Modelling Inter-Ethnic Partnerships in New Zealand 1981-2006: A Census-Based Approach

Lyndon Walker

A thesis submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy,
The University of Auckland,

June 2010

Abstract

This thesis examines the patterns of ethnic partnership in New Zealand using national census data from 1981 to 2006. Inter-ethnic partnerships are of interest as they demonstrate the existence of interaction across ethnic boundaries, and are an indication of social boundaries between ethnic groups. A follow-on effect of inter-ethnic marriage is that children of mixed ethnicity couples are less likely to define themselves within a single ethnic group, further reducing cultural distinctions between the groups.

The main goals of the research are to examine the historical patterns of ethnic partnership, and then use simulation models to examine the partnership matching process. It advances the current research on ethnic partnering in New Zealand through its innovative methodology and its content. Previous studies of New Zealand have examined at most two time periods, whereas this study uses six full sets of census data from a twenty-five year period. There are two key components to the methodological innovation in this study. The first is the use of log-linear models to examine the patterns in the partnership tables, which had previously only been analysed using proportions. The second is the use of the parallel processing capability of a cluster computing resource to run an evolutionary algorithm which simulated the partnership matching process using unit-level census data of the single people in the Auckland, Wellington and Canterbury regions.

The European group showed a much lower rate of same ethnicity partnering than that suggested by the proportion of homogamous couples. European individuals and Maori individuals showed similar rates of same ethnicity partnering, with little change over time. The Pacific group was the only one to see an increasing tendency for same-ethnicity partnerships, whilst the rate for Asian people decreased dramatically. Individuals with dual ethnic affiliations were more likely to have a partial match of ethnicity than none at all, and there was evidence of gender asymmetry amongst some ethnic combinations. The evolutionary algorithm showed that age and education similarities were the dominant matching factors for recreating ethnic patterns. The rate of same-ethnicity and mixed-ethnicity partnerships also contributed to the matching algorithm, providing some evidence of a micro-macro link.

Acknowledgements

Firstly, I want to acknowledge the valuable advice and support of my supervisors, Professor Alan Lee and Professor Peter Davis.

I would not have been able to complete the simulation component of my thesis without the programming and grid-related support of Yuriy Halytskyy and Nick Jones at the Centre for eResearch at The University of Auckland, the simulation advice and feedback from Babak Mahdavi and David O'Sullivan, and general IT support from Stephen Cope.

I want to acknowledge the Marsden Fund for their research funding, and Statistics New Zealand for providing access to the census data I needed to complete this thesis.

I also want to acknowledge my mother for helping with the proof-reading of the final draft.

Finally, I want to acknowledge the support and encouragement of my wonderful wife Stephanie. Without her love I would only be half a person.

Statistics New Zealand Disclaimer

1. The results presented in this study are the work of the author, not Statistics New Zealand.
2. Access to the data used in this study was provided by Statistics New Zealand in a secure environment designed to give effect to the confidentiality provisions of the Statistics Act 1975.
3. I acknowledge Statistics New Zealand as the source of the Census data used in this thesis.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

In broad terms, this dissertation aims to examine and model the influences, prevalence and underlying drivers of inter-ethnic cohabitation in New Zealand through the period of 1981 to 2006. It forms part of the Modelling Social Change (MoSC) project, funded by a Royal Society of New Zealand Marsden grant and uses statistical modelling and social simulation techniques on New Zealand Census data, provided by the official statistics agency of New Zealand, Statistics New Zealand.

There are a number of elements to this research which make it unique. The statistical analysis and the simulation modelling are conducted on a complete set of unit level Census data. By comparison, local studies have had to rely on published tables of counts (Callister, 1998, 2003; Callister, Didham, & Potter, 2005), whilst overseas studies have typically had to rely on small samples of census data (Blackwell & Lichter, 2000; Qian & Lichter, 2007). The analysis methodology also extends previous research in the area. Log-linear models are used to replicate the analyses seen in the international literature for the New Zealand case and then extended through alternative parameterisations in order to gain a better understanding of the changes in inter-ethnic cohabitation.

The statistical analysis is followed by a series of abstract and empirical social simulation models which are used to examine aspects of the partnership process. These differ from the more common simulation models of partnership which tend to rely on simple fixed probability ("roll a dice") estimates for the purpose of extrapolation data (van Imhoff & Post, 1998; White, 1999; Yue et al., 2003) rather than an examination of the partnership process itself. Social simulation is also used to examine sociological processes, such as emergence, that are introduced in Section 1.6. The literature review of simulation modelling examines both microsimulation and agent-based simulation models, with aspects of each being incorporated into this study.

The unique demographic composition of New Zealand also adds to the value of the research, with the opportunity to compare the indigenous Maori population with the majority European population and the large Pacific and Asian immigrant populations. Being able to examine the interaction between indigenous, majority and new settler population groups in the New Zealand context adds to the international literature where many studies focus on two of these groups (i.e. immigrant vs. non-immigrant, black vs. non-black (America), etc). Locally, the study adds quantitative analysis to the existing body of knowledge on inter-ethnic partnerships.

This chapter introduces the idea of social stratification and homogamy and defines the key research questions for the study. It begins to address some of the details of the study such as why it is a study of cohabitation rather than just marriage and why inter-ethnic cohabitation is of interest. The chapter concludes with a summary of the content of the remaining chapters in the thesis.

1.1. Definition of Social Stratification and Homogamy

Social differences become social stratification when people are ranked hierarchically along some dimension of inequality. Members of the various layers or strata tend to have common life-chances or lifestyles and may display an awareness of common identity, and these characteristics further distinguish them from other strata.
(Abercrombie, Hill, & Turner, 2000)

Parsons (1954) describes social stratification as "*the differential ranking of the human individuals who compose a given social system and their treatment as superior and inferior relative to one another in certain socially important respects*". The idea of social stratification is that groups within society tend to form some kind of hierarchy which is indicated by the predominant homogamy of partnerships in most settings (Blackwell, 1998).

The primary market of social hierarchy is the distribution of scarce and desired resources such as income, status and power, and homogamy is also reflective of that (Becker, 1973). The degree of social stratification in a society is an indicator how well various groups interact, with higher levels of segregation indicating a reluctance for the groups to interact and suggesting that the groups do not view one another as equals. Stratification can occur across numerous dimensions and may be affected by factors such as religion, socio-economic status and education. The focus of this study is an examination of stratification by way of ethnic difference.

By modelling the cohabitation patterns of the New Zealand population over time, we can observe any changes in the social stratification of the population with regards to ethnicity. This can be seen as a change in the social distance between the groups of interest. For example, if there is an increased rate of partnering between the European people and Asian people over the census periods, then this could be viewed as a decrease in the social distance between these two ethnic groups.

Social homogamy refers to a level of similarity between two or more parties. In this case the focus of the study is on the homogamy of cohabitating couples with regard to ethnicity. Homogamy can be considered in a binary sense; where the partners have the same ethnicity (homogamy) or they have different ethnicities (heterogamy). These terms are sometimes referred to as endogamy and exogamy, particularly within the area of Anthropology. Homogamy can also be extended from a binary measure to one with varying degrees of closeness. For example, a couple may have exactly the same ethnicity or ethnicities, or they may share a partial match of ethnicity, or they may have no common ethnicity.

1.2. Central Research Questions

This examination of the inter-ethnic cohabitation in New Zealand will focus on the following two questions:

1. What changes can be seen in inter-ethnic cohabitation patterns in the period 1981 to 2006?
2. What factors and/or social processes, other than ethnicity, influence the patterns and matching process of ethnic partnership formation?

These research questions provide the basis for the study, looking at the "what has changed" and the "how/why has it changed" of New Zealand's social structure and the pattern of inter-ethnic cohabitation.

Question one focuses on a description of the data using tables and proportions, followed by odds ratios and log-linear models. By examining and describing the changes in the data, inferences can be made about whether the distribution of inter-ethnic cohabitation and the choice of cohabitation partners in New Zealand have become more highly stratified and segregated over the period 1981 to 2006.

Question two uses a combination of logistic regression and social simulation routines to look at how the patterns in inter-ethnic cohabitation can be modelled and what factors may influence them. The simulation routines provide the opportunity to examine abstract and empirical sociological models of partnership choice and examine social processes such as emergence and downward causation in order to identify the key underlying mechanisms. These mechanisms focus on the linkages between the micro-level and macro-level, where changes in the behaviour of the agents (micro-level) impact on the society as a whole (macro-level), which in turn affect the decisions of future individual agents in the system. This "feedback loop" is at the core of the simulation investigation.

1.3. Why is Inter-Ethnic Cohabitation and Marriage of Interest?

"*… what makes intermarriage sociologically relevant lies in its inherent dynamic: It is not just a reflection of the boundaries that currently separate groups in society; it also bears the potential of cultural and socioeconomic change."* (Kalmijn, 1998)

Inter-ethnic marriage demonstrates the existence of interaction across ethnic boundaries and is an indicator that cohabiting individuals from different ethnic groups consider each other to be social equals (Kalmijn, 1998). A follow-on effect of inter-ethnic marriage is that children of mixed ethnicity couples are less likely to define themselves within a single ethnic group, further reducing cultural distinctions and social boundaries between the groups (Stephan & Stephan, 1989). This can become a recursive process, with the cohabitation decisions of one generation helping to influence the next. As the rate of inter-ethnic marriage increases it helps to normalise it, which will in turn see it become more prevalent.

It is also important to note that inter-ethnic partnerships may be an indicator of a lack of availability of potential partners of the same ethnicity (Blau, 1977). This is an effect due to relative group size and can be referred to as a demographic effect, where patterns of inter-ethnic marriage vary due to the numerical availability of potential partners rather than changing preferences or social norms. This distinction is addressed in the analyses that are able to identify patterns that are independent of the marginal distributions of ethnicity. This separates out the purely "demographic effects" in the data. Arguably this separates the numerical "availability" (or opportunity) from "attraction" (or normative) components of cohabitation patterns.

Gaining an understanding of the micro and macro social processes that drive partnership choice will help provide an insight into the way in which society is shaped. The simulation of partnership choice through these processes will give an indication as to the future structure of New Zealand society. As well as academic interest, this will be of interest to local and central government planners, demographers, market research groups, epidemiologists and anyone else who is directly or indirectly concerned with how the social structure of New Zealand will look in the future.

1.4. Cohabitation or Marriage

The focus of this thesis is on cohabitation rather than marriage. Cohabitation patterns form a basic unit of social structure. They can also be seen as an indicator of the association between social groups. Whilst cohabitation is often viewed as a weaker form of relationship than marriage (Manning & Smock, 1995; Schoen & Weinick, 1993), it has become a common substitute for marriage in New Zealand, particularly for first unions and amongst younger couples (Dharmalignam, 2002). Schoen and Weinick (1993) write that cohabitations are seen as “informal marriages where, in large part, couples follow established patterns of behaviour”. If this study was to only examine new marriages, it would miss many equivalent, newly formed cohabiting relationships. Therefore, to best capture the social processes which generate the social hierarchy in New Zealand, the population of interest is defined as couples who are married or living in a de-facto relationship at the same residence.

1.5. Census Data as a “Test Bed” for Inter-Censal Change

The New Zealand Census of Population and Dwellings is recorded every five years and provides cohabitation and demographic information about the entire population. Analysing this data provides the opportunity to model the entire population rather than just a sample, which otherwise tends to be the case in most of the published literature. The six sets of census data provide a series of cross-sectional snap-shots of New Zealand. However, the data can be thought of as a kind of perturbed time-series which allows for the examination and analysis of inter-census changes over time. It also provides the opportunity to test and validate empirical propositions. Although individuals can’t be tracked through time, changes in cohorts can be. For example, the eighteen to thirty year-old cohort in one census will become the twenty-three to thirty-five age-group in the following one.

Statistics New Zealand is bound by the Official Statistics Act and so in providing the data must ensure that certain confidentiality requirements are met. Although the data is relatively complete in the sense that it includes the whole population of New Zealand, it does have some shortcomings. These are discussed in Chapter 4.

1.6. Sociology, Statistics and Simulation

The Modelling Social Change project lies at the intersection of a number of disciplines, combining theories of sociology with statistical techniques and computer-based simulation methodologies. This requires consideration of the concepts and theories from each of the disciplines. In order to answer the research questions, this dissertation considers the relevant material from the fields of Sociology, Statistics and Simulation.

Statistical analysis provides us with the ability to answer the first research question; that is, whether the distribution of inter-ethnic cohabitation is changing over time. Regression techniques are used to examine how variables such as age and education affect the likelihood of people having inter-ethnic partnerships. However, they do not provide an in depth insight into the "why" and "how" questions; that is why, and how, have such changes occurred in the way described.

This is where simulation provides further answers. Simulation can be used to model complex scenarios that cannot easily be addressed analytically. Simulation provides the opportunity for abstract and empirical models to be examined to see how various parameters and roles can affect inter-ethnic cohabitation rates and what effect social processes such as emergence and downward causation have on those rates.

The social phenomenon of emergence can be thought of as the creation of a macro-level process or pattern which "emerges" from the culmination of potentially more complex micro-level processes (Sawyer, 2005). In the case of cohabitation patterns, the simple macro-level occurrence would be the changing structure of inter-ethnic partnerships which would come about through the complex micro-level interaction and subsequent partnering of individuals. Downward causation is the theory which suggests a top-down model of causation and is sometimes referred to as macrocausation. If the macro patterns of partnership are caused by the interaction of the micro-level choices of individuals, and then the choices of the individuals are influenced by the macro environment in which they exist, then this would be an instance of emergence followed by downward causation. This is a special case of what is studied in Sociology as a “micro-macro linkage” (Coleman, 1990).

While conducting the statistical and simulation modelling it is vital to consider the relevant sociological theory detailed in the literature review, so that the results are based on both the data and the sociological theory.

1.7. Analysis and Modelling

Two-way frequency tables provide a good starting point for the analysis and modelling of partnership data with homogamy as a special case. They are a convenient way of displaying the frequencies of categorical data - in this case the number of cohabitating couples - and provide the ability to examine partnership patterns, including homogamy, amongst matched pairs of data representing attributes of cohabitating couples.

Table 1.1 shows a trivial hypothetical table which matches the eye colours of couples. The diagonal of the table shows couples who have the same eye colour (i.e. are homogamous) and the off-diagonal entries show the heterogamous couples (i.e. those who have different coloured eyes from one another). This table is used as a sample case for illustrative purposes to demonstrate the analytical techniques on fictitious data, prior to applying them to the “real-world” ethnicity data from the census.

Husband's Eye Colour	Wife's Eye Colour			
	Blue	Brown	Green	Total
Blue	15 (0.15)	3 (0.03)	2 (0.02)	20 (0.20)
Brown	3 (0.03)	18 (0.18)	9 (0.09)	30 (0.30)
Green	12 (0.12)	9 (0.09)	29 (0.29)	50 (0.50)
Total	30 (0.30)	30 (0.30)	40 (0.40)	100

Table 1.1 - Hypothetical two-way frequency table of eye colour

In addition to displaying the counts of the different combinations, two-way tables can be used to generate descriptive statistics, such as the proportion of couples in the diagonal cells of the table. For example, Table 1.1 shows a strong degree of homogamy amongst the couples, in the sense that a large proportion of the couples share the same eye colour. Beyond this, it also shows that the proportion of men with green eyes who have a partner with non-green eyes is greater (21/50) than the corresponding proportions of blue-eyed (5/20) and brown-eyed (12/30) men with non-blue eyed or non-brown eyed partners. However, it is also evident that these proportions are likely to be affected by the marginal (row and column) distributions since the proportion of women with blue eyes or brown eyes is greater than the proportion of men with blue eyes or brown eyes. This suggests that care must be taken in interpreting simple proportions of this kind without taking account of the marginal distributions.

More sophisticated inferences about the data can be drawn by applying advanced statistical methods. One of the main tools is log-linear modelling, a technique that is commonly used to model cell frequencies. This technique has the advantage of generating analytic results that are invariant to the marginal distributions. Log-linear models and their extensions are discussed further in Chapter 2 and 4.

The simulations are a way of exploring how different decisions at the individual level might affect various summaries of the population, such as the two-way table mentioned above. As the agents in the simulation apply the empirical or theoretical rules of partnership that they are programmed with, they will shape two-way tables of their virtual world. At the end of each simulation run, the pattern of partnerships from the simulation can be compared to those that actually occurred. These simulations are approached in two different ways.

The first is following an abstract methodology, where the agents produce decisions based on some sort of theoretical criterion. Although this is an abstract simulation using an artificially generated population, it provides insights into the behaviour and outcomes of a real-world population that might be subject to similar constraints and decision rules.

The second approach is more data driven, with the decisions of the agents being based on patterns and probabilities extracted from the data. Ideally, this approach should complement the first, but also provides greater relevance since it refers to the New Zealand population rather than a generic or artificial one.

1.8. Simulation and Parallel Computing

One of the other methodological steps forward in this thesis is the use of parallel processor computing to run an evolutionary optimisation algorithm. The Auckland cluster of the BeSTGRID (www.bestgrid.org.nz) computer network was used to run the simulation multiple times in parallel in order to use an evolutionary algorithm to search the parameter space of the simulation scoring function. Although evolutionary algorithms are not a new concept, they are typically applied to engineering problems rather than social data. The evolutionary algorithm and the BeSTGRID computer system are described in detail in Chapter 7.

1.9. Introduction to the Chapters

Following on from this chapter, Chapters 2 and 3 examine the literature that is relevant to this study. This encompasses articles about homogamy and cohabitation from sociological sources, together with literature about the statistical methods and simulation methods which are used. Chapter 4 provides a discussion of the strengths and weaknesses of using the Census data for the project. The construction and selection of the variables and their inherent advantages and disadvantages are also explained. This is followed by a set of descriptive and analytical statistics using graphs and tables, followed by log-linear models and logistic regression in Chapter 5. Chapters 6 and 7 will detail the simulation side of the study, first with abstract simulation models and then with ones that are empirical. This leads to a concluding section, which in turn is followed by appendices, providing the partnership cross-tabulations by ethnicity and the computer code for the analysis and simulations.

Chapter 2 – Literature Review: Sociology and Statistics

To examine the literature relevant to this thesis it is important to draw upon a number of threads from different fields which all relate to the project. This literature review aims to encapsulate the sociological, statistical and computer simulation literature relevant to the study.

There is a significant set of literature in the field of sociology which discusses social patterns of marriage, particularly the subject of inter-marriage across education, religion, occupation and ethnicity. The literature in this area can be roughly divided into two tendencies – theorists and empiricists; that is, there are those authors who use an explicit theoretical framework with illustrative data and there are those who are less theoretically oriented and rely instead on a more quantitative approach with descriptive statistics and statistical modelling. In order to fully discuss the social relevance of findings from this study, it is important to consider both the quantitative findings and the underlying social theory.

The quantitative side of the project has two facets to be reviewed. To examine the changes in stratification of New Zealand society and identify relevant demographic variables that may have led to these changes, traditional statistical techniques can be applied. Whilst these kinds of techniques may provide sets of probabilities which describe how the structure of society is changing, they don't provide any explanation of how and why these changes are occurring. Social simulation provides a tool to examine this idea of underlying processes that generate the changes in the societal structure.

This chapter draws together these threads, examining the methodology and findings of national and international partnership literature and discussing how it relates to the study of inter-ethnic cohabitation in New Zealand.

2.1. Sociological Literature

The sociological literature on the subject of cohabitation and homogamy spans long standing theoretical models of how various social processes occur, through to more recent applied studies which focus on descriptive statistics and data analysis. This section reviews the sociological literature about ethnic homogamy and contrasts homogamy research in New Zealand with the rest of the world. Homogamy and social stratification can be observed across numerous demographic attributes, of which ethnicity is just one, so, following a discussion of emergence and the micro-macro link, literature on religious, educational and occupational homogamy are also discussed.

2.1.1. Social Patterns of Marriage

"In modern sociological research marriage patterns are regarded as the result of an interplay between three social forces: the preferences of the individual for a specific marriage candidate, influences from a person's own social group, and structural constraints in the marriage market."
(Bull, 2005)

Early work on the theories of social structure and marriage patterns such as those by Robert Merton and Talcott Parsons tended to focus on theories rather than analysis. Merton (1941) wrote about how interclass marriage acts as a means of social mobility. He states that social norms affect the degree and type of social contact one has with other groups, whether it be race, religion or class, and "non-normative conditions" will affect the proportions of intermarriage. This concept suggests that one is more likely to marry/cohabit with those who are similar since their social network is more likely to be made up of people like themselves. This idea was echoed by Parsons (1954) who discussed the idea of "occupational hierarchy" and affluence as factors that influenced the structure of an individual's social network and their subsequent partnership choices. During this time, anthropologists were also writing about homogamy, although they used the terms endogamy (within group marriage) and exogamy (out of group marriage) (Levi-Strauss, 1969). These articles led the way for more specific studies which examined partnership similarity by attributes such as religion and education rather than generalising across multiple attributes.

The release of Peter Blau's book, *Inequality and Heterogeneity: A primitive theory of social structure* (Blau, 1977), reignited interest in the area of homogamy and partnership choice. In particular, Blau introduced a new macrosociological theory of social structure which discussed how the demographic make up of an individual's social network would place constraints on who they were actually able to choose as a partner, despite what their personal preferences might be. These social constraints of the decision making process in partnership choice form an important part of what will be modelled in the New Zealand context, helping to provide a crucial link between the micro and macro effects that take place.

Merton, Parsons and Blau are still regularly referenced in current work in the area of homogamy (Best, 2005), but it is now more common to see empirical evidence accompanying the sociological theories (Kalmijn, 1993; Qian & Lichter, 2001; Schwartz & Mare, 2005). Although the analysis of the New Zealand data considers the ideas of opportunity and preference that are described in the earlier sociological literature, the focus of the research is on the empirical findings rather than sociological theorising.

2.1.2. Ethnicity and Marriage Patterns

Nationally and internationally there is a considerable body of literature examining ethnicity and marriage. Although this study examines cohabitation rather than marriage, these studies are still highly relevant. From a statistical perspective the studies can be divided into those that rely on descriptive statistics and tables of proportions, and those that go beyond the use of descriptive statistics, adding strength to their inferences through the use log-linear models (discussed in Section 2.2.1).

American articles dominate the literature, particularly with studies that focus on patterns of marriage cutting across the white majority and one or more other minority ethnic or racial groups. These have traditionally examined the extent of marriage across the "colour"/racial line of blacks and whites (Crowder & Tolnay, 2000; Kalmijn, 1993; Kelly Raley, 1996; Tucker & Mitchell-Kernan, 1990). However, recent immigration patterns have seen more articles on the contribution of those of Asian heritage to the dynamic of ethnic patterns of marriage (Aguirre, Saenz, & Hwang, 1995; Hwang, Saenz, & Aguirre, 1997; Tzeng, 2000). Bratter and Zuberi (2001) expanded their study to incorporate intermarriage between the white majority and Asian, American Indian, Hispanic and African American groups. They used American Census data from 1960 to 1990 to examine how the levels of racial diversity and interracial marriage have changed over that time and found that the changes varied greatly between ethnic groups.

Similar work, particularly of a descriptive nature, has been seen in Britain (Muttarak, 2003), China (Stephan & Stephan, 1989), Australia (Parimal & Hamilton, 2000; Roy & Hamilton, 1997), Fiji (Richmond, 2003) and Germany (Klein, 2001). These studies have all primarily looked at intermarriage between one particular minority group and the majority ethnic group through two-way frequency tables and the graphs/tables of the proportion of intermarriage.

Jones (1991) examined Australian inter-ethnic marriage data from 1950 through to 1982 and applied log-linear models rather than indices to examine the change in rates of inter-ethnic marriage over time. He advocated log-linear models since these separated the marginal effects and provided tests of significance, features absent from previous research based on index models. Giorgas and Jones (2002) followed up this work in Australia with another study of intermarriage, this time focussing on second and third generation European immigrants. They found that the rate of intermarriage increased over the generations and argued that social factors such as age, residence and language strongly shaped patterns of intermarriage.

More recently, Khoo *et.al.* (2009) and Heard *et.al.* (2009) have examined intermarriage in Australia. Khoo *et.al.*, like Giorgas and Jones (2002) before them, examined intermarriage by ancestry, examining first-, second-, and third-generation Australians.

They also found that the rate of intermarriage increased over the generations, although for more recent migrant communities, particularly those from Asian and Africa, the second generation are not yet of marriageable age and there is no third generation. Heard *et.al.* took a slightly different approach, examining intermarriage between indigenous and non-indigenous Australians. They found lower levels of exogamy in non-metropolitan areas, possibly due to limited opportunities for social mixing. Exogamy was also more likely to occur amongst Indigenous individuals (both male and female) with higher levels of education and male partners with higher levels of income, suggesting an association between exogamy and upward mobility.

Working with data from the 1979 Current Population Survey of America, Alba and Golden (1986) found that the relative size of each ethnic group influenced the rate of intermarriage, with smaller non-European ethnic groups tending to have higher in-marriage rates. The number of ethnic groups which a person identified with was also shown to have an influence, with people who identified with more than one ethnic group being less likely to form a homogamous partnership. This is highly relevant to the New Zealand Census data, as in several of the censuses people have been able to select multiple ethnic affiliations.

Kalmijn (1993) used log-linear analysis to examine factors which increased or decreased the likelihood of black-white intermarriage in America. He found that education and socioeconomic status both correlated with intermarriage patterns, and that it was much more common for the white partner in a mixed-marriage to be "marrying up" than it was for the reverse. Kalmijn published a follow up paper (1998) which reviewed literature on intermarriage, and again found education to be a factor in intermarriage patterns. It summarised the findings of several other papers that had found that people of higher education levels were more likely to intermarry. Kalmijn theorised that this could reflect both opportunity (particularly interaction at colleges) and preference (a more universalistic view of life and other ethnicities).

Harris and Ono (2005) took a different approach from many previous authors when examining interracial marriage. They suspected and subsequently demonstrated, that previous log-linear modelling at a national level was biased due to strong regional differences in marriage markets and rates of interracial marriage. New Zealand has vastly differing levels of ethnic diversity across the country, particularly when comparing Auckland to much of the South Island. Therefore, where possible it is important to consider the impact of geography on partnering opportunities when working with New Zealand data.

In the same piece of work Harris and Ono (2005) also discussed the importance of differentiating between prevalence rates and incidence rates. This is echoed by Schwartz and Mare (2005), who highlighted the fact that the area of interest is in the formation of new partnerships (incidence) rather than existing ones (prevalence). For New Zealand ethnicity data this is particularly important as the prevalence rates show that European/European partnerships dominate all others. The solution in both the Harris and Ono and Schwartz and Mare studies was to use the partnerships of young people, for example eighteen to thirty year-olds, as a proxy for new relationships (since actual relationship duration data was not available).

A common thread through the ethnic intermarriage articles was the use of log-linear models to summarise the frequency table patterns. Section 2.2 will introduce the types of log-linear models that will be used for this study and relate them back to previous research in the area. Another important idea taken from the articles was the distinction between incidence and prevalence rates of inter-ethnic partnerships. By focussing on new partnership formations only, a clearer picture of the current "marriage market" can be established, whereas if all partnerships were considered throughout the analysis, the figures would be dominated by the large proportion of established European-European partnerships. This creates a focus on the current partnering patterns and dynamics rather than the historical ones.

2.1.3. Research in New Zealand

Research on inter-ethnic partnerships in New Zealand has been largely of a historical (Anderson, 1991; Riddell, 2000) or descriptive nature (Callister, 2003, 2004; Callister & Blakely, 2004; Callister, Didham, & Potter, 2007; Howard & Didham, 2007), compared to the more analytically advanced international articles. As with the international articles, the focus of most studies has been on marriage rather than cohabitation.

Callister (2003) provided basic descriptive statistics of Maori/non-Maori intermarriage from the 1996 New Zealand Census and determined that about half of partnered Maori had non-Maori partners. This rate of intermarriage declined to one third of partnered Maori once only those of a single ethnic group were considered. This introduces one of the issues that will be discussed later in the thesis, regarding the definition of ethnicity. Each Census has used a different method for collecting ethnicity and defining ethnic groups. This becomes particularly problematic with the Maori ethnic group, as provision is provided for both ancestral and self-definitions of ethnic affiliation. In addition, by not controlling for the marginal distributions (group sizes), the rate of intermarriage for Maori will appear higher than that of Europeans due to the smaller relative size of the group.

Howard and Didham (2007) followed up this research by examining Maori/non-Maori marriages in the 2001 census. This study delved deeper into the definition of ethnicity, seeking to explain the high rates of intermarriage involving Maori as being due in part to ethnic transference and the changing definition of what it means to be Maori. Again this used frequencies and percentages and so did not try to control for the relative sizes of the ethnic groups.

This examination of intermarriage in New Zealand was expanded upon by Callister, Didham and Potter (Callister et al., 2007) to include the ethnic groups of Pacific, Asian and Other. The analysis of these groups remained descriptive and only used data from the 2001 census, providing a snapshot in time but no information about how the rates of intermarriage for the different ethnic groups were changing over time. This thesis extends Callister *et.al.*'s work by examining six sets of census data in order to examine changes over time. It also extends the analysis from simple proportions to log-linear modelling in order to separate the effects of the marginal distributions from changes in marital association. Callister's paper also provided some discussion on the problem of individuals with multiple ethnicities, coding separate groups for single and multiple ethnicities, which are further discussed in Chapter 4.

The only real use of analytical statistics in the local literature was the application of a log-linear model in another Callister article (1998) which uses census data from 1986 to 1996 to examine educational assortative mating patterns in New Zealand. It concluded that people were more likely to marry someone of similar educational attainment. If Kalmijn's (1998) findings on the relationship between education and ethnic patterns of marriage are true for the New Zealand population, then this would indicate that there is potential for education to be a confounding factor in the analysis of ethnic patterns.

Examining the New Zealand literature, we find that log-linear models are not as prevalent as in the equivalent international studies, with most studies relying on the analysis of proportions of inter-ethnic partnership. The focus of earlier studies tended to be on Maori/non-Maori partnerships, in part because of the historical relevance, but also in part due to the difficulties in establishing reliable ethnic groupings. One of the key points taken from the review of the New Zealand literature is the discussion surrounding the classification of ethnicity and the inherent difficulties involved.

2.1.4. **Emergence and the Micro-Macro Link**

Agent-based simulation is often used to demonstrate emergent patterns amongst a set of individuals (Macy & Willer, 2002; Sawyer, 2000; Silverman & Bryden, 2007). Simão and Todd (2003) used agent-based simulation to simulate emergent patterns of mate choice in a small artificial population, but the concepts that they discuss have equal validity when examining a real population. Agent-based simulation is discussed further in Section 3.1.

The micro-macro link is an extension of the idea of emergence. It is the idea that the relationship between the micro and macro levels of a social process are actually recursive, where the actions of individuals at the micro level are influenced by existing macro structure, but they in turn will impact on the future macro structure, which will affect future micro decisions. A schema for this relationship was drawn up by Coleman (1990) and is commonly referred to as "Coleman's Boat". An example of the Coleman's Boat can be seen in Figure 2.1.

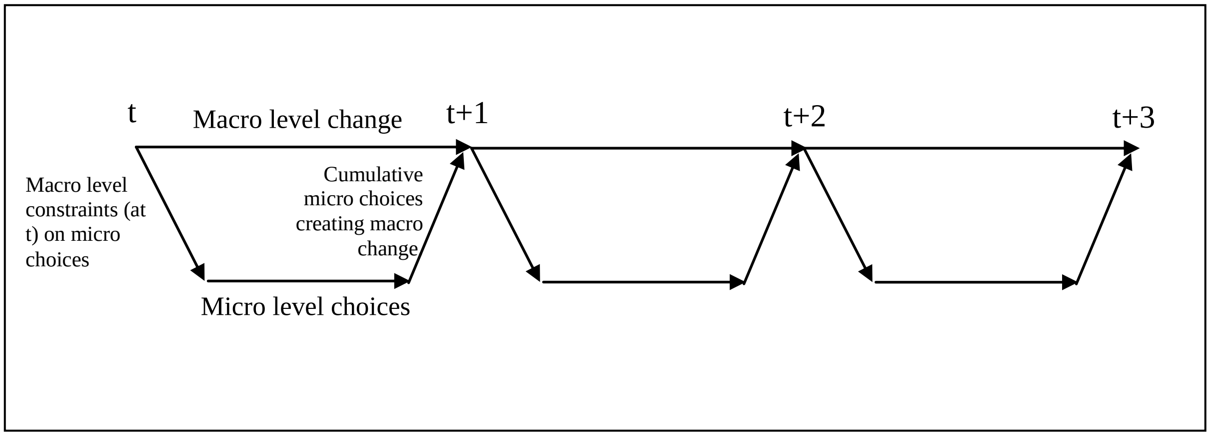

Figure 2.1 - Coleman's Boat

In the case of partnership formation in New Zealand, a micro-macro link may exist whereby the partnering choices of individuals are influenced, or possibly even constrained, by what they observe in the society around them. This macro effect may influence who they choose as a partner, which will in turn shape the future structure of the society and its influence on the following generation in the marriage market. Although this is difficult to test through traditional statistical models, recursive computer simulation models[1] allow us to produce models with feedback loops that can be used to examine possible micro-macro patterns.

[1] A recursive function is one whose current state is determined by referring to its previous states.

2.1.5. Educational Homogamy

The link between education and ethnic patterns of marriage in New Zealand and American data was mentioned in Sections 2.1.2 and 2.1.3. Given the apparent relationship between the two variables, it seems prudent also to examine the international literature on educational homogamy. As with ethnicity, the literature is dominated by American studies, although one article (Smits, 2003) does provide an inter-country comparison.

Mare (2001) used US Census data from 1940 to 1987 in order to examine whether the level of educational homogamy in the US had changed over that period. He documented increasing levels of homogamy, attributing it in part to rising levels of educational attainment over that period. The study did not look for the factors that may have been driving the changes, preferring to limit the analysis to a description of changing levels of homogamy.

A more detailed examination of this data was conducted by Schwartz and Mare (2005) using data from 1940 to 2003. Using log-linear models they found that, whilst inter-marriage between those in the middle section of their educational distribution increased, this was overshadowed by increased homogamy within two groups – those with low education and those with high education. They theorised that growing economic inequality occurring across the education levels over the last 50 years could be contributing to increased homogamy at the extremes of the educational distribution. In particular, it was argued that an increasing amount of time between graduation and marriage contributed significantly to high marital homogamy for those with college degrees (Schwartz & Mare, 2005).

Smits (2003) also examined educational homogamy, although he narrowed the education levels down to "highly educated" and other. The study included data from 55 different countries and found that in most countries rates of marriage within education levels was generally high. However, there were lower rates of educational homogamy in Protestant countries, countries with higher rates of higher education, and also among younger age groups.

The lowest rates of intermarriage between those of different levels of educational attainment were recorded in predominantly Catholic, Muslim and Confucian countries (Smits, 2003). These findings suggest that the examination of ethnic homogamy in this thesis should consider the impact of education and religion as social factors that may condition ethnic-specific patterns. However, Smits did show New Zealand to have one of the lowest rates of educational homogamy in the world once size, dominant religion and modernisation had been accounted for.

The findings of the Schwartz and Mare (2005) and previously discussed Kalmijn (Kalmijn, 1998) articles would suggest that when analysing New Zealand ethnicity, educational attainment should be included in the analysis. Although New Zealand has a relatively low rate of educational homogamy compared to the rest of the world (Smits, 2003), Callister's (1998) (see Section 2.1.3) findings on educational homogamy indicate that there are education-related patterns of partnership. These should therefore be considered whilst examining the patterns of ethnicity.

2.1.6. Religious Homogamy

Religious affiliation has been shown to correlate with marital homogamy in other attributes such as ethnicity and education (Smits, 2003). To this end we must consider trends in religious homogamy and the effect of religion on ethnic patterns.

Robert Johnson's book, *Religious Assortative Marriage in the United States* (1980), was one of the early analytical works in the area of religious homogamy. The study used basic log-linear models to measure religious homogamy among married couples in the US. Since the decennial US census does not include a question on religion, the study had to rely on a disparate selection of social and regional surveys for data. Johnson concluded that religious homogamy in the United States had declined over the 1960s and

1970s. However, inconsistencies in the data sources made it difficult to control for other variables.

More recent American studies have still had to rely on survey rather than census data, but have tended to use more consistent longitudinal sources. Religion has been shown to still generate high levels of marital homogamy, particularly within the Catholic faith and other more conservative sects (Sherkat, 2004). Kalmijn (1991) found increasing rates of intermarriage between Catholics and Protestants in the United States since the 1920s, but reported that there was still a strong tendency towards homogamy (i.e. within-religious group marriage).

Internationally, studies in England (Voas, 2003), Ireland (O'Leary, 2001) and Australia (Dempsey & De Vaus, 2004; Hayes, 1991) have all found similar results; that is, strong initial religious homogamy, but decreasing over time.

Sherkat (2004) used a probit regression model, rather than the more standard log-linear models, to examine partnership data and found that sharing proximate geographic location and similar educational attainment both increased the likelihood of intermarriage across religious boundaries.

Although no quantitative studies of religious homogamy have been conducted in New Zealand, the decreasing tendency towards religious homogamy seen internationally, and the moderate religious views of most New Zealanders (Statistics New Zealand, 2006a), would suggest that religion might not play a large part in partnering patterns.

2.1.7. Occupational Homogamy

Marriage patterns across occupation and socio-economic levels has been examined by a number of authors. Some focus on marriage (Kalmijn, 1994; Rytina, Blau, Blum, & Schwartz, 1988; 2003; Smits, Ultee, & Lammers, 1999), whilst others focus instead on inter-generational social mobility. These typically focus on the idea of "marrying up" in order for an individual to gain upward social mobility.

The argument for marriage as a means of social mobility takes its roots in the earlier work based on class structures and is commonly based on an economic approach towards analysing marriage patterns. See (Becker, 1973, 1974; Duncan, 1979; Thomas & Sawhill, 2002; Ultee & Luijkx, 1990; Waters & Ressler, 1999; Xie, Raymo, Goyette, & Thornton, 2003) for a selection of economic-based arguments for marriage patterns. The increasing participation rates of females in tertiary education (Ministry of Education, 2007) and the labour force in general (Statistics New Zealand, 2007b) tend to weaken the economic arguments that marriage can be treated as a path towards upward social mobility for women, particularly across ethnic boundaries.

Social patterns of marriage can be examined across numerous demographic variables but the main focus of this thesis is on ethnicity, using other factors as a way of conditioning and clarifying the patterns. The way in which people form partnerships with those of the same or different ethnicities provides insight into the ethnic dimension of New Zealand's social structure and the degree of interaction between ethnic groups.

2.2. Statistical Methodology

In order to examine how the rate of inter-ethnic partnerships is changing over the 25-year period and to determine which variables have a significant effect on the probability of entering into an inter-ethnic partnership, log-linear and logistic regression techniques are used.

2.2.1. Log-Linear Models

Over the last ten years there has been a shift away from the descriptive reporting of the frequencies of mixed-marriage partnerships, with researchers now tending to make use of analytical log-linear models instead (Harris & Ono, 2005). This trend is seen across local (Callister, 1998) and international studies (Blackwell & Lichter, 2000; Mare, 2001; Qian & Lichter, 2007).

Log-linear models provide a number of advantages over a solely descriptive approach to analysing the data. The analysis removes the effect of the marginal distributions (row and column totals), which means that the variation in the relative sizes of the ethnic groups is removed from the analysis of the rates of homogamous partnership between ethnic groups. Uunk et.al. (1996) describe this as disentangling the effects of the marginal distributions (i.e. structural homogamy) and relative chances of marital association (i.e. relative homogamy). The parameters of the model allow researchers to quantify the different effects that are present within the data. Leo Goodman, who is considered to be one of the founding fathers of log-linear models for social categorical analysis, provided a non-technical review of log-linear models and other categorical analyses in the 2007 Annual Review of Sociology (Goodman, 2007).

Log-linear models can also be parameterised in numerous different ways. Parameters can be added to measure different effects and components of the table. For example, this allows a more precise examination both of homogamy, using diagonal dominance parameters, and patterns of heterogamy via crossings parameters.(Mare, 2001; Qian & Lichter, 2007)

The two-dimensional case of a log-linear model, as defined by Agresti (2002), is as follows:

A multinomial[2] sample of size n on an I x J contingency table[3] takes probabilities π_{ij} for the joint distribution of the two categorical responses.

These are independent if: $\pi_{ij} = \pi_{i+}\pi_{+j}$ where $i = 1, \dots, I; j = 1, \dots, J$.
The expected frequencies under independence are: $m_{ij} = n\pi_{ij} = n\pi_{i+}\pi_{+j}$
which take the additive form: $\log m_{ij} = \log n + \log \pi_{i+} + \log \pi_{+j}$
where π_{i+} represents the probability for row i and π_{+j} represents the probability for column j[4].

This is commonly represented by the log of the cell frequencies, m_{ij} being modelled by the base category value μ and the row and column effects, λ_i^X and λ_j^Y:

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y$$

where

$$\lambda_i^X = \log \pi_{i+} - \log \pi_{1+} \quad \text{and} \quad \lambda_j^Y = \log \pi_{+j} - \log \pi_{+1}$$

This method of defining the parameters is known as the "corner constraints" method and is the default parameterisation in the R software programme (http://www.r-project.org). Under this parameterisation $\lambda_1^X = \lambda_1^Y = 0$ and μ represents the row one, column one cell estimate.

This composition is known as the *log-linear model of independence* (Agresti, 2002) and can be extended for tables of more than two dimensions. It treats the row and column effects as being independent of one another and is mathematically equivalent to

[2] A discrete probability distribution used for random categorical data (generalisation of the binomial distribution).
[3] Joint frequency distribution of two or more categorical variables.
[4] The + notation indicates a total, i.e. *i*+ is the total for row *i* across all of the columns.

estimating the cell frequencies by calculating the product of the row and column totals for that cell and then divided by the overall total.

	Wife's Eye Colour			
Husband's Eye Colour	Blue	Brown	Green	Total
Blue	12	2	5	17
Brown	7	23	3	33
Green	4	6	22	32
Total	23	31	30	84

Table 2.1 - Hypothetical example table

For example, using the hypothetical table above we could estimate the frequency of couples who both have green eyes, assuming independence, as (32x30)/84 = 11.43, which is a little over half of the actual frequency. Using the standard parameterisation of the independence model, and treating the blue/blue combination as the base cell frequency, results in the set of coefficients shown in Table 2.2. The Z and p-values in the table represent statistical tests to examine whether each of the parameters are significantly different from zero.

	Coefficient	**Std Error**	**Z-Value**	**P-Value**
Intercept (base cell)	1.6491	0.2902	5.683	0.0000
Husband brown effect	0.5521	0.2880	1.917	0.0552
Husband green effect	0.5213	0.2896	1.800	0.0719
Wife brown effect	0.2985	0.2752	1.085	0.2781
Wife green effect	0.2657	0.2771	0.959	0.3377
Null deviance	49.926			
Residual deviance	43.869			

Table 2.2 - Coefficients for independence model: Hypothetical example

The model parameters are calculated using a method known as Maximum Likelihood (ML) estimation. They represent the "most likely" set of parameters, given the observed data. The model can also be assessed as a whole, using the residual deviance (defined and discussed further in Section 5.2). Table 2.2 shows that the residual deviance (total error of the model) is lower than the null deviance (total error when no variables are used), indicating that the independence model is an improvement over just using a constant and no variables, i.e. assuming all of the cell probabilities are the same.

The estimated frequency for the number of couples who both have green eyes can also be calculated using these parameters:

$$\log m_{green,green} = \mu + \lambda_{green}^{husband} + \lambda_{green}^{wife}$$
$$\log m_{green,green} = 1.6491 + 0.5213 + 0.2657$$
$$m_{green,green} = e^{1.6491+0.5213+0.2657}$$
$$m_{green,green} = 11.43$$

Although this result could be determined through simple arithmetic, the log-linear form becomes more important once additional parameters are added. In addition, exponentiating the row and column parameters provide multiplicative factors relative to the base category. For example, the cell frequency of a husband with brown eyes is estimated to be 1.74 ($e^{0.5521}$) times greater than the cell frequency of a husband with blue eyes (the base category). When you are comparing eye colour, the parameters are not of great value. However, when comparing patterns of ethnicity, being able to compare to a base group, for example, the European Only group, allows for inter-ethnic comparisons.

