 2	0.75	0.00	0.00	0.25	124 ± 1
Inter-Racial Degree	0.30	0.00	0.30	0.40	55 ± 2	0.75	0.10	0.00	0.15	96 ± 1

We also tested the generalised odds ratio as an alternative measure of homogamy but it gave similar error to the use of a simple proportion. This indicates the importance of distinguishing between ethnicities.

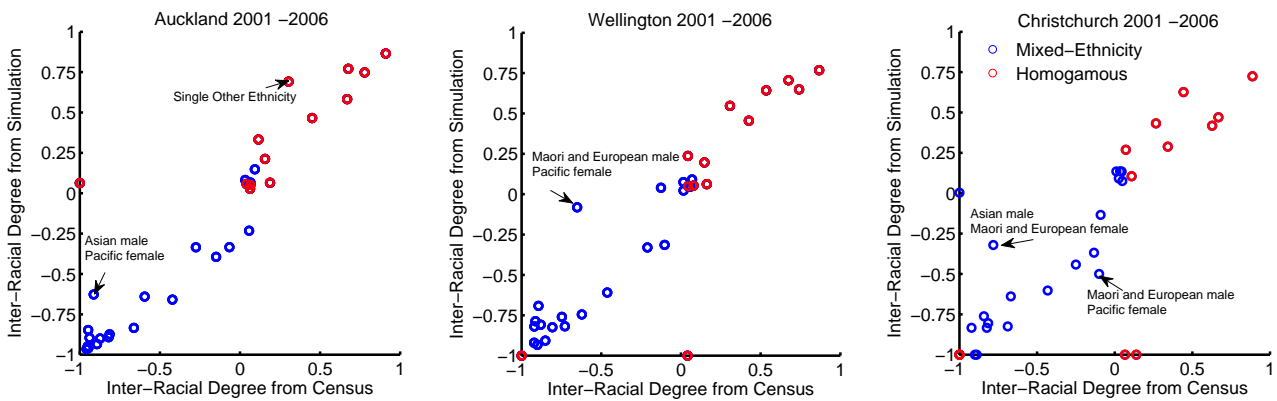
3.4 Ethnicity Outcomes from Simulations with Optimal Weights

Using the optimal weights with the inter-racial degree model found above we look at the ethnicity outcomes for the simulation compared to the census data. The census data used here is the 23 to 35 year old couples at the end of the five year period. For comparison the number of simulated couples is added to the number of 18 to 30 year old couples from the census data at the beginning of the five year period.

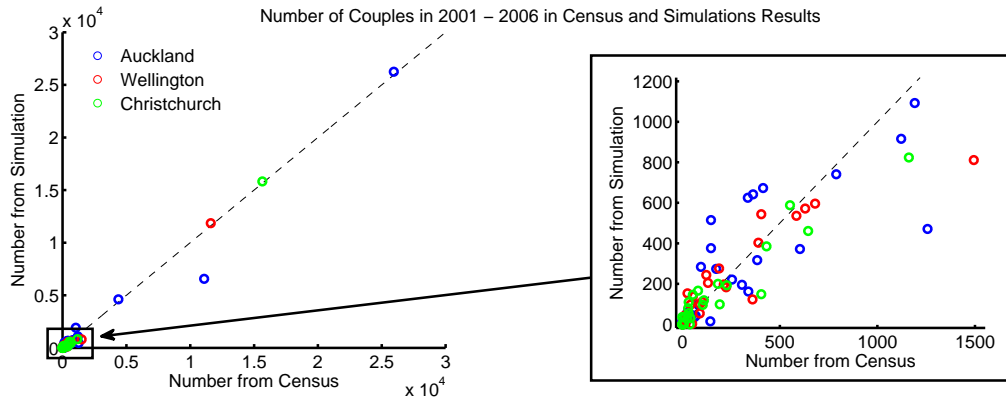
Comparing the inter-racial degree for the simulated couples and the census couples there is a strong correspondence with the values lining up along the diagonal. From the census data we saw that the homogamous couples occur more frequently than they would under random matching and this pattern is maintained, and slightly over emphasised in the simulation producing the S bend that is more evident in Auckland.