In the previous example the estimated frequency was only about half of the actual frequency, indicating that the independence model did not fit that set of data very well. However, the log-linear model can be extended to include other parameters, with models such as quasi-independence, quasi-symmetry and crossing parameters. These different forms of log-linear model use additional parameters to improve fit and enhance interpretability, and have been used to examine educational and religious homogamy and propensity for people to cross educational and religious boundaries (Blackwell & Lichter, 2004; Kalmijn, 1991; Qian & Lichter, 2007; Schwartz & Mare, 2005).

2.2.2. Quasi-Independence Models

The quasi-independence model has the log-linear form (Agresti, 2002):

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \delta_i I(i=j),$$

where $I(.)$ is the indicator function

$$I(i=j) = 1, \quad i = j$$
$$= 0, \quad i \neq j$$

and tests whether the statistical relationship between the row and column variables (partner characteristics) are confined to a particular section of the table (Blackwell & Lichter, 2004). In this case we are interested in the δ_i parameters for the diagonals cells as they reflect the degree of homogamy between the partners of each group by showing whether the diagonal cells have a greater observed frequency than the expected frequency under statistical independence. They are computed by fixing the estimated diagonal values using the actual diagonal values, and then using the independence model to calculate the estimated frequencies for the off-diagonal cells.

In the case of the hypothetical table from earlier in this section (see Table 2.1), the exponentiated λ parameters represent the multiplicative effect, under the independence model, of the husband's eye colour and the wife's eye colour on the frequency of couples in every cell. The exponentiated δ_i (quasi-independence) parameters would be the multiplicative factors for each of the diagonal cells only, indicating the effect of the diagonals (green eyes with green eyes, blue eyes with blue eyes and brown eyes with brown eyes) above and beyond the values fitted under the independence model. The greater the value of δ_i, the stronger the inclination of that group towards homogamous partnerships.

The parameters for a quasi-independence log-linear model using the eye colour frequencies from Table 2.1 are displayed in Table 2.3. The coefficients correspond to the "base" case (male blue eyes and female blue eyes), the effect on log cell frequency of brown eyes and green eyes for males and females, and the diagonal dominance or quasi-independence parameters which indicate the additional frequency values on the diagonal of the table beyond those predicted under the independence model.

Parameter ("Effect")	"Base" Case	Male Brown	Male Green	Female Brown	Female Green	Blue/ Blue	Brown/ Brown	Green/ Green
Coefficient	1.4534	0.2513	0.2513	-0.2007	-0.2007	1.0315	1.6314	1.5870
Exp(Coefficient)	4.2776	1.2857	1.2857	0.8182	0.8182	2.8053	5.1110	4.8891

Table 2.3 - Quasi-Independence parameters for eye colour table

For example, using the exponentiated parameters, the cell frequency for a couple who both have green eyes is modelled as:

$$\begin{aligned} Cell\ \ frequency &= e^{baseline} \times e^{male\ brown\ effect} \times e^{female\ brown\ effect} \times e^{diagonal\ \ effect} \\ &= 4.2776 \times 1.2857 \times 0.8182 \times 4.8891 \approx 22 \end{aligned}$$

The exponentiated Green/Green parameter value of 4.8891 indicates that the predicted frequency for a couple who both have green eyes is about 4.9 times greater than what would have been expected under the independence model. The quasi-independence model provides a far better fit for tables that have a large proportion of their observations on the diagonal (recall that the independence model estimated 11.4 for the number of green/green couples). This is because the model effectively holds the diagonal frequencies constant and models the remaining cells. It then creates the parameters for the diagonal cells. This means that the fitted estimates for the diagonal cells will always be the same as the actual ones since the parameters are calculated using the fixed diagonal frequencies. The quasi-independence fits the data much better than the independence model did. The residual deviance has reduced from 43.87 for the independence model, to 3.18 for the quasi-independence model due to the large proportion of couples in the diagonal cells of the table that were not being well modelled by the independence model.

2.2.3. Quasi-Symmetry Models

Although the focus of homogamy amongst partnerships, and hence the diagonal dominance of the frequency tables, is the main point of interest, the relationships seen in the off diagonals - i.e. the heterogamous partnerships - can also be examined through a quasi-symmetry form of log-linear modelling. The model for the cell frequencies takes the form of:

$$\log m_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij} \quad \text{where } \lambda_{ij} = \lambda_{ji} \text{ for all } i < j$$

This model implies that the odds ratios for the terms on one side of the diagonal are equal to the odds ratios for their counterparts on the other side of the diagonal. In the context of examining partnership tables of ethnicity, this means that the fit of the quasi-symmetry model provides a test of symmetry for the different male-female ethnicity combinations. For example, do we expect the counts for Asian males with European partners to be the same as that of European males with Asian partners? A number of studies of partnership homogamy have incorporated quasi-symmetry models into their analysis (Blackwell & Lichter, 2004; Kalmijn & Vermunt, 2007; McCaa & Schwartz, 1983), although Blackwell and Lichter (2004) state that "the quasi-symmetry model is not as theoretically interesting as the quasi-crossing parameter model, in that it tests for the presence of a particular pattern in the data, but not movement *per se*". What they mean by this is that the quasi-symmetry model will provide an indication of asymmetry in a frequency table but it will not indicate the existence of social boundaries in the same way that a cross parameter model will.

2.2.4. Crossing Parameter Models

Further to the quasi-independence and quasi-symmetry models, a crossing parameter model is also used to attempt to parameterise social distance in several recent studies (Blackwell & Lichter, 2004; Kalmijn, 1991; Mare, 2001). The crossing parameter model treats a movement between two categories as a barrier to cross, with the parameters indicating the degree of difficulty in crossing a particular barrier. For example, the barrier faced by a person with a school qualification marrying someone with a tertiary qualification. Estimated crossing parameters are symmetrical by definition (Blackwell & Lichter, 2004). However, for some variables such as education this requirement is not met, so an asymmetry parameter is fitted to allow the crossing parameters to vary; for example, by gender.

From Mare's (1991) model, the crossing parameter model for education could be written:

$$\log m_{ij} = \mu + \lambda_i^H + \lambda_j^W + \Sigma_k \beta_k^C d_k^C$$

where

$\lambda_i^H / \lambda_j^W$ = parameter for husband/wife in education category i/j

d_k^C = 0/1 indicator of a marriage is k cells from the diagonal

β_k = the crossing parameters (see Table 2.4).

The model still follows a similar form. It models the log of the cell frequencies using a constant, an effect for the education level of the husband, and an effect for the education level of the wife. However, it how also adds these crossing parameters (β) which measure shifts between different groups within the frequency table.

	Wife's Education Level			
Husband's Education Level	None	High School	Vocational	Tertiary
None	0	β_1	$\beta_1 + \beta_2$	$\beta_1 + \beta_2 + \beta_3$
High School	β_1	0	β_1	$\beta_1 + \beta_2$
Vocational	$\beta_1 + \beta_2$	β_1	0	β_1
Tertiary	$\beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_2$	β_1	0

Table 2.4 - Hypothetical education table showing crossing parameters

Table 2.4 shows how the crossing parameters cumulate for partnerships the further they are from the diagonal of the table. Each β parameter measures the odds or difficulty of shifting one cell further away from homogamy (the diagonal). For example, the exponentiated sum of all three β parameters in the top right and bottom left corners of the table give the odds of an individual without a qualification partnering with someone with a tertiary qualification. The odds of crossing an educational barrier are functions of the β values. The odds of crossing barrier k, net of the marginal distributions of the spouses' education, are found by $\exp(\beta_k^C)$, with each parameter corresponding to a single move across adjacent levels in the table.

This method generally requires orderable categories, a feature more commonly seen in articles on educational homogamy (Blackwell & Lichter, 2004; Mare, 2001). Within the New Zealand ethnicity data, the different combinations of ethnic matches (no matching ethnicity, one common ethnicity for a couple with two nominated ethnicities each, one common ethnicity for a couple where one partner has only one ethnicity, full match of the same ethnicities) provide this kind of ordering.

The crossing parameters measure the changes in the likelihood of people to partner across each “boundary”, i.e. whether people are more likely to form partnerships with partial homogamy or some homogamy or no homogamy rather than forming a partnership with full homogamy. For the ethnicity data, the boundaries become the likelihood of people partnering someone who has a partial match of ethnic group compared to someone with no common ethnicity, and then the likelihood of people partnering someone with a full match of ethnic group compared to someone with a partial match. These parameters are an extension of the quasi-independence parameters as they move the analysis from a binary situation that only considers partners to have the same or different ethnicity to allowing for different levels of ethnic matching.

2.2.5. Logistic Regression

Logistic regression is used to model a binary (0,1) response variable using the following linear relationship (Agresti, 2002):

$$\log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta x$$

In this relationship, π is the probability of the response being a “1” and *x* is a set of covariates that are being used to model the response. Logistic regression complements log-linear analysis by incorporating the effect of other factors on comparisons between homogamous and heterogamous partnerships. For this study, the variable of interest is the probability of a homogamous partnership (compared to a non-homogamous one). The β coefficients provide the log odds ratio for the effect of each *x* variable on the probability of a homogamous partnership, and the p-values provide an indication of the significance. The logistic regression is presented in Section 5.3.

Tzeng (2000) applied logistic regression to a random sample of marriages to examine the impact of education, duration of immigration and employment status on the probability of Asian Canadians having a spouse of a different ethnicity. She found that the younger age groups were more likely to have heterogamous marriages than the older age groups. Those with a greater number of years of education and the ability to speak English and/or French (the official languages) were also more likely to be in a heterogamous marriage.

Mamet *et.al.* (2005) took a similar approach, applying logistic regression to Chinese census data to explore the factors that affected the probability of marrying a partner of a different ethnicity. They found a number of social, political, cultural, linguistic, and religious variables to have a significant effect on the probability of marrying a partner of a different ethnicity.

It should be noted that logistic regression parameters from larger (particularly census) studies can show results that are statistically significant due to the large sample sizes (Simonoff, 2003), but are not necessarily of practical significance. This study will examine some logistic regression models to compare the characteristics of the homogamous and heterogamous couples in the data, but will mainly use log-linear models to analyse the partnership frequency tables.

Chapter 3 – Literature Review: Simulation

3.1. Social Simulation and Modelling

"Artificial society modelling allows us to "grow" social structures in silico demonstrating that certain sets of microspecifications are sufficient to generate the macrophenomena of interest." (Epstein, 2006)

Simulation provides the opportunity for further examination of partnership choice. Whilst the regression parameters from the log-linear and logistic regression models indicate the effect that variables have on the probability of existing partnerships being homogamous, simulation models can be used to examine the partnership formation process as a complex system. Of particular interest is the ability to mimic the choices of individuals at the micro level and then assess the impacts of this pattern of choice both on the wider society (macro level) and in shaping later micro level choices.

There are many methods of simulation (for a diagram of current and historical methods see Gilbert and Troitzsch, 2005; pg 7). However, the focus of this study will be on a series of methods known as microsimulation and agent-based simulation.

This section contrasts microsimulation and agent-based simulation, and then examines the literature which uses each to model partnership formation. The use of simulation models to examine the macro-micro link is followed by discussion about different algorithms that have been used to model the formation of partnerships. An alternative approach that uses network models (graph theory) to connect people to one another concludes the section. In addition to summarising the existing work in the area of the simulation of partnership formation, each section will highlight the components of other simulations which will contribute to the construction of the simulation model for this study.

3.1.1. Introducing Microsimulation and Agent-Based Simulation

Although Orcutt's seminal article on microsimulation (Guy H. Orcutt, 1957) is often referred to as the beginning of microsimulation, it is really only in the last twenty years that simulation has become a widely used technique. This is largely due to the availability and increased power and sophistication of computer hardware and software over this period. Microsimulation involves simulating a system from the level of individual units rather than at the aggregate level.

To date most microsimulation models have been constructed to predict the effects of social or financial policy (Gilbert & Troitzsch, 2005). This approach is typically used for forecasting purposes and so tends to focus on outcomes of a process (such as the fiscal outcomes of a new taxation policy) rather than trying to examine and understand the process itself. For example, Zhao (2000) uses microsimulation to examine residential patterns in historical China and is able to verify the simulation results by comparing them to empirical studies. By comparison, Rephann (2004) examines the demographic and economic impact of immigration in Sweden, using microsimulation models to evaluate the level of immigration which the country could comfortably handle. In the case of both studies, individual actors or agents are given a set of instructions or rules to apply to their situation in order to simulate the complex scenario of interest. There is an obvious parallel where a set of rules could be provided to a population of agents for choosing a partner.

However, in a discussion paper about microsimulation, van Imhoff and Post (1998) write that despite the flexibility of simulation, partnership matching tends only to form a small part of larger demographic simulations, with the choice of partner often determined from a fixed set of life table probabilities with little contemplation of the processes underlying the matching of partners. This would suggest there is significant scope for sociological and methodological advances in the area of partnership simulation.

By definition, the key difference between agent-based simulation and microsimulation is the way that the agents interact with one another (Gilbert & Troitzsch, 2005). In agent-based simulation the actions of one agent can have a direct influence on the actions of other agents, whereas in microsimulation the agents tend to behave independently of one another. This property is relevant to partnership choices since once an agent is partnered, they can't take on additional partners, thus impacting on the choices of the other agents. Beyond this, agents may exert some sort of peer or societal influence on the partnering choices of one another. Beyond the behaviour of the agents, another difference is the intention of the models. Agent-based models are more commonly used for experimental simulation where the behaviour of the agents is of interest, while microsimulation tends to be used for predictive purposes (van Imhoff & Post, 1998).

One of the other advantages of agent-based simulation is that it requires explicit recognition and specification of interactive heterogeneous behaviour. It can also provide dynamic feedback from individuals to groups and vice-versa (Åström & Vencatasawmy, 2001). These qualities make it a useful tool for examining cohabitation choices, with the literature suggesting there are potentially feedback loops between an individual's decisions and the historical decisions of the community they exist within.

Although it may seem that agent-based simulation models are more suitable for modelling partnership choice, the distinction between microsimulation models and agent-based simulation models is not necessarily clear cut, particularly for dynamic microsimulation models. A dynamic microsimulation is one where the parameters controlling the agents may change over time. This blurs the line between the approaches because a micro-macro linked variable may create an indirect relationship between the actions of agents over time, producing a model which straddles the distinction between the two methods.

3.1.2. Microsimulation Models of Partnership

During the last ten years, a number of large scale microsimulation projects have incorporated partnership matching into their population and policy simulation models, and they have done this with a greater degree of detail than those which were critiqued by van Imhoff and Post (1998). This section reviews the microsimulation-based literature on partnership matching and discusses how different components of some of those models could be useful for simulating partnership in the New Zealand context. The section is concluded with a table (Table 3.1) that summarises the different simulation models.

Bouffard *et.al.* (2001) compared the standard life table method with an alternative stochastic algorithm. This was based on a logistic regression using both the age differences of the potential partners and economic factors to find better matches. Validation testing demonstrated that the stochastic algorithm provided better predictions against American census data than the fixed probability method. However, the investigators also note that further improvements could be made, particularly by dividing the broader marriage market into sub-markets, thus improving the accuracy of the matches. The lack of independence between partner choices (you can't have the same partner as someone else) makes the assumptions of a logistic regression-based method somewhat suspect, but it does show the strength of a stochastic algorithm compared to matching via a fixed table. The New Zealand census data provided for the simulation models in this study (see Section 7.3.1 for details of the simulation data) has a limited amount of detail, thus ruling out logistic regression on practical grounds. Bouffard's comments about dividing the marriage market into regional sub-markets seem relevant in terms of improving the accuracy of the matches and also in terms of the theoretical aspect of the model which requires opportunities for interaction and choice. The regional sub-markets are incorporated into the simulations via separate simulation models for Auckland, Wellington and Christchurch.

Similar findings on stochastic algorithms compared to fixed probabilities were reported by Perese (2002) who also applied a logistic regression-based algorithm to marriage matching within a microsimulation model. Using the American Congressional Office's Long Term (CBOLT) microsimulation model he was able to replicate joint distributions of spousal age differences, education and earnings that were similar to collected survey data. The logistic regression modelled a set of data where the actual partnership matches for each male were coded as one, and then a set of hypothetical alternative partnerships to each other female were coded as zero. These were modelled against the age differences, education levels and earnings of the males and females to create the parameters for the simulation model. Again, the logistic regression-based model is not suitable for the New Zealand data and the assumption of independent observations does not seem to be met, but Perese does mention an alternative methodology based around the DYNASIM model.

The DYNASIM model was first developed by the Urban Institute (http://www.urbaninstitute.org) as an income model in the 1970s (G. H. Orcutt, Caldwell, & Wertheimer II, 1976) but has seen updates in the 1980s (Zedlewski, 1990) and 2000 (Favreault & Smith, 2004). The DYNASIM matching algorithm, as described in Perese (2002), is a Monte Carlo styled method based around the probability function below:

$$P(Union) = e^{-0.5\sqrt{(Age\ difference\ )^2+(Difference\ \ in\ years\ \ of\ \ education\ )^2}}$$

The bachelors and bachelorettes in the model are randomly queued and then the probability of the match between the first bachelor and the first bachelorette is calculated. A random number is drawn and a match is made if the probability is greater than the random number. If not, the process is repeated up to nine more times. If no match is made after these ten trials, then the bachelor is matched to the bachelorette with the highest probability score. This methodology is appealing in several ways. By creating a probability function rather than relying on regression models, the algorithm is not bound by assumptions of independence or other statistical factors. It is a simple heuristic which fits with the theories of homophily seen in Sociology (McPherson, Smith-Lovin, & Cook, 2001), is easily extensible, and can be applied to the data that is available. Other researchers have also used the same basic matching function, including the Australian Dynamic Population and Policy Microsimulation Model (APPSIM) (Bacon & Pennec, 2007; Harding, 2007).

The APPSIM model was constructed by the National Centre for Social and Economic Modelling (http://www.natsem.canberra.edu.au) in Australia to extrapolate the population to the year 2050 for the purpose of planning and policy in areas including taxes, wealth, housing, health status and service usage (Harding, 2007). The partnership matching algorithm of the APPSIM model was based on that of the DYNASIM model, but with the addition of an alignment process limiting the number of the couples formed at each iteration. This meant that couple formation was in line with the frequencies projected by an independent macrosimulation model (Bacon & Pennec, 2007). The APPSIM model shows that the DYNASIM algorithm has been successfully applied to a different population than the one on which it was developed and originally applied. The alignment process can be easily implemented into simulations of New Zealand data since they will be based on historical data, thus allowing the results of each simulation to be compared to actual Census data for the following period. The DYNASIM/APPSIM model is also appealing in its fit with the simplified unit-level census data that is available for the simulation.

Two European models, the Microsimulation Model of Family Dynamics (FAMSIM) (Spielauer & Vencatasawmy, 2001) and the Simulating Social Policy in an Ageing Society (SAGE) model (Cheesbrough & Scott, 2003), represent a more robust methodology to that of the regression approach. Each uses retrospective accounts of partnership histories, collected from stratified random samples of women in their respective regions, to inform their models, rather than using matrices of actual and hypothetical partnerships as logistic regression outcomes. Unfortunately, such data is not available in New Zealand, although this could provide an avenue for future research.

Spielauer and Vencatasawmy (2001) also introduce the idea of microsimulation, multilevel models, and context driven agent-based simulation, and extol the virtues of synthesising the important components of each. They explain, quoting Troitzsch (1996), that whilst the microsimulation and agent-based simulation have “evolved in almost total ignorance of each other”, there is an increasing crossover in the techniques and philosophy of the two approaches and that microsimulation “has the potential to improve the accuracy of forecasting and provide new insights into underlying individual behaviour”. Although it would not be practical to apply their methodology to the New

Zealand census data, the philosophy of combining microsimulation and agent-based methods and taking the best of both worlds reflects one of the goals of using the census data as a "test-bed" and trying to produce simulations for forecasting and theory testing.

Not all microsimulation models of partnership formation focus on predictive outcomes. Chen (2005) uses an alternative paradigm, electing to examine possible strategies for the process of how people find their partners rather than focussing solely on an extrapolation-based outcome. He uses simulation models to examine five different theoretical matching strategies:

- "choosing for the best", where agents will continue to search for a partner until they find the one that they consider the best across all of their multiple search criteria in the marriage market.
- "well rounded", where agents will seek a balance of performance across multiple search criteria.
- "differential preference", where the male and female agents will each focus on several specific criteria.
- "compensatory", where the agents will allow potential partners to draw on the strength of one criterion to offset weakness in another, with no particular criterion being considered more important than any other with the differential preference.
- "immediate matching", where the first potential partner is matched.

Results were measured as the proportion of couples who were successfully matched together and the "cost" of the search within an artificially generated population of one hundred. Chen showed that under these conditions, the most effective search method was the "compensatory" method, since it would successfully match without an excessive amount of searching. By comparison, the "choosing for the best" method performed poorly, exposing the agents to the risk of losing all chances of finding a partner.

Although several of the assumptions made in some of Chen's simulations seem unrealistic, such as having no competition between agents for the same partner or permitting potential partners to appear randomly (ignoring the fact that marriage markets are generally segmented), the models show how the simulation of partnership choice can be used to examine the dynamics of marriage markets within a population. The simulations allow a range of theoretical positions to be tested and compared, even if it is only in abstract. Applying Chen's ideas to real census data provides scope for examining possible underlying social processes that may be taking place within the New Zealand population. His work suggests that simulations of partnership formation should include some form of satisficing behaviour, so that agents will partner in a more realistic manner.

A number of important points can be found in the microsimulation literature. Although the logistic matching routines used by Bouffard *et.al.* (2001) and Perese (2002) are not practical with the New Zealand data, both studies provide other useful information that can be applied to the New Zealand case. Both highlight the value of stochastic methodology over deterministic models, while Bouffard *et.al.* also suggests that simulating segmented marriage markets through the use of independent sub-markets will improve accuracy. The DYNASIM (Zedlewski, 1990) and APPSIM (Bacon & Pennec, 2007; Harding, 2007) models both utilised a simple probability function which was successfully implemented on two different populations and could be applied to the New Zealand data. Spielauer and Vencatasawmy (2001) argue that microsimulation and agent-based simulation methods have strengths that can be borrowed from one another and that researchers should not feel constrained by using one or the other. With this in mind, we will now examine the literature that uses agent-based simulation models to model partnership choice.

Model/Reference	Partner Matching Variables	Partner Matching Process
APPSIM (Bacon & Pennec, 2007; Harding, 2007)	Age difference Years of education difference	Similar to DYNASIM. Singles randomised. First pair evaluated using exponential probability function. Paired if random number less than probability, otherwise repeat for up to 10 potential partners. If no matches are made and total number of couples has not been met, then pair those with highest probability.
Bouffard *et.al.* (2001)	Age difference, correlation of husband-wife earnings.	Data is set up with each male paired with every possible female partner. Logistic regressions are run with outcome 1 for actual partner and every other woman with the same characteristics, and 0 otherwise. Transition probabilities from this model are used in Monte Carlo simulation.
CBOLT (Perese, 2002)	Age, education, average lifetime earnings quintile, marriage number	Data is set up with each male paired with every possible female partner. Logistic regressions are run with outcome 1 for actual partner and every other woman with the same characteristics, and 0 otherwise.
Chen (2005)	Two generic rankable traits	Trials 5 different methods of matching: best only, well-rounded, differential, compensatory, and immediate.
DYNASIM (Zedlewski, 1990)	Age difference Years of education difference	Singles randomised. First pair evaluated using exponential probability function. Paired if random number less than probability, otherwise repeat for up to 10 potential partners. If no matches, then pair those with highest probability.
FAMSIM (Spielauer & Vencatasawmy, 2001)	Children, age, education, pregnant.	Monte Carlo simulation based on logistic regression models using retrospective partnership histories from survey.
SAGE (Cheesbrough & Scott, 2003)	Age, marital status, education, pregnant.	Monte Carlo simulation comparing transition probabilities based on logistic regression models using retrospective partnership histories from survey.

Table 3.1 - Summary of microsimulation partnership models

3.1.3. Agent-Based Simulation Models of Partnership

"… the crucial difference between it (agent based simulation) and other techniques … is that agent based simulation attempts to represent explicitly the interactions between agents in a population and corresponding changes in internal states." (Chattoe, 2006)

Agent-based models provide a different approach and philosophy to the modelling of partnership choices. The size and goals of the simulations are two of the key differences setting these models apart from those in the previous section. Whilst the microsimulation models were often working at a city or population level with thousands or even millions of agents, most of the agent-based models use very small populations. Instead, they focus on the processes that are occurring, such as the emergence of particular patterns, rather than on purely seeking to produce population projections for the future. This section examines the relevant agent-based models of partnership choice that have been published in the last twenty years, summarising the methodology and findings and seeking further elements that can contribute to a New Zealand simulation of partnership. As with the previous section, this section concludes with a summary table of the simulation models.

Some of the earliest agent-based simulations of partnership were conducted by Kalick and Hamilton (1986, 1988) in the mid to late 1980s. Their programming was coded in the FORTRAN language as dedicated social simulation packages were close to a decade away. The main focus of their 1986 paper was to compare matching schemes where preference was based on a rankable trait and agents were attracted to those with a similar trait value, or agents were attracted to those with the highest value of the trait. Although it is not reasonable to treat ethnicity as an ordinal set, chapter 6 presents some precursor abstract simulations conducted using the Netlogo programme where a rankable characteristic is used to compare some different matching schemes. They also incorporated the idea of declining preferences over time, which has been replicated in some way by all of the subsequent studies that will be discussed. Simão and Todd (2003) cited one critique of this early work being that agents had to "date" a large number (over forty) of individuals before a significant percentage of the population would mate.

Despite this, the Kalick and Hamilton work sets the scene for much of the future abstract simulation of partnership choices.

During the ten years that followed the Kalick and Hamilton articles there were major advances in computing power and many new programming languages and simulation tools, including Java, (Sun Microsystems), Netlogo (http://ccl.northwestern.edu/netlogo) and Repast (http://repast.sourceforge.net/). However, there was a surprising paucity of articles about partnership simulation during this period. Since 2000, the number of publications has increased, in particular thanks to a number of different collaborations involving Peter Todd of the Max Planck Institute for Human Cognitive and Brain Sciences (http://www.cbs.mpg.de/index.html) (Hills & Todd, 2008; Miller & Todd, 1998; Simão & Todd, 2001, 2002, 2003; Todd & Billari, 2003; Todd, Billari, & Simão, 2005; Todd & Miller, 1999, 2002).

These articles have used a variety of matching heuristics to simulate partnership matching in artificial populations. Rather than building these matching models for prediction, one of the main goals of this work has been to actively search for emergent patterns in the simulated populations, such as the shape of the aggregate age-at-marriage distribution (Todd & Billari, 2003; Todd et al., 2005). In both of these articles, the authors demonstrated the effectiveness of a simple search strategy based on a two-sided (mutual) process where matches were based on the combined evaluations of the male agents and the female agents. They found that by combining a bottom-up agent-based modelling approach with top-down demographic constraints, they could reproduce similar patterns in their variable of interest (age at marriage) to the age distributions of several European countries. The natural extension of this work would be to apply it to a population of singles in a real population rather than an artificial one, and benchmark it against that same population at a later time period. In this study, this will be achieved by using census data to both populate the model of New Zealand and as a benchmark for the results.

By comparison, the Simão and Todd articles (2002, 2003) focussed less on the emergent outcomes of their partnership simulations and more on the simulation process itself. Age-at-marriage distributions were used to check on the validity of the models, but in each case the focus was on the inputs to the process. In the earlier of the two articles, the authors used their simulation model to test the utility of courtship. They experimented with scenarios where a courtship period would allow agents to adjust their aspiration levels or to swap partners if a superior match was found. The disadvantage of this process comes with the increased number of iterations required to scale up from a small population (50 agents) to a census-sized one. The later of the two Simão and Todd articles examined the effects of a skewed sex ratio. They found that within their simulation model an unbalanced sex ratio led to the larger of the two groups becoming more likely to pair down in order to find a match. Although this result is not at all surprising, it provides a good example of using a simulation model to examine a process-based hypothesis.

Hills and Todd (2008) developed an agent-based model known as MADAM (Marriage and Divorce Annealing Model). MADAM built on earlier work by Miller and Todd (1998), Todd and Billari (2003) and Todd, Billari and Simão (2005), using a "homophilic trait matching model" which would gradually relax the expectations of the matching preferences as the agents aged. The main matching process involved a randomly generated world of agents, each of which have k mate-relevant traits from a set of N possible traits. Each will initially seek a mate who has a perfect match of the same k traits but, as they age, they will settle for a mate with j ($j < k$) traits. This satisficing level was built into the simulation by using the function:

$$j = ke^{-\lambda t}$$

where j is the current threshold required to partner, k is the initial number of matched traits required, and λ is the rate of decay of expectation. Much like the DYNASIM model discussed in the previous section, this is a decaying exponential function of the form $e^{-f(x)}$.

Alam and Meyer (2008) used an agent-based network simulation to study the spread of HIV/AIDS in an African village. They developed two matching algorithms to model the sexual networks in the village. Although sexual networks do not restrict agents to a single partner like cohabitation/marriage models, the idea of agents interacting and accepting/rejecting partnership offers is similar. Two different choice mechanisms were tested. The first one involved each agent having an attraction score and an aspiration value. When agents encountered one another, they would partner if the attraction score for each was greater than the aspiration value. The aspiration values would decay as the agents aged, much like the expectations of the agents in the model used by Hills and Todd (2008). The second scheme used an "endorsement mechanism", where the decision to partner was based on a potential partner having sufficient positive feedback from others in the network, akin to an explicitly communicated social norm. The authors found that both methods gave broadly similar results, although the endorsement mechanism had less variability than the attraction/aspiration one. They did not attempt to integrate the two schemes to see how the partnering patterns might change when the agents used both micro and macro information to make decisions.

The literature on the agent-based modelling of partnership choice is sparser than that of microsimulation. The models typically use small artificial datasets rather than real data, although some of the more recent literature has incorporated some level of comparison to actual data (i.e. Hills & Todd, 2008 comparing to real age-at-marriage distributions). One of the reasons for this is that the agent-based models tend to focus on process-related factors or on the demonstration of emergent properties, rather than prediction. One such property of note is the micro-macro link, which is discussed in the next section. One of the common themes of the agent-based models, which is also evident in the microsimulation models, was the need for some kind of decaying aspiration level or expectation over time, effectively increasing the chances of remaining (older) agents finding a partner by making them less discriminating.

Model/Reference	Partner Matching Variables	Partner Matching Process
Alam & Meyer (2008)	Attraction scores and endorsement scores.	Partner selection occurs either through attraction scores of each agent being greater than the aspiration value of the other agent or through high cumulative endorsement scores from others in the network.
Hills & Todd (2008) MADAM model	*N* generic mate relevant traits.	Agents partner when they find *k* matches out of *N* generic traits in another agent. The number of matches required decays exponentially as the agents age.
Kalick & Hamilton (1986)	A single generic rankable trait.	Agents are either attracted to the agent with the highest trait, the agent with the most similar trait, or a combination of both.
Simão & Todd (2002, 2003)	A normally distributed generic trait and aspiration level for evaluating partners.	Agents encounter each other randomly. They compare each other's traits with their aspiration levels, then make "offers" to one another if traits are both greater than aspiration. Otherwise, one will reject the other and they adjust their aspiration levels.
Todd & Billari (2003), Todd *et.al.* (2005)	A normally distributed generic trait and aspiration level for evaluating partners.	Agents encounter each other randomly. They compare each other's traits with their aspiration levels, then make "offers" to one another if traits are both greater than aspiration. Otherwise, one will reject the other and they adjust their aspiration levels.

Table 3.2 - Summary of agent based partnership models

3.1.4. Simulation Modelling of the Micro-Macro Link

"…advancing simulation technology offers some advantages, particularly the modelling of macro-micro links too complex to deal with linguistically or mathematically." (Halpin, 1999)

In principle, multi-agent systems provide sociologists with a powerful tool for empirically assessing theoretical propositions in a practical way (Sawyer, 2003). The concept of a micro-macro link (discussed in 2.1.4) is one area of sociological interest which can be implemented through the use of computer simulation. Examples of this can be seen in settings as diverse as modelling micro-macro links in military coups (Saam,

1999) through to stock markets (Hoffmann, Jager, & von Eije, 2006). It should be noted that emergence and the emergence of macro phenomena and patterns from micro-level events in simulation models are not the same as the sociological term “the micro-macro link” (Coleman, 1990). The micro-macro link is a long-standing point of debate in sociological theory. However, in this study the emergent properties of the simulation are being used as a way to examine this analytical perspective.

The modelling of these micro-macro links is not limited to human populations. Conte, Paolucci and Di Tosto (2006) worked forward from emergent patterns towards modelling micro-macro links in the mating of vampire bats. Stolk, Hanan and Zalucki (2007) worked along similar lines, examining how the micro behaviours of butterflies in a field created macro patterns. In each case, the simulation models were informed by animal behaviour theories but were based in artificially created environments. Although it is likely that human populations behave differently from animal ones, the idea of simulating individual behaviour to search for emergent patterns is comparable.

To date, much of the modelling of the macro-micro links within animal and human populations has been based on small, artificially-generated societies. Although this provides some insights into the macro-micro relationships in each case, the natural extension of this work is to empirically test and validate it with real data. By examining possible recursive relationships between micro and macro level behaviour, the simulation models in this study extend beyond other similar partnership models in the literature. They not only use real data and get validated against subsequent data sets, but they also incorporate the potential recursive relationship between the rate of inter-ethnic partnering in one time period and the choices of people in future time periods. Constraints in the level of detail in the data (see Chapter 4) limited the investigation into other recursive relationships, but the modelling is still a step beyond the small artificial populations that have been used for previous investigations of this type.

3.1.5. Algorithms of Mate Selection

The algorithms for mate selection are pivotal to a successful model. In microsimulation models there are typically two approaches – the stable marriage approach, and the stochastic approach (Bacon & Pennec, 2007) – whilst agent-based models are more likely to have some sort of rules or heuristics for decision making (Spielauer & Vencatasawmy, 2001). Although Sections 3.1.2 and 3.1.3 have already discussed microsimulation and agent-based simulation, this section will provide further detail on the algorithms that are used.

One of the earlier algorithms for partnership matching came about from an applied mathematical problem which aimed to match students to colleges such that the utility of individuals could not be increased by swapping any pair of individuals (Gale & Shapley, 1962). It was shown that there was always a combination which would meet this requirement. The application of this algorithm to marriage problems became known as the "stable marriage problem" (Gusfield & Irving, 1989). The application of the Gale-Shapley algorithm worked as follows. At each iteration, every single male proposes to the highest female on his preference list who is available (that is, one he has not proposed to already). If a woman is "free" then she will accept, otherwise she will compare the new proposal to her current match and reject the less favoured of the two. This process repeats until all of the couples are matched. The one-way nature of the algorithm, and the assumption that the men have some pre-existing knowledge of all of the women in the population, make it an algorithm that, while it offers an elegant solution to a mathematical puzzle, does not provide a good reflection of the social reality of the "marriage market".

Knuth (1997) wrote a short book summarising the algorithm, proofs and several variations of the problem. He presented the fundamental algorithm as follows:

> Variables: k, X, and x Constants: n, Ω
> n = number of men = number of women;
> k: number of (trial) couples already formed;
> X: suitor;
> x: Woman toward whom the suitor makes advances;
> Ω: (very undesirable) imaginary man.

$k \leftarrow 0$; all the women are initially engaged to Ω;
while $k < n$ **do**
    begin $X \leftarrow (k + 1)^{st}$ man;
    while $X \neq \Omega$ **do**
        begin $x \leftarrow$ best choice remaining on X's list;
        if x prefers X to her fiancé **then**
            begin engage X and x;
            X $\leftarrow$ preceding fiancé of x
            end;
        if $X \neq \Omega$ **then** withdraw x from X's list
        end;
    $k \leftarrow k + 1$
    end;
celebrate n weddings
(Knuth, 1997)

For example, Table 3.3 shows the preferences for males {A,B,C,D} and females {W,X,Y,Z}. Applying the stable matching algorithm with the men proposing to the women, and working in alphabetical order:

- A proposes to Z, Z accepts
- B proposes to X, X accepts
- C proposes to X. X prefers C to B so rejects B and accepts C.
- B proposes to Y, Y accepts
- D proposes to Y, Y prefers B to D so declines D.
- D proposes to W, W accepts.

So the final set of couples is: {(A,Z), (B, Y), (C, X), (D, W)}. This is a stable solution, as none of the couples are able to swap to mutually improve their positions.

Males	Preferences				Females	Preferences			
A	Z	W	X	Y	W	D	A	C	B
B	X	Y	W	Z	X	A	C	B	D
C	X	Z	Y	W	Y	A	B	C	D
D	Y	W	Z	X	Z	D	A	C	B

Table 3.3 - Stable marriage algorithm example

Knuth provided proofs demonstrating that this algorithm is stable; that is, each man has the best possible marriage match and no rearrangement or swapping would improve the match for any man. He also showed that through the course of the algorithm a woman's situation never worsens, and that no two women can have the same fiancé.

Although this algorithm is relatively simple and mathematically elegant, it is not suitable for the empirical simulation of the New Zealand census data for a number of reasons. The first problem is that this model assumes that all of the participants can see all of the other participants. They can also judge and rank all of the other participants. For a small group of people this is not unrealistic, but for a population of 100,000 individuals it is not. In addition to this, there is the question of the amount of computation that would be required for scaling the example up to a city level. Knuth (1997) states that in the possible worst case the algorithm would go through $n^2 - n + 1$ iterations. For a city with $n = 100{,}000$ men and 100,000 women, this would mean 10^{10} iterations of the algorithm. The final issue is that the algorithm requires the participants to be able to explicitly rank one another. For some variables such as ethnicity, it is difficult to assign a ranking structure. Even for variables such as education where it is possible to rank, possible partners can be ranked in different ways (e.g. prefer the highest level education, prefer most similar level of education etc). In real life, traits such as education are also not immediately obvious, and neither is information about a person's potential pool of partners. This means that one of the problems that must be considered for the matching algorithm is that of incomplete information.

One final problem with the stable marriage algorithm was highlighted by Bouffard *et.al.* (2001). They showed that the stable marriage algorithm would tend to produce an excess number of central marriages and also an excess number of extreme marriages. Figure 3.1 shows a comparison of the stable marriage algorithm relative to the census data for the same period. There is a huge spike in the centre but there is also an increase at the top of the right tail of the graph, where an excessive number of extreme matches are created.

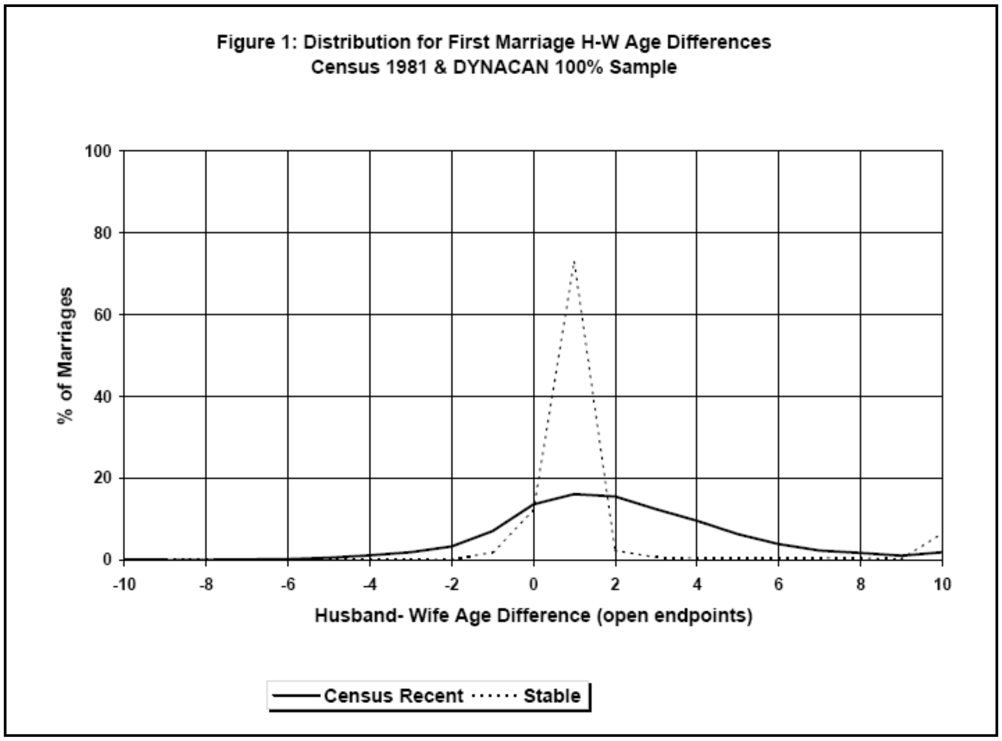

Figure 3.1 - Actual vs stable marriage algorithm from Bouffard et.al. (2001)

One of the more famous matching problems which deals with the problem of incomplete information is the "Secretary Problem" (also known as the "Dowry Problem") (Freeman, 1983). The basis of this "problem" is that, because of the way in which the interview and hiring process is set up, the "best" candidate might not be selected. It functions as follows. The potential secretaries are interviewed for a position. Each one is assessed and ranked in order of preferences. They can be ranked without ties. However, once a secretary has been interviewed, an immediate decision to hire or not has to be made. Once a potential secretary is rejected, the rules require that this decision cannot be reversed. Therefore, if it turns out that a particular secretary, once rejected, is in fact better than a later choice, they cannot be brought back. The same problem has also been referred to as the "dowry problem", where a suitor's family wish to maximise the dowry they receive, examining the dowry offer from each potential wife in a random order, with no idea of the dowries that might be offered in the future.

A summary of work on the Secretary/Dowry problem demonstrated that the optimal strategy for success, given that each secretary/wife arrived in a random order, was to view e^{-1} (or approximately 36.8%) of the population then choose the next candidate/dowry who ranked first relative to those already observed (Ferguson, 1989). By incorporating the concept of incomplete information, the Secretary/Dowry strategy is a more realistic approach to partnership matching than the stable marriage algorithm.

However, this approach also has a number of drawbacks. Although this strategy is more realistic than the stable marriage algorithm, it still makes some unrealistic assumptions. The first is that it assumes that the encounters with potential partners occur completely at random from the population. Studies have shown that social networks and contact between individuals is not a completely random process (McPherson et al., 2001), so drawing potential partners at random reduces the realism of the approach. The larger problem is that it is not reasonable to assume that people in the population will wait until they have viewed or dated nearly 37% of potential mates before making a decision. Any reasonable algorithm must include some form of satisficing in order for it to be considered practical as a partnership choice model. Although examining 37% of the population may be the mathematically optimal solution, it is far from realistic, or even computationally practical. It can be demonstrated that satisficing algorithms are not only more pragmatic but can still provide a good framework for matching (Todd, 1997). Even working with this figure of 37%, these combinatorial style algorithms become unreasonably computationally expensive once they are applied to hundreds of thousands of agents.

An alternative to the combinatorial-based approaches such as the stable marriage algorithm, is to use some sort of stochastic strategy. This is often referred to as Monte Carlo simulation and generally involves drawing random numbers and comparing them to some existing probability value. This type of methodology is often used in predictive microsimulation models, particularly large-scale government policy models, such as those looking at potential tax or pension reforms. In a number of marriage studies (Bouffard et al., 2001; O'Donoghue, Lennon, & Hynes, 2009; Perese, 2002; Spielauer &

Vencatasawmy, 2001) these probabilities were generated via logistic regressions, but they may be drawn from other sources such as actuarial life tables or suitable probability distributions.

Two contrasting examples of stochastic methodologies are the DYNASIM (Zedlewski, 1990) and APPSIM (Bacon & Pennec, 2007) models and the Congressional Budget Office's Long-Term (CBOLT) model (Perese, 2002). The DYNASIM/APPSIM model randomly sorts the bachelors and bachelorettes. The probability that the first available bachelor will partner the first available bachelorette is calculated by: $P(union_{mf}) = e^{-0.5\sqrt{(age_m - age_f)^2 + (edu_m - edu_f)^2}}$, taking the differences in age and years of education. This figure is compared to a random uniform number. If the random draw from a uniform distribution is less than the calculated probability then a match is made.

By comparison, the CBOLT model uses information about *n* couples to create an *n*-squared size dataset, where each male is paired to his real partner and then each other woman. A new binary variable is created, which takes a value of one for each male's actual partner, and each other woman with identical characteristics to his partner, and a zero otherwise. This binary variable is modelled by a series of covariates, using a logistic regression model. The resulting model is used to generate the probabilities of each match in the simulation. These probabilities are compared to randomly drawn values to determine which couples are matched together in the simulation.

The advantage of these kinds of stochastic approaches is that the simulation can be informed by empirical evidence, and it can incorporate variability through the use of random numbers. The downside of a stochastic approach is that it often requires more detailed information about the population or system that is being simulated, and can subsequently require further detailed data. From the examples above, the CBOLT model required much more information than the DYNASIM/APPSIM models. It requires information on a number of covariates for the singles and couples in the population, and would be difficult to conduct under the privacy restrictions of the New Zealand data. By comparison, the DYNASIM/APPSIM model only requires the age and years of education for each agent, but could be easily extended to include other available variables.

An additional problem with the logistic regression methods is that the way the data is formulated invalidates the regression assumption of independence. The algorithm works by creating a set of "hypothetical partners" for each male, assigning a one to the observation for that male and his actual partner and each hypothetical partner with the same characteristics, and a zero to each other hypothetical match. It then uses logistic regression to generate the probability of the match. In reality, a person can only take one partner, so the additional "hypothetical partners" who are created are not independent observations, and therefore this has the potential to create bias in the logistic regression estimates.

A third approach that is seen more often in agent-based models is to use a set of sociological rules or heuristics for the agents in the model to make their partnership choices. These rules may be deterministic, or they may incorporate some kind of stochastic element into the decision-making process.

Simão and Todd (2001, 2002, 2003) have progressively developed a set of matching heuristics for a one-dimensional, normally distributed "quality" measure. At each time step in the simulation, each agent randomly encounters another agent who they have not previously encountered. If each agent is single, they are matched as a "date", provided their quality scores are higher than each other's "threshold" value. These threshold values reduce over time to mimic satisficing behaviour. If either of the agents is already "dating", the new suitor must not only have a quality score that exceeds their threshold, but must also have a quality score that is greater than that of their current partner (referred to as their aspiration level). Once a pair of agents has been dating for a certain number of time steps then they are permanently matched and removed from the system. There is still a probabilistic element to this simulation in that the agents encountered one another randomly, but the matching process itself is determined by the threshold and aspiration values of the agents. Hills and Todd (2008) reconsidered the matching heuristic and instead decided to have agents try to match a certain number of common traits in order to make a match, rather than levels of a single trait. The number of matching traits decays over time in the same way as the threshold values of the earlier simulations.

Chen (2005) sought to extend Todd's models by examining different variations of the decision-making process. Chen's results (discussed in Section 3.1.1) demonstrated the value of "fast and frugal" satisficing algorithms. These are algorithms that match partners more quickly and efficiently, even if not all matches are optimal, as a preferred strategy to a slower searching approach. Chen was able to take the step of demonstrating algorithms in a simulation setting, rather than just relying on mathematical theorising for proof.

Kalick and Hamilton (1986) showed that a probabilistic function could be combined with a search heuristic. Their simulation model randomly paired couples for "dates". The outcome of the date would be determined by one of three heuristics; either agents had a preference for a partner with the highest level of a randomly assigned "attractiveness" trait, or they preferred a partner who was similar to themselves, or attraction was based on the average of both preferences. However, each of these heuristics also had a probability function attached to it, and a random number was drawn to make the decision to partner or not. Previous empirical work in a number of studies had shown a strong correlation in the level of attractiveness between individuals in many couples. Kalick and Hamilton's simulation showed that universal preferences for the highest level of attractiveness (as opposed to the most similar) would see their simulated populations self-organise into couples with realistic patterns of attractiveness. The different matching preferences used by Kalick and Hamilton are explored further with some abstract simulation models in Chapter 6.

Each of the three main types of matching algorithms (stable marriage/combinatorial, stochastic, and heuristics) have advantages and disadvantages. The stable marriage algorithm was mathematically elegant, but unrealistic and computationally expensive.

The stochastic approaches were empirically-based and more realistic, but some required detailed datasets. Using a set of heuristics for agents to make decisions provided flexible rules but did not take advantage of empirical information. The simulation algorithm for this study will combine elements of the stable marriage, stochastic matching and sociological heuristic algorithms. Couples will be paired from most attracted to least attracted but in a way that also combines stochastic scoring elements, availability and ordered matching. The empirical simulation modelling is introduced in Chapter 7.

3.1.6.   Network Models of Cohabitation

An alternative approach is to consider the structure of an individual's social network and then simulate it as a network of points. This idea is based on the theories of Levi-Strauss (1969), who stated that human beings cannot cognize the complexity of the possible network patterns and therefore formulate rules of social structure for their "kinship networks". This idea of networks translated conveniently into mathematical models of networks and what is known as ring structures. Network structures can form the basis for the analysis of patterns of social relationships (Pujol, Flache, Delgado, & Sanguesa, 2005).

These models have been applied to data (White, 1999, 2004; White & Jorion, 1996). However, shortest path algorithms across these interlinked networks become computationally expensive even with relatively few nodes. If it is to work with population level data, an ideal approach to modelling inter-ethnic cohabitation should try to incorporate the concept of social networks, but within a more pragmatic methodology that can deal with large data sets.

A slightly different approach to the simulation of partnerships can be seen in an article about the controlled simulation of marriage (White, 1999). This paper used the simulation of nodes and shortest path graph algorithms to simulate marriage within a small population. The simulation of these methods is too computationally intensive to be practical for any reasonable sized population such as the New Zealand Census datasets.

3.2. Incorporating the Threads

In order to examine the rates of inter-ethnic cohabitation, the changes in inter-ethnic cohabitation and the underlying social processes that may be driving these changes, a combination of the discussed methods must be employed. In order to address the research questions posed, this study combines sociological theory, statistical analysis and simulation methods.

Although some of the social theories relating to homogamy have changed over time, there are still a number of important contributions which should be included in the analysis. The concept of intermarriage rates being closely related to the demographic structure of a society, as introduced by Peter Blau (1977), is an important idea to be incorporated in the data analysis. The application of log-linear models, logistic regression and other modelling techniques provide useful insights when appropriate variables are included. By providing relative proportions of the ethnic groups and other demographic variables in an individual's geographic region, Blau's idea of constrained choice can be modelled. The inclusion of the geographic variables is also important for capturing variation across different regions, a factor which has been demonstrated to have significant effect on marriage and cohabitation patterns (Harris & Ono, 2005).

The simulation modelling, although informed by the statistical modelling, is also driven by the relevant social theory. The use of micro-macro linkages (van Imhoff & Post, 1998) demonstrates the evolution of the social process and shows how individual decisions influence the structure of society, which in turn generate change in decision making at an individual level (Coleman, 1990).

By examining the sociological, statistical and simulation literature on partnership formation some clear patterns emerge, and so do some gaps in the research. The patterns of ethnicity in partnership have been observed and analysed statistically through survey and census data using log-linear models. However, the coefficients of these models are often not interpreted, with authors relying on summary measures of goodness-of-fit rather than examining the actual relationships that the parameters measure. Little work has been done in this area in New Zealand, where the research has tended to be done by the official statistics agency using descriptive measures. The focus on most studies, both nationally and internationally, has also been on marriage rather than looking at all cohabitating couples.

The simulation models of partnering fell into three broad categories. The first was combinatorial methods such as the stable marriage algorithm. These methods had their origins in mathematical theory, and although they were mathematically elegant, they assumed that every individual had prior knowledge about every other individual to use in their decision making. It was also a computationally expensive algorithm, which would potentially require billions of iterations for a city-sized dataset. The second category was microsimulation models that tended to use empirical data to match their agents. These models tended to be either very data intensive or use very simplistic matching algorithms. The third category was agent-based models. These were more likely to use artificial data but work with more complex algorithms and matching heuristics. There seems to be a clear research gap for a model which is empirically based and uses real people, but that also gives thought to the matching process, rather than just relying on a Monte Carlo "roll of the dice" to match agents.

Chapter 4 - Data

Chapter 4 examines the census data that was used for the analysis and the provisions under which it was provided. It discusses the strengths and weaknesses of the data and describes how the couples were matched from the data. Statistics New Zealand is New Zealand's national statistical office. It administers the Statistics Act 1975 and is the country's major source of official statistics (www.stats.govt.nz). Amongst its main data collection activities is the New Zealand Census of Population and Dwellings which is conducted every five years. A census form is completed for every individual and each household in the country, detailing a variety of individual and household demographic information. This study draws on aggregate and unit-level data from the 1981 through to 2006 census data sets. The provision and usage of the census data sets and the variables provided are governed by the Statistics Act, 1975.

4.1. Statistics Act 1975

Statistics New Zealand is required to operate under the authority of the Statistics Act 1975. The act has a number of key provisions, with those of most significance to this study listed below (Statistics New Zealand, 2006b):

- Security of information provided [Section 37].
- Information furnished under the Act to be used only for statistical purposes [Section 37(1)].
- No information from an individual schedule is to be separately published or disclosed [Section 37(3)], except as authorised by the Statistics Act (the act permits others to see information from an individual schedule, but only when it is in a form that prevents identification of the respondent concerned, and then only under strict security conditions).
- All statistical information published is to be arranged in such a manner as to prevent any particulars published from being identifiable by any person as particulars relating to any particular person or undertaking [Section 37(4)].

The privacy component of the act has several implications for this research. The unit-level data is accessed only within a secure data laboratory environment and all output is scrutinised by Statistics New Zealand prior to release. The requirement of confidentiality also means that certain regional variables that could make individuals identifiable have either been concatenated or randomised. In addition, all frequencies and totals on all frequency tables and other output are required to be independently randomly rounded to base three.

4.2. Variables

Table 4.1 and Table 4.3 show the key variables for this study and their Census dataset codes. Although many variables are included in the dataset, only a limited number of variables are required for much of the analysis. Two datasets were used for the research. The first is the data that was provided for use in the secure data laboratory. This dataset is described in Section 4.2.1. The second dataset was the data that was used on the BeSTGRID computer network (outside of the data laboratory, see Section 7.2.1 for more details regarding BeSTGRID) and is described in Section 4.2.2.

4.2.1. The Data Laboratory Dataset

Table 4.1 shows the key variables that were provided by Statistics New Zealand in the data laboratory datasets. One of the most important variables for this study is the family id variable. This is the variable which is used to match the couples using the process described in Section 4.4. The family id variable is a unique family identifier that identifies each different family group[5]. The Statistical Standard for Partnership Status in Current Relationship (Statistics New Zealand, 2008a) and the Statistical Standard for Relationship Between Member in a Private Dwelling (Statistics New Zealand, 2008b) show that this variable is derived from the “relationship between members in a private dwelling”, “living arrangements” and “usual residence indicator” variables and allows for family units to be identified. When combined with the family code variable, which

[5] Family group is a subset of household group. There may be several families living within one household.

describes the role within a family group, married and de-facto couples can be identified and extracted.

Variable	Census Codes	Description/Construction
Sex	sex_code	Gender of participant, straight from census.
Ethnicity	ethnic_origin (1981) ethnic_origin1-3 (1986) primary_ethnic_group, second_ethnic_group, third_ethnic_group (1991) primary_ethnic_grp, secondary_ethnic_grp, third_ethnic_grp (1996) ethnic_grp1_code (grp1-grp6) (2001) ethnic_rand6_grp1_code (grp1-grp6) (2006)	Grouped ethnicity variable, see Section 4.3.
Age	age_code	Age in years on census night.
Highest Qualification	school_qual (1981) highest_school_qual, tertiary_qual1-3 (1986) highest_school_qual, tertiary_qualA-D (1991) highest_qualification_gained (1996) highest_qual_code (2001) highest_qual_code (2006)	Grouped by none/unknown, school, trade, tertiary.
Territorial Region	cn_reg_council01_code (1981-2001), CNRegC06 (2006)	Territorial regional council on census night.
Anonymised Area Unit	id_area_unit	Anonymised census area unit codes.
Family ID	family_id_nbr (constructed from family_code and dwelling_id, 1981-1986) id_family (1991-2006)	Unique family identifier
Family Code	family_code	Position within family
Country of birth	Birthplace_2d	Country of birth

Table 4.1 - Data laboratory variables

The key variable of interest to be examined once the couples have been matched is ethnicity, which is discussed separately in Section 4.3. A number of variables that were identified in the literature were also available for use in the data laboratory, with several also made available for the simulation model as well (see Section 4.2.2).

Geographic location can have a strong impact on ethnic patterns of partnership, through differing levels of availability of some ethnic groups and potential social factors such as the acceptance of mixed ethnicity partnerships (Harris & Ono, 2005). Due to regulations on data privacy, the geographic regions could not be so specific as to make individuals identifiable in the data. Regions were provided at two levels. At a higher level the regional territorial authority (council) on census night for each individual was provided. This provided sufficient detail to identify centres such as Auckland and Wellington. Geographic information was also provided at the "census area unit" (CAU) level, which are clusters of about 3000-5000 people. However, these were anonymised for privacy reasons so did not provide any additional information.

Age is an important variable for identifying the new (emergent) relationships (Dempsey & De Vaus, 2004). It is measured consistently across every data set and measured in whole years. Education is also of interest as a number of studies have shown that some of the variation in patterns of ethnic partnering can be explained by education (Blackwell, 1998; Callister, 1998). As with a number of other variables (see Table 4.2 for a summary) the collection of educational attainment information has not been performed consistently over the censuses. A measure of highest qualification was possible and was consistent across each census from 1986 onwards. The 1981 census included information about school-level qualifications, but not higher qualifications, so it had a more limited set of categories than the other five.

Although socio-economic factors are considered to have an impact on partnership choice (Rosenfeld, 2005), the five year "snapshot" nature of the census made the income variables unsuitable for analysing partnership formation. Since there was no way to be able to measure the income of two individuals prior to their relationship forming, it is not possible to relate individual income to partnership formation and the subsequent ethnic patterns. For example, one of the partners may have worked full-time prior to the relationship but only works part-time now. Their income and possibly their socio-economic status will have changed but this information will not be captured in the census data. However, education has been shown to be correlated with income and socio-economic status, and is often seen as a key contributor to indicators of socio-economic status (Xie et al., 2003). This means that there is still some information about socio-economic status in the data, with education acting as a proxy for socio-economic position, as well as being a variable in its own right.

Although most of the variables in each of the census data sets provided by Statistics New Zealand record the same kind of information, there is some variation over time due to different phrasings of questions and different formats for answering the questions. The degree of inter-census consistency of the variables is examined in Table 4.2. The comparability comments are based on the information on variable consistency published in the Family Wellbeing Indicators 2006 report (Milligan, Fabian, Coope, & Errington, 2006) from the Family and Whanau Wellbeing Project published by Statistics New Zealand. Solutions to the minor comparability issues are also listed in the table.

Variable	**Inter-Census Comparability**	**Actions Taken**
Sex	Identical in each dataset	None
Age	Identical in each dataset	None
Ethnicity	Changes to how the question was asked and option create some impact on comparability across time.	Appropriate grouping to rectify.
School Qualification	Broadly comparable for presence/absence of a qualification, limited comparability on attainment.	Examine highest qualification with consistent grouping.
Post-school Qualification	Broadly comparable for presence/absence of a qualification, limited comparability on attainment.	Examine highest qualification with consistent grouping.
Income variables	Broadly comparable over time.	None
Region	Comparable over time.	None
Country of birth	Identical in each dataset	None

Table 4.2 - Variable consistency over time

4.2.2. The Simulation Dataset

A second smaller dataset was produced as the input for the simulation. Permission was granted for this dataset to be stored on a secure server within the BeSTGRID computer network. This meant that it could be used for more complex simulations than would have been possible in the data laboratory environment. Datasets were created for the Auckland, Wellington and Canterbury regions for each of the census years from 1981 to 2006. Table 4.3 shows the variables used for the simulation data sets. A more detailed breakdown of the contents of this dataset is provided with the description of the simulation in Section 7.3.1.

Variable	Census Codes	Description/Construction
Sex	`sex_code`	Gender of participant, straight from census.
Age	`age_code`	Age in years on census night.

Ethnicity	`ethnic_origin (1981)`
	`ethnic_origin1-3 (1986)`
	`primary_ethnic_group,`
	`second_ethnic_group,`
	`third_ethnic_group (1991)`
	`primary_ethnic_grp,`
	`secondary_ethnic_grp,`
	`third_ethnic_grp (1996)`
	`ethnic_grp1_code (grp1-grp6)`
	`(2001)`
	`ethnic_rand6_grp1_code`
	`(grp1-grp6) (2006)`
		Grouped ethnicity variable, see Section 4.3.

Highest Qualification	`school_qual (1981)`
	`highest_school_qual,`
	`tertiary_qual1-3 (1986)`
	`highest_school_qual,`
	`tertiary_qualA-D (1991)`
	`highest_qualification_gained`
	`(1996)`
	`highest_qual_code (2001)`
	`highest_qual_code (2006)`
		Grouped by none/unknown, school, trade, tertiary.

--------------- below used to differentiate files but not explicitly reported ---------------		
Territorial Region	`cn_reg_council01_code (1981-` `2001), CNRegC06 (2006)`	Territorial regional council on census night.
Single	`id_family and family_code`	Derived from family id and code.

Table 4.3 - Simulation variables

4.3. Definitions of Ethnicity

"There has been an explosion worldwide of research on the construction of identity, of which ethnicity is just one part" (Callister, 2009).

The definition of ethnicity is a difficult issue which has a considerable, often contradictory, literature. It is still a topic of debate, as highlighted by Callister (2009) in the quote above, which comes from the foreword of the August 2009 issue of the Social Policy Journal of New Zealand dedicated to the theme of measuring ethnicity.

For the examination of ethnic patterns of cohabitation, there are several requirements. The first requirement is that the definition of ethnicity has to remain as consistent as possible over time, otherwise an examination of the changes over time are meaningless. The census question for ethnicity and the ability to choose one or multiple ethnicities has varied over the census questionnaires and will generate some variation between periods. This is especially important as self-identified ethnicity has its own level of variation over time, particularly for individuals with Maori, Pacific, Asian or multiple ethnic identities (Carter, Hayward, Blakely, & Shaw, 2009). This is compounded by the variations in the way that the ethnicity question has been asked in each census. For example, the ethnicity question in the 1981 census was written as an open-ended question where people were given a space to write their own ethnicity. Once the responses to this question were compiled, the question had over 100 different response categories.

The second requirement is that in order to conduct the statistical analysis on the data, the ethnicity categorisation needs to result in mutually-exclusive groups since any overlap in groups would invalidate conventional categorical data analysis methods. Since ethnicity categories are being generated from census data, it makes sense to examine Statistics New Zealand's ethnicity-related documentation.

At the time of making decisions about the ethnic groupings for this study, the most recent document relating to ethnic categories in census data was the Guidelines for Using Ethnicity Data: 2006 Census (Statistics New Zealand, 2007a). However, the main focus of this document is the treatment of the "New Zealander" category (see Section 4.3.1). The Report of the Review of the Measurement of Ethnicity (Statistics New Zealand, 2004) indicated there was a continued relevance and demand for ethnicity-based data. It suggested discontinuing the prioritised ethnicity format for standard output and led to the Statistical Standard for Ethnicity 2005 (Statistics New Zealand, 2005).

The Statistics New Zealand's Statistical Standard for Ethnicity 2005 (Statistics New Zealand, 2005) provided several different options which provided an acceptable degree of consistency across censuses whilst providing mutually exclusive categories. One of the main methods of categorisation involved either eight, fifteen or forty-four different categories. The set of fifteen categories provided the best balance of ethnic detail without an excessive number of categories, and is listed below.

Ethnicity Categories	Used in Analysis
1. European Only	✓
2. Maori Only	✓
3. Pacific Peoples Only	✓
4. Asian Only	✓
5. MELAA Only (Middle Eastern, Latin American, African)	✓
6. Other Ethnicity Only	
7. Maori & European	✓
8. Maori & Pacific Peoples	✓
9. Pacific Peoples & European	✓
10. Asian & European	✓
11. Two Groups Not Elsewhere Included	
12. Maori & Pacific Peoples & European	
13. Three Groups Not Elsewhere Included	
14. Four to Six Groups	
15. Not Elsewhere Included	

Table 4.4 - Ethnic groupings

The first five groups are for individuals who nominated a single ethnicity, or only ethnicities from a single group. The MELAA (Middle Eastern, Latin American, African) group is treated as a "catch all" of a number of areas. Due to its small size, and the broad range of ethnic groups that it covers, it is only used for some of the analysis. The four largest dual ethnicity groups are also included with the five single ethnicity ones. The use of the Asian Only and Pacific Only groups does have the disadvantage of not being able to distinguish specific sub-groups within them. For example, a "heterogamous" partnership between a Samoan person and a Tongan person would be reported as a "homogamous" one between two Pacific Only people. However, the next level of ethnic detail would have created privacy concerns due to small cell counts in some of the groups, and would have also significantly reduced the parsimony of the tables and the models.

The frequency tables showing the ethnicities of the couples are shown in Appendix A. To keep these tables relatively parsimonious, groups 13 to 15 were concatenated into the "Not Elsewhere Included" category. The statistical analysis required a further reduction of the data because the frequencies were too low in some of the groups. When examining the patterns of homogamy with proportions and log-linear models, the "other" categories were also removed, since it was not possible to tell whether a partnership where each partner is in an "other" category is homogamous or not. Overall, the individuals in the "other" categories made up only a small part of the total data set, comprising less than 5% of the data in five of the six census datasets (see Table 4.5).

Year	Total Frequency "Other" Categories	Total Frequency	Percentage of Total Frequency
1981	17,538	1,325,412	1.32%
1986	44,268	1,422,450	3.11%
1991	44,310	1,461,690	3.03%
1996	92,460	1,555,596	5.94%
2001	61,852	1,559,160	3.97%
2006	84,954	1,735,236	4.90%

Table 4.5 - Individuals in the "other" categories

Other studies have addressed the issue of "other" ethnicities in one of two ways. They have either omitted the individuals in those groups, rationalising that they cannot be interpreted as homogamous or heterogamous, or they have retained the individuals in an "other" category, but not deliberated too much over their findings. Callister *et.al.* (2005), working with New Zealand census data, chose to include an "other" group in their examination of ethnic intermarriage. However, they do acknowledge that the patterns that involve the "other" category are not readily interpretable since there is no way of knowing whether the matches of "other" with "other" are two people of the same ethnicity. By comparison, studies such as those by Qian and Lichter (2007), and Khoo *et.al*. (2009), conduct their analyses on explicitly stated ethnic groups only, and exclude anyone who is listed in an "other" category because of the lack of interpretability of such categories. For this study, the "other" category is removed. It represents a small proportion of the total data, and does not provide a group that is of practical use in the analysis.

4.3.1. The "New Zealander" Category

The "New Zealander" category became a new ethnicity option in the 2006 census. Statistics New Zealand produced a reference document regarding statistics on this category (Statistics New Zealand, 2007c). It found that individuals identifying themselves as New Zealander were more likely to be male, had higher regional proportions in the South Island than in the North Island, and were most often born in New Zealand. They were also less likely to be of Maori descent. Figure 4.1, from the 2009 Draft Report of a Review of the Official Ethnicity Statistical Standard (Statistics New Zealand, 2009), shows the change in European and "Other" categories as a result of adding the New Zealander category to the 2006 Census. The Statistics New Zealand guidelines for the 2006 Census (Statistics New Zealand, 2007a) suggest that the optimal treatment of the New Zealander category is to merge it with either the European or Other Ethnicity grouping, producing either a "European and Other Ethnicity (including New Zealander)" or a "European (including New Zealander)" group. To achieve the goal of studying ethnic partnership patterns and identifying ethnic homogamy, the latter is chosen.

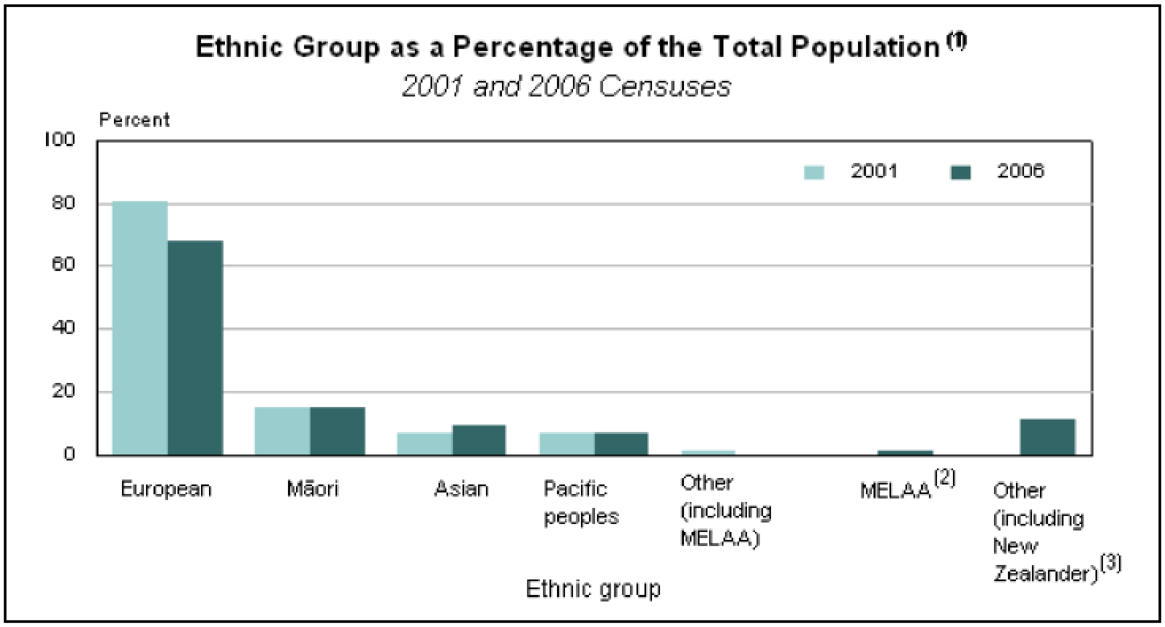

Figure 4.1 - New Zealander ethnic group (Statistics New Zealand, 2009)

4.4. Constructing the Couples Data

It is important that the data remains as consistent as possible over time. In order to generate the sets of couples, the following process was used for each Census dataset:

1. All individuals must be adults living at a private dwelling
    a. Remove those aged under 16.
    b. Remove those at non-residential dwellings.
2. Couples are identified by matching family id codes.
3. Same sex couples are removed. Analysis is currently limited to heterosexual couples.
4. Emergent couples are considered to be those with a male partner aged 30 and under.
5. Couples where both partners were born overseas are flagged as immigrant couples.

Figure 4.2 shows a flow chart documenting how the couples were matched up. Each census had slightly different variable names but the same process was followed for each to ensure consistency over the censuses.

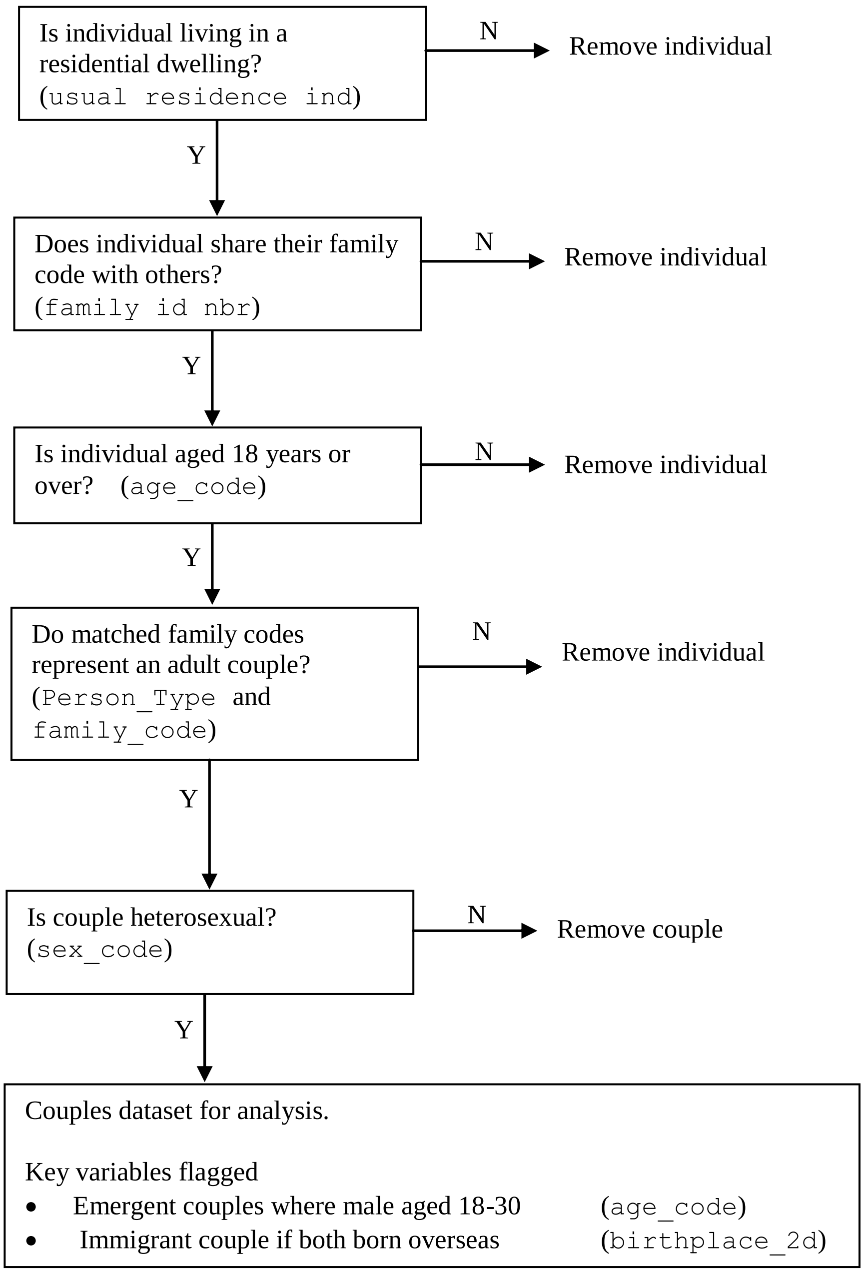

Figure 4.2 - Diagram of couples dataset construction

Since the census is a self-complete survey there is the possibility that relationships may be misreported, either unintentionally due to respondents being confused by how to answer the living arrangements question, or intentionally if they want to conceal their relationship. For example, a beneficiary may not want to report that they are living with a partner (Statistics New Zealand, 2008b). Unfortunately, there is no solution to this problem, other than to note that the number of couples may be underreported.

An initial examination of the partnership patterns showed some clear immigration-based patterns, caused by couples of the same ethnicity immigrating to New Zealand. This was particularly noticeable in the Asian Only and MELAA groups. Since the study is focused on partnership formation in New Zealand, it was important to develop criteria for excluding couples where it was likely that the relationship was formed overseas. The more recent censuses have included a question on how long an individual has lived in New Zealand, but the length of time that people have been in their current relationship is not measured. One variable that has been collected in all six censuses is country of birth. Although this does not provide a perfect indicator of where a partnership was formed, it allows couples where both partners were born overseas to be identified. This means that the couples where at least one partner was born in New Zealand can be analysed separately. The disadvantage of this process is that it may eliminate some couples where, even though both individuals were born overseas, the relationship was formed in New Zealand. However, this is balanced out by the need to separate the relatively large number of immigrant couples whose relationships were formed overseas.

The other variable that is used for separating some of the couples is age. Section 2.1.2 introduced the idea of patterns of emergence compared to patterns of prevalence (new relationship formations compared to the existing ones). Since the length of relationships is not recorded in the census, couples where the male partner was aged between 18 and 30 is used as a proxy for emergent couples.

The SAS code for matching the couples is shown in appendix B.1.

Chapter 5 – Descriptive Statistics & Statistical Modelling

This chapter uses the census frequency tables in Appendix A to address research question one (see Section 1.2): what changes have occurred in inter-ethnic cohabitation patterns in New Zealand over the six census periods, 1981-2006. It uses a series of descriptive graphs and tables to show the changes in the proportion of homogamous couples over time. Log-linear models are then applied to the tables, to examine the patterns in the data independently of the size of the ethnic groups.

The analysis in this chapter is conducted on a reduced form of the ethnicity tables, with the "other" categories removed. This is done because a couple with each partner in the "other" category is not necessarily homogamous. It could be that a couple who appear on the diagonal of the table with both ethnicities as "other" could have completely different ethnicities, meaning that the analysis would classify them as homogamous even though they are not. There is a similar issue for the Pacific Only and Asian Only categories. For example, a Pacific Only/Pacific Only couple does not necessarily represent a precisely homogamous couple since, for example, one could be Samoan and the other could be Tongan. However, in order to maintain relatively parsimonious tables and analyses, these groups have not been expanded. The "other" category is also a relatively small group, so removing it has a negligible impact on the analysis.

5.1. Descriptive Statistics

This section provides an initial examination of partnership patterns by examining the changing number of partnerships over time and the percentage of these partnerships which are homogamous. The data is broken down by other variables to isolate where the changes in partnership patterns are occurring. The proportion of homogamous partnerships for each ethnicity are graphed for the six census periods and the frequency tables are included in Appendix A for the verification of the figures.

5.1.1. Number of Partnerships

Table 5.1 shows the number of partnerships (married and de-facto) at each time period. The total number of couples with at least one New Zealand-born partner and the total number of couples where both partners were born overseas has increased at every census since 1981. The number of couples where the male partner is aged between eighteen and thirty has decreased over this period.

Year	Number of couples (at least one partner NZ born)	Number of couples (non-NZ born)	Number of couples (male partner 18-30, NZ born)	Total Number of Couples
1981	583104	79599	124746	662706
1986	621288	89937	114777	711225
1991	632580	98265	103944	730845
1996	660891	116907	98529	777798
2001	650787	128793	80865	779580
2006	691827	175791	80850	867618

Table 5.1 - Number of couples in New Zealand

One of the few published articles that can be used to compare to these figures, and the tables in Appendix A, is the Official Statistics Research Series paper: *Ethnic Intermarriage in New Zealand* (Callister et al., 2007). In this paper, Callister *et.al.* examine interethnic marriage using data from the 2001 census, although they do not compare these figures to other census periods. As with this research, Callister *et.al.* use the social definition of marriage, that is, couples who are legally married, and those living in de facto relationships. As expected, their key findings, which are discussed later in this chapter, are very similar. However, there are some differences between the frequency table counts, due to slightly different treatments of the data. The Callister *et.al.* article shows a total of 730,335 couples (Callister et al., 2007, pg 29), compared to 779,580 in this research. This difference is because Callister *et.al.* removed couples with undefined ethnicities, rather than categorising them as “not elsewhere included” as they are here. There are also some variations in some of the cell frequencies. The very small differences can be attributed to the random rounding that is required by Statistics New Zealand. The slightly larger differences have come about due to a different process of separating the ethnic groups. The groupings described in Section 4.3 and used throughout this dissertation are mutually exclusive, whereas a number of the tables in the

Callister *et.al.* paper have double counting for people who nominated more than one ethnicity (i.e. someone who nominated Maori and European ethnic groups would be counted once in each group).

5.1.2. Proportion of Homogamous Partnerships - Total

The initial investigation of homogamy will examine the percentage of partnerships that are homogamous within each ethnic group. These percentages have been graphed over time for the nine key ethnicity groups identified in Section 4.3. They were calculated by dividing the number of couples in each of the diagonal cells of the tables in Appendix A by the row totals (male proportions) and column totals (female proportions). The single ethnicity groups are plotted on a separate graph to the dual ethnicity groups for clarity of display. It should be noted that the relative group sizes have an impact on the size of the proportions. For example, most European individuals would be expected to have a European partner simply because this is a far larger group than any other. This technical issue will be addressed with log-linear models in Section 5.2.

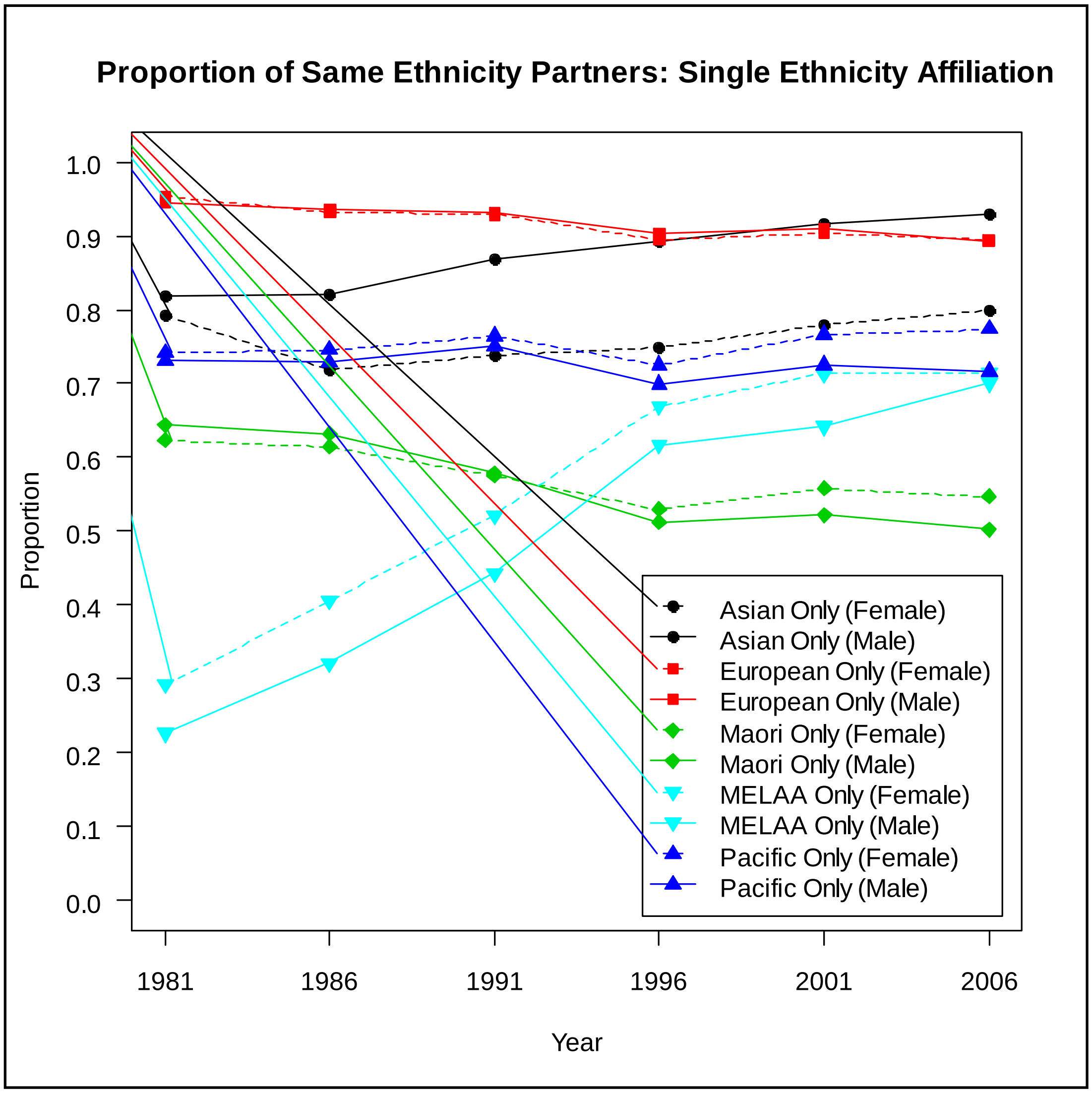

Figure 5.1 - Proportion of homogamous partnerships (single ethnicity)

Figure 5.1 shows the proportion of homogamous partnerships for each of the single ethnicity groups. Overall, the European ethnicity has the highest proportion of within-ethnicity coupling, followed by Asian, Pacific and Maori. Over the six census periods, the proportion of homogamous partnerships for the European, Pacific and Maori groups have decreased. The Asian and MELAA groups both appear to have an increasing percentage of homogamous partnerships over time. However, a closer examination of the frequency tables in Appendix A suggests that the increases actually represent immigration patterns, with the proportions increasing over time through the migration of couples of the same ethnicity. The proportions for males and females in most of the ethnic groups are similar, and follow similar patterns over time. The exception to this is the divergence in the proportions for the Asian male and Asian female groups. Although the proportions for both groups are increasing, the proportion for the Asian males is increasing at a greater rate than the proportion for the Asian females. This divergence will need to be confirmed when the data is re-examined without the immigrant couples.

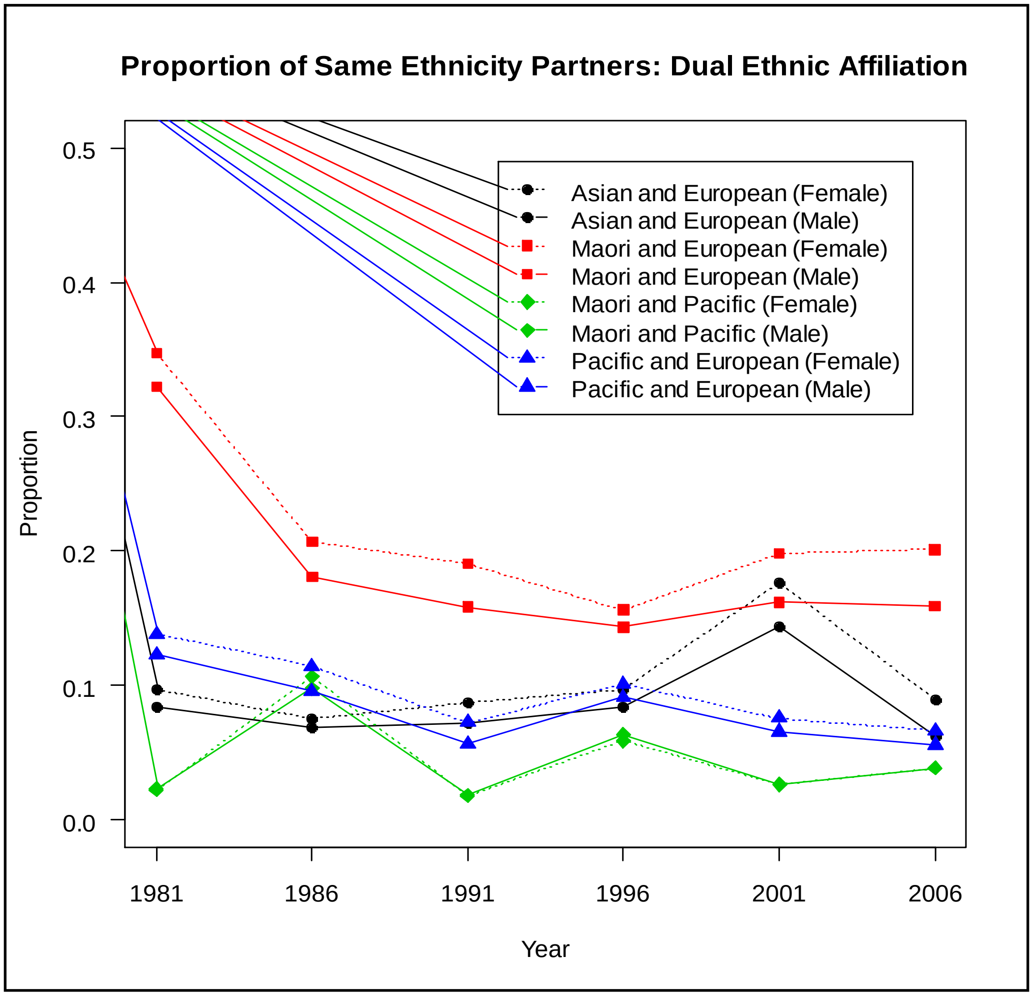

Figure 5.2 - Proportion of homogamous partnerships (dual ethnicity)

The plot of the percentage of homogamous partnerships for dual-ethnicity individuals, seen in Figure 5.2 shows that the proportions of individuals with a partner of the same ethnicity tend to be much lower than those reported in the previous figure for individuals with a single ethnic affiliation. Most of the groups have stayed fairly consistent over time, although the Maori & European group saw a decrease between 1981 and 1996.