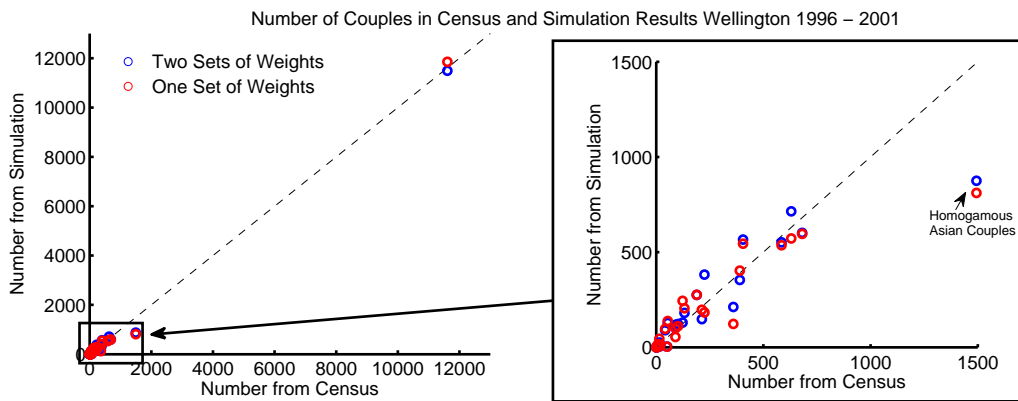
The points lying on either the x or y axis correspond to an inter-racial degree of -1 in either the census or simulation results; this occurs when there are 0 couples of the given type and is more common in Christchurch where there is less ethnic diversity and smaller group sizes. Aside from the values lying on the axes the majority of the outliers involve Pacific females suggesting a particular issue with how this group is modeled.



When we look at group size instead of degree we see that the size of the homogamous European partnership means the level of error is optimised when this group is matched more accurately and smaller groups are less realistically modelled. This may mean the weights are not appropriate for other ethnic groups, particularly in Auckland where there are large non-European groups.



As an alternative we tried using one set of weights in the scoring function for the European males and another set for all other males. With two sets of weights the evolutionary algorithm has a lot more combinations to try and so we only ran a preliminary test. Using the Wellington 2001 - 2006 data the single set of weights with the inter-racial degree had an error of 149 ± 1 while with two sets of weights a limited search found weight values that gave an error of 136 ± 1 . With one set the weight on the inter-racial degree term was 0.20 while with two sets it was 0.25 for non-European males and 0.10 for Europeans. This reflects the larger number of Europeans and so with greater chances to meet other Europeans the European males did not need such a high weighting on the inter-racial degree term to still maintain high levels of homogamy.



The graphs above show the contrast between the number of couples when we use the two sets of weights, one for the European males and one for everyone else, compared to using the same set of weights for everyone. In particular while the census records 11607 homogamous European couples there are 114 too few in the simulation with two sets of weights but 245 too many with only one set. This is an improvement although it does not alter the largest outlier which is the homogamous Asian couples. The census records 1494 homogamous Asian couples but the simulation with two sets of weights has 619 too few and the simulation with one set of weights has 683 too few. As the Asian group is the second largest extending the model to have three sets of weights, one for Europeans, one for Asians and one for everyone else may deal with this outlier.

4 Conclusion and Recommendations for Future Work

The inter-racial degree is a useful measure to look at the changing pattern in inter-racial partnerships while controlling for changing demographics over time. While there is a tendency towards having homogamous relationships in New Zealand this tendency differs between groups with a very high tendency for the Asian group and only a moderate tendency for the Maori group. Breaking the pattern down into ethnic groups the level of the inter-racial degree did not change much over time and any changes that did occur still maintained the order of most likely to least likely combination once group size is controlled for.

From the simulation results the static inter-racial degree performed the best which shows the relevance of distinguishing partnership matching processes between ethnic groups. One of the limitations of our simulation is that it was based on a short five year period and relied on fixing the number of couples that could form and assigning the inter-racial degree from census data. By relying on the census data for these key structures in the model it is harder to claim that the pattern simulated emerges from the partnership matching algorithm we use to score and match agents.