The difficulty with examining this data for dual ethnic affiliation is that they do not show individuals who have a partner with a part ethnicity, for example, the proportion of those with dual Maori & European affiliation having a Maori only or European only partner. Table 5.2 shows that despite a low proportion of individuals with dual Maori & European affiliation having a partner with the same ethnicity, they are much more likely to have a partner who is either Maori only or European only than, say, Asian. Section 5.2.5 introduces crossing-parameter models as a way of measuring these partial matches of ethnicity.

Maori & European		**Ethnicity of Partner**		
Dual Ethnicity		Maori & European	Maori	European
1981	Male	34.7%	10.1%	51.4%
	Female	32.2%	12.3%	50.5%
1986	Male	20.7%	11.4%	62.1%
	Female	18.1%	13.9%	59.2%
1991	Male	19.1%	10.8%	63.1%
	Female	15.8%	14.8%	59.9%
1996	Male	15.6%	10.2%	63.3%
	Female	14.3%	13.7%	59.0%
2001	Male	19.8%	9.1%	63.1%
	Female	16.2%	13.9%	58.1%
2006	Male	20.1%	9.3%	60.6%
	Female	15.9%	14.0%	57.2%

Table 5.2 - Partner proportions for the Maori & European dual ethnicity group

The next step in the exploratory analysis of the data is to break the data down using other variables of interest, such as country of origin, age, and legal marital status, and then re-examine the proportions of homogamy to see if there are any changes in the patterns.

5.1.3. Proportion of Homogamous Partnerships – New Zealand Born

One of the variables that could potentially obscure partnership patterns in the aggregate figures is whether or not a partnership was formed overseas. If a couple form a partnership in another country and then immigrate to New Zealand together, their partnership is more likely to be ethnically homogamous, but is not a reflection on the New Zealand "marriage" market. Since there is no indicator in the census of where a partnership was formed, country of birth is used as a proxy for immigrant status. By examining only the couples where at least one partner was born in New Zealand, the bias created by same ethnicity couples whose partnership was formed in another country, and therefore under different social conditions, is reduced. Figure 5.3 and Figure 5.4 show the proportion of homogamous couples for the different ethnic groups for couples where at least one partner was born in New Zealand. These couples represent between 80% (2006) and 88% (1981) of the total number of couples.

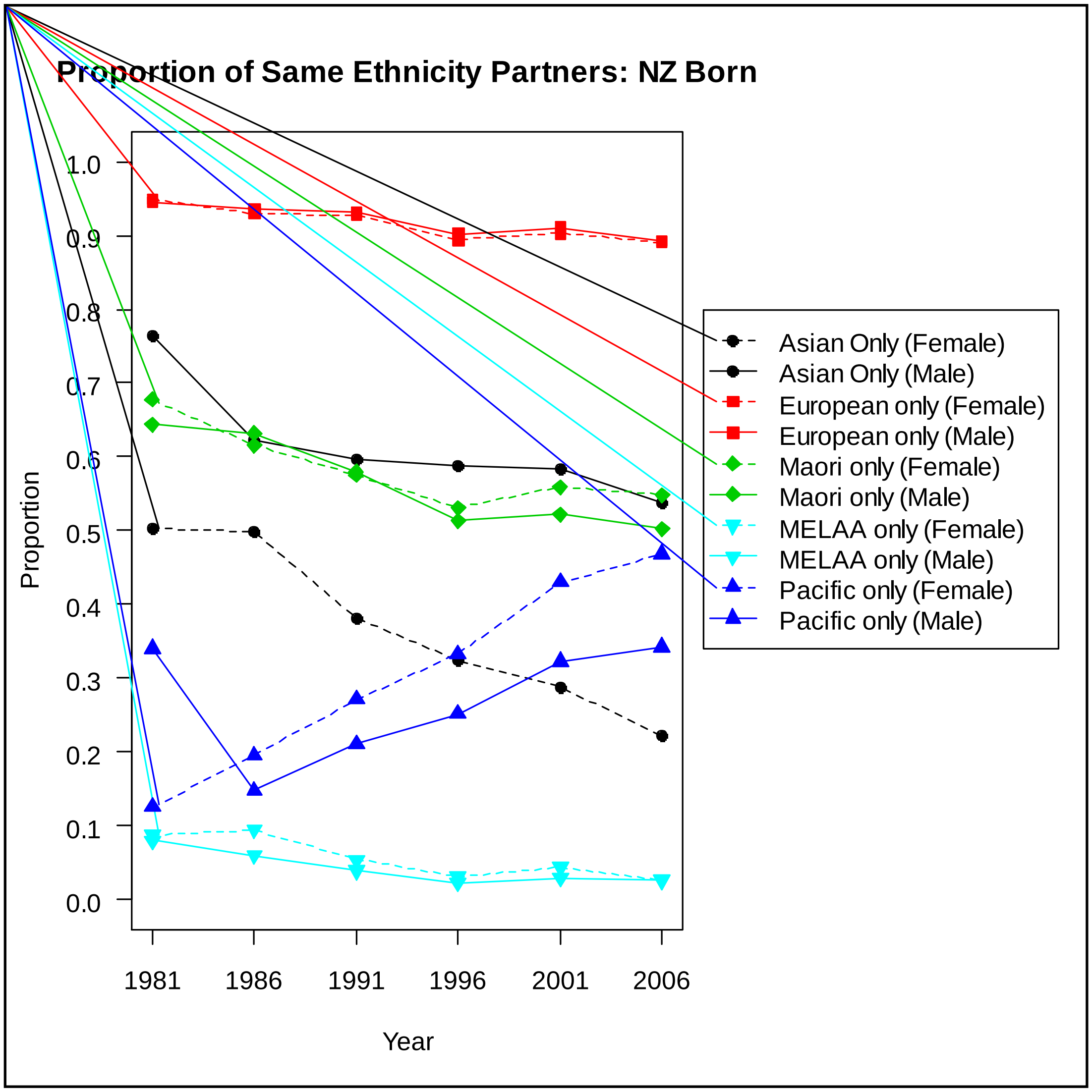

Figure 5.3 - Proportion of homogamous partnerships (single ethnicity, NZ born)

The patterns for the European Only and Maori Only groups in Figure 5.3 are very similar to those in the previous graph. There was a very low number of Maori Only born overseas, so the frequencies for this group were largely unaltered. For the European Only group the frequencies were smaller, but the high proportion of homogamous couples and the slight decrease in this over time remained the same. After removing the couples who were born overseas, the proportion of homogamous couples for the Asian Only males and females now show a decreasing trend over time. There is still a divergence in the proportions, with the proportion of Asian Only females in homogamous partnerships decreasing at a much faster rate than their male counterparts. One of the key drivers of this divergence is the asymmetry in the rates of European/Asian partnerships (Callister et al., 2007), with the tables showing that the number of European males partnered with Asian females was about five times that of Asian males partnered with European females in 2006. The effect of removing the foreign-born couples is even more pronounced in the MELAA Only group, with the steep increase in the proportion of homogamous couples for this group being replaced with a slight decrease. A large change is also seen in the Pacific Only group. Once the foreign-born couples are removed, the pattern for the proportion of homogamous couples changes from a fairly constant proportion of between 70-80% homogamous couples to a much lower initial proportion that is consistently increasing over time. This indicates that over time there has been a shift in the Pacific population towards a greater rate of homogamous partnering. However, Callister *et.al.* (2007) notes that care must be taken in this interpretation, as it represents the "average" trend across all the Pacific ethnicities.

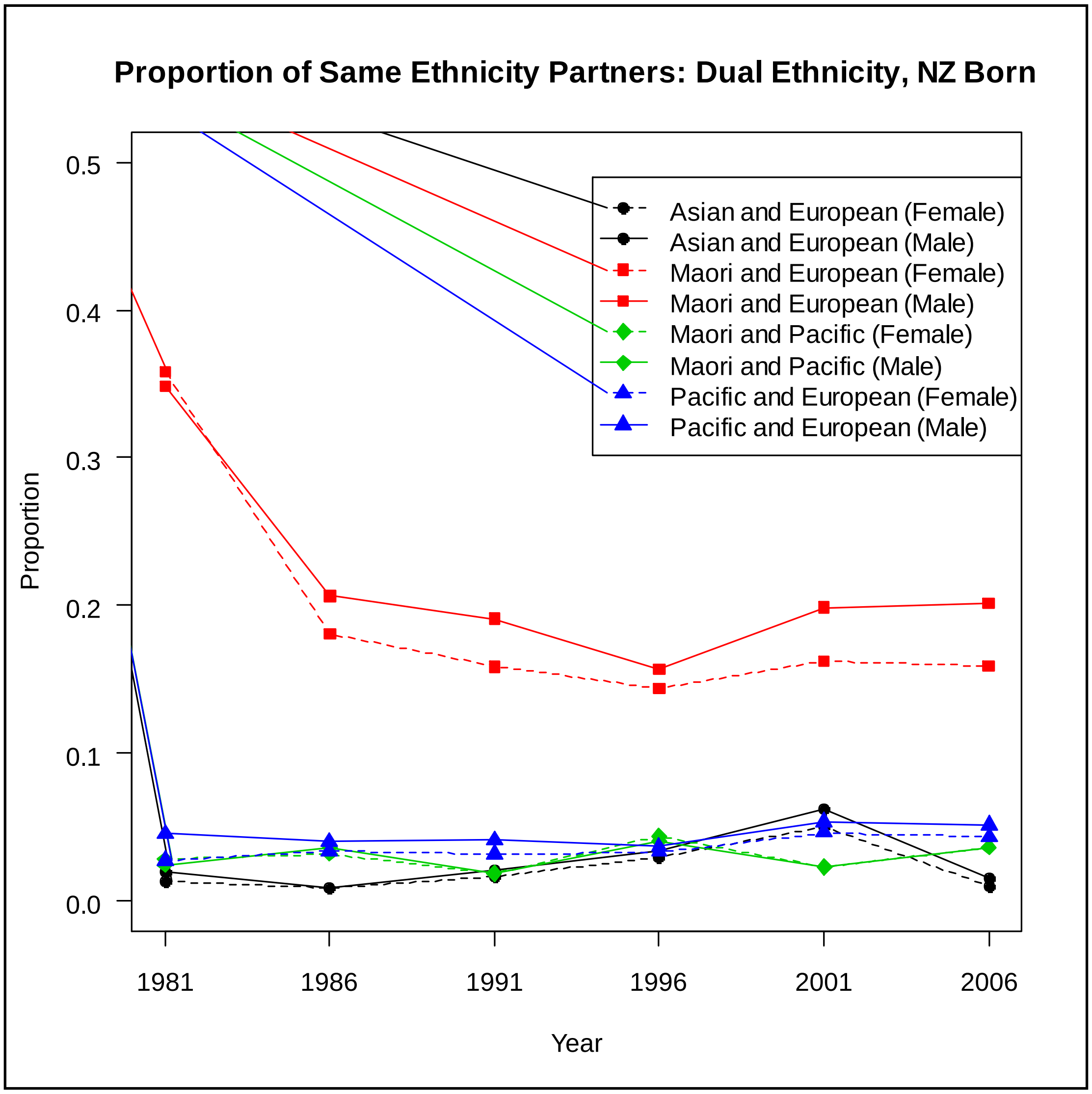

Figure 5.4 - Proportion of homogamous partnerships (dual ethnicity, NZ born)

By comparison, Figure 5.4 shows little change in the patterns for the dual ethnicity groups once the foreign-born couples are removed. In order to separate the effect of immigration from any other effects, the remainder of the analyses are conducted on the data with foreign-born couples removed.

5.1.4. Proportion of Homogamous Partnerships – Emergent Partnerships

An important consideration in the analysis of partnership data is the difference between new partnerships and existing partnerships. These emergent partnership patterns, which are an area of interest as they tell you about social change, are most likely to be detected in new (incident) rather than existing (prevalent) partnerships. The census does not record length of partnership, so a proxy must be used. Harris and Ono (2005) addressed this problem by creating a dataset which only included young couples, as a way of measuring the emergence of partnership trends. For this analysis, couples where the male partner is aged between 18 and 30 are treated as emergent. These couples represent between 11.7% (2006) and 24.0% (1981) of the total number of couples with at least one partner born in New Zealand. The MELAA group is not included in these groups due to the very small frequencies once the data is divided by age and country of origin.

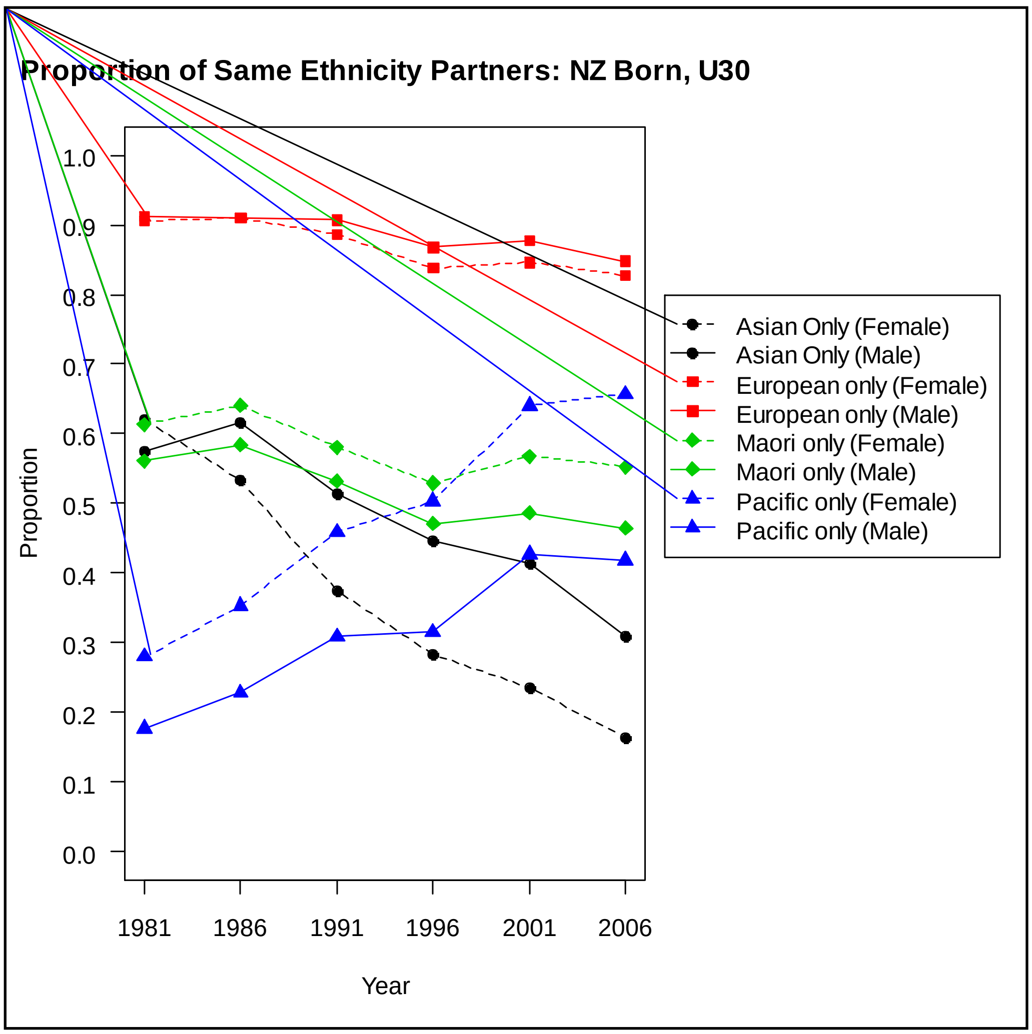

Figure 5.5 - Proportion of homogamous partnerships (single ethnicity, NZ Born, Under 30)

Figure 5.5 and Figure 5.6 show that there are some differences in the proportions of homogamous partnerships between the younger couples (male partner 18-30) and the older couples. The patterns for most of the single ethnicity groups and all of the dual ethnicity groups (not shown) are similar between emergent partnerships and existing partnerships. The biggest difference is in the Pacific Only group, where the proportion of homogamous partnerships for younger couples is consistently higher than it is for the older couples. Amongst the younger Pacific Only couples the gap between male proportion and the female proportion for the Pacific only group is also much greater, with females having a much higher proportion of homogamous partnerships.

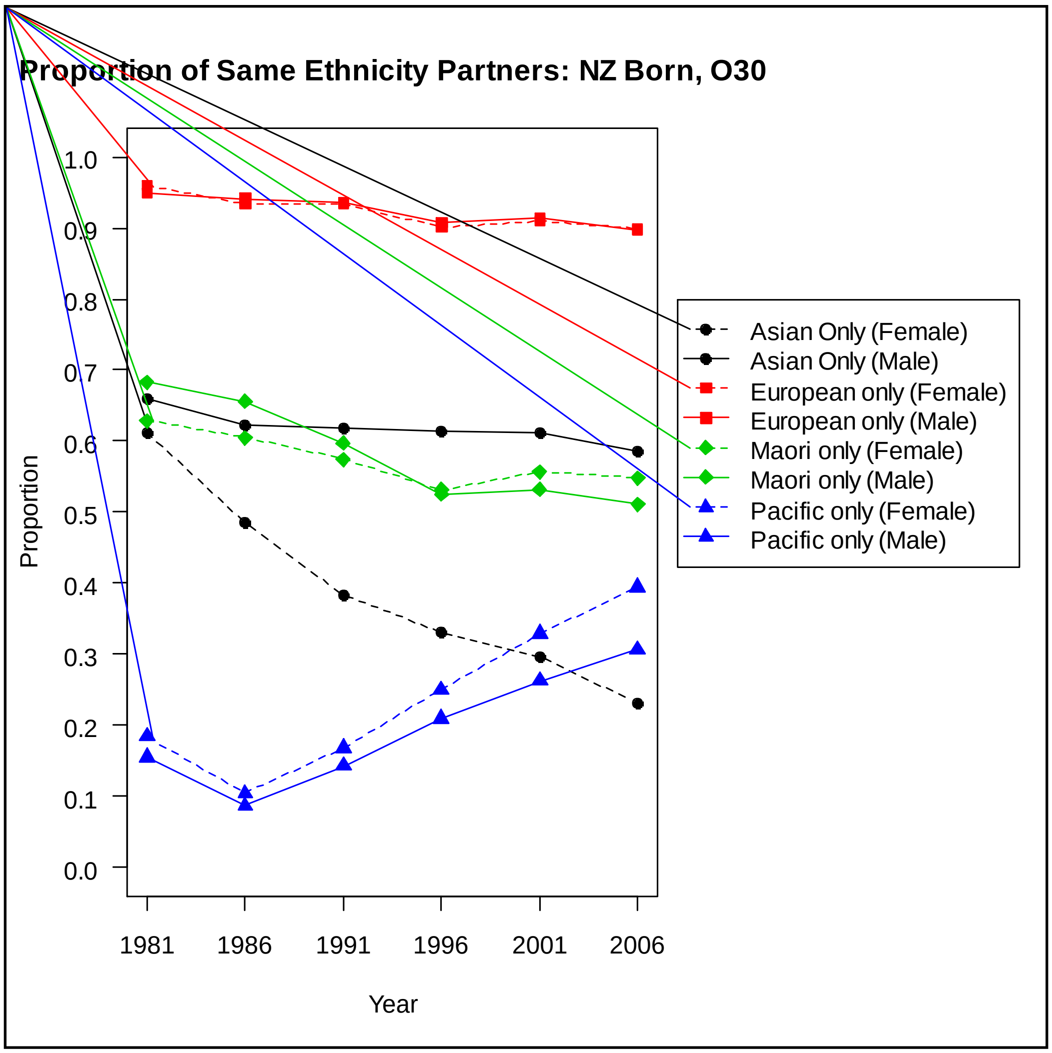

Figure 5.6 - Proportion of homogamous partnerships (single ethnicity, NZ born, Over 30)

5.1.5. Proportion of Homogamous Partnerships – Married vs. De-Facto

Although the focus of this study is on all cohabiting couples, the differences between the proportion of married couples and the proportion of de-facto couples in homogamous relationships is also of interest. Figure 5.7 shows the proportion of marriages that are homogamous amongst the single ethnicity groups. Figure 5.8 shows the proportions for the de-facto couples.

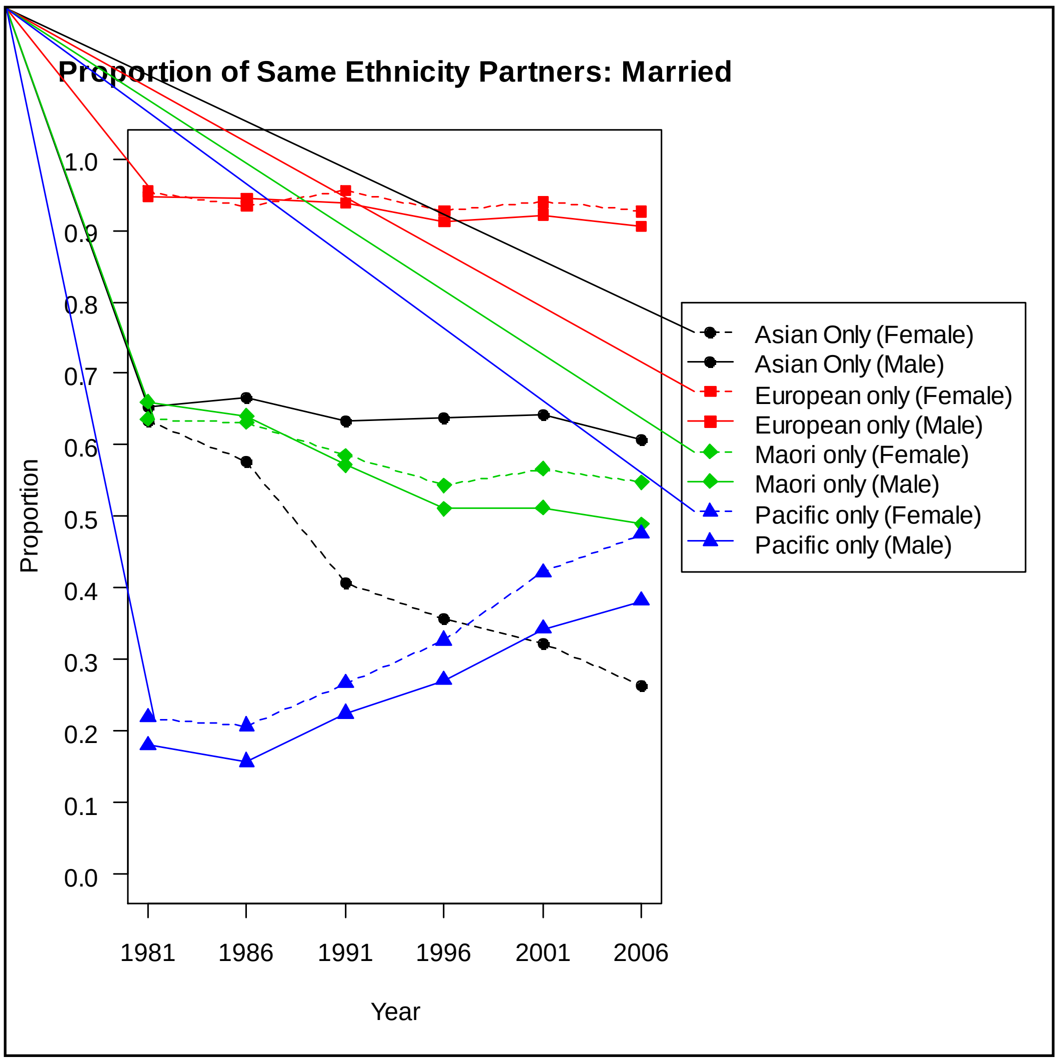

Figure 5.7 - Proportion of homogamous marriages (single ethnicity, NZ born)

Figure 5.7 shows that the patterns in the proportion of same ethnicity partnerships for married couples are very similar to those seen in the general population. By comparison, the proportions for de-facto couples in Figure 5.8 show that the de-facto couples have lower proportions of same ethnicity partnerships than the married couples, particularly for the Asian Only group. There is also more variation in the patterns over time for the de-facto couples. However, it should be noted that in the earlier census periods, de-facto relationships comprised a very low proportion of total couples. For example, in 1981 only 6% of cohabitating couples were in de-facto relationships, with the remainder being married. There is the possibility the change in the social acceptability of cohabitation outside of marriage has seen not only an increase in the number of de-facto relationships, but also in the reporting of them in the census.

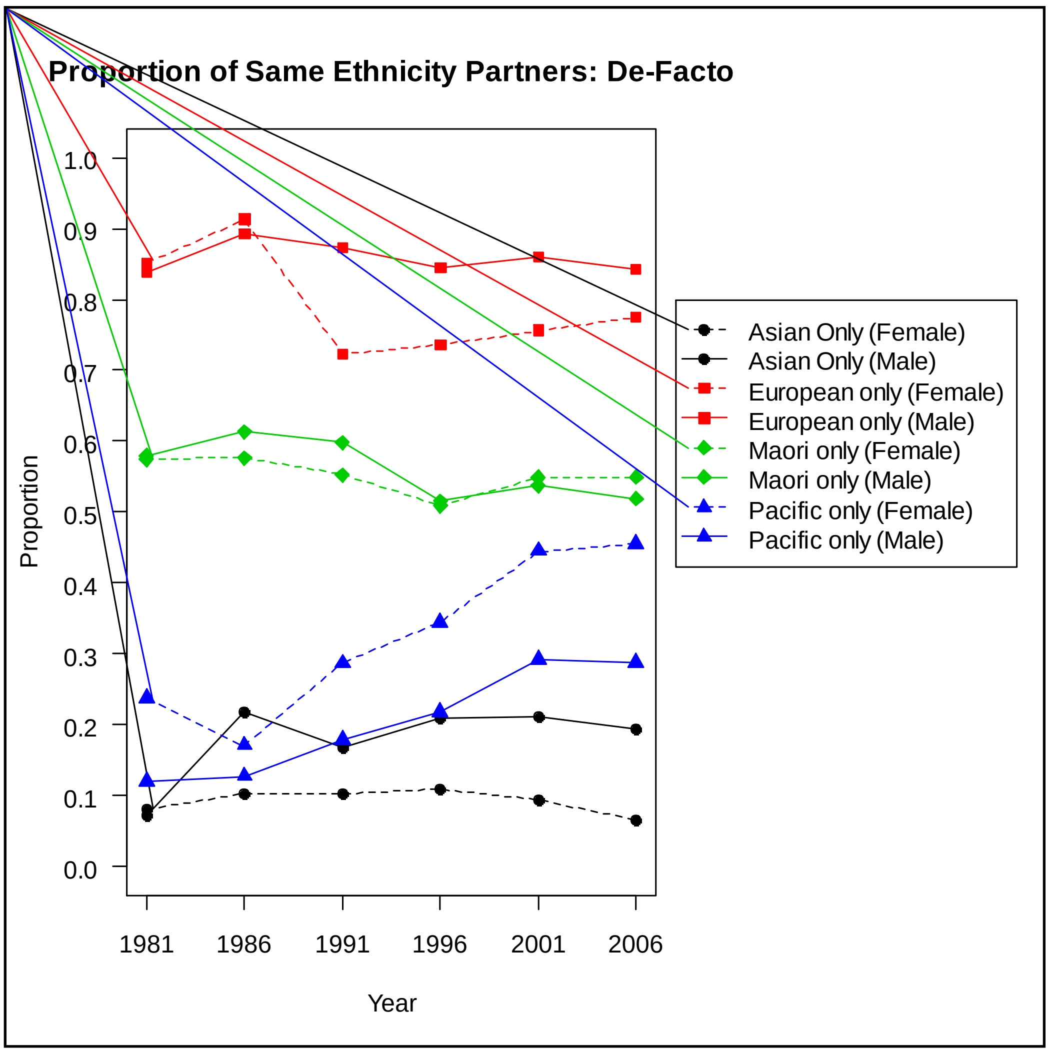

Figure 5.8 - Proportion of homogamous de-facto partnerships (single ethnicity, NZ born)

In summary, the plots of the proportion of homogamous partnerships show that both males and females in the European Only group have the highest proportion of homogamous relationships. Although this has decreased slightly over time, it is a pattern that occurs independently of age group, immigrant status and marital status. The smaller ethnic groups all had lower proportions than European Only. The steep increase in the proportion of homogamous couples in the MELAA group prompted the need to separate the couples whose relationship may not have formed in New Zealand. Examining only the couples who had at least one New Zealand born partner showed similar proportions decreasing over time for most of the ethnic groups. However, the Pacific Only group showed an increase. Individuals with two ethnic groups had lower proportions of homogamous partnerships than those with a single ethnicity, but this did not take into account partial matches of ethnicity. The patterns for couples where the male partner was aged between eighteen and thirty years old were similar to those where the male partner was over thirty, although there was a divergence in the proportions for males and females in the Pacific Only group for the under thirty group. The proportions of homogamous partnerships for all groups were lower for de-facto couples than married couples.

Using the percentage of homogamous partnerships for each ethnicity as the measure of homogamy means that the ethnicities are not directly comparable with each other. This is because these percentages are directly proportional to the size of each ethnic group. Section 5.2 uses log-linear modelling to control for the relative sizes of the different ethnic groups in order to examine the patterns of ethnic partnership independently of group size.

5.2. Log-Linear Modelling

Section 2.2 introduced the concept of log-linear modelling, where the log of the cell frequencies in a table are modelled using the row and column variables of the table, and other parameters, such as whether a cell is on the diagonal of the table.
Log-linear models provide the opportunity to examine the relationships between the ethnicities of cohabitating couples in a way which is not affected by the marginal distributions. This means that meaningful comparisons can be made between groups without the row and column totals influencing the parameters of the model (as they do where proportions are used, in Section 5.1). For example, if a population which consisted of 100 European people and 10 Maori people had five inter-ethnic couples, this would suggest a large proportion of Maori but a low proportion of Europeans form mixed-ethnicity couples. However, the difference in these proportions is due to the relative total size of each group, and therefore the number of available partners, rather than necessarily suggesting differences in the propensity of each group to form mixed ethnicity couples. Log-linear models allow the propensity of each group to be examined by modelling the log cell frequencies rather than the proportion of couples in the diagonal cells (which are related to the row and column totals).

Log-linear models are commonly used in the sociological literature to measure homogamy across various different variables (Kalmijn, 1998; Mare, 2001; Qian & Lichter, 2007; Schwartz & Mare, 2005). Qian and Lichter (2007) refer to them as the gold standard for analysing intermarriage partners. They provide a way of examining the levels of homogamy, and tracking how these change over time, independent of the marginal distributions of the ethnicities. This allows for inter-group comparisons as well as intra-group comparisons since the rates of homogamy are not dictated by the size of the group. Logistic regression models are then used to examine what variables have an influence on the probability that a partnership will be homogamous.

The basic form of a log-linear model (as described in Chapter 2) examines the effect of the row and column variables on the log frequencies of the cells as shown below:

$$\log m_{ij} = \mu + \lambda_i^{mEth} + \lambda_j^{fEth}$$

where μ is the overall mean, the m_{ij} values are the cell frequencies and the λ_i^{mEth} and λ_j^{fEth} parameters represent the effect of the male ethnicity (row) and female ethnicity (column) variables. This form of the model is often referred to as the independence model as it assumes that the variables are not associated.

When evaluating a log-linear model, the residual deviance is used as a measure of goodness of fit (Agresti, 2002, p. 452). It provides a score based on the difference between the observed and expected log cell counts. The residual deviance scores of different models can be compared to one another to compare how well each model fits the data, with lower residual deviance scores indicating a better fit to the data. It can also provide insights by allowing you to examine the residuals cell by cell. The Akaike Information Criterion (AIC) is also used as a measure of goodness of fit for log-linear models. The main difference between the two measures is that the AIC also incorporates the complexity of the model into the score, with more complex models having a higher (and therefore less desirable) score. Another measure that is seen in some of the literature is the Bayesian Information Criterion (BIC) (Mare, 2001; Qian & Lichter, 2007). The BIC is a maximum likelihood measurement similar to the AIC, but it also includes a stricter penalty term for the number of parameters in the model to overcome the overfitting problems that can be encountered with AIC.

Most goodness-of-fit measures are, in part, a function of the number of observations in the data. This can create problems when dealing with very large frequencies, such as those observed in census data. Simonoff (2003) explains that problems can occur with hypothesis tests and goodness-of-fit measures for very large datasets, as virtually any null hypothesis that is tested will be rejected, even when the observed deviation is very small and may be of little or no practical significance. This is illustrated by Kuha (2004), who examines AIC and BIC scores for a number of published data sets, including occupational mobility data (Hazelrigg & Garnier, 1976), and concludes that for large data sets, all reasonably parsimonious models may be rejected for a lack of fit. For this reason the log-linear results presented do not include AIC or BIC scores, but instead focus on the model coefficients.

The independence model can be thought of as a random mixing model or a model without homogamy or heterogamy. It can be extended to form more complex models such as the quasi-independence and crossing parameter models. These particular models add additional parameters to measure the homogamy within a table over and above the independence model.

	Female Ethnicity								
Male Ethnicity	European Only	Maori Only	Pacific Only	Asian Only	MELAA Only	Maori & European	Maori & Pacific	Pacific & European	Asian & European
European Only	50.6	-139.04	-51.97	-22.88	-0.27	-44.38	-23.56	-20.59	-4.82
Maori Only	-131.78	238.76	12.87	-22.39	-2.12	41.25	21.5	8.32	-1.09
Pacific Only	-68.08	42.32	122.04	-10.23	-2.05	23	27.48	36.42	2.64
Asian Only	-44.18	-9.16	-1.23	120.3	-0.47	-7.35	0.31	-0.57	10.6
MELAA Only	-0.98	-2.6	0.87	-0.9	11.06	2.19	-1.39	2.68	1.64
Maori & European	-33.62	17.2	4.24	-7.84	-1.39	83.64	9.6	18.69	5.24
Maori & Pacific	-19.94	21.74	15.29	-0.64	-1.14	14.79	8.61	7.68	-1.34
Pacific & European	-14.73	6.06	18.91	-0.14	2.43	20.16	5.39	20.16	3.16
Asian & European	-3.62	-2.93	-1.67	8.57	2.34	3.37	-1.2	3.29	15.29

Table 5.3 - Deviance residuals for the independence model fitted to 2001 (NZ born) data

Table 5.3 shows the set of deviance residuals for the independence model fitted to the 2001 census data and highlights the need for additional parameters to be used in order to better fit the data. The generalised form of the deviance residual is given as:

$$\operatorname{sign}(y_i - \hat{\mu}_i)\sqrt{|d_i|} \quad \text{where} \quad d_i = 2\omega_i\left[y_i\left(\tilde{\theta}_i - \hat{\theta}_i\right) - b\left(\tilde{\theta}_i\right) + b\left(\hat{\theta}_i\right)\right]$$

where d_i is the contribution of the ith cell to the total residual deviance.

Poisson model: $$d_i = 2\left(y_i log\left({}^{y_i}/_{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right)$$

where y is the number of observations and $\hat{\mu}$ is the fitted mean.

The table features large positive residuals down the diagonal of the table and patterns of positive and negative residuals off the diagonal. The positive residuals on the diagonal indicate that the independence model is under-estimating the diagonal terms by producing predicted values for the diagonal which are much lower than the observed ones, i.e. underestimating homogamy. Since every diagonal cell is being under-estimated, it would indicate that a separate parameter or parameters are required for the diagonal cells.

The contrasting patterns in the off-diagonal cells suggest some varying patterns of under and over-estimation amongst the heterogamous ethnicity combinations. The European Only group has mainly negative residuals with the other ethnic groups, indicating that the independence model is over-estimating the number of mixed ethnicity partnerships for the European Only group. The Asian Only group also has predominantly negative residuals, although of smaller sizes than the European Only group. By comparison, the Maori Only and Pacific Only groups have more positive residuals, other than the ones with a European Only or Asian Only partner.

5.2.1. Quasi-Independence Models to Examine Homogamy

Quasi-independence models, as discussed in Section 2.2, can be used to examine the diagonal dominance of tables of male and female ethnicity. The model for the cell frequencies m_{ij} can be stated as:

$$\log m_{ij} = \mu + \lambda_i^{mEth} + \lambda_j^{fEth} + \delta_i I(i=j)$$

where $I(.)$ is the indicator function for the diagonal of the frequency table.

$$I(i=j) \quad = 1, \quad i = j$$
$$\qquad\quad = 0, \quad i \neq j$$

The δ_i parameters represent the dominance of the cells on the diagonal of the table above and beyond the effect of the row and column variables. Exponentiating the δ_i parameters provide a multiplicative factor which indicates how many more times greater (or less) the frequency of couples expected to have a homogamous partnership, over and above the independence model (Goodman, 2007).

5.2.2. Results from the Quasi-Independence Models

Table 5.4 shows the residual deviance and e^{δ_i} parameters, calculated separately for each census, using the full data sets (including foreign-born couples). The residual deviance is the sum of the squared deviance residuals that were explained in Section 5.2; thus, a model with a lower residual deviance has a better fit to the data. It can also be compared to the null deviance, which is the sum of the deviance residuals for a model which is a constant and does not have any variables. A large difference between the null deviance and the residual deviance indicates that the addition of parameters for the variables in the model have improved the model.

Ethnicity	**1981**	**1986**	**1991**	**1996**	**2001**	**2006**
MELAA Only	386.89	332.76	266.94	545.58	416.65	366.22
Asian Only	364.97	245.07	224.71	301.62	279.63	267.65
Pacific Only	106.79	88.97	94.62	98.91	91.71	91.32
Asian & European	17.72	23.84	22.86	13.20	44.91	19.88
European	16.10	9.51	8.35	8.14	7.58	7.16
Maori Only	15.55	9.63	7.97	9.41	9.72	9.63
Maori & Pacific	10.07	54.98	8.81	15.62	6.54	8.19
Pacific & European	7.73	14.01	10.71	7.87	5.77	5.03
Maori & European	1.55	3.03	3.29	1.28	1.66	1.65
Null Deviance	4592320	4776877	4787302	4567405	4616814	4867134
Null Degrees of Freedom	80	80	80	80	80	80
Residual Deviance	6125.2	4124.1	4671.2	9192.6	9985.5	13101
Residual Degrees of Freedom	55	55	55	55	55	55

Table 5.4 - Exponentiated diagonal dominance parameters 1981-2006

The residual deviance of the quasi-independence model for each census is still quite high, but there is an improvement in each census period over the independence model (χ^2 <

0.001). It is important to note that for tables with large frequencies, such as the ones being analysed here, there will always be a large residual deviance, which would lead to all models being rejected (Simonoff, 2003). What is of more importance is to observe the change in deviance, rather than just the total. The residual deviance has increased at a much faster rate than the null deviance over the census periods, indicating a worse fit for the model.

The multiplicative diagonal dominance factors for each ethnicity in Table 5.4 show how many times higher the estimated frequency is for each ethnicity over and above that of the independence model. Since log-linear models control for group size, these parameters provide a better basis for comparisons between ethnic groups than the proportions in Section 5.1.

The Asian Only and MELAA Only categories have the largest parameters, indicating a high degree of homogamous partnerships in each of these groups. For example, the value of 267.65 for the Asian Only group in 2006 indicates that the number of homogamous Asian relationships is 267.65 times greater than what would be expected under the independence model. The Pacific Only category also has a high factor, ranging between 87 and 109. As with the proportions in Section 5.1, the parameters for ethnicity groupings which contain two ethnic groups are in general much lower. There is some evidence of a decreasing trend in the parameters for each of the single ethnic groups, but as previously observed, the immigration of homogamous couples has an impact on the patterns in the data. Again, the analysis is repeated for non-immigrant couples only, as the focus of the investigation is on the New Zealand “marriage” (marriage and cohabitation) market.

The deviance residuals for the initial quasi-independence model in Table 5.5 show that the model now has residual values of zero on the diagonal since the data is modelled with the diagonal estimates fixed at the observed values. The absolute value of the error terms for the other cells have decreased and are now a mix of positive and negative errors, indicating that the quasi-independence model is providing a better fit to the data (both on and off the diagonals) than the independence model did. Collecting the residuals together sees the total residual deviance decrease from 512,749 (table not shown) to 9,986 (Table 5.4).

	Female Ethnicity								
Male Ethnicity	European Only	Maori Only	Pacific Only	Asian Only	MELAA Only	Maori & European	Maori & Pacific	Pacific & European	Asian & European
European Only	0.00	-2.40	2.15	-4.67	-18.30	24.61	6.64	-9.82	-8.25
Maori Only	11.30	0.00	-4.16	-13.13	0.12	-31.08	-10.19	1.44	-2.22
Pacific Only	3.21	-5.63	0.00	-2.95	-2.40	6.70	-0.10	-0.28	-7.07
Asian Only	-22.01	-1.33	-2.36	0.00	21.18	-15.96	-4.17	28.79	34.61
MELAA Only	-13.63	3.87	-2.70	10.48	0.00	-5.74	-2.66	3.39	15.99
Maori & European	3.03	-16.31	13.29	17.04	-1.67	0.00	4.30	-2.40	-9.15
Maori & Pacific	5.36	-10.90	-1.32	1.13	-3.23	7.63	0.00	-2.31	-10.15
Pacific & European	-3.36	2.20	-0.48	17.41	1.72	-8.35	-1.43	0.00	-1.83
Asian & European	1.45	12.14	-7.35	4.32	14.57	-36.05	-8.78	-2.22	0.00

Table 5.5 - Deviance residuals for quasi-independence model, 2001 data

5.2.3. Quasi-Independence Models Controlling for Immigrant Status

To better examine the New Zealand "marriage" market the analysis from Section 5.2.2 is repeated after removing couples where both partners were born overseas, as these relationships were most likely to have been formed in another country. Ideally, these individuals would be identified by comparing the length of their relationship to the number of years they had lived in New Zealand, but this information is not available in the census. Instead, place of birth is used as a proxy for immigrant status. Couples where both partners were born overseas were removed from the data and the quasi-independence model was refitted.

Ethnicity	1981	1986	1991	1996	2001	2006
Asian Only	180.11	106.57	56.78	52.88	34.74	22.30
MELAA only	179.18	53.91	17.78	11.28	10.92	5.59
European Only	15.33	9.3	7.89	7.36	7.56	6.99
Maori Only	14.76	8.81	6.95	8.45	8.44	8.76
Pacific Only	10.72	6.95	10.43	17.09	18.73	21.41
Maori & Pacific	10.39	12.38	5.96	10.4	5.50	7.28
Asian & European	2.64	2.61	7.32	7.41	13.86	3.51
Pacific & European	2.60	5.79	5.72	3.74	3.97	3.74
Maori & European	1.42	2.82	3.12	1.28	1.56	1.50
Null Deviance	4102737	4294116	4508862	4282799	4286146	4203753
Null Degrees of Freedom	80	80	80	80	80	80
Residual Deviance	3777.8	2895.2	3765.2	6717.8	7987.4	10226
Residual Degrees of Freedom	55	55	55	55	55	55

Table 5.6 - Exponentiated diagonal dominance parameters 1981-2006 (NZ Born)

Table 5.6 shows the quasi-independence parameters for the New Zealand-born couples. The residual deviance values in each period have decreased dramatically relative to the independence model for the same period. This indicates that the quasi-independence model has a better fit to the data than the baseline (independence) model, which does not have any diagonal cell parameters.

With the foreign-born couples removed, the patterns in the diagonal dominance parameters are clearer, and are plotted (with parameter confidence intervals) in Figure 5.9. In 1981, the Asian Only and MELAA Only groups still have a much higher propensity to homogamy than the other ethnic groups. However, both groups have had a dramatic decrease in these rates over the six census periods (note the use of a log scale to allow for easier comparisons). Overall, the Asian group still has the highest rate of homogamy, but their diagonal dominance factor has reduced from 180 to 22. The confidence intervals for the parameters show more variation in the MELAA estimates, which are due to the smaller group size.

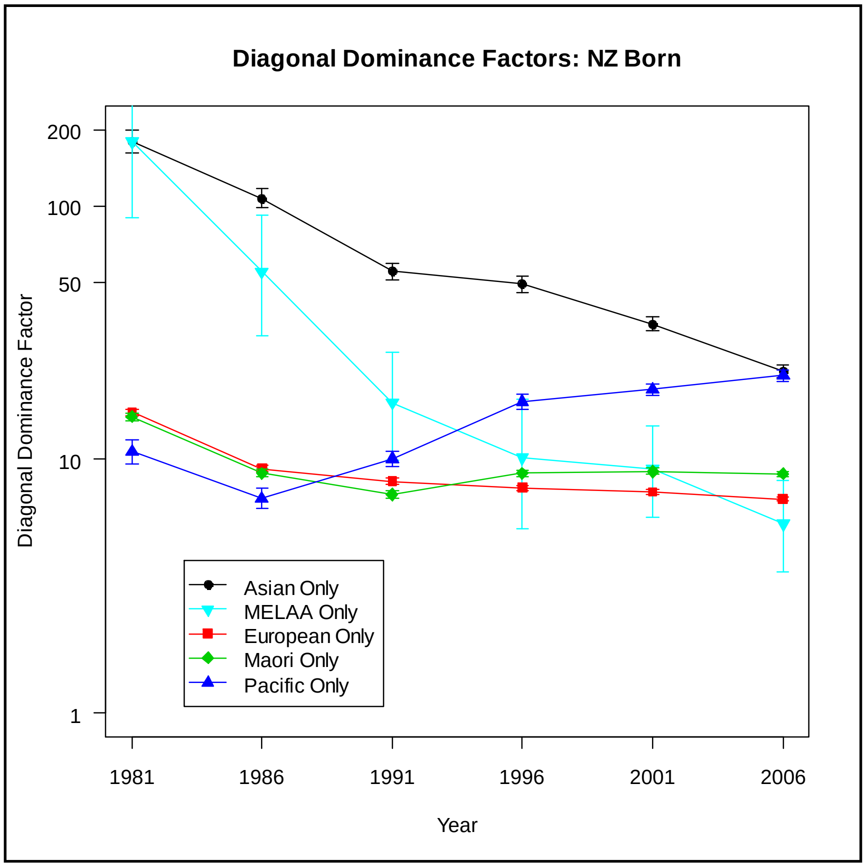

Figure 5.9 - Diagonal dominance factors: Single ethnicities only

The decrease in the diagonal dominance - representing homogamy - would suggest that there has been a loosening of ethnic stratification over the period 1981 to 2006 for these two groups. Of the other single ethnicity groups, there have been slight decreases in the rates for New Zealand Europeans and Maori and an increase in the rate for Pacific peoples. The increase in the homogamy rate for Pacific peoples does not consider the homogamy and heterogamy of partnerships within the Pacific group (i.e. where the partners may be of a different Pacific ethnicity from each other), but does show an increasing tendency towards homogamous partnerships for Pacific people relative to the other ethnic groups. The MELAA Only group was removed from the remainder of the log-linear models as the frequencies in some of the sub-groups were too small.

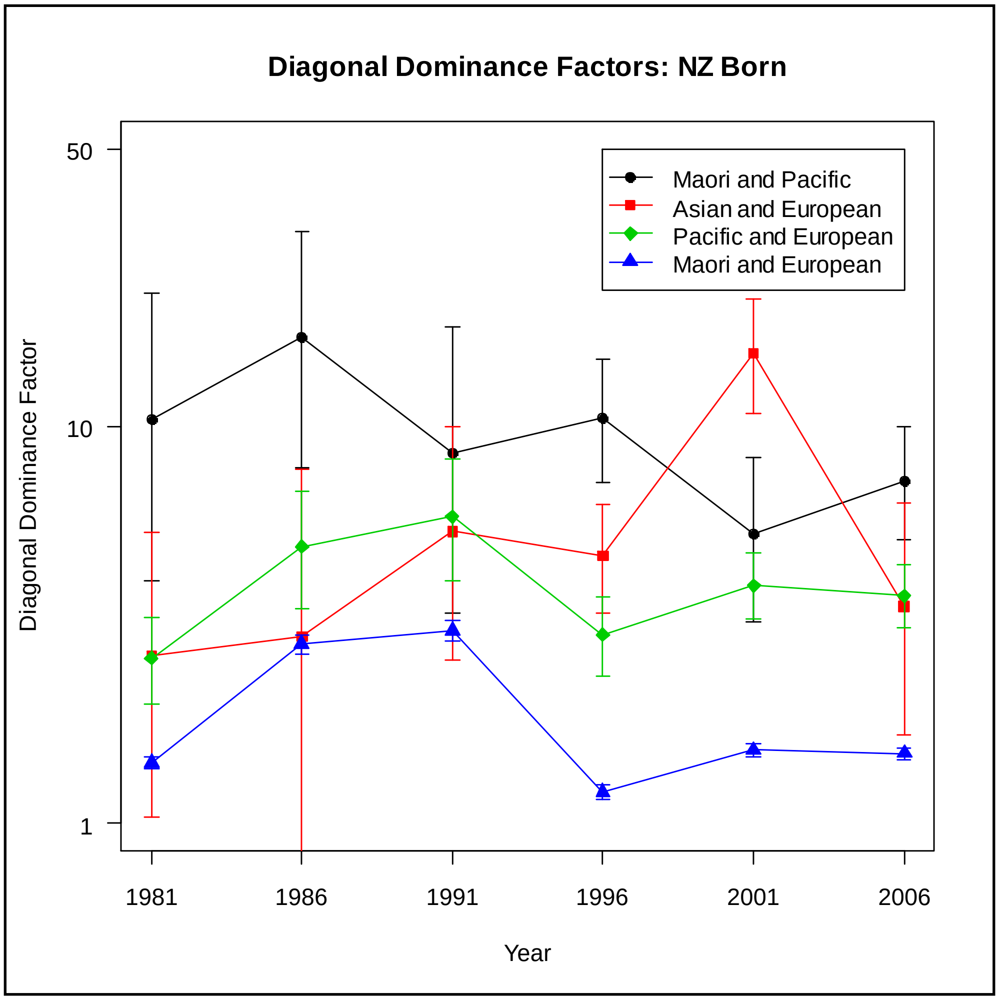

Figure 5.10 - Diagonal dominance factors: Dual ethnicity, NZ Born

Figure 5.10 shows that as with the earlier results, the mixed ethnicity groups (European and Maori, European and Pacific, Asian and European, Pacific and Maori) had a lower propensity for homogamy. Over the five Census sets, there was a decrease in the homogamy rate for the Maori & Pacific group and an increase for the Asian & European group. The Pacific & European and Maori & European groups both fluctuated. These groups will be examined further in using a crossing parameter model in Section 5.2.5 where they can be examined on partial as well as full ethnicity matches.

5.2.4. Quasi-Independence Models – Emerging and Existing Partnerships

Figure 5.11 and Figure 5.12 show the diagonal dominance factors for emerging and existing partnerships, as defined by partnerships with the male partner aged between 18 and 30 (emerging) or over 30 (existing).

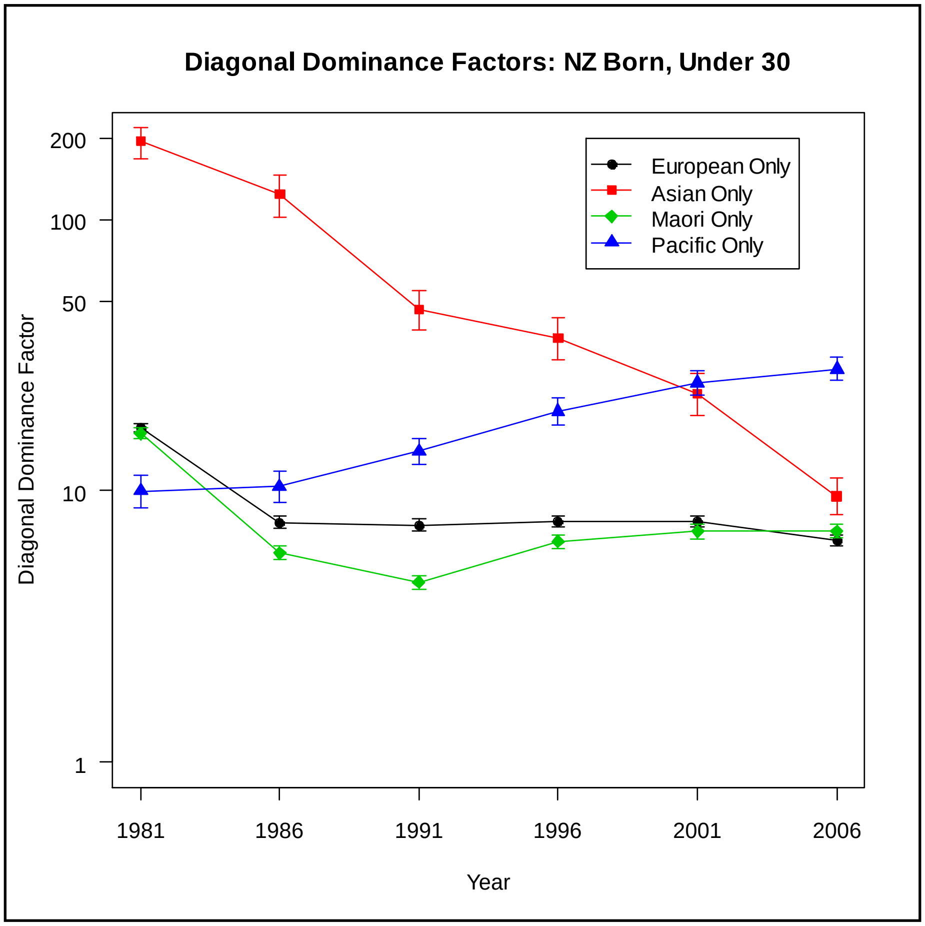

Figure 5.11 - Diagonal dominance factors: Emerging partnerships

The emergent relationships for the Maori Only and European Only groups have higher diagonal dominance factors than the existing cohort in 1981. This indicates that the number of homogamous relationships for the emergent group, above and beyond the independence model, is greater than that of the existing group. However, the diagonal dominance factors for the emergent relationships for both ethnic groups decrease in 1986 to values similar to those for the existing partnerships and remain at a similar level to one another for the remainder of the census periods. The diagonal dominance factors for the emergent partnerships in the Asian Only group show a steeper decrease than those for the existing partnerships. Conversely, the emergent partnerships in the Pacific Only group show slightly higher factors than the existing partnerships.

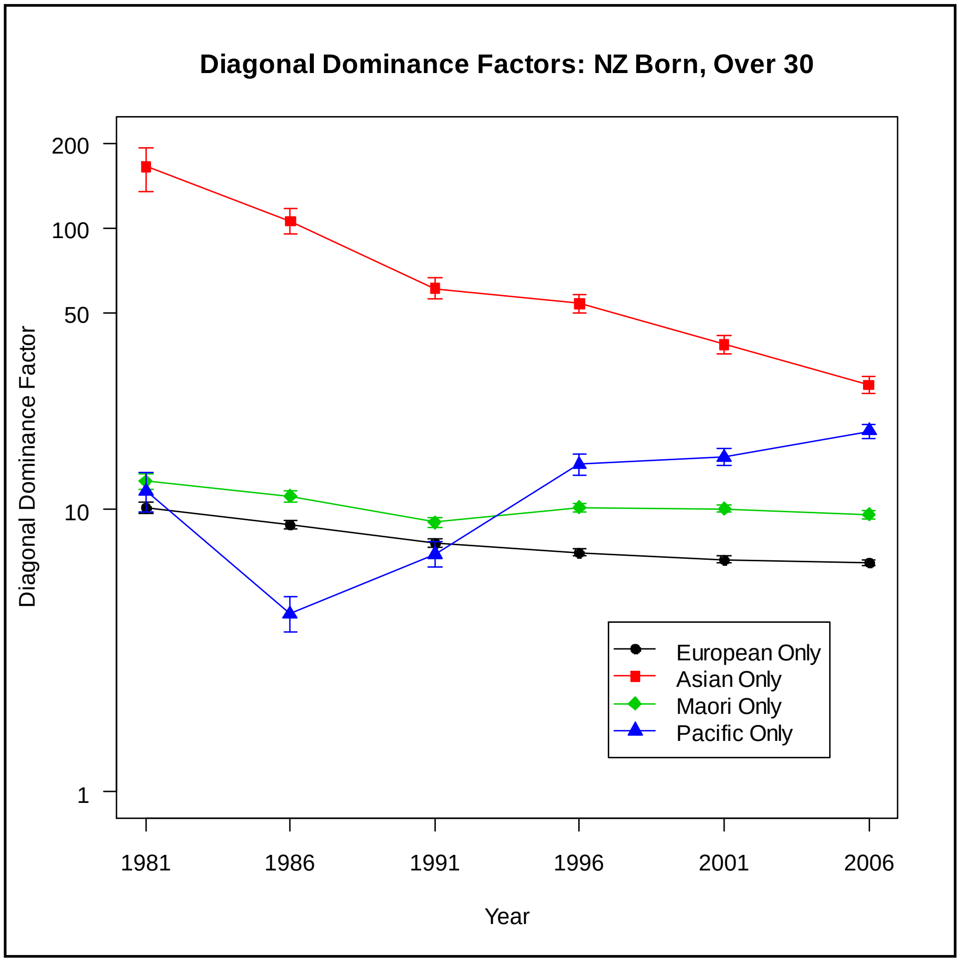

Figure 5.12 - Diagonal dominance factors: Existing partnerships

The next step in the analysis is to use crossing-parameter models to examine whether there are patterns of partial homogamy by incorporating partial matches of ethnicity between partners.

5.2.5. Crossing Parameter Models

As described in Section 2.2.4, crossing parameter models create an alternative measure of social distance, by examining the difficulty of crossing a “barrier” in a table of counts. With partnership data, it provides a way of examining the ethnic patterns of the dual ethnicity groups by creating parameters for partial ethnic matches, rather than just dividing the table into homogamous (on the diagonal) and heterogamous (off the diagonal) partnerships. For example, although a relatively low proportion of individuals who were in the Maori and European group had a partner who was also in the Maori and European group, many had a partner who was in the Maori Only group or the European Only group.

The partnership ethnicity data is divided into four groups for the purpose of examining these crossings by ethnicity. The partnerships are ordered from no homogamy to complete homogamy below:

1. No common ethnicity (no homogamy).
2. Two partners of mixed ethnicity who share a common group (some homogamy).
3. One partner of a single ethnicity and one with mixed ethnicity who share a common group (partial homogamy).
4. Two partners with the same ethnicity (full homogamy).

Group two and group three can be thought of as degrees of partial homogamy and the barriers that are being crossed as the difficulty of a partnership occurring in a particular group relative to an adjacent group. Table 5.7 shows the crossing parameters for each partnership combination.

	Female Ethnicity								
Male Ethnicity	E	M	P	A	MELAA	E/M	E/P	E/A	M/P
European Only	$\delta_1 + \delta_2 + \delta_{3_1}$	0	0	0	0	δ_1+δ_2	δ_1+δ_2	δ_1+δ_2	0
Maori Only	0	$\delta_1 + \delta_2 + \delta_{3_2}$	0	0	0	δ_1+δ_2	0	0	δ_1+δ_2
Pacific Only	0	0	$\delta_1 + \delta_2 + \delta_{3_3}$	0	0	0	δ_1+δ_2	0	δ_1+δ_2
Asian Only	0	0	0	$\delta_1 + \delta_2 + \delta_{3_4}$	0	0	0	δ_1+δ_2	0
MELAA Only	0	0	0	0	$\delta_1 + \delta_2 + \delta_{3_5}$	0	0	0	0
European & Maori	δ_1+δ_2	δ_1+δ_2	0	0	0	$\delta_1 + \delta_2 + \delta_{3_6}$	δ_1	δ_1	δ_1
European & Pacific	δ_1+δ_2	0	δ_1+δ_2	0	0	δ_1	$\delta_1 + \delta_2 + \delta_{3_7}$	δ_1	δ_1
European & Asian	δ_1+δ_2	0	0	δ_1+δ_2	0	δ_1	δ_1	$\delta_1 + \delta_2 + \delta_{3_8}$	0
Maori & Pacific	0	δ_1+δ_2	δ_1+δ_2	0	0	δ_1	δ_1	0	$\delta_1 + \delta_2 + \delta_{3_9}$

Table 5.7 - Parameters for crossing effects.

The model has a similar form to the quasi-independence model. In addition to the male and female ethnicity parameters, it also includes the δ_1,δ_2, and δ_{3_i} values for the levels of homogamy, giving:

$$\log m_{ij} = \mu + \lambda_i^{mEth} + \lambda_j^{fEth} + \delta_1^{some} + \delta_2^{partial} + \delta_{3_i}^{full}$$

The δ parameters in this model build up from no homogamy (0) to full homogamy ($\delta_1 + \delta_2 + \delta_{3_i}$), which is the opposite of the example shown in Section 2.2.4, which was parameterised from full homogamy to no homogamy. Once exponentiated, the crossing parameters (δ_1, δ_2 and δ_3) show the multiplicative factor for the number of relationships at that level of homogamy, net of the marginal distributions, and relative to the number of relationships with no homogamy.

Group	**Parameter**	**1981**	**1986**	**1991**	**1996**	**2001**	**2006**
Two mixed ethnicity partners with partial match (some homogamy)	e^{δ_1}	4.77	7.11	4.50	3.79	4.18	4.37
One single ethnicity partner, one mixed ethnicity partner with matching ethnicity (partial homogamy)	$e^{\delta_1+\delta_2}$	3.81	2.57	1.84	2.55	2.46	2.52
Two partners of the same ethnicity (full homogamy)	$e^{\delta_1+\delta_2+\delta_3}$						
	European Only	20.91	10.13	8.73	9.32	8.83	8.38
	Maori Only	13.72	8.25	6.94	8.01	8.24	8.09
	Pacific Only	5.95	6.02	9.12	12.90	14.77	16.61
	Asian Only	97.47	91.29	49.85	36.74	26.15	16.73
	MELAA Only	96.93	46.91	15.16	7.58	7.07	4.25
	Maori & European	14.92	16.79	9.60	6.36	7.52	7.56
	Maori & Pacific	22.54	36.31	13.89	17.96	9.63	13.42
	Pacific & European	16.84	25.05	17.13	12.56	16.45	16.02
	Asian & European	16.27	14.47	15.45	19.48	61.85	14.83
Residual Deviance		1528.6	2151.0	3209.6	4475.6	6223.4	7994.5
Residual Degrees of Freedom		53	53	53	53	53	53

Table 5.8 - Exponentiated crossing parameters

Table 5.8 shows the crossing parameters for each census. These parameters are multiplicative factors that measure the number of relationships at each level of homogamy, relative to the number of partnerships with no homogamy, whilst controlling for the size of each ethnic group. Since all of the parameters are greater than one, there is strong evidence that homogamous partnerships of any kind are more likely than partnerships with no match of ethnicity at all. Other than in 1986, the δ_1 parameters are quite consistent, taking values between 3.8 and 4.8. This means that net of the relative group sizes, at each census there are about four times as many couples with a match with "some homogamy" than couples with no homogamy. The δ_2 parameters are slightly lower than the δ_1 parameters. This indicates that the partial homogamy match is more likely than one with no homogamy but slightly less likely than the broader "some homogamy" match.

Although the δ_3 parameters measure the diagonal dominance in the tables of the various ethnic groups, they are not directly comparable to the parameters for the quasi-independence models in Section 5.2.1 due to the additional parameters in the model. However, they do measure the same general patterns. The δ_3 parameters are all greater than the δ_1 and δ_2 ones. This means that the homogamous partnerships with a full match of ethnicity are more likely than those with a partial match. The patterns within the δ_3 parameters are generally consistent with the results from the quasi-independence model in Section 5.2.3. However, the additional parameters in this model have the effect of reducing the scale of the parameters for the diagonal. The residual deviance is also lower than the quasi-independence model.

5.2.6. Quasi-Symmetry Models

The quasi-independence models used in Sections 5.2.1 to 5.2.4 showed a strong fit for the diagonal terms of the tables but a lack of fit for some of the off-diagonal terms, due to asymmetry in some of the off-diagonal relationships. Although this asymmetry can generally be identified visually (for example, comparing the frequencies for the male European, female Asian couples, to the male Asian, female European couples), quasi-symmetry models can be used to confirm these patterns. Quasi-symmetry models were fitted to the tables of non-foreign-born couples in each census and the residual deviance values were compared to those from the crossing parameter and quasi-independence models. Lower values for the residual deviance indicate a better fit and confirm the asymmetric patterns in the data.

Model	Measure	1981	1986	1991	1996	2001	2006
Independence	Residual Deviance	160743	157325	145395	140490	158502	168976
	Residual df	64	64	64	64	64	64
Quasi-Independence	Residual Deviance	3777.8	2895.2	3765.2	6717.8	7987.4	10226
	Residual df	55	55	55	55	55	55
Crossing Parameter	Residual Deviance	1528.6	2151.0	3209.6	4475.6	6223.4	7994.5
	Residual df	53	53	53	53	53	53
Quasi-Symmetry	Residual Deviance	635.0	475.9	642.6	708.2	856.1	897.0
	Residual df	28	28	28	28	28	28

Table 5.9 - Log-linear goodness-of-fit summary

Table 5.9 and Figure 5.13 show the residual deviance for the independence, quasi-independence, crossing parameter and quasi-symmetry models. As the complexity of the models, as measured by the number of parameters, increases, the residual deviance decreases. Since the main focus of this research is on homogamy, the diagonal dominance parameters from the quasi-independence model are of the most interest. However, beyond the homogamy in the tables, there are clear patterns of asymmetry by gender for some ethnic groups, as observed in the frequency tables and residuals (see Table 5.3). The deviance values for the quasi-symmetry model, which are significantly lower than all of the other models, confirm that the asymmetrical patterns in the data are statistically significant.

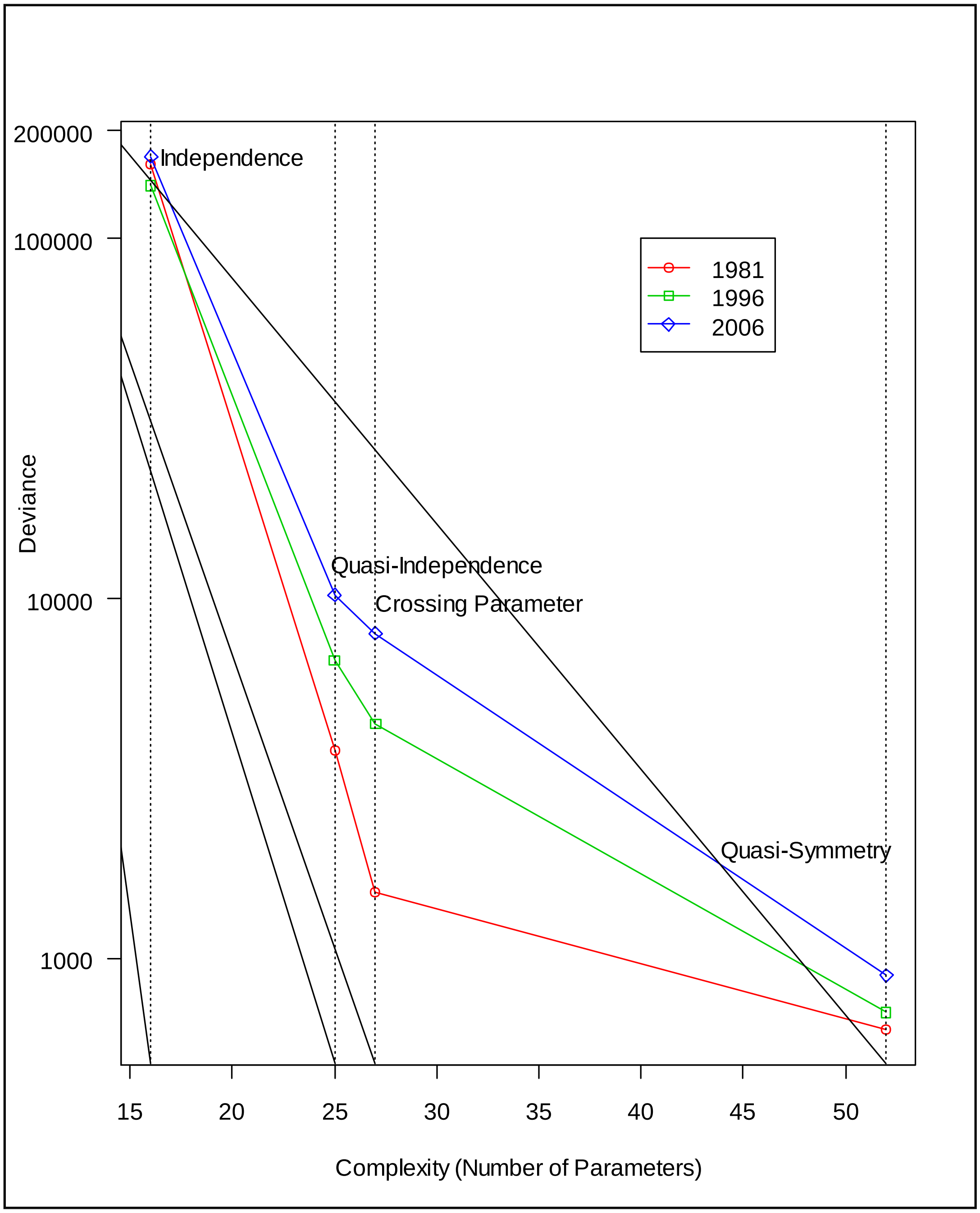

Figure 5.13 - Goodness-of-fit by complexity

5.3. Logistic Regression Modelling

Logistic regression is used to examine the effect of a set of continuous and/or categorical variables on a binary response. In this case the binary response variable will be homogamous partnership (yes/no). It will examine the effect of each explanatory variable on the probability of a partnership being homogamous.

The basic form of the logistic regression is shown below: (Agresti, 2002, p. 85):

$$\pi(x) = \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}}$$

It shows that the probability of a "yes", in this case a homogamous partnership, is a function of a set of covariates (x). This is sometimes written using a logit transformation, where the log odds of a "yes" are written as a linear combination of the covariates ($\alpha + \beta x$).

Variable	Description
Auckland	Dummy variable indicating if couple live in Auckland
Age	
Male (18-30)	Dummy variable indicating male partner aged 18-30
Female (18-30)	Dummy variable indicating female partner aged 18-30
Education	
Male (School)	Dummy variable indicating highest qualification is school (vs none)
Male (Tertiary)	Dummy variable indicating highest qualification is tertiary (vs none)
Female (School)	Dummy variable indicating highest qualification is school (vs none)
Female (Tertiary)	Dummy variable indicating highest qualification is tertiary (vs none)

Table 5.10 - Logistic regression explanatory variables

The explanatory variables for the logistic regression are listed in Table 5.10. The key variables of interest are the age groups of the male and female partner, the highest qualification of each partner, and whether the couple live in Auckland or not. Confidentiality concerns limited the geographic comparisons to comparing couples in Auckland and out of Auckland. Ethnicity was also included in the model as a control variable, but the parameters are not shown as they are in the model to capture the patterns already described in this chapter.

Table 5.11 shows the odds ratios (exponentiated model parameters) and R^2 values from the logistic regression models. Due to the size of the data sets, all of the parameters were statistically significant, with p-values below 1%, so the focus of the results is on the exponentiated parameters (odds ratios) rather than the significance. It should also be noted that the rescaled R^2 figure, which is produced by the standard SAS output, does not measure the proportion of variance explained, but instead measures the percentage change in deviance between the intercept only model and the model with covariates (Menard, 2000).

	Odds Ratios (exponentiated model parameters)					
Variable	1981	1986	1991	1996	2001	2006
Auckland	0.952*	0.813*	0.810*	0.681*	0.601*	0.593*
Age						
Male (18-30)	1.045	1.029	0.979	0.991	0.879*	0.933*
Female (18-30)	1.485*	1.362*	0.704*	1.188*	1.071*	1.079*
Education						
Male (School)	X	1.141*	0.939*	1.168*	1.081*	1.073*
Male (Tertiary)	X	0.999	1.018	1.011	0.940*	0.995
Female (School)	X	0.989	0.982	1.163*	0.965*	1.109*
Female (Tertiary)	X	1.039	0.860*	1.281*	0.972*	1.010
Maximum rescaled R^2	52.20	46.35	45.71	53.23	56.25	56.75
(% reduction in deviance)					* significant at 1% level	

Table 5.11 - Logistic regression results

The parameter for couples in Auckland shows that odds of a couple having the same ethnicity are lower than for couples outside Auckland. In 1981, the odds of an Auckland couple having the same ethnicity as one another were 5% lower than those for a couple living outside of Auckland. By 2006, the odds for an Auckland couple were 40% lower. By comparison, age and education have little impact on the odds of a partnership being homogamous. The odds for males aged 18 to 30 of having a homogamous partnership are slightly lower than for older males, with a small decrease over time. The odds ratios for males and females with either a school or tertiary qualification, relative to having no qualification, were very close to one, with no major increase or decrease over time. This may relate to the very low rate of educational homogamy in New Zealand (Smits, 2003).

5.4. Summary of Statistical Analysis

The key goal for the statistical analysis of the patterns of ethnic partnering was to observe what patterns of ethnic partnering exist, based on the census data from 1981 to 2006, and then to examine how those patterns have changed over that period. Callister, Didham and Potter (2007) provide a benchmark for confirming the basic patterns in the 2001 census. However, this research extends their work, by examining the data with log-linear models and looking at the trends over time.

The patterns in the proportion of homogamous couples in the different ethnic groups were all similar to those presented by Callister *et.al*. The European Only group had the largest proportion of homogamous partnerships, but it is also by far the largest ethnic group. All of the other ethnic groups had smaller proportions. Dual ethnicity groups had lower proportions than single ethnicity groups, although this did not take into account partial ethnic matches. The initial proportions showed a clear immigration effect, particularly in the Asian Only group, so the data was analysed with foreign-born couples removed. The proportions for the European, Asian and Maori groups were all declining over time, whilst the proportion for the Pacific Only group showed a slight increase. The disadvantage of analysing proportions was that they are a function of the size of the group they are calculated from, so the next step in the analysis was to re-examine these patterns independent of the size of the ethnic groups by using log-linear modelling.

The main log-linear model used in the analysis of homogamy is the quasi-independence model. The quasi-independence model used diagonal dominance parameters to investigate the number of homogamous partnerships, net of the effect of the marginal distributions. The analyses were conducted on the data with the foreign-born couples removed. The quasi-independence parameters showed a consistent rate of homogamy for the European Only and Maori Only groups over time. The Asian Only and MELAA Only groups both had dramatic decreases, going from very high levels of homogamy in 1981 to much lower levels in 2006. The quasi-independence model also confirmed the increasing rate of homogamy for the Pacific Only group. This increase was slightly greater for the emergent partnerships (where the male partner was aged 18 to 30) than the overall population.

The crossing-parameter model broke the partnership tables down further than the quasi-independence model. The model showed that partnerships with a part match of ethnicity were more likely than those with no common ethnicity, after the sizes of the ethnic groups had been accounted for. This pattern was consistent over time. Fully homogamous partnerships were still shown to have the greatest number of partnerships, above and beyond the independence (random matching) model. The quasi-symmetry model showed the strongest fit to the data, with the lowest residual deviance values. Although it did not have specific parameters that could be interpreted in terms of the partnerships, it did reinforce the asymmetric nature of the tables.

The logistic regression examined the effect of age, education and location on the odds of a couple having a homogamous partnership. Couples in Auckland were less likely to have a homogamous partnership than couples outside of Auckland. This difference became greater over time. Education and age (18-30 versus over 30s) had little impact on the odds of a couple having the same ethnicity as one another.

The examination of the proportion of homogamous couples and the log-linear modelling of the frequency tables showed the patterns of the ethnic partnering that has already occurred and can be observed in the census data. An alternative and complementary way to examine the patterns of ethnic partnering is to focus on the partner matching process itself. Since this process cannot be observed directly, computer simulation is used to match singles based on different factors and rules, and to examine the subsequent ethnic patterns.

Chapter 6 – Abstract Simulation

This chapter introduces the idea of social simulation as a method for examining partner choice. It uses a simple abstract model applied to a small artificial population using the Netlogo programme to examine some properties of partnership selection. This provides the first steps in investigating the second research question, examining partnership patterns within the matching process itself rather than just historical patterns.

6.1. Netlogo

The Netlogo programme (http://ccl.northwestern.edu/netlogo) is a freely downloadable, programmable modelling environment for simulating social phenomena. It was created in 1999 by Uri Wilensky and is in continuous development at the Center for Connected Learning and Computer-Based Modeling.

Netlogo was chosen as the initial software programme since it provided a simple programming interface and because there are existing templates available that can be used as a starting point for building simulation models. The Netlogo model used to examine partner choice is based on a template by Robert Mare of the University of California Los Angeles. It was sourced from http://cgi.stanford.edu/dept/anthsci/cgi-bin/rlab/doku.php?id=lab:exercises and adapted to include further parameters and a more efficient algorithm.

6.2. Abstract Simulation Models

Abstract simulation models provide the opportunity to examine a social system without being constrained by the limitations of real world data (Gilbert & Troitzsch, 2005). Although there may be some relation to the real world, abstract simulation models do not need to be based on or informed by real data in the same way that an empirical simulation model does. They can be used to test ideas and examine scenarios in a simple abstract setting. The abstract model presented in this chapter was developed prior to the release of the census data for the empirical simulations in Chapter 7. It shows some simple properties of different partnership choice mechanisms and population distributions.

Figure 6.1 illustrates the algorithm that was used in the abstract simulation. The Netlogo code for the simulation can be seen in Appendix C.1.

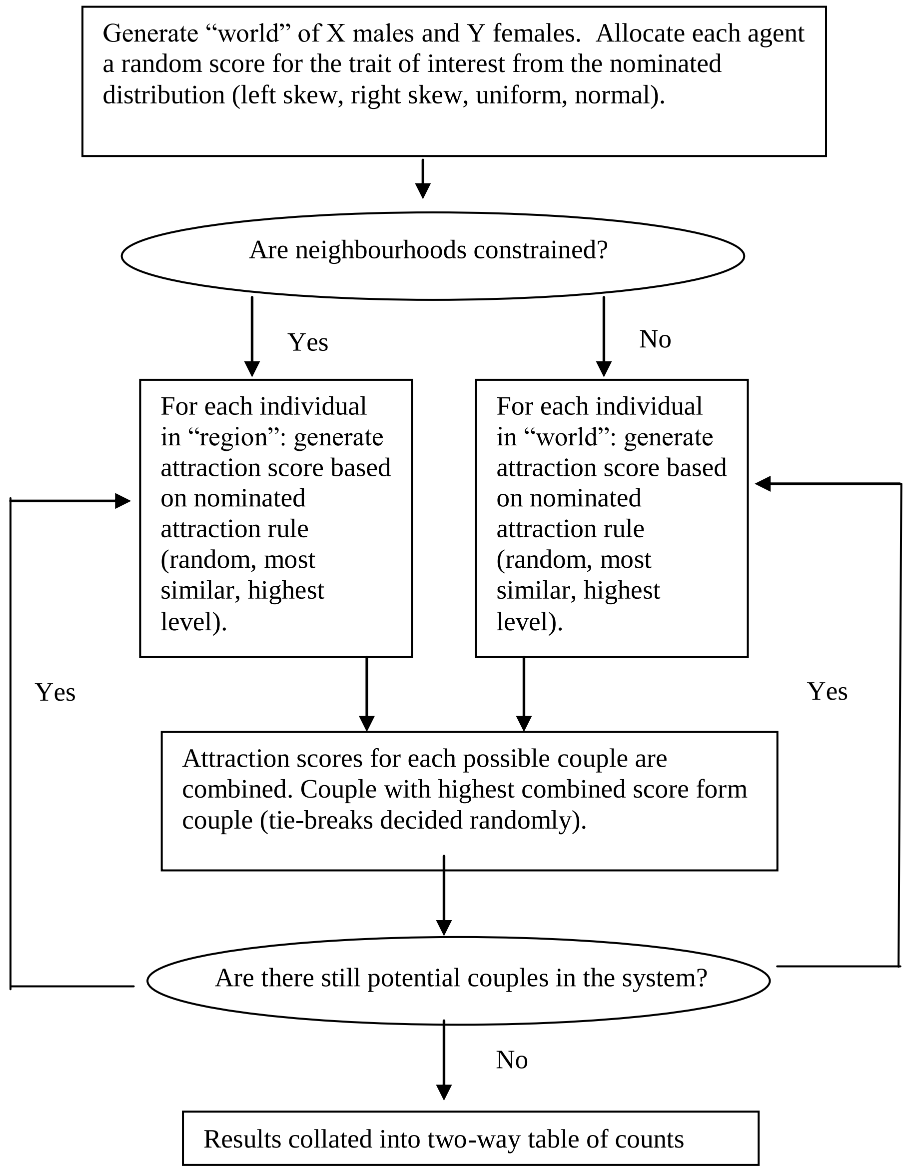

Figure 6.1 - Netlogo abstract simulation diagram

The abstract model examines a generic hierarchical trait of interest which is present in all males and females in the population. In the Netlogo programme the trait is labelled as "education", but it could represent any hierarchical trait. At each iteration, each individual in the population examines the “education” score of each potential partner in their field of vision (neighbourhood or world), and then allocates them an attraction score based on the attraction method which has been selected. For this exercise the method of attraction and the scoring were done identically for each individual. To mimic a competitive marriage market, the algorithm proceeds in such a way that the pair with the highest mutual attraction are partnered first. Then the algorithm is repeated with the next remaining pair having the highest mutual attraction, continuing until there are no unpartnered pairs remaining. Where there is a tie for highest mutual attraction, one of the top ranked couples will be chosen at random to partner. For a simulation with no neighbourhood constraint, the algorithm functions such that every individual becomes partnered. If the agents only have a limited range of sight – i.e. there is a neighbourhood constraint - then some individuals may remain unpartnered due to the inability to “see” another unpartnered, potential mate.

6.3. Simulation Parameters

The adjustable parameters of the model are listed in Table 6.1.

Parameter	**Options**
Neighbourhood (agents can see all or only near neighbours)	On/Off
Partner selection trait (e.g. levels of “education”)	0 – n
Attraction	Random/Most Similar/Highest Level
Distribution of levels	Uniform/left skew/right skew/normal
Size of neighbourhood	1-10 units

Table 6.1 - Abstract simulation parameters

The main parameters of interest are the constraint of a neighbourhood, the method of attraction, and the distribution of different levels of the trait. The neighbourhood variable controls whether the agents in the model can see the levels of all other agents or whether their sight of other agents is constrained within a certain distance (i.e. within their neighbourhood). The size of the neighbourhood can also be adjusted, although in practice this works only to amplify any effect of the neighbourhood constraint, rather than have any quantitatively distinct effect of its own.

The method of attraction is scored in three different ways: random selection, selection of another agent with the most similar level of selection trait, or attraction to the agent with the highest level of the selection trait. The random attraction model can be considered a baseline against which to compare the other two methods. It should be noted that the "most similar" and "highest" methods are quite similar to one another because of the symmetrical nature of the matching. The key difference is that under the "most similar" method, two individuals who have low, but similar, levels of education have an equal chance of being matched together as two individuals who have similar high levels of education, whereas under the scheme where individuals are attracted to those with high levels of education, the couples with high education levels will be paired off before those with lower levels.

One thousand iterations of the simulation were run for each combination of the parameters using a population of fifty males and fifty females and the data collected into cross-tabulations of the male "education" against female "education". The overall model criterion was the level of homogamous selection made in the population. Although the differences in homogamy between the models could be seen by simply examining the tables, the social homogamy index provided a more concise way of measuring the overall homogamy present in each of the tables.

6.4. Social Homogamy Index

Romney (1971) introduced the idea of trying to capture in a single measure the degree of subgroup endogamy in a population of finite size. He examined hypothetical cases of intermarriage and examined the ratios of the row and column totals to the cell counts.

Romney's model was adapted by Robbins (1981) to incorporate distance into the measure of homogamy. Robbins' measure of social homogamy provides an index number (*H*) between 0 (minimum homogamy) and 1 (maximum homogamy) based on the ratio of the current level of homogamy to the possible extremes given the marginal totals for each category. It is calculated as:

$$H = 1 - \frac{H_{\max} - H_0}{H_{\max} - H_{\min}}$$

In the equation H_0 is the sum of the off-diagonal frequencies multiplied by the distance (number of cells) that each frequency is from the diagonal in the table of interest. H_{max} uses the same measurement, but for a hypothetical table that holds all of the marginal (row and column) totals the same but has the largest possible frequencies on the diagonal. H_{min} calculates the measurement for the hypothetical table with the largest possible frequencies on the off diagonals. These hypothetical tables are generated by a heuristic that is akin to a linear programming algorithm. For the H_{min} table, the cell frequencies are filled in from the top right (furthest from the diagonal), working right to left and ensuring that the row and column totals remain the same as the original table and that the greatest possible frequencies are used in the cells that are further from the diagonal. The H_{max} table uses the same heuristic but starts from the top left corner so that the largest frequencies are in the diagonal cells of the table. By calculating the homogamy index for each of the simulated tables, the degree of homogamy in each can easily be compared in a single table.

Table 6.2 shows the Social Homogamy Index applied to the eye colour example from earlier chapters. It shows the original frequency table and the hypothetical minimum homogamy and maximum homogamy tables. The hypothetical minimum homogamy table was generated by starting at the top right cell. The largest number that could be allocated to the top right cell (blue/green) was 20, since the same row and column totals have to be retained. This meant that the remainder of the values in the top row must be zero. The algorithm then returns to the far right cell in the second row. This value can be set to 20, producing the required column total of 40 (and forcing the bottom right cell to be zero). To keep the second row total as 30, a total of ten must be allocated across the rest of the row. Since the aim is to produce the table with the lowest amount of homogamy possible, the ten is allocated to the off-diagonal cell rather than the on-diagonal one. This process is continued, with the values being added to the cells, whilst holding the row and column totals fixed, until the hypothetical minimum homogamy table is produced. In this particular case, the table has been produced with counts of zero down the diagonal. The same process is applied in order to create the hypothetical table with maximum possible homogamy. However, it starts in the top left cell instead of the top right and aims to maximise the frequencies on the diagonal instead of off the diagonal, whilst retaining the same row and column totals. The H_0, H_{max}, and H_{min} values are computing by multiplying each cell frequency by its distance from the diagonal (0,1 or 2) and adding them all up. For example the H_0 value is computed by:

$$0 \times (15 + 18 + 29) + 1 \times (3 + 3 + 9 + 9) + 2 \times (2 + 12) = 52$$

The final homogamy index (H) is then computed as:

$$\frac{140 - 52}{140 - 20} = 0.733$$

Original Table	Wife's Eye Colour			
Husband's Eye Colour	Blue	Brown	Green	Total
Blue	15	3	2	20
Brown	3	18	9	30
Green	12	9	29	50
Total	30	30	40	100

Minimum Homogamy	Wife's Eye Colour			
Husband's Eye Colour	Blue	Brown	Green	Total
Blue	0	0	20	20
Brown	10	0	20	30
Green	20	30	0	50
	30	30	40	100

Maximum Homogamy	Wife's Eye Colour			
Husband's Eye Colour	Blue	Brown	Green	Total
Blue	20	0	0	20
Brown	10	20	0	30
Green	0	10	40	50
	30	30	40	100

H_0	52
H_{max}	140
H_{min}	20
H	0.733

Table 6.2 - Social Homogamy Index example

This methodology provides a single summary measure which is useful for quick comparisons of the degree of homogamy between different tables. Although less rigorous than measures derived from log-linear models, it allows a simple and accurate comparison of the homogamy of numerous abstract simulation tables short of the need to generate and interpret multiple regression parameters. It is also not suitable for the ethnicity tables as it relies on ordinal data in the columns and rows, which would rely on some kind of ordering of the ethnic groups.

6.5. Abstract Simulation Results

The homogamy index values are shown in Table 6.3 below. Although there are twenty-four different combinations of the input parameters, time constraints meant that only ten were able to be tested. However, the ten listed below provide sufficient opportunity to compare the impact of each of the parameters on the homogamy index as they cover all of the key comparisons of the parameters. The unlimited neighbourhood and the random attraction method were the two least realistic parameters, so were only used in two of the simulations. The proportion of couples on the diagonal (homogamous relationships) is also reported. The slight variation in the order of the homogamy index numbers compared to the proportions is due to the homogamy index numbers incorporating the distance from the diagonal, whilst the proportions only measure the relative proportion of the on-diagonal values.