Future work could consider doing the simulation over a longer period of time when the dynamic updating of the inter-racial degree may have more relevance. To do this would require having data on the 28 to 40 year old couples to compare against. Alternatively the model could include all 18 to 30 year olds in the model and allow a warm up period when time is fixed during which the number of initial couples forms and the resulting inter-racial degree is used to drive the model over the next five year period.

References

- [1] Birkelund, G. and Haldal, J., 2003. Who Marries Whom? Educational Homogamy in Norway. *Demographic Research* 8(1), pp 1-30.
- [2] Bozon, M. and Heran, F., 1989. Finding a Spouse: A Survey of How French Couples Meet. *Population* 44, pp. 91 - 121.
- [3] Callister, P. and Didham, R. and Potter, D., 2005. *Ethnic Inter-marriage in New Zealand*. Statistics New Zealand Working Paper.
- [4] Epstein, J. M. and Axtell, R., 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press.
- [5] Fu, X. and Heaton, T., 2008. Racial and Education Homogamy 1980 to 2000. *Sociological Perspectives* 51(4), pp. 735 - 758.
- [6] Goodyear, R. K., 2009. *The Differences Within, Diversity in Age Structure Between and Within Ethnic Groups*. Wellington: Statistics New Zealand.
- [7] Howard, S. and Didham, R., 2005. *Ethnic Inter-marriage and Ethnic Transference Amongst Maori*. Wellington: Statistics New Zealand.
- [8] Kalmijn, M., 1998. Inter-marriage and Homogamy: Causes, Patterns and Trends. *Annual Review of Sociology* 24, pp 395-421.
- [9] Kalmijn, M. and Tubergen, van F., 2009. A Comparative Analysis of Inter-marriage of 94 National Origin Groups in the United States. Forthcoming in *Demography*.
- [10] Liu, H. and Lu, J., 2006. Measuring the Degree of Assortative Mating. *Economics Letters* 92, pp 317 - 322.
- [11] Logan, J. A., Hoff, P. D., and Newton, M. A., 2008. Two-sided estimation of mate preferences for similarities in age, education, and religion. *Journal of the American Statistical Association*, 103(482), pp. 559-569.
- [12] Perese, K., 2002. *Mate Matching for Microsimulation Models*. Technical Paper Series, Congressional Budget Office. [Online] <http://www.cbo.gov/ftpdocs/39xx/doc3989/2002-3.pdf>. [Accessed 15 May 2011].
- [13] Ryan, M., Boddington, B., and Dunstan, K., 2005. Age differences between partners at the time of marriage: 1963, 1983 and 2003. *Demographic Trends 2004*. Wellington: Statistics New Zealand.
- [14] Schwartz, C. and Nikki, L., 2009. Assortative matching among same-sex and different-sex couples in the United States, 1990-2000. *Demographic Research* 21(28), pp. 843-878.
- [15] Smits, J., 2003. Social closure among the higher educated: trends in educational homogamy in 55 countries. *Social Science Research*, 32(2), pp. 251-277.
- [16] Squazzoni, F., 2008. The Micro-Macro Link in Social Simulation. *Sociologica* 1/2008.
- [17] Statistics New Zealand, 2006. *Ethnic Groups in New Zealand* Available at: <http://www.stats.govt.nz/Census/2006CensusHomePage/QuickStats/quickstats-about-a-subject/culture-and-identity/ethnic-groups-in-new-zealand.aspx> [Accessed 3 June 2011].
- [18] Statistics New Zealand, 2006. *Quick Stats About Culture and Identity* Available at: <http://www.stats.govt.nz/Census/2006CensusHomePage/QuickStats/quickstats-about-a-subject/culture-and-identity/religious-affiliation.aspx> [Accessed 12 June 2011].