Attraction method	Neighbourhood	Population	Homogamy Index (H)	Proportion in Diagonal Cells
Highest Level	Limited	Left Skew	0.9153	83.0%
Most Similar	Unlimited	Uniform	0.8868	84.6%
Most Similar	Limited	Right Skew	0.8317	78.8%
Most Similar	Limited	Left Skew	0.7967	64.4%
Most Similar	Limited	Uniform	0.7965	68.5%
Most Similar	Limited	Normal	0.7307	80.1%
Highest Level	Limited	Right Skew	0.6761	55.4%
Highest Level	Limited	Uniform	0.6346	38.7%
Highest Level	Limited	Normal	0.5215	58.9%
Random	Limited	Uniform	0.3351	19.9%

Table 6.3 - Homogamy index values for abstract simulation

The initial comparison is between limited and unlimited neighbourhood simulations. Table 6.4 shows results from two simulations for agents in a population with uniformly distributed levels and the “most similar” partner selection heuristic. However, in one simulation the agents are constrained by having a "neighbourhood" limiting their field of vision for potential partners, while the other simulation does not have this restriction.

Attraction: Most similar level, Uniform Population, Limited Neighbourhood							Attraction: Most similar level, Uniform Population, Unlimited Neighbourhood						
	Female Levels							**Female Levels**					
Male Levels	**0**	**1**	**2**	**3**	**4**	**Total**	**Male Levels**	**0**	**1**	**2**	**3**	**4**	**Total**
0	5744	1143	429	301	354	7971	**0**	8432	544	289	296	474	10035
1	1221	5826	1021	352	331	8751	**1**	506	8410	455	232	262	9865
2	470	1054	5891	1033	445	8893	**2**	290	518	8535	527	324	10194
3	350	358	1020	5649	1181	8558	**3**	246	260	513	8442	540	10001
4	333	333	453	1171	5966	8256	**4**	430	237	235	501	8502	9905
	8118	8714	8814	8506	8277	42429		9904	9969	10027	9998	10102	50000

Table 6.4 - Abstract simulation results

Visually, the degree of homogamy can be seen by the size of the frequencies in the diagonal cells of the table relative to those in the off-diagonal cells. Beyond this, the social homogamy index values for each set of parameters can be compared. A limited neighbourhood constrains the number of possible partners for each agent. As a result, agents are forced to choose from a smaller pool of possible partners, which creates more sub-optimal matches. This is most noticeable with the "highest level" method of attraction, where those with the highest level of the trait partner quickly, leaving the remainder of the population to find a match amongst the lower levels.

Random attraction, where each level is equally desirable, can be considered a baseline against which to test the impact of conscious partner selection procedures. The random attraction method resulted in an even distribution of frequencies across every combination of "education" levels. The "most similar" method generally had the highest level of homogamy. This was because in this method the utility of the agents was maximised by homogamy (that is, having a partner of the most similar level to themselves). The only exception to this was when the "education" levels were left skewed, the "highest level" method created the highest level of homogamy of any of the simulations. This was because with left skewed levels of the trait, most agents have a high level of the trait. Therefore, most matches are between two agents with high levels, and there are fewer pairings between an agent with a high level and an agent with a low level of the trait.

Within each method of attraction, the different population distributions of “education” level resulted in small changes in the homogamy index, with the exception of the left skew data in the “highest level” method, which had a higher homogamy index than the “highest level” results for the other distributions. This was in part due to the way that the “highest level” method pairs agents from the top education levels downwards. In practical terms, the main difference was seen in the marginal distributions of the trait amongst the couples, where the row and column totals resembled the population distribution. Beyond this, the only noticeable effect was to amplify the patterns created by the decision methods.

6.6. Abstract Simulation Summary

The abstract simulation model provided an initial means of investigating the second research question. It examined patterns of partnership, by simulating partnership formation. Rather than using ethnicity, a generic hierarchical trait, labelled as “education” was used. Attraction to an agent who was “most similar”, and attraction to an agent with the “highest level” of the trait, were contrasted to random pairing and were both shown to generate higher levels of homogamy. Constraining agents to a “neighbourhood” decreased the level of homogamy because agents had less choice. The distribution of the traits made little difference to the level of homogamy, other than the interaction between the “highest level” method and when the trait was left skewed.

Although the abstract simulation model provides some insight into choice mechanisms, one of the difficulties with applying it to empirical ethnicity data is that there is not a natural, rankable hierarchy for ethnicity, unlike the case of education. The following chapter examines an empirical simulation of partnership choice using ethnicity data from the New Zealand census. However, it focuses on the “most similar” case, rather than trying to arbitrarily assign an ethnic hierarchy.

Chapter 7 – Empirical Simulation Modelling

The empirical simulation modelling follows on from the abstract simulation models shown in Chapter 6. By applying simulation models to Census data, social models can be evaluated in a real setting rather than an abstract one. This chapter describes the simulation of partnership formation using a model populated with unit-level census data. Whilst the log-linear modelling is a way of examining the historical patterns, the empirical simulation is a more dynamic method, tackling the idea of emergence by looking at how the patterns form from the matching process itself.

The simulation model was programmed in Java and applied to real, unit-level micro data from the New Zealand Census. In order to deal with the large amount of data and the associated number of iterations involved in the matching algorithms, the code was run across a secure grid computer system. In addition to the increased processing power, the grid computer system also provided the high level of security that was necessary for the unit-level census data that was used as the main input for populating the simulation.

This chapter outlines the simulation goals and describes the data that was used for the simulation. Details of the grid-based computing resources and the simulation algorithm that were used in the simulation are followed by the results of the simulations, and then a discussion of the findings. The chapter is concluded with a section on possible future simulation work in the area.

7.1. Simulation Goals

As with the statistical analysis, the key measures of interest in this chapter are the patterns of inter-ethnic cohabitation. The simulation modelling of these patterns uses non-ethnicity based micro variables, observation of the macro environment, and a random stochastic factor to simulate the partnering process. The simulation is run using the eighteen to thirty year olds who are single at one census, and then the ethnic partnering patterns are observed and compared to the actual patterns of cohabitation or partnership in the appropriate age group at the following census.

There are three main goals of the empirically based simulation model:

1. To examine the individual effect of each of the scoring variables (age, education, macro, random) on ethnic partnering patterns.
2. To find the weighted combination of the scoring variables that produces the most similar set of inter-ethnic cohabitation patterns to those that actually occurred, as measured by the sum of squared errors for the table proportions.
3. To investigate the possibility of a micro-macro relationship within ethnic partnering patterns.

The first goal assesses each scoring variable individually - using a one hundred percent weighting on that variable - to see what patterns of ethnic partnership occur in these circumstances. Although these scenarios do not provide as much explanatory power, they are of sociological interest, examining attraction using a single factor. They also provide a form of sensitivity testing for the model.

The second goal focuses on the prediction of the patterns of inter-ethnic partnership from one census to the next. It uses an evolutionary optimisation method to find the combination of variables, as observed by the variable weights, that generates results most closely reproducing those observed in the actual census cohorts. Since each census is treated as a separate cohort, five sets of weights are produced for the Auckland, Wellington and Canterbury regions. The weights are examined for patterns of convergence within the iterations of the algorithm. The differences in patterns between the different census periods are also compared.

Beyond examining the effect of different combinations of the scoring variables, the final goal focuses on the other phenomenon of interest; whether a relationship between micro-level behaviour and macro-level patterns can be observed. This is where the decisions of micro-level agents are influenced by macro level observation. When these decisions are collated they form the macro-level at the next time step, which can then be assessed for their impact on the decisions of the following set of micro agents, as discussed in Section 3.1.4.

7.2. Computer Resources for the Simulation

7.2.1. Enabling the Simulation Using Grid Technology

"Grid computing can be defined as coordinated resource sharing and problem solving in dynamic, multi-institutional collaborations. More simply, Grid computing typically involves using many resources (compute, data, I/O, instruments, etc) to solve a single, large problem that could not be performed on any one resource.
(Nabrzyski, Schopf, & Weglarz, 2004)

Simulating the interactions between hundreds of thousands of individuals requires a significant amount of computing power. Although it was possible to run trial simulations on a desktop PC, the Auckland cluster of the BeSTGRID computer network provided the high-end computing power required to run the census-based simulations. The BeSTGRID system allowed the stand-alone simulations to be supplemented with an evolutionary optimisation routine, where multiple simulations were run in parallel.

The BeSTGRID computer network (https://www.bestgrid.org) is a national eResearch project which was started as a Tertiary Education Commission-funded project. It includes shared computational resources made up of powerful computing clusters at several New Zealand universities, including the University of Auckland. The Auckland cluster features five systems of two nodes, each of which is powered by two quad core Xeon 2.8 GHz processors, providing a total of 80 cores, together with 250 gigabytes of disk space and 16 gigabytes of memory for each node. The 80 cores allow for appropriately written code to be processed in parallel, greatly improving computational performance.

Moving to a grid-based computer system provided two key advantages over working with a single machine. Firstly, even when working with a single core (CPU) of the grid there was a significant increase in performance relative to the desktop PC and the server for running the equivalent set of code. Secondly, and more importantly, the multiple processors of the grid allowed for parallel processing. This is where code can be split up and run across a number of CPUs in order to provide reduced running times and increased computational efficiency (Parry, Evans, & Heppenstall, 2006). In the case of this research, it meant that multiple simulations could be run simultaneously through an evolutionary algorithm in order to efficiently search for optimal sets of weights (see Section 7.3.5 for more on the optimisation method).

The grid is accessed using a secure shell client (http://www.ssh.com) and a secure file transfer protocol (FTP) programme. A proxy is set via the Grisu client (http://grisu.arcs.org.au/downloads/beta/webstart/) so that the grid can be accessed from multiple machines. The simulations are written in Java and operated via JDK 1.5 on the Auckland cluster of the grid. Unix shell scripts are used to enable the parallel processing of the search strategies, and can be viewed in Appendix C.3.2.

7.2.2. Data Security

Statistics New Zealand allowed a limited unit-level dataset to be securely stored on the BeSTGRID system in order to run the simulation models. The data was kept secure using the following contracted set of security procedures (taken from the Statistics New Zealand – researcher contract):

- Files (data and scripts) were securely copied using the Secure Copy (SCP) programme from the OpenSSH project (http://www.openbsd.org/cgi-bin/man.cgi?query=scp&sektion=1).
- Files were encrypted as they were loaded onto the research data storage, and remained encrypted while stored on the research data storage.
- Files for use on the BeSTGRID Grid were securely stored and password protected on The University of Auckland's research data storage server.

- File permissions were set so that only the system administrator and researcher could read and use the files.
- All variables in the files were coded numerically, and without headers, so that in the event of them being seen, there would be no obvious association to the data they described.
- When processing was run, the Grid would arrange for the script to be run on a cluster under its control. It arranged for a private scratch area, into which the data file could be securely copied from the research data storage. Data stored in the private scratch area on the cluster was securely erased when each job completed.
- While the job ran, the name of the data file, the name of the script file, and the name of the Grid job could be seen by other logged-in users. However, the input data files and output files were not viewable at any time.
- After the script had loaded the data file, it would then be securely erased. Results were returned by the Grid to the researcher's account on the research data storage server at the end of the job.
- Once the data are no longer required, the data will be securely removed from the research data storage. Secure Remove (SRM) was used for securely erasing (http://srm.sourceforge.net/).

7.3. The Simulation Model

The following sections discuss the formulation of the simulation model by examining the input variables, the simulation algorithm and scoring function, and the method for optimising the weightings of the different factors. The model is loosely based on the DYNASIM model (Zedlewski, 1990) which was used to simulate partnership as part of a taxation model on American data and has subsequently been applied to a national level simulation (APPSIM) of Australian data (Bacon & Pennec, 2007). It seeks to improve the sociological interpretability of the simulation by also incorporating facets of several other microsimulation and agent-based simulation models of partnership.

7.3.1. Simulation Input Data

The simulation is populated with unit level data from the New Zealand Census of Population and Dwellings in such a way as to create a simulation environment which closely resembles regions of New Zealand at each census date. One of the strengths of the study is that the agents in the simulation are the actual unit-level records for the entire population of the selected regions. They are not scaled up from a census sub-sample (Pennec & Bacon, 2007), nor have they been reverse-engineered from frequency tables in order to match marginal distributions (Bouffard et al., 2001). They are the actual unit level records for the sub-population that is being studied. The creation of the variables and the census variable codes was described in Chapter 4, and the security of the data provided by Statistics New Zealand has been outlined in Section 7.2.2. Census data from 1981 to 2001 are used as inputs for the simulation, but the 2006 data is not. As the evolutionary algorithm component of the simulation (explained in Section 7.3.5) uses data from the following census for predictive purposes, this leaves the 2006 census without a succeeding census benchmark.

The agents in the model represent all those individuals who are unpartnered and aged between eighteen and thirty in the Auckland, Wellington and Canterbury territorial regions. These regions were chosen for two reasons. Firstly, they are three of the largest territorial authorities, thus alleviating privacy concerns with the use of unit-level data.

Secondly, the three regions represent three different levels of ethnic diversity, from the high level of ethnic diversity in Auckland to the lower levels in Canterbury and Wellington. The regions were identified using the `cn_reg_council01_code` census variable which represents the territorial council region on census night. The territorial council regions were also the most detailed level of geographical data that could be released at a unit level (due to the privacy constraints discussed in Section 4.1). As a result, the construction of the simulation algorithm and scoring method was constrained by the availability of data. Thus, only age, education and ethnicity covariates for each region were available. This meant that only matching using combinations of these variables could be used.

The age grouping of eighteen to thirty year olds was used in order to create a fixed cohort that could be compared from one census to the next. The eighteen to thirty year old group at one census would become the twenty three to thirty five year old group at the next, allowing inter-census comparisons to be made and validation of the results to be performed. This age grouping was also used in the log-linear models in Section 5.2.4 as a way of providing a reasonably close approximation to the cohort of recently formed partnerships.

Agents were identified as single or unpartnered using the same process as that for identifying cohabitating couples, as described in Chapter 3. The `id_family` variable which indicates people living in the same family group, in the same household, was used to identify individuals who were not living with a partner.

Each agent in the simulation has age, gender, ethnicity and highest qualification (as described in Chapter 3) values assigned from the census data. Agents can see the traits of the other agents with which they interact, and they are also aware of the macro properties of their region (such as the proportion of homogamous partnerships in their area). It is assumed that there is no interaction across the three regions, given the degree of geographical separation. It should also be noted that there is no data available on any social interactions across these regions.

Due to privacy regulations with Census data, only one-way tables of frequencies can be shown and cells with counts of less than one hundred have been replaced with an X. Table 7.1 to Table 7.3 show the marginal distributions of ethnicity, age and education levels for each region, for the single people aged eighteen to thirty in each census (who are to be matched in the simulation).

	1981			1986			1991			1996			2001		
Ethnicity	**A**[6]	**W**[6]	**C**[6]	**A**	**W**	**C**	**A**	**W**	**C**	**A**	**W**	**C**	**A**	**W**	**C**
European Only	29225	17635	19414	34484	19592	20865	37392	20649	23951	34907	18884	24655	34747	18982	22912
Maori Only	2544	1205	330	6549	3028	1202	7666	3134	1547	5847	2285	1272	5961	2350	1217
Pacific Only	3161	910	167	5039	1245	293	6603	1401	314	5774	1285	313	7393	1405	364
Asian Only	966	647	444	1270	842	645	3858	1407	643	5311	1215	1924	11930	1878	2521
MELAA only	X	X	X	X	X	X	200	X	X	334	157	113	638	241	155
Other only	X	X	X	371	220	221	163	X	X	8714	2654	1857	8991	2052	1663
Maori and Euro	3239	1784	977	1569	731	427	1863	811	495	3394	1529	1207	4096	1853	1329
Euro and Pacific	104	X	X	122	X	X	202	X	X	365	101	X	508	147	X
Euro and Asian	556	166	X	381	124	X	422	166	X	924	308	156	991	327	148
Maori and Pacific	X	X	X	X	X	X	156	X	X	440	179	134	227	156	115
Other multiple	823	400	252	323	117	X	419	148	X	5628	2392	2085	4064	1539	1263

Table 7.1 - Ethnicity distributions for simulation

Table 7.1 shows the ethnicities of the single people in each of the three regions. Although Auckland remains the most ethnically diverse region, there has been an increase in the ethnic diversity in all three of the regions. The European Only group forms the majority in each region in each time period. However, as a proportion of the total population, several groups have changed over time. Figure 7.1 shows the proportions of the main four ethnic groups. The Auckland data has the highest level of ethnic diversity, and data for the Canterbury region has the lowest. The proportion of Asian Only singles is much higher in the later census periods for all three regions, and the proportion of European Only singles has declined.

[6] A=Auckland, W=Wellington, C=Canterbury

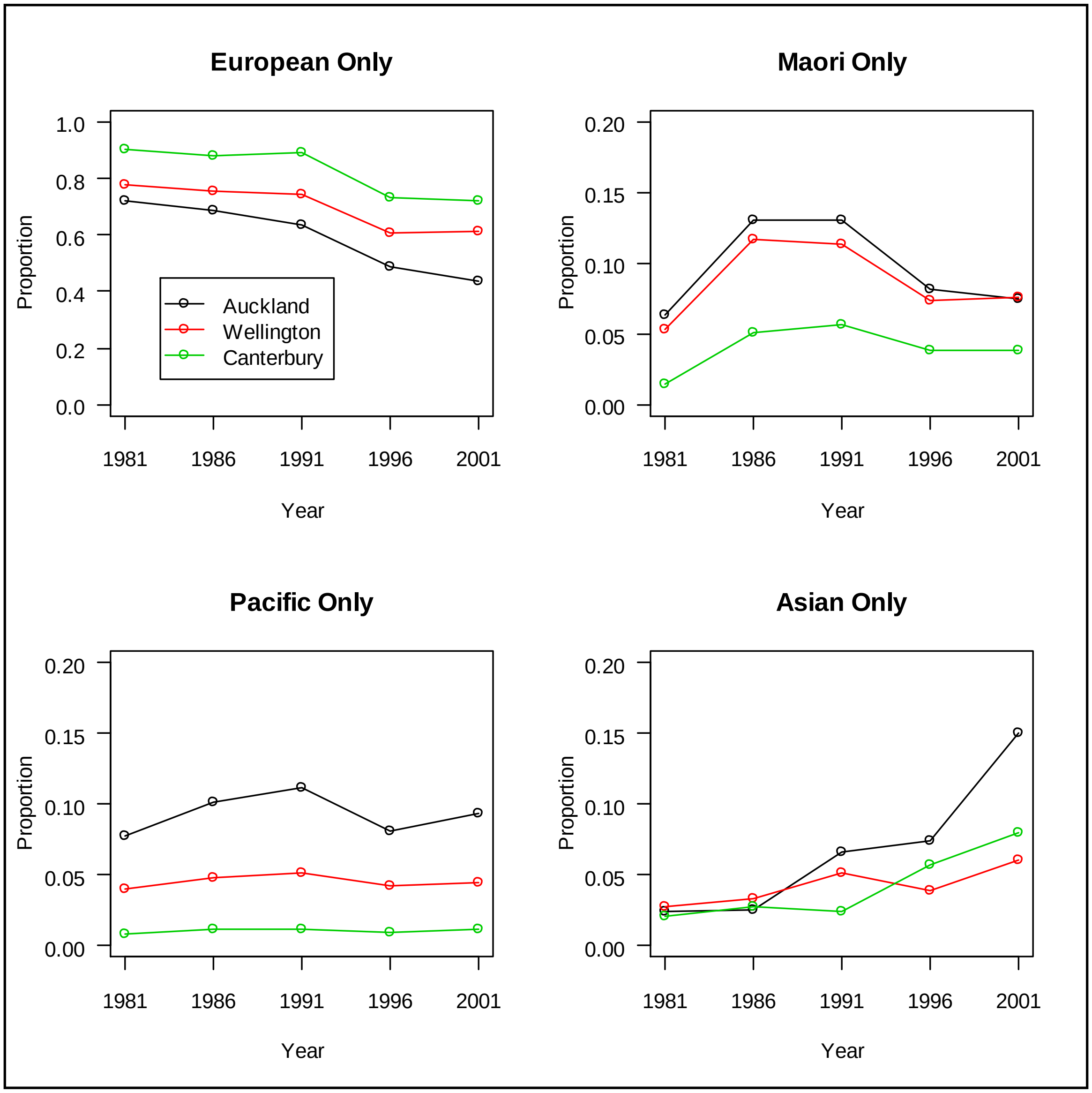

Figure 7.1 - Main ethnicity groupings

Table 7.2 shows the highest qualifications for each of the agents. In the 1981 census, the education classifications were limited to "no education" and "secondary education". Subsequent censuses also had tertiary and vocational options. Over time there has been an increase in the proportion of people in all three regions with either tertiary or vocational qualifications, with a slightly higher proportion in Wellington than in the other two regions.

Education Level	1981			1986			1991			1996			2001		
	A	W	C	A	W	C	A	W	C	A	W	C	A	W	C
No Qualification /Unknown	15056	6816	6735	15771	6437	6169	16277	6156	6492	30161	10324	10848	27089	7814	8586
School Qualification	25677	16027	14964	20246	10867	10872	25424	11978	13095	23301	10320	14080	29887	11861	14139
Vocational Qualification				8337	3845	3766	10074	4417	4329	10317	4804	5213	11747	4950	5273
Tertiary Qualification				5942	4837	3040	7169	5460	3367	7859	5541	3607	10823	6305	3724

Table 7.2 - Highest qualification distribution for simulation

Table 7.3 shows the age distribution of the agents in each region. The table shows how the distribution of single people across the ages of 18 to 30 has changed since 1981. As the population has aged, the number of single people in their mid to late twenties has increased but this has not been matched by an increase in the younger half of the cohort. For example, in Auckland, the 18 to 24 year olds account for 62% of those aged 18 to 30 in 1981, but by 2001 this had reduced to 52%.

	1981			1986			1991			1996			2001		
Age	A	W	C	A	W	C	A	W	C	A	W	C	A	W	C
18	2841	1678	1541	2791	1394	1417	2487	1118	1478	3272	1233	1775	4763	1581	1908
19	3602	2249	2285	3657	2011	2129	3756	1877	2406	4228	1878	2767	5507	2265	2807
20	4002	2516	2683	4306	2267	2374	4552	2366	2739	4853	2200	3239	5906	2417	3091
21	4134	2487	2521	4518	2435	2528	5041	2693	2880	5362	2534	3333	6180	2611	3027
22	3893	2351	2153	4795	2673	2455	5160	2648	2540	5987	2735	3102	6404	2628	2648
23	3549	2127	1857	4769	2555	2163	5107	2613	2326	6345	2846	3067	6465	2643	2531
24	3298	1789	1596	4593	2355	2046	5000	2446	2108	6488	2952	2888	6458	2635	2427
25	2941	1654	1480	4248	2253	1847	4968	2333	1981	6367	2835	2619	6384	2515	2265
26	2799	1433	1232	3839	2001	1626	4803	2106	1896	6265	2705	2472	6085	2440	2215
27	2557	1274	1151	3597	1748	1441	4773	2189	1848	5995	2415	2293	6257	2344	2139
28	2496	1156	1113	3236	1551	1375	4592	1986	1778	5755	2345	2235	6227	2252	2276
29	2386	1062	1005	3074	1450	1301	4497	1892	1707	5319	2174	2081	6464	2284	2207
30	2235	1067	1082	2873	1293	1145	4208	1744	1596	5402	2137	1877	6446	2315	2181

Table 7.3 - Age distribution for simulation

There have been a number of changes in the demographic makeup of the 18 to 30 year old single population used for the simulation input data. Although these changes may have an effect on the simulation results, they may also be as a result of the changing patterns of partnering in the previous period. With the exception of the 1981 education levels, the demographic information has been collected in a consistent manner and subsequently reflects the real changes in the demographic makeup of these regions of New Zealand (see Section 4.2 and Table 4.2). The variable which has shown the biggest change across the three regions is ethnicity. This may mean that changes in ethnic patterns may be due to changing availability, rather than changing attitudes towards other ethnicities.

7.3.2. Simulation Algorithm

This section provides an explanation and justification for the partnering algorithm that is used to match up the agents. The algorithm used for the simulation builds on the ones used in the DYNASIM (Zedlewski, 1990) and APPSIM (Bacon & Pennec, 2007) simulation models. The simplicity and intuitive reasoning used in the algorithms of these two models was one of the appealing factors for using a similar style of algorithm. In addition, the algorithm also appealed because it could be applied to the census datasets without requiring additional information or adaptation of the data. Many of the other algorithms that were also considered used data above and beyond what was available in the unit-level census datasets (see Table 3.1 and Table 3.2).

The basic premise of the model is that each single male agent is initially assigned a random "social network" of potential female partners. Each of these potential partners are assigned a score using an exponential function that is based on the DYNASIM and APPSIM models and is described further in Section 7.3.3. Starting with the couple with the highest score, partnerships are formed and those agents are removed from the system. The total number of partnerships formed at each time step in the simulation is determined by dividing the total number of partnerships that were actually formed over the five-year census period by the number of time steps used in that simulation. The social network of each remaining male agent is then appended by a certain number of additional female agents as the male agents "meet" new people.

One full execution of the simulation represents a five-year census period. It is run for a pre-determined number of time steps, where each time step represents a period of five years divided by the number of time steps that are used. The default number of time steps used is five, meaning that each time step represents one year. Once the simulation completes the final time step a cross-tabulation of partnerships by ethnicity and a calculation of the deviation from the actual census figures are generated. The simulations for each of the three regions are run independently of one another.

Figure 7.2 shows a diagram of the partnering algorithm.

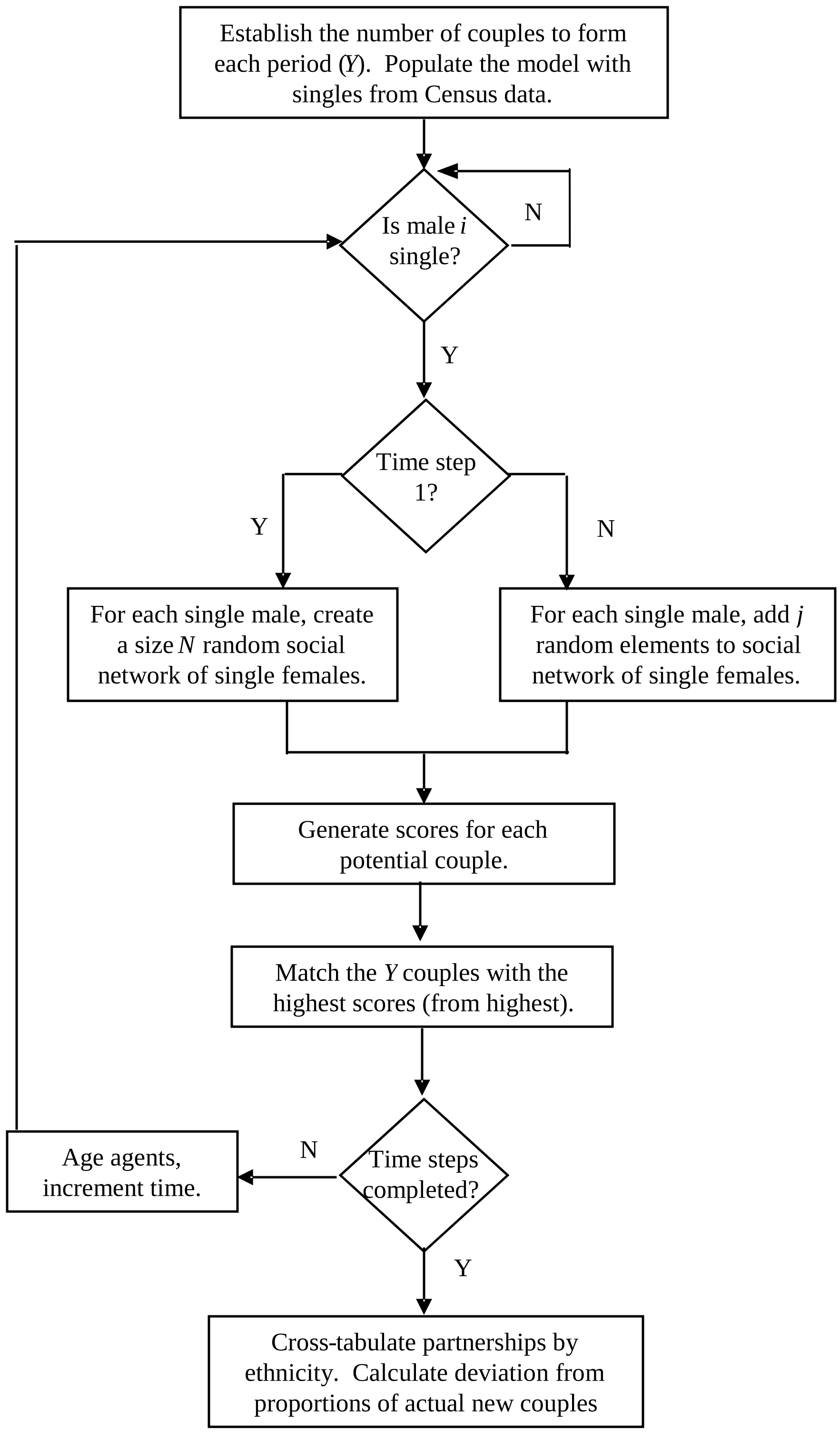

Figure 7.2 - Simulation algorithm

The simulation starts by having each male agent form a randomly assigned social network of female agents. This is the initial pool of women that will be evaluated as possible partners. For simplicity each network is generated to be the same size, although the effect of varying this size is examined. Although there is evidence that social networks often exhibit homogamy (McPherson et al., 2001), most simulation models still use randomly allocated social networks (Bacon & Pennec, 2007; Hills & Todd, 2008; Todd & Billari, 2003; Todd et al., 2005). The two main reasons for this are that either a more complex method of allocating networks would complicate the model and possibly confound any effects that are seen, or there is insufficient information about the structure of social networks in the population of interest for any significant improvements to be made in the allocation of the networks. For this simulation there was no information about the social networks of New Zealanders and there was only limited information available in the simulation dataset that could have been used for allocating the networks. Beyond this, a more complex method for allocating the social networks could confound any patterns that are generated.

Although the simulation forms the social networks around the male agents, iterations of the simulation were also run with male and female roles reversed in order to check for any asymmetry in the choice mechanism. Although having the agents of one gender being randomly allocated to the agents of the other gender could leave some agents who are not part of any network, a one-way matching process produced a simpler algorithm than a two-way process. However, no significant differences were seen in the patterns of ethnicity generated by the simulation models when they were run with the male and female roles swapped.

The potential partnerships of each male agent with each female agent in his network are scored using an exponential function which is based on the DYNASIM and APPSIM models and is described in detail in Section 7.3.3. The scoring mechanism incorporates the similarity of the age and education levels of the agent, a stochastic “attraction” factor, an observed macro factor, and an age dependent component. The age dependent component incorporates the idea that as agents age they become more willing to accept a suboptimal match. In some partnership simulations (Alam & Meyer, 2008; Hills & Todd, 2008; and others) an attraction "threshold" is used to make this adjustment. This simulation incorporates the decaying expectations into the main scoring function rather than matching all couples above a certain threshold score. This is because the simulation sets the number of couples that will partner in each time period at the outset of the simulation and then matches the couples in order, starting from the highest attraction score. This is further discussed in the scoring functions Section (7.3.3).

At each time step, once the scores have been allocated, $N/_T$ couples are formed, where N is the net change in the number of couples in the age cohort over that five year census period and T is the number of time steps for that particular simulation. This methodology was one of the improvements that APPSIM made over DYNASIM, ensuring that the simulation will produce the correct number of couples seen in the census data. For this simulation it allows for a much easier comparison of the ethnic patterns within the simulation output.

One of the improvements that this simulation makes over the DYNASIM and APPSIM models is the use of a competitive marriage market, where the couples with the strongest level of attraction are paired first. Since the number of comparisons required in a competitive partnership model make it computationally expensive, marriage models have tended to match agents as they go. This creates the problem that the agents that get matched first are not necessarily the most attracted to each other or the best match, but are in part just a by-product of being first on a randomly sorted list. Once an agent has been paired with a partner, they are no longer available to be partnered again and will not get added to any new social networks.

At the end of each time step the agents are aged by one time unit and then the social network for each male who is still single gets a certain number of new single females added to it. The purpose of this expansion of the social networks is to simulate the expanding social circles of each of the agents over time. Each social network grows by the same number of new agents, but the actual agents are allocated randomly so the agents are heterogeneous with regard to their social network. However, the growth of each social network is uniform. Although the uniformity of the growth is not realistic, it is done in order to simplify the coding of the algorithm.

The Java code for the simulation can be found in Appendix C.2.

7.3.3. The Scoring Function: Description

As with the simulation algorithm, the scoring function is adapted from the one that is used in the DYNASIM/APPSIM models that were discussed in Section 3.1.2 (Bacon & Pennec, 2007; Zedlewski, 1990). The age difference between potential partners, the difference in their educational qualifications and the ethnic patterns of recently formed partnerships are combined with a random stochastic variable to form the scoring function that agents use to evaluate one another. This section details the theoretical and pragmatic reasons for selecting these particular variables for scoring the potential partnerships. It examines each of the components of the scoring function, shows an example score, and provides references and justification for their inclusion.

From a practical point-of-view the simulation scoring function had to work with the variables that were available in the unit-level datasets; that is, the age, ethnicity, education and sex of the single eighteen to thirty year-olds in the Auckland, Wellington and Canterbury regions. This limitation meant that a logistic regression-based simulation scoring function, such as those seen in some of the simulations discussed in Chapter 3 (Bouffard et al., 2001; O'Donoghue et al., 2009; Perese, 2002; Spielauer & Vencatasawmy, 2001) could not be used. In contrast, the exponential scoring function in the DYNASIM and APPSIM models, that had previously been shown to work well for both the Australian and American populations, could easily be adapted for the data that was available. Each of these simulations uses an exponential function of the form:

$$P(Union_{mf}) = e^{-0.5\sqrt{\left[\left(Age_m - Age_f\right)^2 + \left(Edu_m - Edu_f\right)^2\right]}}$$

to calculate the probability of a match. In the DYNASIM model this is compared to a random uniform number, with a match made if the probability is greater than the random number attempts. In the APPSIM simulation, this process is altered to incorporate the number of partnerships that are expected to occur in each time period. This simulation follows a similar line of logic. However, it incorporates the stochastic component into the score itself and then ranks the scores to mimic a competitive environment rather than just comparing each possible match independently to a random number. By ranking the scores, this overcomes the issue of the time-dependency effect within each time period. Further details on the algorithm were previously discussed in Section 7.3.2.

The decision to use those variables was also based on their repeated appearance in the partnership choice literature (see Table 3.1 and Table 3.2 for a summary of partnership simulation models, and Section 2.1 and 2.2 for discussion on partnership literature). By limiting the function to four variables it remains relatively parsimonious and works within the limitations on the level of detail that could be provided with unit-level sets of census data. Ethnic preference is not included as an explicit micro-level term in the scoring function. This is in order to simulate ethnic patterns as a function of non-ethnic factors, rather than having the simulated ethnic patterns only reflecting the ethnic preferences incorporated in the model. However, it does incorporate ethnicity at a macro-level with a variable that is based on examining the number of homogamous and non-homogamous partnerships formed in the previous time period.

The score between male *i* and female *j* is shown below:

$$Score_{i,j} = e^{-\sqrt{threshold + \pi_1 (age_i - age_j)^2 + \pi_2 (edu_i - edu_j)^2 - \pi_3 macro_{ij,t} + \pi_4 random}}$$

$$\text{where}$$

$$threshold = \left(80 - age_i - age_j\right)$$

The function has obvious similarities to the DYNASIM and APPSIM ones, featuring an exponentiated square root, and the squared differences of the age and education level of the two agents. It also adds a macro term, a random term, an age decaying threshold component and a set of weights (π_1 to π_4) that can be used to control the relative contribution of each variable to the total score. The function can be thought of as a probability or scoring function, but it can also be thought of in economic terms as a utility function. It combines the utility derived from educational homogamy, partner age difference, conforming to social norms, and random attraction. Table 7.4 provides a summary of the variables that are used in the function, followed by a brief discussion of each. This is followed by Section 7.3.4 where justifications beyond the data availability are provided for each of the variables.

The exponential and square root components of the function are not strictly necessary to a system for scoring and ranking. There are, however, two advantages to incorporating them into the function. By taking the exponential of the square root of the sum of the terms in the equation, the function more closely resembles the DYNASIM and APPSIM models that it is based on. The exponential decay property is also seen in several other partnership models such as the MADAM model (Hills & Todd, 2008). One disadvantage of the square root term is that it means the sum of the scoring terms must be non-negative. This problem is overcome by incorporating a constant value of 80 in the "threshold" term so that no combination of variable values can generate a negative term for which a square root cannot be calculated.

Table 7.4 shows each component of the scoring function, together with a brief explanation and the range of possible values.

Variable	Explanation	Range
$80 - age_i - age_j$	Age decaying "threshold" factor	20 to 44
$(age_i - age_j)^2$	Age similarity factor	0 to 144
$(edu_i - edu_j)^2$	Educational similarity factor	0 to 9
macro	Macro term: Social pressure rate	-10 to 0
random	Random Uniform(0-9) term	0 to 9
π_i	Weight for variable *i*	0 to 1

Table 7.4 - Scoring variables

The education, macro and random terms are all of a similar range to one another and, although the maximum possible value for the age difference term is 144, this term will also typically take on values between zero and ten. By keeping the variables to within a similar range of one another, the optimised weights generated by the evolutionary algorithm (see Section 7.3.5) are more meaningful as they are not being influenced by any kind of scale effect (i.e. if the education scores ranged from zero to nine, but the macro scores were measured from zero to one thousand, then a 5% weight on macro could equate to much more than a 90% weight on education).

There are two age terms in the formula. The first age term serves several purposes. It increases the score/probability for older agents, mimicking a decaying threshold on their expectations. The constant term in it also ensures that the sum of the variable scores cannot be negative. The second age term squares the difference between the ages of the two agents. At its extreme, this would be 144, where one agent was thirty and the other was eighteen. The education term takes the difference between the highest qualification scores for each agent. Since the individual scores range from one to four (one to two in 1981) the squared difference of these scores will range from zero to nine. The macro term can range between zero and negative ten. It acts as a positive social reinforcement, with higher scores for couples whose ethnic pattern matches the more commonly formed partnerships in the last time period. It is scored by the number of couples (as a rate per 10 formations) who have the same ethnicity pattern (homogamous or non-homogamous) as the couple being scored. The random term is generated from a Uniform(0,9) distribution and provides a random element to the attraction scores. Further discussion and justification of these variables can be seen in Section 7.3.4.

The π_i terms represent the weights for each of the components. They must be non-negative and sum to one. The model can be run with a specified set of weights, or using the evolutionary optimisation routine described in Section 7.3.5. In addition to the optimised set of weights, other combinations are run to explore "what if" situations. For example, what would the patterns of inter-ethnic cohabitation look like if partnership choices were made based purely on age differences (i.e. one hundred percent weight on the age difference term)? Although running the simulation with a one hundred percent weighting on one variable will produce a sub-optimal set of pairings relative to the evolutionary algorithm, it helps to build a better picture of the patterns created by partnership matching.

In applying the scoring formula we might, for example, see a male agent who is evaluating two female agents. In this simulation, perhaps the weights are equally distributed as 0.25 for each variable. Suppose that the male agent is Maori, 26 and has an education level of 2 and the first female agent is also Maori, 24 and has an education level of 2. If the random "attraction" score was 5, and 7 out of every 10 partnerships in the last time step were homogamous, then their mutual attraction score from the function would be:

$$e^{-\sqrt{80-26-24+0.25(26-24)^2+0.25(2-2)^2-0.25\times7+0.25\times5}} = 0.003995$$

By comparison, if the same male also had the opportunity to evaluate another female agent as a potential partner and she was Asian, 22, had an education level of 1, and 3 out of every 10 partnerships in the last time step were not homogamous, and their random score was 2, then their score would be:

$$e^{-\sqrt{80-26-22+0.25(26-22)^2+0.25(2-1)^2-0.25\times3+0.25\times2}} = 0.002479$$

If these were the only two options, then the algorithm would allocate the man to partner the first of the two females since they generated a higher score together. In the simulation, the social networks start with 50 partnering possibilities, and then 10 further possibilities are added after each time period to simulate new acquaintances made over the period. Section 7.4.1 discusses the sensitivity analysis of the social network size and growth.

7.3.4. The Scoring Function: Justification

Although the structure and variables that were provided in the unit-level census datasets provided one of the key drivers for the selection of each of the variables, the literature on partnership choice provided further reasons for their selection. This section provides a discussion and justification for each of the variables in the scoring function based on other literature on the simulation of partnership matching. It elaborates on the specifics of this simulation model and relates them back to the models described in Table 3.1 and Table 3.2 of the literature review.

As previously mentioned, the “age decaying threshold” (80 – male age – female age) component serves two functions. The first is to keep the function non-negative. Negative values were not a problem for the DYNASIM and APPSIM models as they only used the sum of the squared age and educational differences. By incorporating the constant into the equation, it was possible to have variables that increased the score or decreased the score without worrying about having to take the square root of a negative number. Secondly, it also increased the scores for the older agents in the model, simulating an age dependent satisficing component, or “decaying aspirations”, in a similar way to a number of other simulation models (Alam & Meyer, 2008; Hills & Todd, 2008; Todd, 1997; Todd & Billari, 2003). Whilst some models have a separate threshold mechanism, this simulation model instead fixes the number of couples that can form per period and adjusts the scoring mechanism. This has the same effect as a decaying threshold by increasing the chances of the older agents forming a couple, even if they are not as strong a match as other variables.

The age difference between partners has been repeatedly shown to have a strong correlation with partnership choice (Bouffard et al., 2001; Cheesbrough & Scott, 2003; Perese, 2002). Figure 7.3 from Logan *et.al.* (2008) shows the relationship between the ages of couples in the American National Survey of Families and Households. A strong linear trend is evident for a large proportion of the data, indicating a strong relationship between the age of the male and the age of the female in most couples. The data points along each axis represent the observations in the study where only the age of one partners was known. This relationship is represented in the scoring function by the squared difference in the ages. This is the same as that used in the DYNASIM and APPSIM scoring functions described in Section 7.3.3.

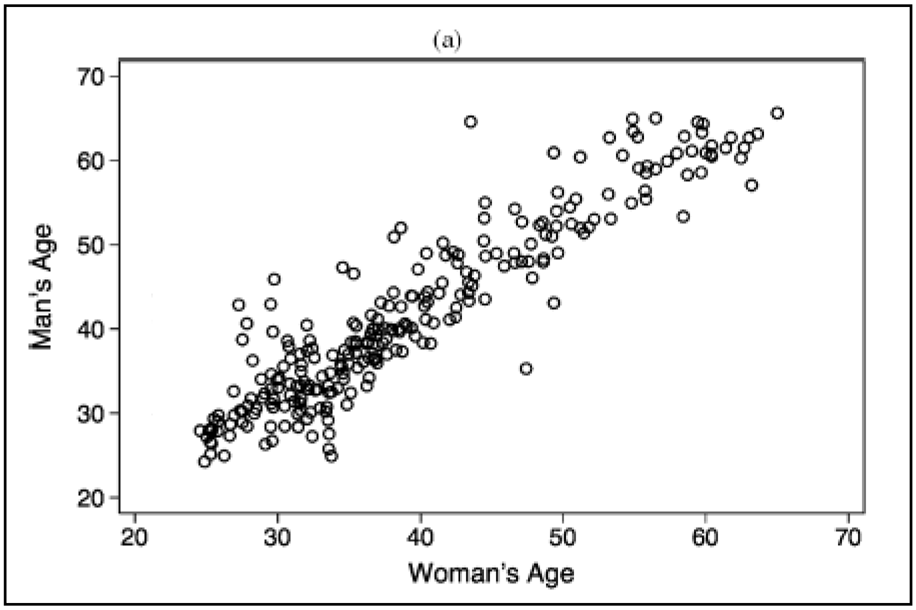

Figure 7.3 - Age scatterplot from Logan et.al (2008)

The assumption of couples of the same age producing an optimal match and the age differences being treated symmetrically is not completely realistic, but helps to simplify the scoring function. It has also been shown to work in the DYNASIM and APPSIM simulations. Historically, the male partner has tended to be slightly older than the female partner in most countries, and the distribution of the age gap is normally slightly skewed in this direction. However, in New Zealand this age gap has been reducing over time, from a median age gap of 2.66 years in 1963 to 2.28 years in 1983 to 1.94 years in 2003 (Ryan, Boddington, & Dunstan, 2005). The symmetrical treatment sees the model correspond more closely to the two models it is based on, although future work could involve a more realistic treatment of the age difference.

Empirical demographic and sociological studies have shown a strong degree of educational homogamy within marriage, often measured by the number of years of education. Although the census does not collect the number of years of education, it does record highest educational qualification. Logan *et.al.* (2008) plotted the years of education (see Figure 7.4) from the American National Survey of Families and Households. Although the relationship is not as strong as it was for age, there is still a clear positive linear relationship between the number of years of education for men and women. Smits (2003) looked at educational homogamy in 55 countries and found New Zealand to have lower levels than many other countries, although still some level of homogamous partnering by education.

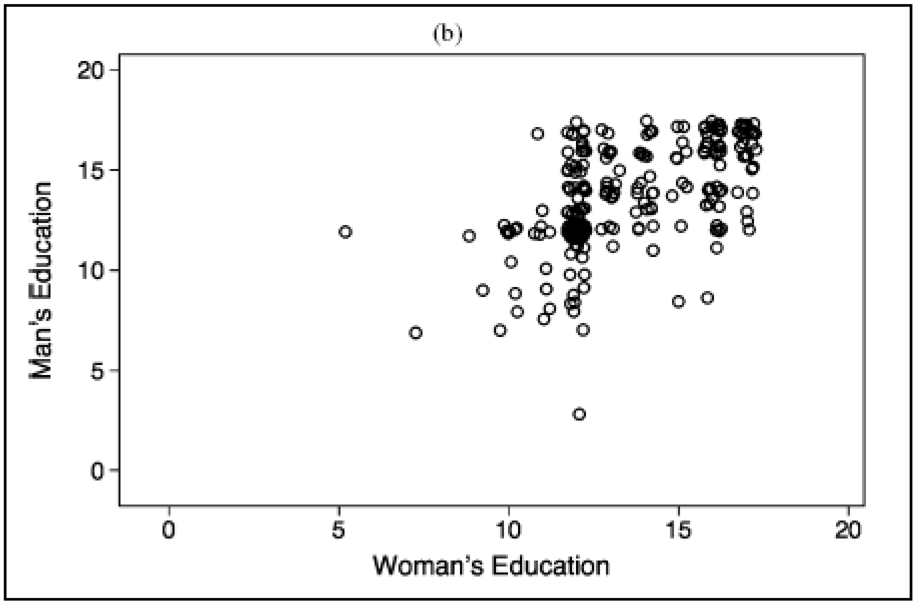

Figure 7.4 - Education scatterplot from Logan et.al (2008)

A model that sees educational homogamy as the optimal match does discount the theory of upward marriage. However, it has been argued nationally (Callister, 1998), and internationally (Kalmijn, 1994), that educational establishments select based on cognitive ability rather than ethnicity, and provide opportunities for people of marrying age to meet others of a similar status. Although it uses attainment categories rather than years of education, the model still provides a good differentiation between individuals - other than in 1981, where the data could only be split into the two categories of “no education” and “high school education”.

This simulation model extends the DYNASIM and APPSIM models by incorporating a macro-level process and a random stochastic factor. The $macro_t$ term is a function of the rate of same ethnicity and inter-ethnic coupling in the agent's region during the previous time period. When a possible match is scored, this term subtracts the number of couples in the previous period, as a rate out of ten, who had the same kind of match (either homogamous or heterogamous). For example, if six out of every ten matches in the previous period had been homogamous, then a European male and a European female would have a macro term of minus six. The term can be thought of as the social pressure to make that kind of match. It is a macro-level term, in that it is a society (or in this case region) level factor, but it influences the micro-level decision making. It gets updated at each time-step, leading to a possible micro-macro link, where the micro-level decision making in one time period creates a macro-level phenomenon (the number of matches of this type), which in turn affects the micro-level decision making in the next time period.

The *random* term represents a positive, uniformly distributed random stochastic value that is used to represent random attraction. The random number is drawn from the Uniform distribution, and is generated by the java code: `*Math.random()`. The range for this value is from zero to nine, keeping it to a similar scale to the other variables. Since a higher random value will decrease the total score, it could be thought of as a negative attraction or repulsion term. Random matching can be thought of as a baseline, where the simulated individuals are not only colour-blind, but have no other specific preferences for matching which could be correlated to ethnicity. As a result, the optimisation method is expected to shift away from the random term, towards the age, education and/or the macro ethnicity term, if they have some kind of effect on the ethnic partnering patterns. A discussion on comparisons between random matching and assortative matching can be found in Jaffe (1999).

7.3.5. Optimisation of the Weights

The π_i terms in the scoring mechanism represent the relative weights for each of the four components. An evolutionary optimisation algorithm is used to determine the combination of weights that minimises the total squared error of the cell proportions in each simulated frequency table relative to the actual proportions seen in the corresponding Census table.

Evolutionary algorithms were originally developed to solve combinatorial optimisation problems, particularly in engineering, but are increasingly used to solve or optimise functions whose evaluation is computationally expensive (Regis & Shoemaker, 2004). Luna *et.al.* (2008) explain that an evolutionary algorithm is a metaheuristic technique designed to iteratively converge to an optimal[7] solution by testing a series of tentative solutions, each of which is assigned some kind of "fitness" value according to its suitability as a solution. From here, a new set of tentative solutions is perturbed around the best of those solutions, and the process repeated until the fitness value has reached an acceptably low (or high) value, or a certain number of iterations have been performed. The operation of the algorithm is analogous to evolution. At each time step or generation, the various solutions are evaluated. The one with the greatest "fitness" gets to propagate whilst the others die off. Over time, this process sees the solutions get stronger as they converge towards the solution that gives the greatest "fitness".

An evolutionary algorithm is described by Jacob (2003) as a method that simulates a "collective learning process within a population of individuals", where the individuals represent a point in the search space of the problem. He describes how after an arbitrary initialisation, the algorithm will move (evolve) towards better and better regions of the search space. Schewfel (2000) writes that evolutionary algorithms can be generalised as a series of robust optimisation methods that are most suitable for computationally-expensive solving functions with highly variable or irregular "fitness landscapes". Computation may be difficult due to noisy non-differentiable, multimodal data, or there may be problems because limited iterations are possible. The main disadvantage to the method is that it may sacrifice an "absolute" best (but possibly incomputable) solution in order to settle at a "reasonable" one.

Luna *et.al.* (2008) introduces a $(\mu + \lambda)$ evolutionary algorithm (notation originally from Schwefel (1981)) in order to solve a mobile communications frequency allocation problem. The μ term refers to the number of initial solutions and the λ term refers to how

[7] Based on some predetermined measure.

many additional, perturbed possibilities are examined at each iteration. They use the following pseudo-code to explain the process:

```
P = new Population(μ);
PAux = new Population(μ + λ);
init(P);
evaluate(P)
PAux = addTo(PAux, P);
for iteration = 0 to NUMBER_OF_ITERATIONS do
     for i = 1 to λ do
          individual = select(P);
          perturbed = perturb(individual);
          evaluate(perturbed);
          PAux = addTo(PAux, perturbed);
     end for
     P = bestIndividuals(PAux, μ);
     PAux = P;
end for
```

from Luna *et.al.* (2008) in Nayak & Stojmenovic (2008)

Using this notation, the optimisation of ethnic partnership patterns is run as ten iterations of a (1, 17) evolutionary algorithm followed by 3 iterations of a (1, 15) evolutionary algorithm.

The process can be seen as analogous to a repeated series of regressions. Ashlock (2006) demonstrates that a simple y = a+bx linear regression can be computed by minimising the sum of squared errors in an $\mathbb{R}^3$ plane (see pg 232) via an evolutionary algorithm. By mutating the parameters in order to minimise the sum of squared errors (SSE), Ashlock demonstrated that the evolutionary algorithm could approximate the results of the least squares regression.

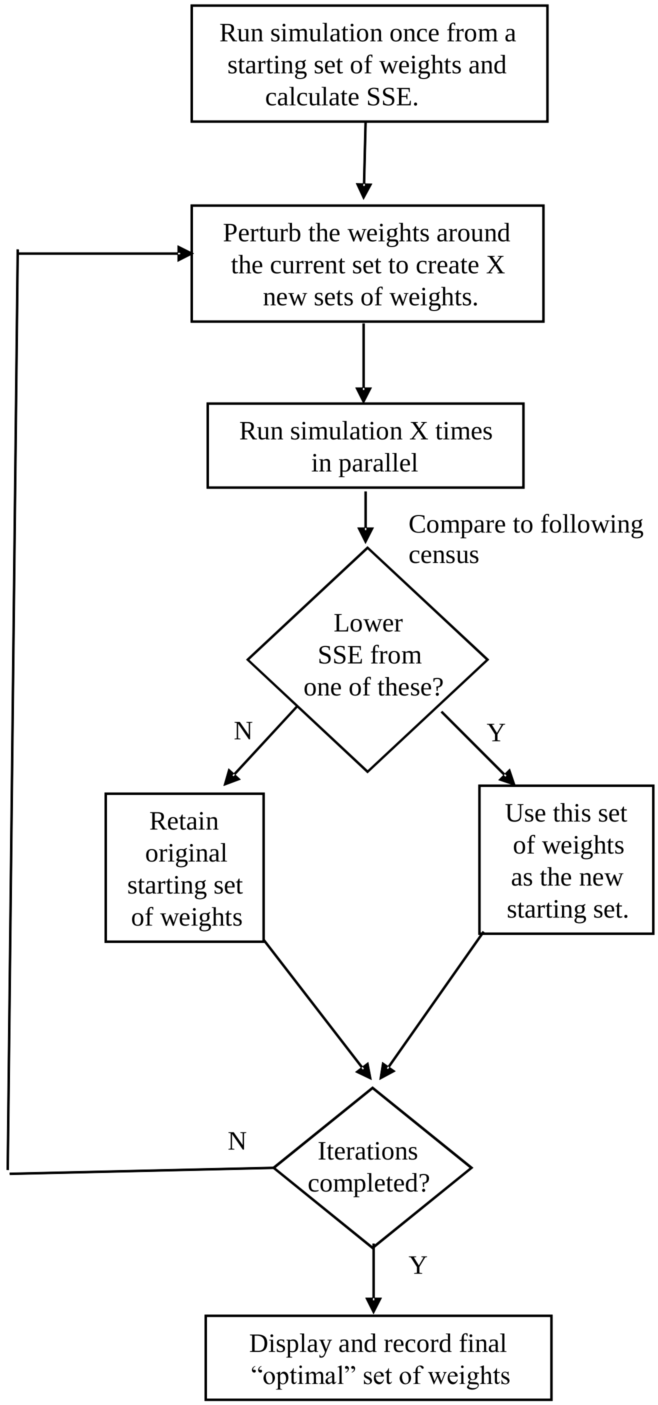

Figure 7.5 - Evolutionary algorithm diagram

Figure 7.5 shows how the evolutionary algorithm loops to find the combinations of weights for the four variables that minimise the total deviation from the actual census results for the next period. It starts by using an initial, user-defined set of weights to run the simulation, and find a starting sum of squared errors (SSE). A series of sets of perturbed weights are then generated using a set of perturbations (see Table 7.6), and the simulation is run simultaneously for each new set of weights. The total squared error (SSE) for each of the simulations with the new weights is calculated, and compared to the initial one. If one of the new sets of weights produces a squared error that is lower than before, then it becomes the new base for the weights to be perturbed in the next iteration. This process is repeated for a set number of iterations (contingent on the available processors and time). It should be noted that the intent of the algorithm is qualitative rather than quantitative. Due to the high degree of variability in the data and the partnership process, the focus of the algorithm is to investigate the patterns in the weights, rather than to determine an exact set of weights.

The formula for calculating the squared error term that the algorithm is trying to minimise is shown below.

$$SSE = \sum_{cell\ i=1}^{n} \left(p_{i,simulated} - p_{i,census}\right)^2$$

Where $p_{i,simulated}$ is the proportion of the total number of couples in cell *i* of the final cross-tabulated table of ethnicity for the paired couples in simulation, and $p_{i,census}$ is the proportion for the same combination in the following census. The sum of squared errors is a standard measure of variation that increases as the simulated proportions vary further from those observed in the census (Urdan, 2005).

For each run of the evolutionary algorithm a different initial sets of weights (π_1–π_4) were chosen. The weights had to sum to one ($\sum_{i=1}^{n} \pi_i = 1$), and be non-negative. Table 7.5 shows the five standard sets of initial weights. An initial set of weights using 25% for each variable, and each of the combinations of 100% for one of the variables, were tested for each census/region combination. In addition, several randomly generated combinations were also used. This was done to confirm that the algorithm would consistently converge to a similar solution set and not settle at some local minima.

Trial	Age (π_1)	Education (π_2)	Macro (π_3)	Random (π_4)
1	0.25	0.25	0.25	0.25
2	1	0	0	0
3	0	1	0	0
4	0	0	1	0
5	0	0	0	1

Table 7.5 - Standard initial weights

The two sets of perturbations for the weights are shown in Table 7.6. For the first 10 iterations of the algorithm, the larger "phase 1" perturbations are applied to the weights. This is followed by three more iterations using the smaller "phase 2" perturbations. The larger perturbation iterations are used to move the weights estimates more swiftly to the approximate area of the solution. The smaller set of perturbations is then used to refine the solution. Any perturbations that created negative weights were not run.

Phase 1				Phase 2			
π_1	π_2	π_3	π_4	π_1	π_2	π_3	π_4
-0.05	-0.05	-0.05	0.15	-0.02	-0.02	-0.02	0.06
-0.05	-0.05	0.05	0.05	-0.02	-0.02	0.02	0.02
-0.05	0.05	-0.05	0.05	-0.02	0.02	-0.02	0.02
0.05	-0.05	-0.05	0.05	0.02	-0.02	-0.02	0.02
-0.05	0.05	0.05	-0.05	-0.02	0.02	0.02	-0.02
0.05	-0.05	0.05	-0.05	0.02	-0.02	0.02	-0.02
0.05	0.05	-0.05	-0.05	0.02	0.02	-0.02	-0.02
0.05	0.05	0.05	-0.15	0.02	0.02	0.02	-0.06
0.15	0	0	-0.15	0.02	-0.02	0	0
0	0.15	0	-0.15	-0.02	0.02	0	0
0	0	0.15	-0.15	0.02	0	-0.02	0
0.15	-0.15	0	0	0.02	0	0	-0.02
0.15	0	-0.15	0	0	0.02	0	-0.02
0	0.15	-0.15	0	0	0.02	-0.02	0
0	-0.15	0.15	0	0	-0.02	0.02	0
-0.15	0.15	0	0				
-0.15	0	0.15	0				

Table 7.6 - Weight perturbations

The number of different perturbations per iteration, the number of iterations of the evolutionary algorithm, and the size of the perturbations were affected by the availability of processors on the grid, and the time that each iteration took to run. With only 80 cores in total on the grid, the number of simultaneous simulation runs – i.e. the number of different sets of perturbed weights – had to be limited for the consideration of other grid users. If a machine with more cores had been available, then more simultaneous perturbations could have been tested. One of the other factors affecting the number of iterations was the time taken per simulation. The larger datasets took approximately thirty minutes per iteration, meaning that the 13 iterations would typically take about five and half hours, provided there were sufficient cores available for the entire period. One final factor that influenced the number of iterations and the number of different perturbations per iteration was a bug in the grid operating software that limited the number of times that a user could access a core within a 12 hour period. According to the system administrators, this number was random, but seemed to normally occur after approximately 700 core uses. An average of 16 cores, which equated to 20% of the total cores, was established as an acceptable level of usage. Slightly more cores (17) were used for phase 1, where the simulation was trying to locate the general vicinity of the weights solution, and then 15 cores were used at phase 2. This resulted in a total core usage for one evolutionary algorithm run of 215 (10 x 17 + 3 x 15), which would allow for three runs of the simulation per day. Utilising 15 to 17 cores at a time, with 13 iterations of the algorithm, provided sufficient parallel processing for the evolutionary algorithm to converge, whilst staying within the usage limits of the resources.

The perturbations in the weights allow for shifts to and from each of the different variables. Some of the perturbations create a change in two variables only, whilst others perturb across the plane of all four variables. Even though it was expected that there would be a shift away from the random variable, opportunity still had to be provided in the weights for a shift towards it. There are two stages to the perturbations. The first sees large changes in the weights (up to ±0.15 at each iteration). The perturbation values of 5% to 15% were chosen after some experimentation. They provided the best combination of being large enough to ensure that the weights converged to a solution, but small enough to provide some degree of accuracy in that solution. After ten iterations the algorithm jumped to the second stage, where smaller sized perturbations were used to improve the accuracy of the solution. However, the smallest perturbations are ±0.02 so the results will not be perfectly precise. This was done to allow more flexibility and movement in the set of weights. Since it was expected that the prediction of partnership patterns would be a particularly noisy function, having very small perturbations would create a false sense of accuracy in the optimised sets of weights.

For example, a simulation might start off with a set of weights (0.25, 0.25, 025, 0.25) for the four weights parameters. The simulation is run once to establish the total squared error for this set of weights. Although this figure will vary slightly from run to run, the variation is fairly minor. See Section 7.4.3 for further discussion on the variation within the algorithm. Once the squared error is calculated for the initial run, the simulation is run simultaneously across multiple cores of the grid with each of the perturbations to the weights that are shown in Table 7.6. The total squared error for each of these simulations is calculated. If any of these simulations produce a lower total squared error than the initial run, then the lowest of these becomes the new starting point for the next set of perturbations. For example, if the (+0.15, 0, 0, -0.15) perturbation produced the lowest total squared error then the perturbations for the next iteration would start from (0.4, 0.25, 0.25, 0.1). The changes in the weights are recorded, and the plots of the evolving weights are shown in Section 7.4.3. Below is a sample of the output of the optimisation method, showing the weights being updated at each iteration of the algorithm.

```
[lwal036@grid1 test]$ ./itter1can01.sh 13 0.25 0.25 0.25 0.25
Submitting job...Done.
Job ID: uuid:2c7e0bf6-98d5-11de-97c3-e2e160fc8bd9
Termination time: 09/04/2009 22:00 GMT
Current job state: Active
Current job state: CleanUp
Current job state: Done
Destroying job...Done.
Cleaning up any delegated credentials...Done.
0.2 0.3 0.3 0.2
Delegating user credentials...Done.
Submitting job...Done.
Job ID: uuid:119990ca-98d6-11de-83a0-e2e160fc8bd9
Termination time: 09/04/2009 22:06 GMT
Current job state: Active
Current job state: CleanUp
Current job state: Done
Destroying job...Done.
Cleaning up any delegated credentials...Done.
0.2 0.45 0.3 0.05
```

Sample output of the evolutionary algorithm in progress

The Unix script which was used to run the evolutionary algorithm can be seen in Appendix C.3. Significant portions of this code were written by Yuriy Halytskyy from the University of Auckland BeSTGRID team.

7.4. Results

This section presents the results of the simulation. There are three subsections, examining the effect of changing some of the internal parameters of the model, the outcomes of running the simulation using only one of the four scoring variables (100% weight on that variable), and the results from evaluating the four scoring variables with the evolutionary algorithm. The first two sections are examined by comparing the patterns in the frequencies of the simulated couples with the net changes in the census frequencies between the simulated period and the following census. The evolutionary algorithm findings are presented with plots of the changes in the scoring variable weights over the iterations of the algorithm.

The number of actual partnership formations is calculated by taking the difference in frequencies between the eighteen to thirty year-olds in the simulated census period and the twenty-three to thirty-five year-olds in the following census. For some groups, this net change is quite small, so the main focus is on the larger ethnic groups. One example of this is the number of homogamous Maori Only couples. Although this group is fairly large overall, the net changes in the tables are relatively low. This is partly due to the small net change in the number of partnerships, but also due to changes in the self-definition of ethnicity (Carter et al., 2009).

7.4.1. Changing Internal Parameters

The first step in testing the model was to examine the impact of varying some of the internal parameters of the model. The effect of varying the initial size of the social network, and the number of time steps were investigated. In addition, the simulation was run with the gender roles reversed to investigate whether there was any gender-based asymmetry in the patterns created by the simulation.

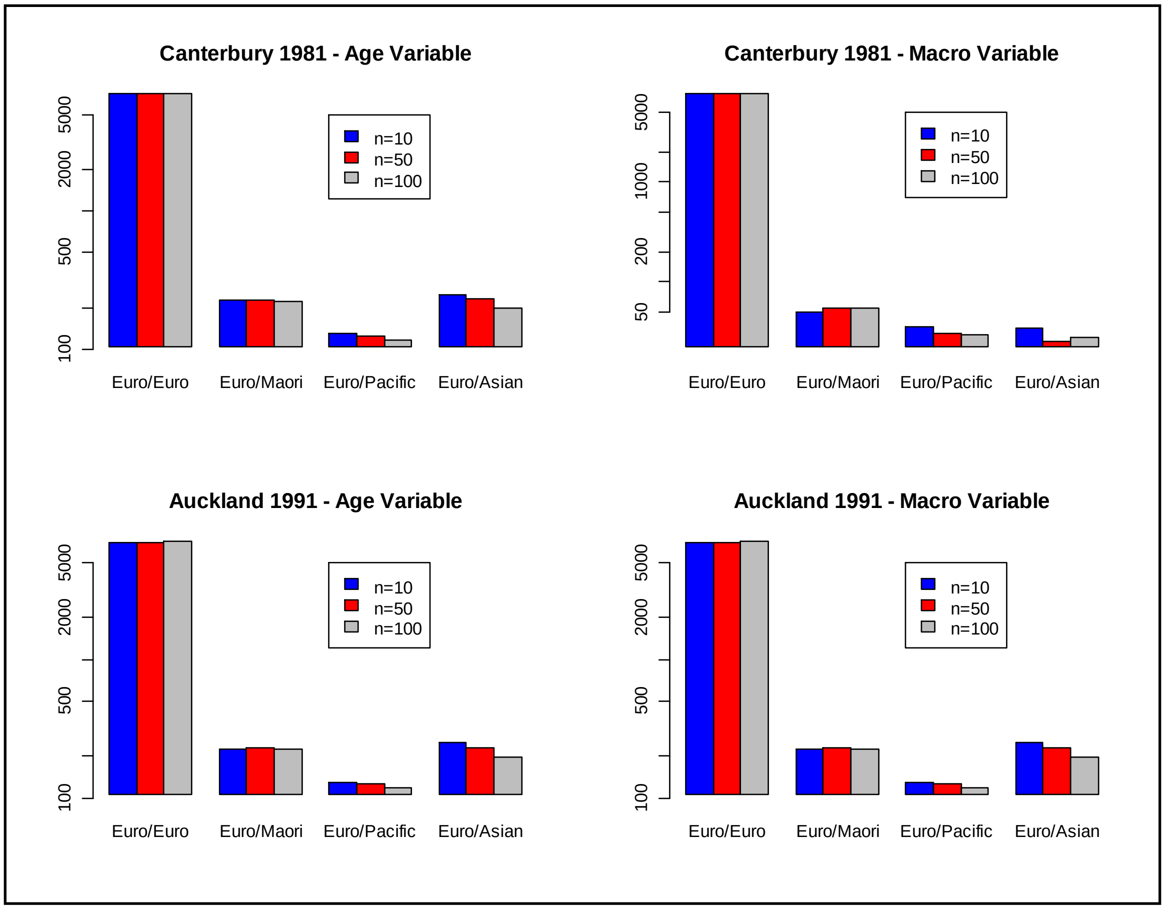

Figure 7.6 - Sensitivity testing: Social network size

Figure 7.6 shows a selection of simulated frequencies for the sensitivity analysis of the size of the social network. It compares starting social network sizes of 10 people, 50 people and 100 people. A social network with 500 people was attempted, but was too computationally expensive to complete. There are some slight reductions in the number of European/Asian and European/Pacific couples as the size of the social network increased, but otherwise the frequencies were very similar. In Chapter 6, it was shown that a smaller network size (neighbourhood) resulted in satisficing behaviour and less homogamy. However, for the empirical simulation, the changes in network size did not result in any major changes to the patterns of ethnic partnering. The key difference is that in Chapter 6 the decision variable ("education") was also the outcome variable, whereas for this simulation, ethnicity is the outcome variable, but the scoring variables are based on age, education, macro-level patterns and a random factor. A strong correlation between any of these variables and ethnicity may have resulted in changes to the ethnic patterns, so the lack of significant changes confirms what was previously seen in Section 5.3, with the logistic regression results, showing little correlation between the probability of homogamy, and age and education.

Figure 7.7 shows the effect of the number of time steps used in the simulation. When the age variable was used for scoring, there was a slight decrease in the frequencies for the mixed ethnicity partnerships. This decrease was more significant when the macro scoring variable was used, particularly for Canterbury. The macro variable adapts itself over time; as the number of time steps increases, so does the strength of the variable. The variation has also come about because the social network of each male grows at each time period. This means that as the number of time periods increases, the single males have increasingly large pools of women to choose from if the size of the social network growth per time step remains constant. The major constraint with increasing the number of time steps for the remainder of the simulations is the additional time the simulations would take to run. Since there are five years between each census, using five time steps equates to having an annual process. Five time steps provide sufficient opportunity for the macro variable to recalculate. Increasing the number of time steps to ten for the sensitivity analysis slowed the programme down to the extent that it was not practical to use it for the later (and larger) Auckland data sets, particularly for the evolutionary algorithm (which sees the simulation run multiple times).

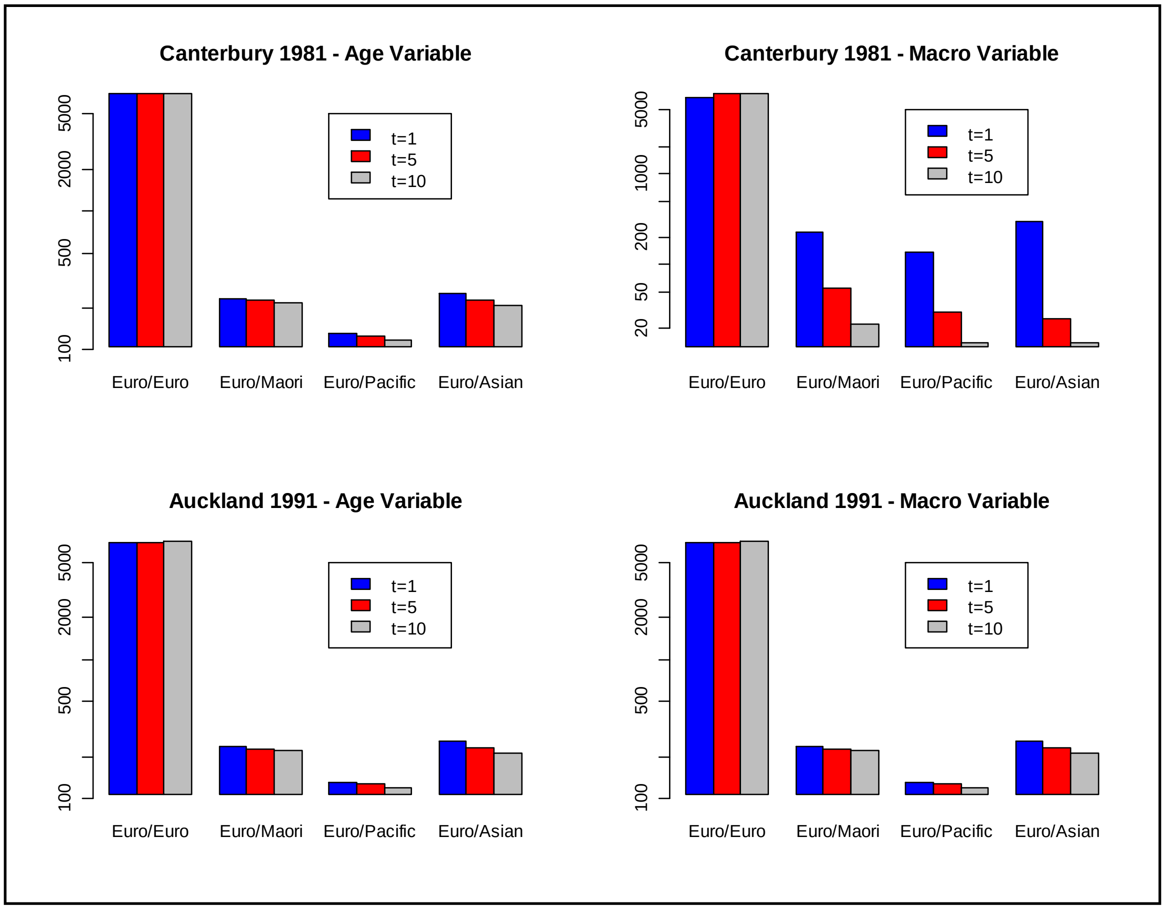

Figure 7.7 - Sensitivity testing: Number of time steps

The simulation algorithm requires the male agents to form social networks of the female agents. To ensure that this does not create different patterns from having the females select the males, the simulation was run with the male and female roles reversed. Figure 7.8 shows a sample of the frequencies from running the simulation with the gender roles reversed. The top row shows different combinations of ethnicity and the selection mechanism for the 1981 Auckland data and the 2001 Auckland data, with the male-led frequencies in blue and the female ones in pink. There was little change in the frequencies and patterns of ethnicity between the male-led and female-led simulations. The plots shown are representative of the effect of reversing the gender roles in the simulation in the other regions and for the other census periods.

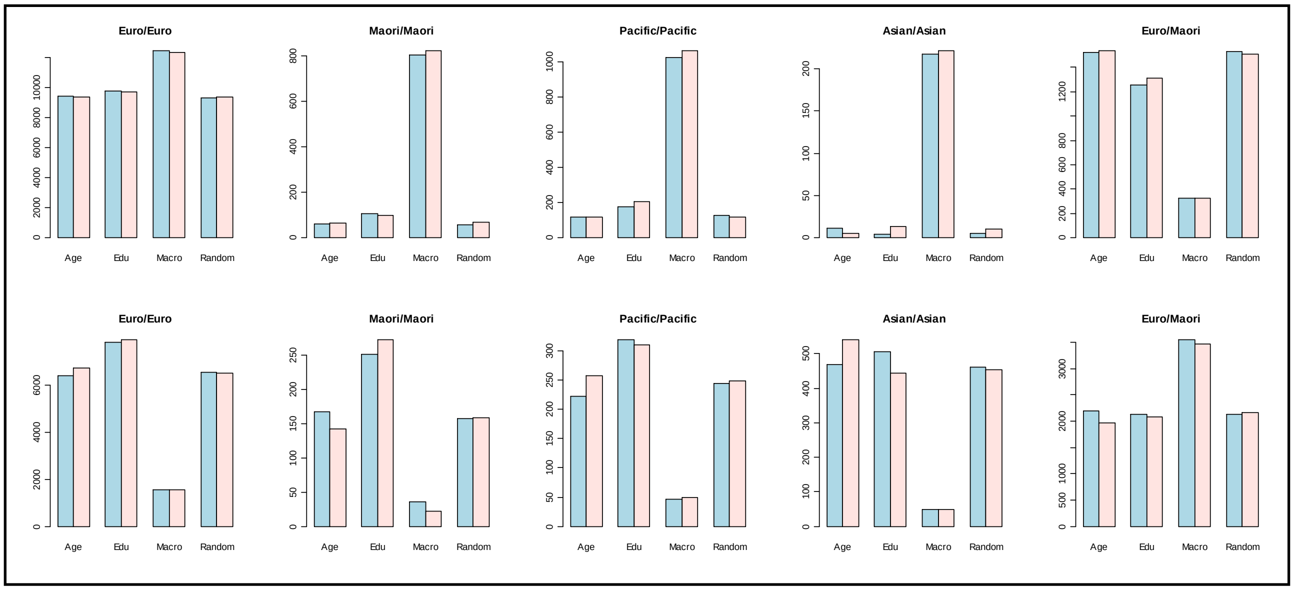

Figure 7.8 - Male versus female simulated frequencies: Auckland 1981 & 2001

7.4.2. Single Parameter Weight Results

After examining the effects of varying the internal model parameters, the next step was to run the simulation using each of the scoring variables individually, i.e. setting one of the variable weights to 100%. Repeated trials of each simulation were run to ensure that the results were stable. However, due to the computational time required per trial, too few trials were run for it to be worthwhile creating standard deviations and confidence intervals.

The results are presented graphically, and compare the actual frequencies (the net changes in the partnership tables) with the four simulated frequencies for the main combinations of ethnicities. Due to small frequencies, and limits on space, the mixed ethnicity results are presented together and are not analysed separately by gender. Some discussion is provided with the graphs, although a more integrated discussion, considering the single parameter and evolutionary results, is presented in Section 7.5.

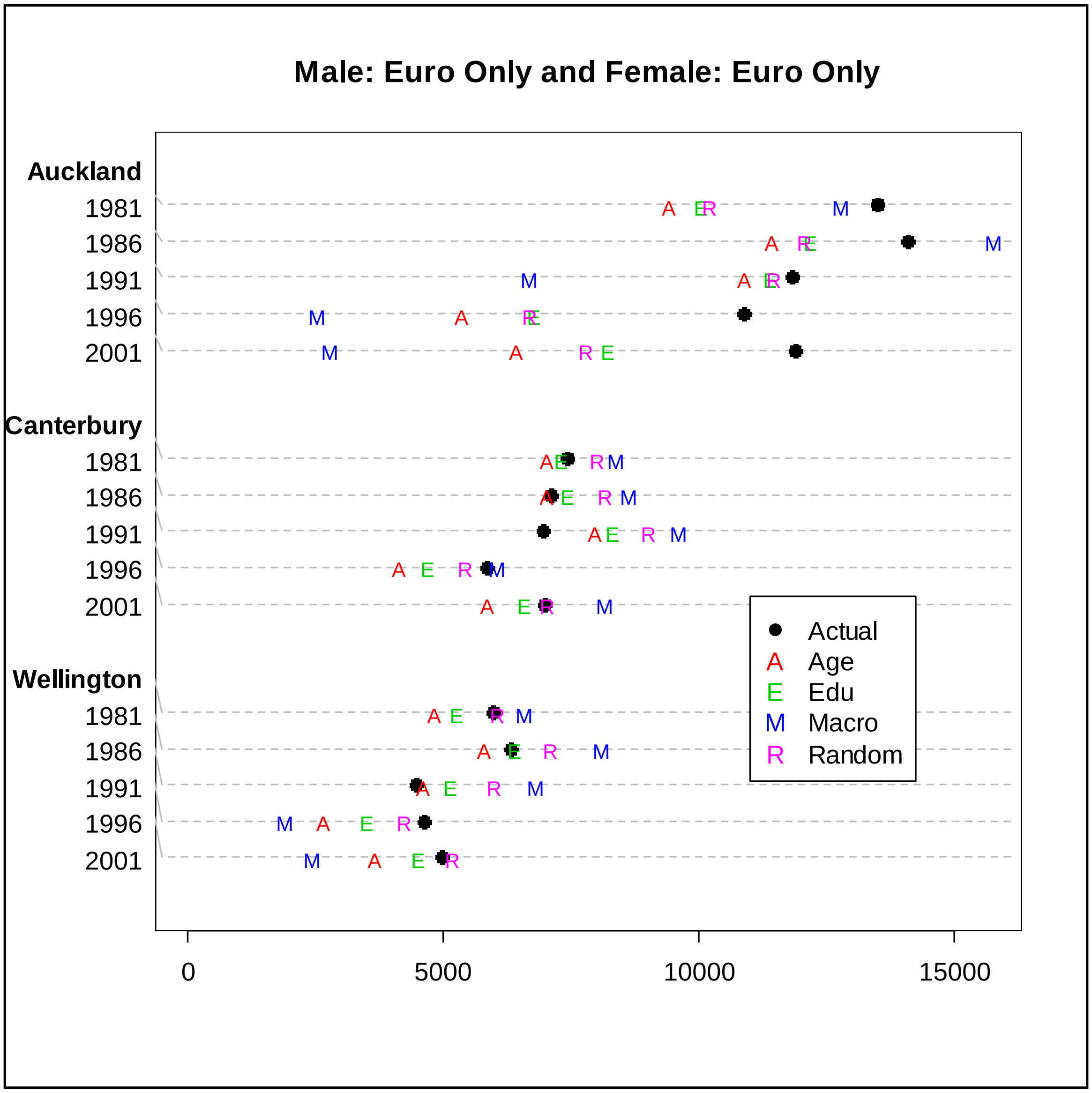

Figure 7.9 - European male and European female estimates using a single variable

Figure 7.9 shows the actual and simulated numbers of homogamous European couples for each of the time periods, in each of the regions. It shows that the number of European/European couples in Auckland was underestimated, particularly in the later census periods where all of the scoring variables produced very poor estimates. By comparison, Canterbury had the most consistent and accurate simulated frequencies for all of the scoring variables. The macro variable was the least consistent of the individual scoring variables for Auckland and Wellington. In the earlier periods it produced higher estimates than the other scoring variables (over-estimates in the case of Wellington), but by 1991 in Auckland, and 1996 in Wellington, it was producing the biggest under-estimates.

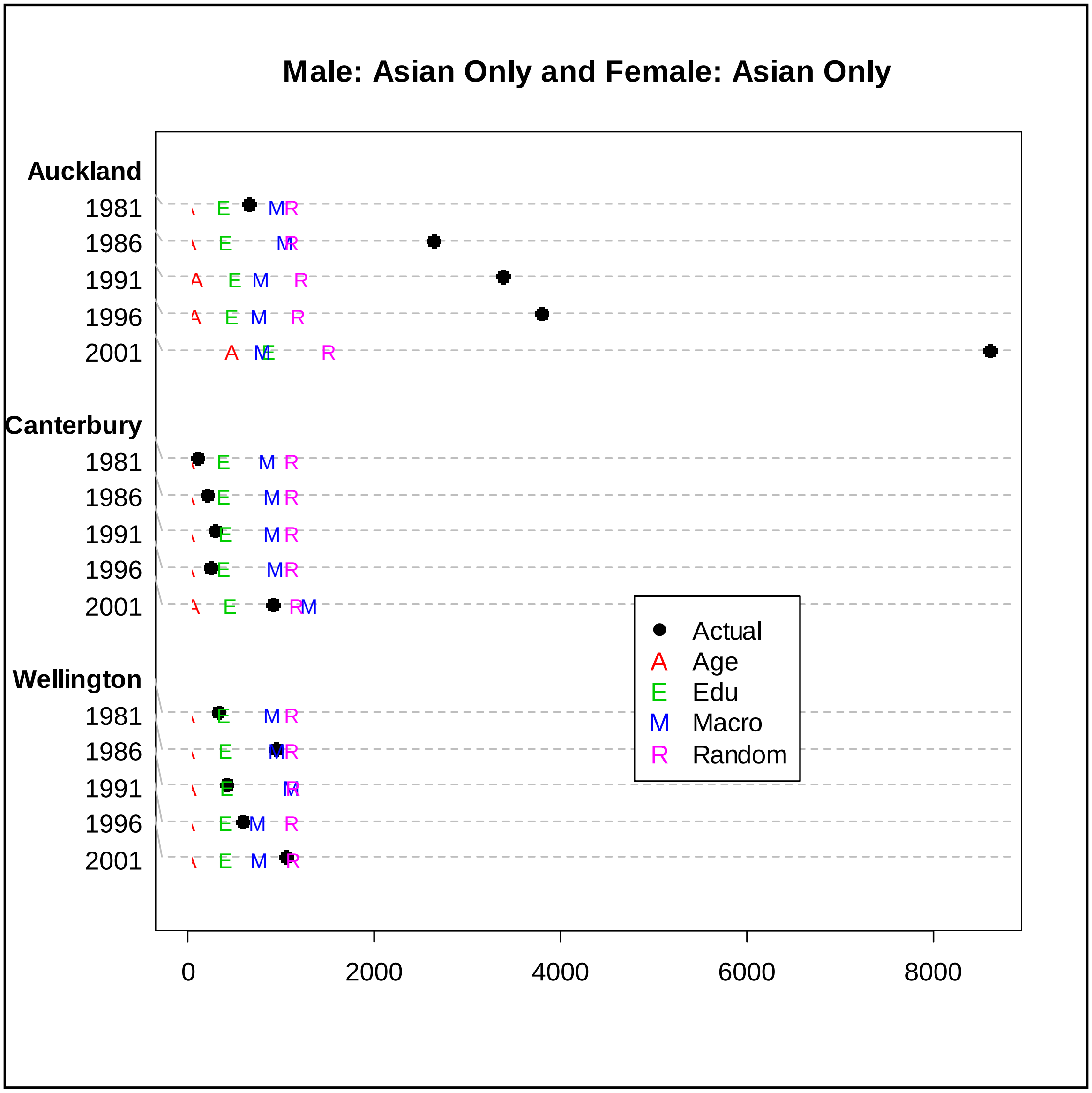

Figure 7.10 - Asian male and Asian female estimates using a single variable

The number of homogamous Asian Only couples estimated via the simulations was quite consistent across all of the regions and census periods. Unfortunately, this led to over-estimation of the frequencies in Canterbury, and progressively worse under-estimation in Auckland. Figure 7.10 shows that the age variable consistently produced the lowest estimates, followed by education, the macro variable, and the random variable.

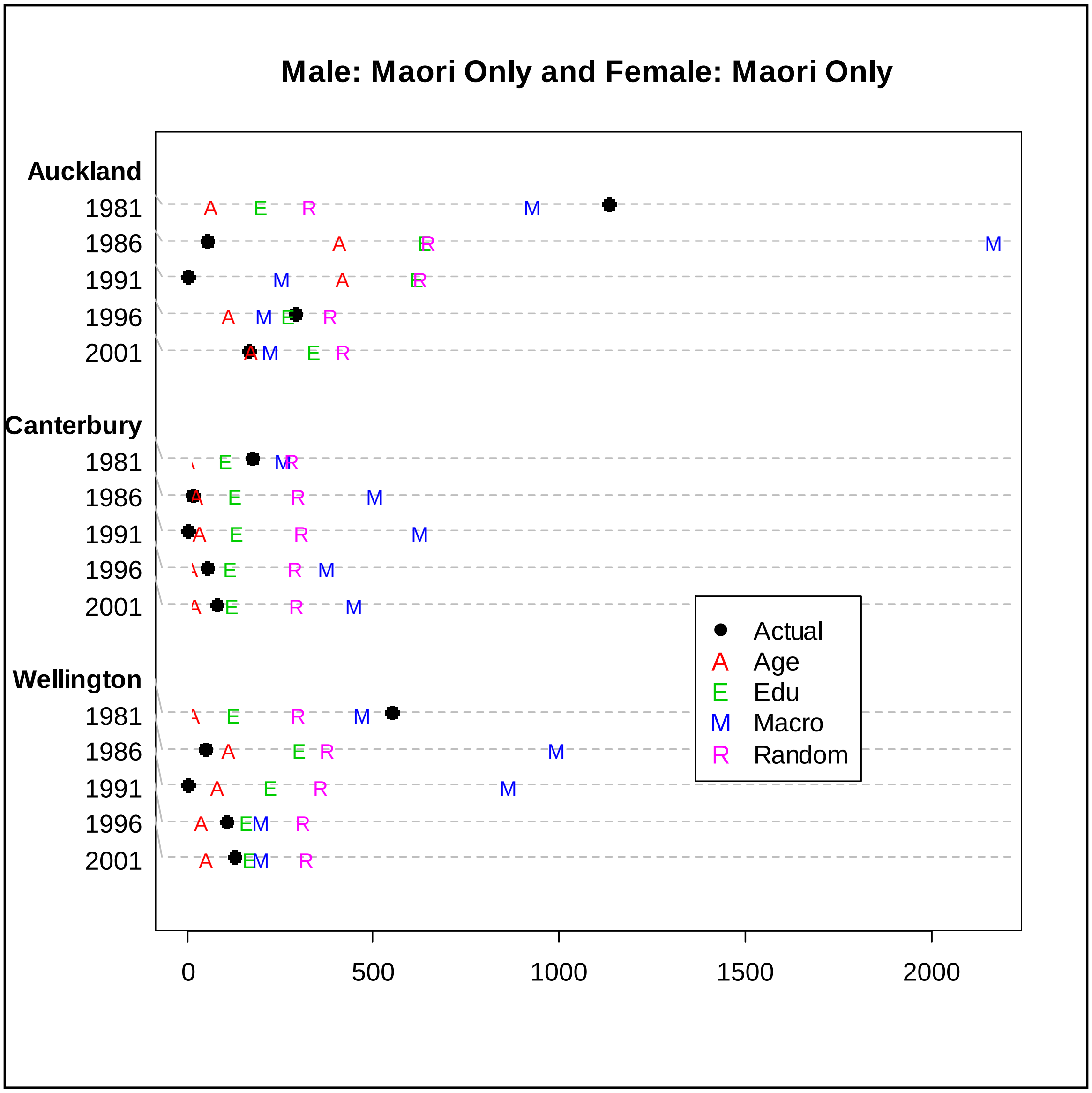

Figure 7.11 - Maori male and Maori female estimates using a single variable

Figure 7.11 shows the frequencies for the homogamous Maori partnerships. As previously mentioned, one difficulty with interpreting these figures is that the net changes between censuses for this group were very small. The macro variable produced the highest estimates for all three regions, and had the highest number of over-estimations, including an outlier for Auckland 1986. The age variable produced the lowest estimates, followed by the education variable.