- [19] Statistics New Zealand, 2006. *Mapping Trends in the Auckland Region* Available at: http://www.stats.govt.nz/browse_for_stats/people_and_communities/geographic-areas/mapping-trends-in-the-auckland-region/education.aspx [Accessed 3 June 2011].
- [20] Statistics New Zealand, 2006. *Mapping Trends in the Auckland Region* Available at: http://www.stats.govt.nz/browse_for_stats/people_and_communities/geographic-areas/mapping-trends-in-the-auckland-region/ethnicity.aspx [Accessed 3 June 2011].
- [21] Walker, L. 2010. *Modelling Inter-Ethnic Partnerships in New Zealand 1981 - 2006: A Census-Based Approach*. Ph.D. thesis, Univeristy of Auckland.

5 Appendix: Pseudo-code

The code uses Ascape, an agent based modeling package for Java. Each agent is assigned various characteristics; age, sex, education level, ethnicity, whether they are single and who their partner is. Male agents are also given a set of female friends which starts empty and an initial threshold to determine a minimum standard while women have a set of proposals which starts empty.

Algorithm 1: Main

```
Set agent properties from file;
foreach  $m$  do
  | add 50 random females to femaleFriends( $m$ );
for  $iteration = 1:5$  do
  | if  $iteration > 1$  then
    | foreach  $male\ m$  do
      | | add 20 new random females to femaleFriends( $m$ );
    | foreach  $male\ m$  do
      | |  $m$  evaluates all  $f$  in femaleFriends( $m$ );
    | foreach  $female\ f$  do
      | |  $f$  randomly chooses a new partner from proposals( $f$ );
calculateError;
```

Algorithm 2: Male m Evaluates Female Friends f

```
foreach  $m$  do
  |  $t = \text{threshold}$ ;
  |  $\text{newInterest} = \text{false}$ ;
  | foreach  $f$  in femaleFriends( $m$ ) do
    |  $\text{score} = e^{-\sqrt{w_{age} \times \text{age}(m,f) + w_{ed} \times \text{ed}(m,f) + w_{att} \times \text{att} + w_{ir} \times \text{interRacial}(m,f)}}$ ;
    | if  $\text{score} > t$  then
      | |  $t = \text{score}$ ;
      | |  $\text{newInterest} = \text{true}$ ;
      | |  $\text{newWoman} = f$ ;
  | if  $\text{newInterest}$  then
    | | add  $m$  to proposals( $\text{newWoman}$ );
```

Algorithm 3: Female f Randomly Chooses New Partner m from Set of Proposals

```
foreach female  $f$  do
  if proposals( $f$ ) is non-empty and ( $couples < N$  or not single) then
    if single or probability  $< 0.05$  then
       $m =$ random male from proposals( $f$ );
      makeSingle( $m$ );
      makeSingle( $f$ );
      makeCouple( $m, f$ );
```

Algorithm 4: makeSingle(agent a)

```
if  $a$  not single then
   $p =$ getPartner( $a$ );
   $p.single =$ true;
  if  $a$  is male then
    simulatedCouplesEthnicity[ $a.ethnicity$ ][ $p.ethnicity$ ]  $- = 1$ ;
  else
    simulatedCouplesEthnicity[ $p.ethnicity$ ][ $a.ethnicity$ ]  $- = 1$ ;
   $p.resetThreshold$ ;
```

Algorithm 5: makeCouple(male m , female f)

```
 $m.setPartner = f$ ;
 $f.setPartner = m$ ;
 $m.setThreshold = t$ ;
simulatedCouplesEthnicity[ $m.ethnicity$ ][ $f.ethnicity$ ]  $+ = 1$ ;
```

Algorithm 6: Calculate Error

```
 $sum = 0$ ;
for  $i = 1 : 5$  do
  for  $j = 1 : 5$  do
     $sum+ =$ (simulatedCouplesEthnicity[ $i$ ][ $j$ ]-censusCouplesEthnicity[ $i$ ][ $j$ ])2;
for  $i = 6 : 13$  do
   $sum+ =$ (simulatedCouplesEthnicity[ $i$ ][ $i$ ]-censusCouplesEthnicity[ $i$ ][ $i$ ])2;
error =  $\sqrt{\frac{sum}{33}}$ 
```