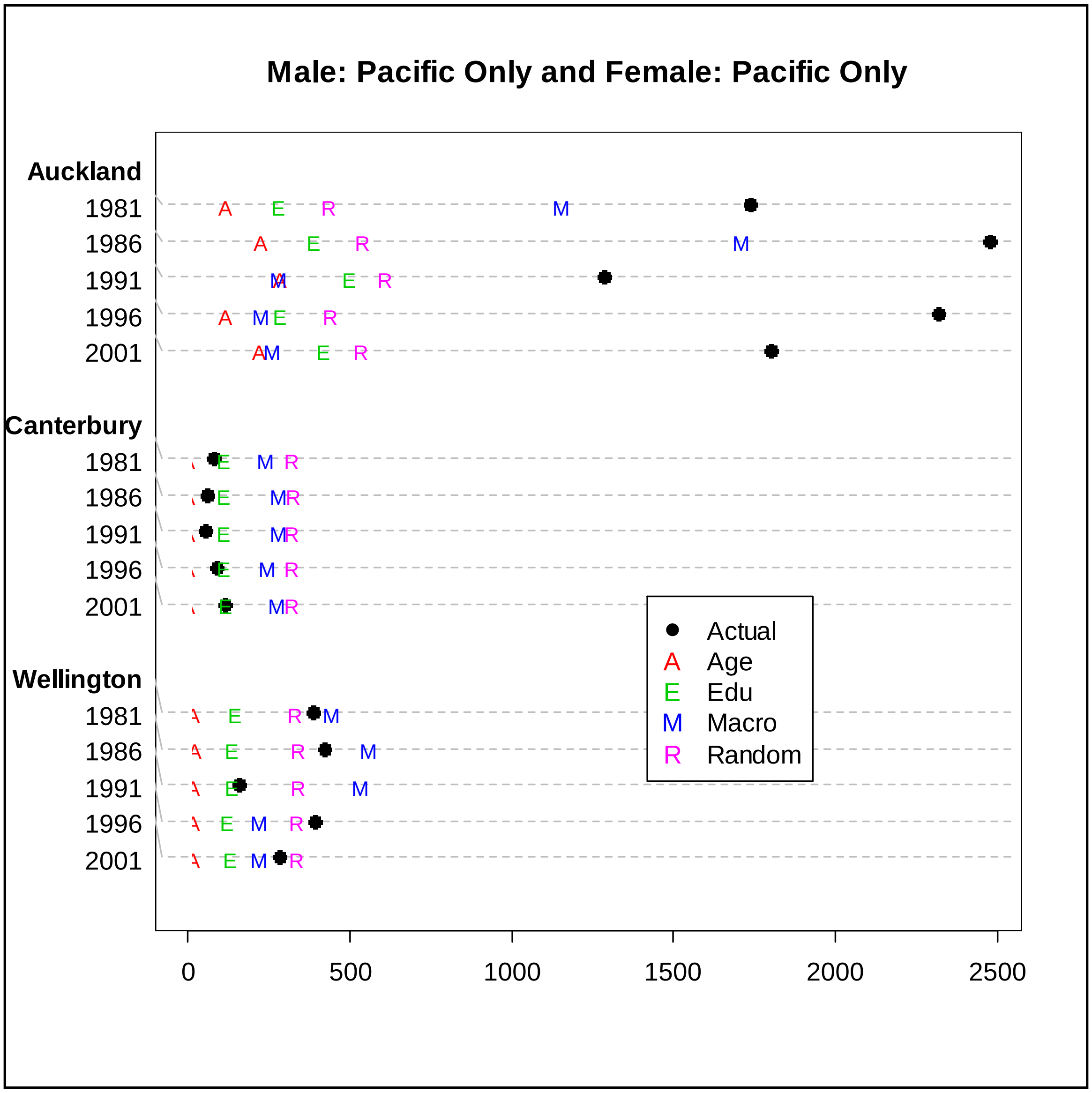

Figure 7.12 - Pacific male and Pacific female estimates using a single variable

As with the Asian/Asian partnerships, the Pacific/Pacific partnerships are significantly underestimated in Auckland. This is most noticeable in Auckland where the simulated frequencies were similar to the other regions, but much lower than the actual values. The actual frequencies in Canterbury are very low, and over-estimated by the macro and random methods, but very closely estimated with the education variable. The macro and random scoring variables produced the closest estimates in Wellington for all of the periods other than 1991.

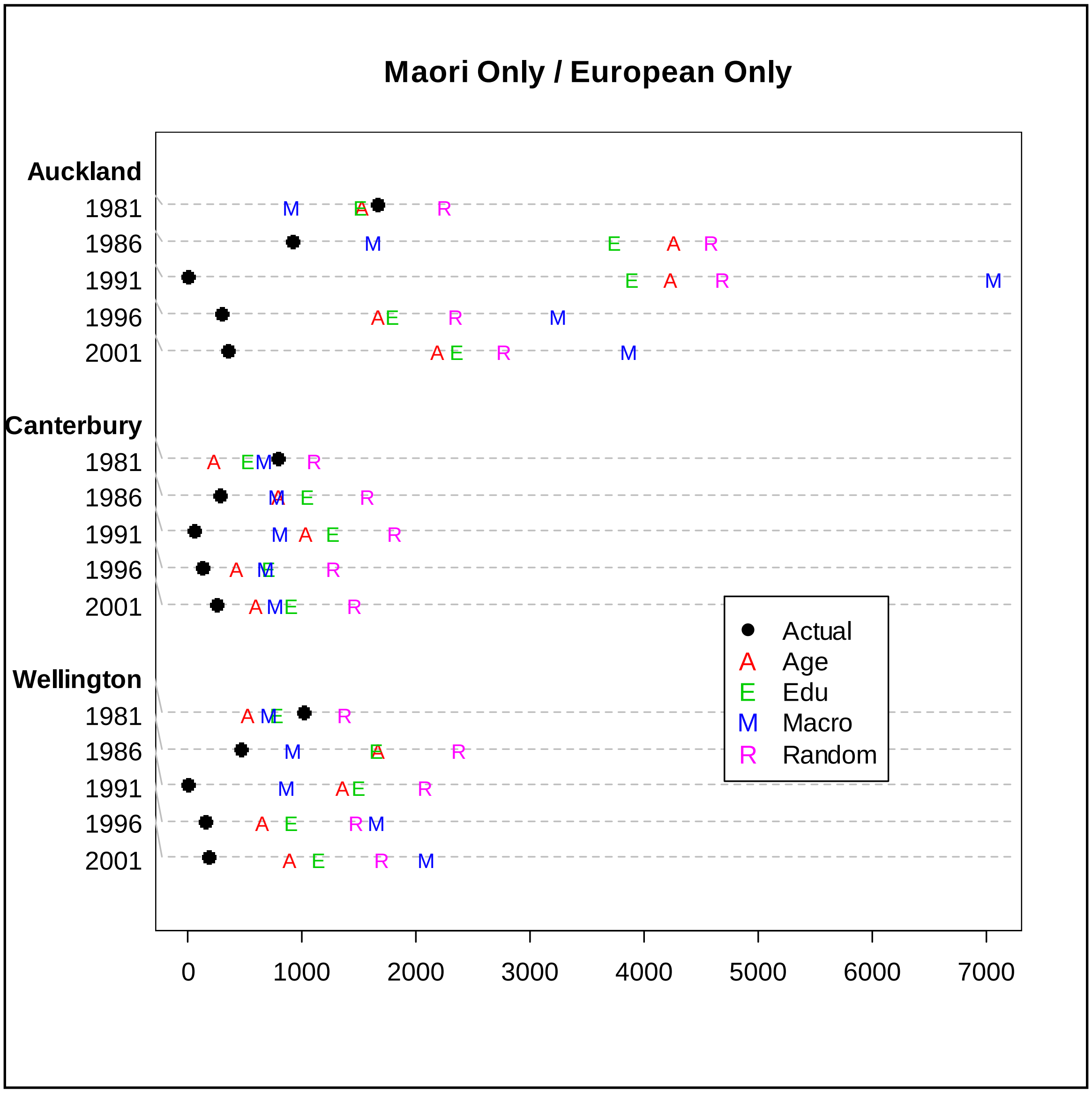

Figure 7.13 - Maori/European mixed partnership estimate using a single variable

Figure 7.13 shows that the number of mixed European/Maori partnerships is over-estimated by most of the scoring variables in most of the periods. The random variable and macro variable produced the largest over-estimates in all three regions.

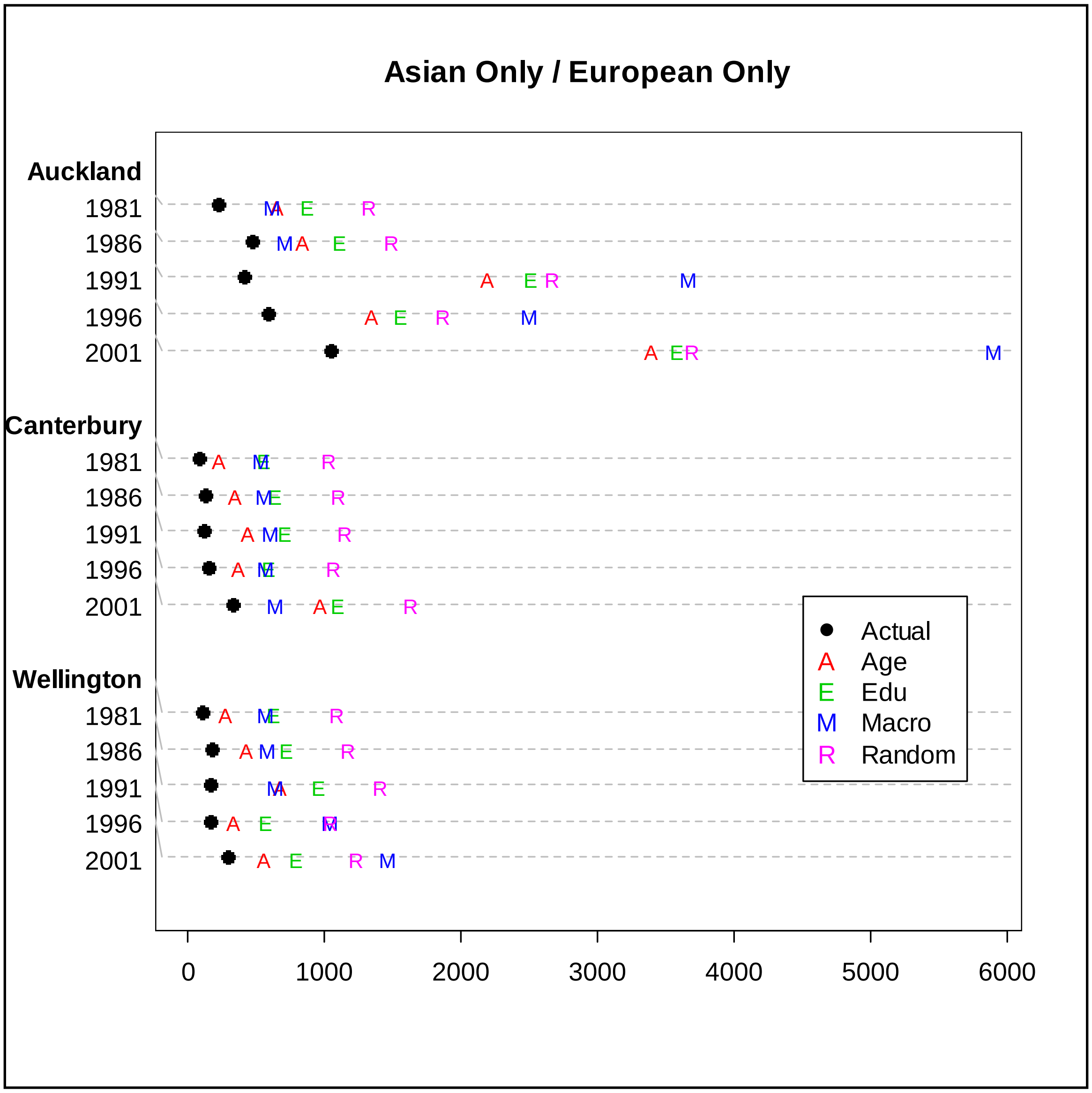

Figure 7.14 - Asian/European mixed partnership estimate using a single variable

The Figure 7.14 shows that the number of Asian/European mixed partnerships was consistently over-estimated by the simulation. The random scoring variable had the highest over-estimates in most of the census periods.

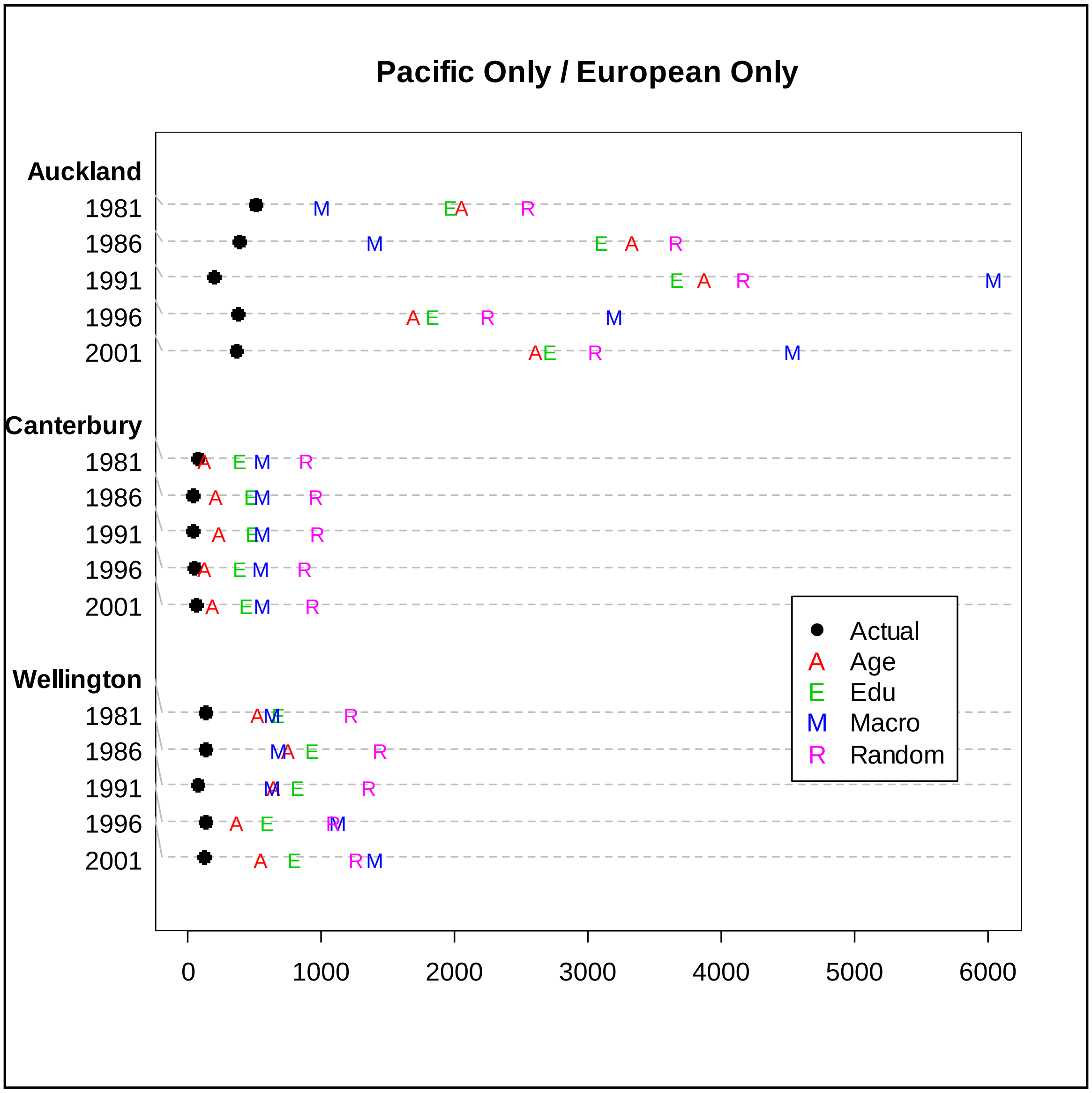

Figure 7.15 - Pacific/European mixed partnership estimate using a single variable

Figure 7.15 shows that the numbers of Pacific/European mixed partnerships are heavily over-estimated by all of the methods, particularly in Auckland. As with the previous mixed ethnicity groups, all of the scoring methods over-estimated the number of couples. The random scoring variable created the largest over-estimates in Canterbury, Wellington, and in the earlier census periods in Auckland. For the later census periods in Auckland, the macro scoring variable creates the largest over-estimates.

In summary, using any of the variables individually tended to produce a mixture of under and over-estimates for the different partnering combinations. This was most evident with the Auckland estimates, where the number of homogamous partnerships for the various ethnic groups were generally under-estimated and the number of non-homogamous partnerships were over-estimated. The age and education variables tended to produce the lowest estimates for the number of couples in each group. There was more variation in the macro variable results, mainly due to the way that the variable worked. Since it would update at each time step, it would create inertia, in the form of social pressure, towards a homogamous or a heterogamous partnership, depending on what had occurred in the region in the previous time step. This sometimes produced high over-estimates and sometimes produced low under-estimates.

The next step is to examine what happens when weighted combinations of the variables are used together. This is done via the evolutionary algorithm described in Section 7.3.5, and aims to find the combination of weights that will minimise the total squared error of the simulated ethnicity tables relative to the actual tables from the following census.

7.4.3. Evolutionary Algorithm Weight Results

The evolutionary algorithm was run for each region and time period. Various starting points were used to check for convergence, although due to space limitations, only one graph is shown for most region/time combinations. All of the regions displayed consistent patterns of convergence of the weights (at a qualitative level). The Auckland 1981 results are shown as an example of how the algorithm would converge from different starting sets of weights. As previously mentioned, it is important to note that due to variation in the process being modelled, the weights and patterns should be considered as more of a qualitative result than a quantitative one. The results are presented with graphs showing the weights for each of the four components (age difference, education difference, macro couples measurement, and the random factor) for each of the iterations (0-12) of the evolutionary algorithm. The variables are represented in the graphs by dark blue, cyan, light blue and beige respectively. The frequencies generated from the evolutionary algorithm are not displayed, for two reasons. Firstly, a bug in the grid machine meant that for many of the runs of the simulation, the final frequencies were not recorded. Secondly, the focus of the evolutionary algorithm results is the relative weights of the four scoring variables, rather than the frequencies that were actually generated.

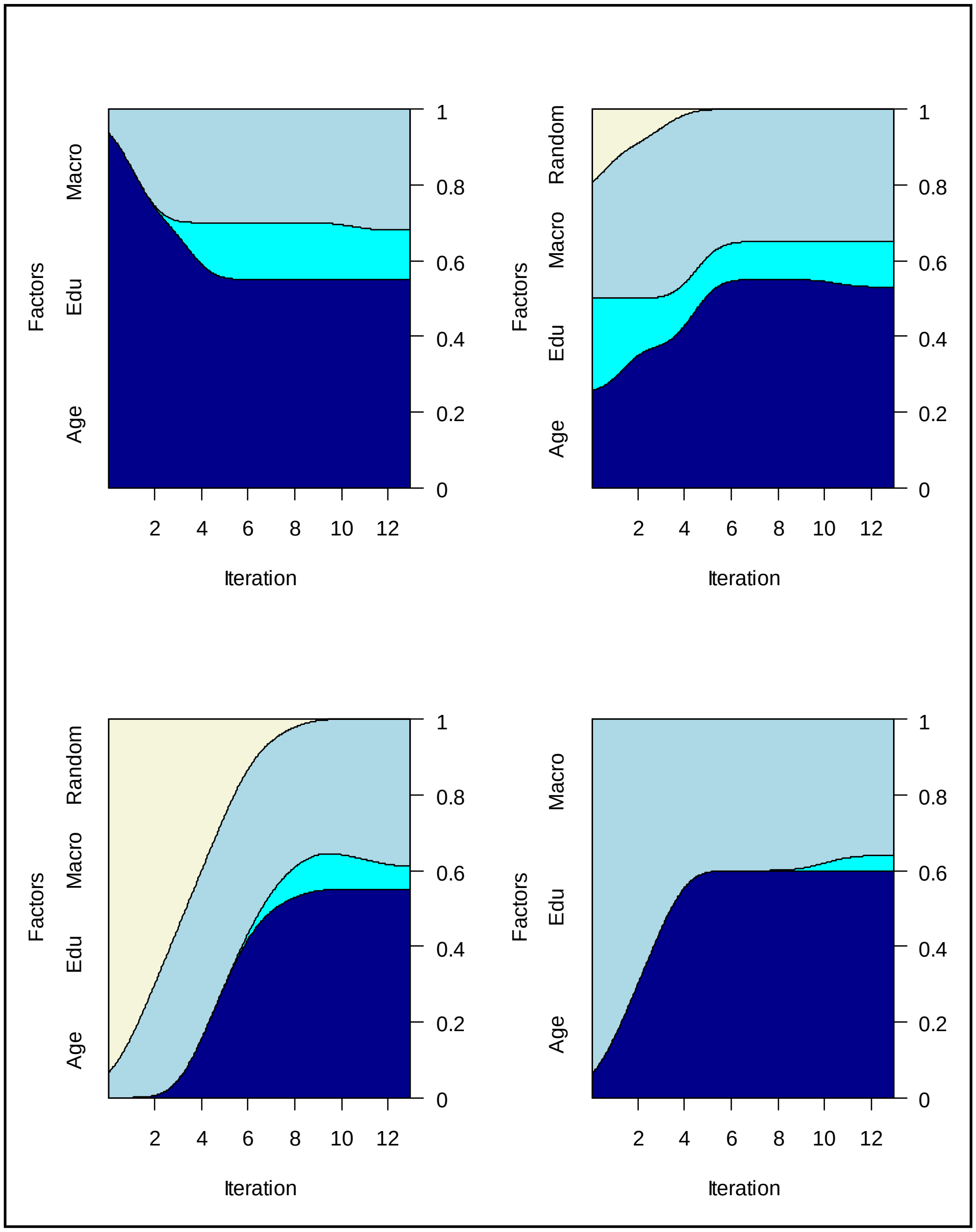

Figure 7.16 - Auckland weights: 1981

Figure 7.16 shows four trials of the evolutionary algorithm for the Auckland 1981 data, each using a different set of starting weights. The graphs show that the weights converged to a similar set of values, irrespective of where they started from. Each graph shows a final set of weights of between a 55-60% for age, a 5-10% for education, 40% weighting for the macro variable, and no weight for the random factor.

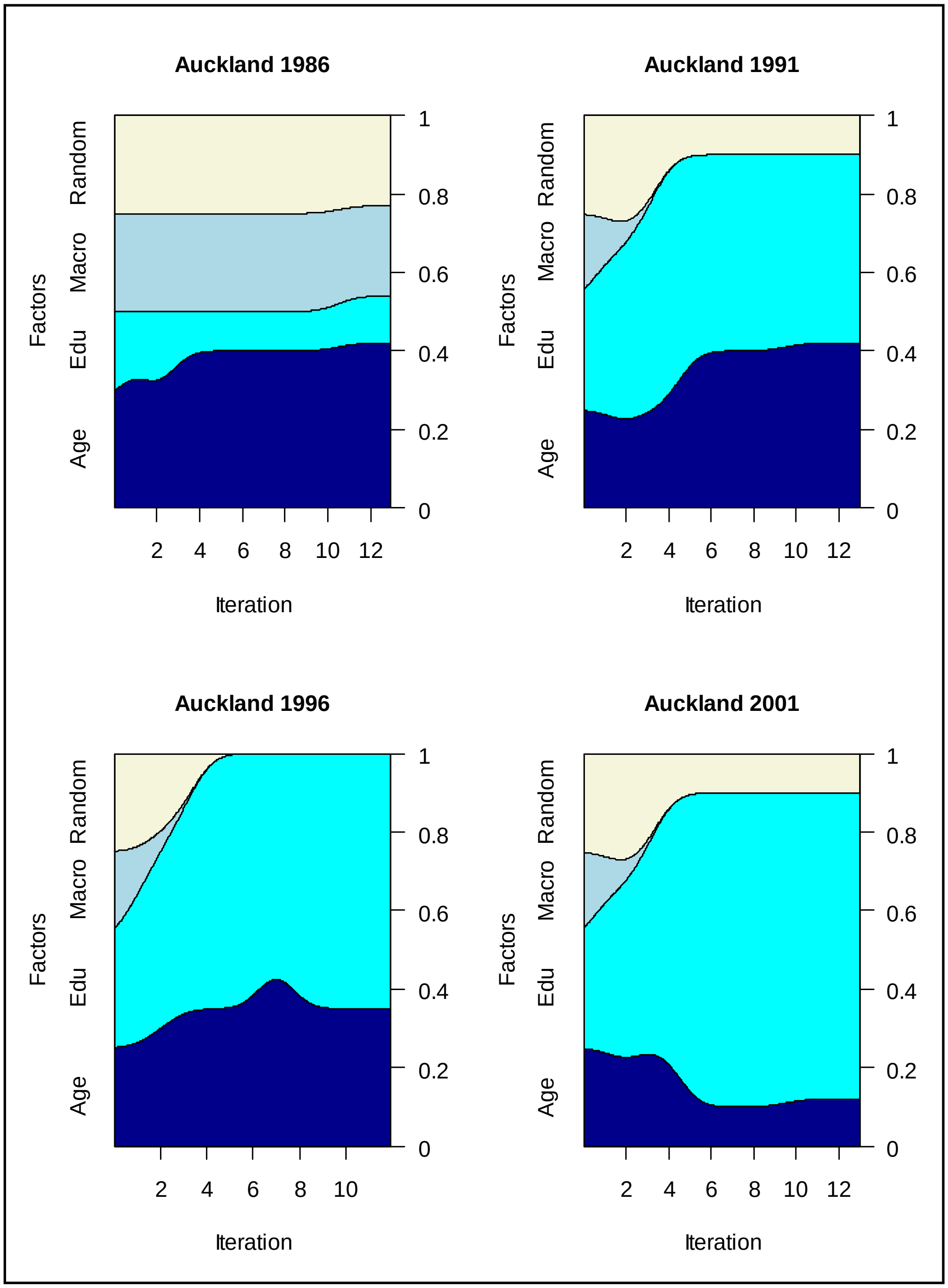

Figure 7.17 - Auckland weights: 1986-2001

Figure 7.17 shows the weights for Auckland in each of the other census periods. As with the 1981 data, education still plays a small role in 1986, and the random factor is completely absent, even when the weights are initialised with 100% of the weighting on the random factor. From 1991 onwards there is a much greater weight on the education variable. There is some weight on the age variable, and the random variable, and no contribution from the macro variable. The results in 1981 are different from the other periods, although this is because the education variable in 1981 had only two possible categories so was not distinguishing between possible partners as effectively as it was in later years (when there were four groups).

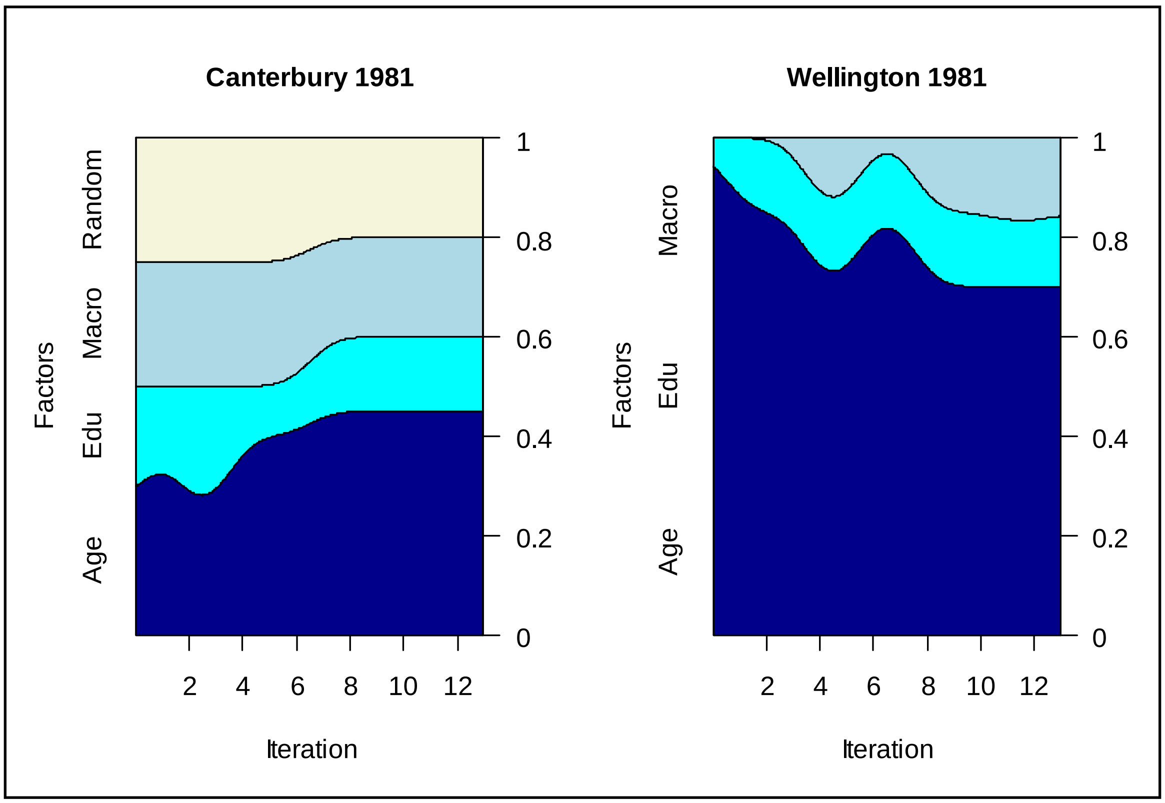

Figure 7.18 - Canterbury and Wellington weights: 1981

Figure 7.18 shows the weights for Canterbury and Wellington using the 1981 census data. They are discussed with the remainder of the weights for their respective regions, but could not be fit into the graphing window. Each follows a slightly different pattern from the 1981 Auckland results. Canterbury has a higher weight on the random variable than Auckland and Wellington, whilst Wellington has a much higher weight on the age variable than the other two regions.

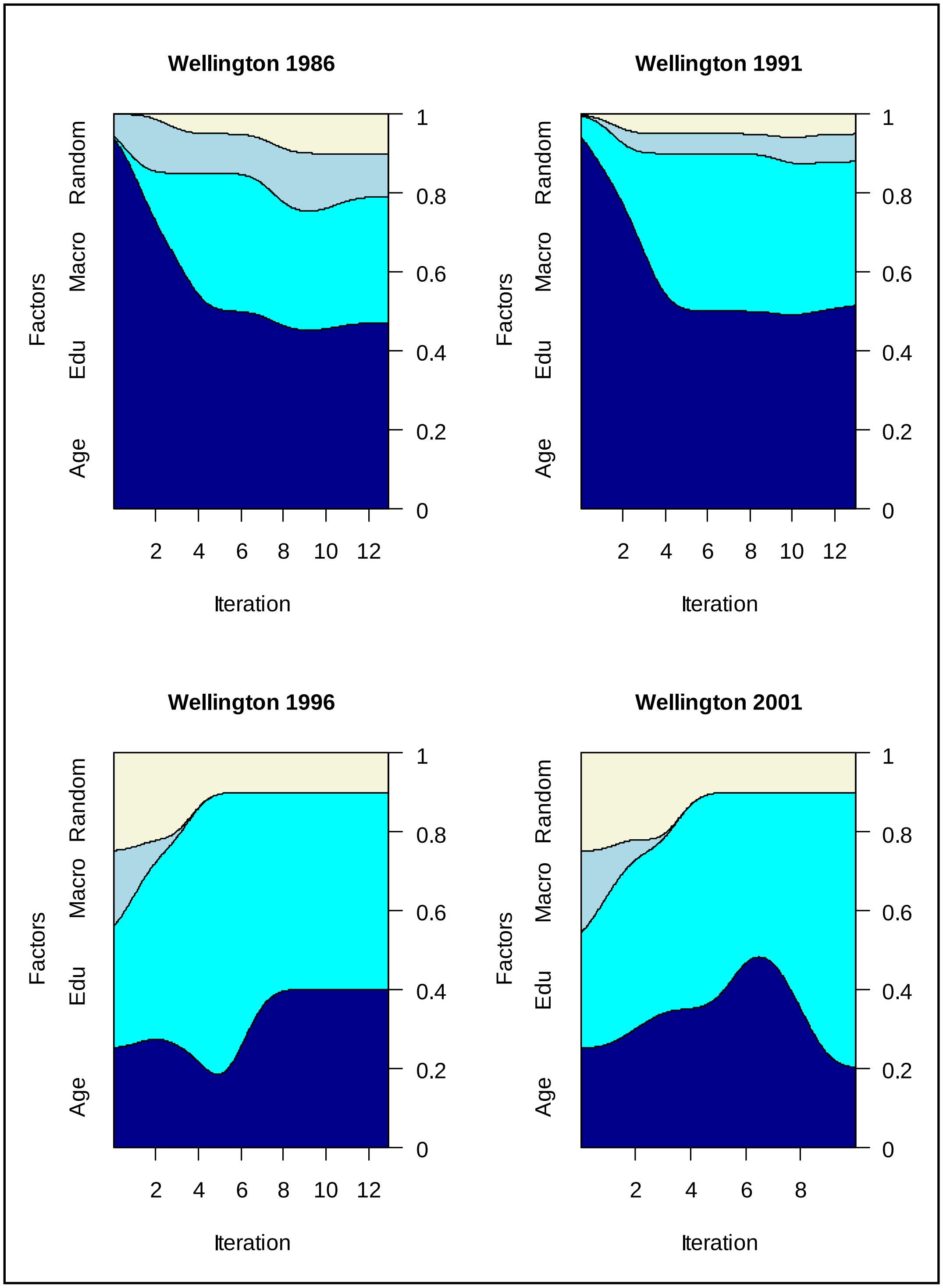

Figure 7.19 - Wellington weights: 1986-2001

Figure 7.19 shows the weights for the Wellington region for 1986 to 2001. The weights are dominated by age and education variables. In the later census periods the age variable became less dominant, with more weight on the education variable. The random variable also retained a weight of 5-10% for each of the census periods other than 1981.

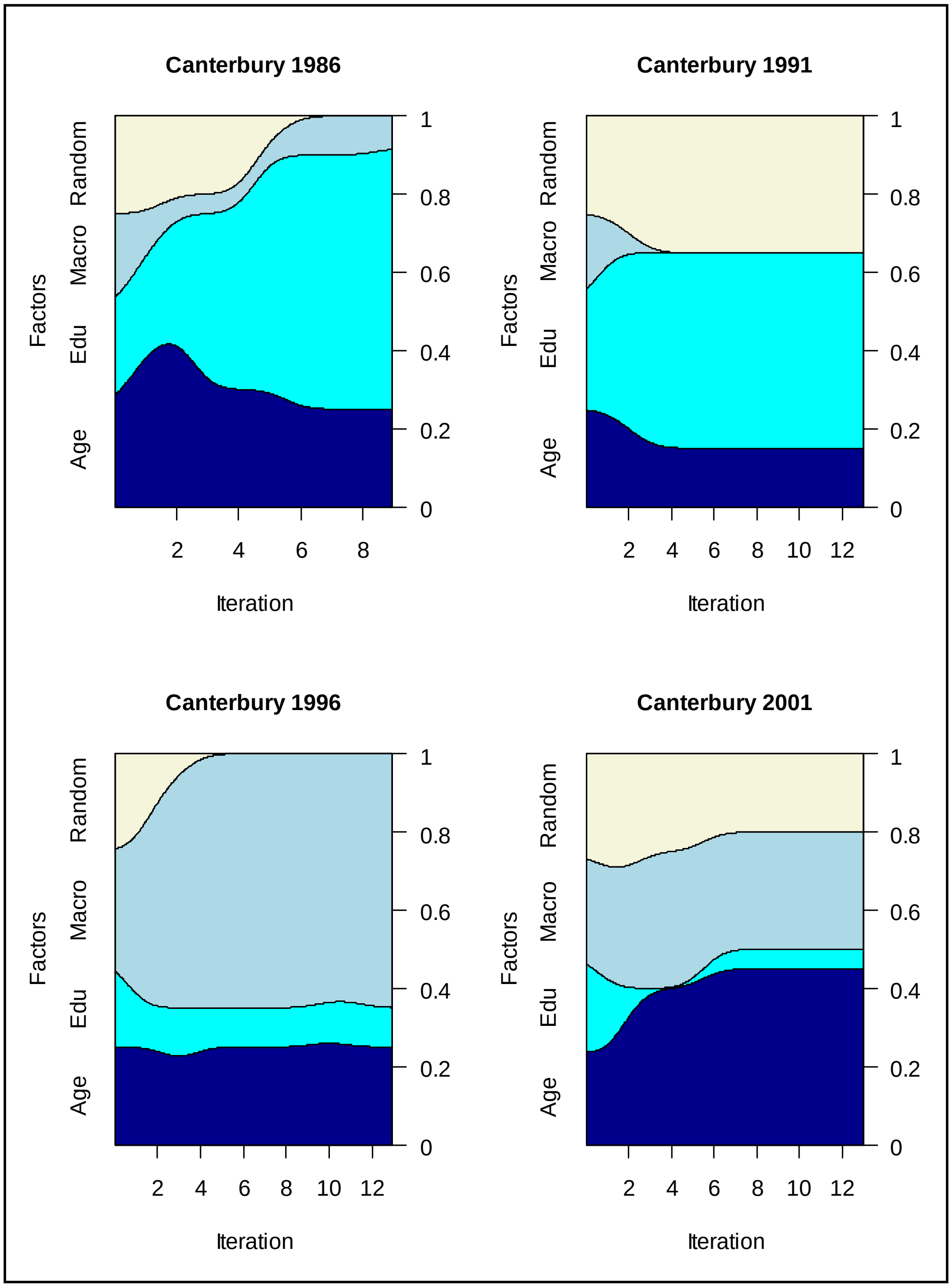

Figure 7.20 - Canterbury weights: 1986-2001

Figure 7.20 shows the weights for the Canterbury region for 1986 to 2001. The Canterbury patterns showed more variation than Auckland and Wellington. The education variable featured prominently in 1986 and 1991, but not in 1996 or 2001. The random variable featured in 1981, 1991 and 2001, suggesting that the other variables were not capturing all of the variation in the matching patterns in these periods. Despite the differences in patterns from one census period to the next, multiple trials of the same census period still resulted in consistent patterns of convergence.

7.5. Discussion of Results

The simulation results provide an alternative way of examining the patterns of ethnicity from the statistical analysis in Chapter 5. Instead of examining the patterns of ethnicity for existing couples using the historical data, social simulation was used to match the single people, and then examine the patterns that emerged from the matching process. The examination of these patterns was broken down into three goals:

1. To examine the effect that each of the scoring variables (age, education, macro, random) had individually on partnership patterns.
2. To find the weighted combination of the scoring variables that produced the most similar set of inter-ethnic cohabitation patterns to those that have actually occurred.
3. To investigate the possibility of a micro-macro relationship within ethnic partnering patterns.

Reversing the roles of the males and the females in the simulation resulted in little change to the simulated patterns of ethnicity. This indicates that for this simulation at least, there was no gender asymmetry in the way that the agents selected mates. Changing the initial size of the social network from which the male agents chose their partners also appeared to make little difference to the ethnic patterns. Although one would expect that if the social network size got too small, the small ethnic groups would have fewer homogamous partnerships by virtue of not being in one another's social networks. Altering the number of time steps had some effect when the macro scoring variable was being used. The macro variable was the most sensitive to changes in the number of time steps because it was an iterative variable. It would change at each time step, based on what had happened at the previous time step. Therefore, as the number of time steps was reduced, the macro variable had less impact on the simulation. This also provides some evidence of the existence of a micro-macro link, since altering the number of time steps altered the behaviour of the variable. However, only limited changes could be made to the size of the social network and the number of time steps, as they were both constrained by computational resources. Increasing either of these variables by too much resulted in unreasonably large processing times.

When the scoring variables (age, education, macro, random) were run individually, there was a strong tendency for the number of same ethnicity partnerships to be under-estimated and the number of mixed ethnicity to be over-estimated. This suggests that there is some level of ethnic preference beyond what was captured by the variables of the simulation model, and reinforces the homogamy shown by the quasi-independence models in Chapter 5. The age and education variables produced more consistent estimates than the random and macro variables. The macro variable produced the most extreme results. In earlier census periods it would produce very large estimates, but in the later periods, particularly in Auckland, it tended to produce much lower ones in the main ethnic combinations, but more pairings in the smaller groups. This indicated that it was successfully mimicking the social trends by shifting from a same ethnicity social pressure to a mixed ethnicity pressure as the regions became more diverse and mixed ethnicity partnerships became more prevalent.

The evolutionary algorithm displayed consistent convergence within the regions and census periods. There was not a single weighted combination of the scoring variables that best reproduced the empirical target, as the patterns in the weights were slightly different for each of the three regions. However, there were similarities between the results. The age and education variables consistently had the highest weights of the four factors, confirming the previous literature (Logan et al., 2008) showing strong correlations between age and education on the one hand and partnership matching patterns on the other. The effect of the education variable became stronger after 1981 when it shifted from having two levels to four levels. The strong presence of age and education would suggest that the correlation between ethnic patterns and these variables in the matching process is stronger than that indicated by the logistic regression in Section 5.3.

The macro variable did not feature as prominently as age or education in most of the evolutionary algorithm results. However, in the earlier Auckland and Wellington periods, and throughout the census periods in Canterbury, it did contribute some weight to the optimal solutions. Although this would suggest that the micro factors, age and education, are the more dominant factors in partnership matching, it also indicated that the macro variable contributed additional value to the simulation, and that further investigation of a micro-macro link would be a worthwhile endeavour. Further investigation of other macro variables could be conducted in the future if more detailed data were to become available. The "random" factor featured the least, particularly in Auckland. It contributed the lowest weight – often zero – to the simulations, and the plots of the weights showed how the algorithm would shift weight from it to the other variables. However, some periods in the Wellington and Canterbury regions retained some random component. This would suggest that the other three variables were not doing a sufficiently good job of simulating the partnership matching for those periods. The random and macro variables were not used in the scoring function of the DYNASIM/APPSIM models on which this simulation was based. Their presence in the results of the evolutionary algorithm indicates that they provide predictive power and produce more accurate results than the age and education variables alone.

Although the simulation results cannot be considered conclusive results, they do add weight to conclusions found by conventional statistical methods and other studies in the literature.

7.6. Future Simulation Possibilities

The simulation models could be extended in the future through improvements to the input data and the simulation process itself. Although one of the strengths of the simulation was that it was simulating the matching of "real" people using the unit-level census data, the security concerns surrounding the data limited the detail of the information that was available. One of the ways that this could be overcome in future research is to look at how other sources of quantitative and qualitative data could be used to improve the amount of information about the simulated agents. More detailed geographic data, such as the location of workplace and residence, at a suburb, rather than city level, would enable more realistic social networks and interactions between the agents.

In addition to adding more detailed data, the other avenue for advancing the simulation models in the future would relate to the simulation itself. Complexity could be added to the way in which the social networks were created and appended. Qualitative information about individual mate preferences would also help to create a more detailed scoring function and matching algorithm. Rather than random networks, structure could be created through individual and geographic information if such information was available. Variations on the algorithm and the scoring system could also be experimented with.

Chapter 8 - Conclusion

This study set out to examine patterns of ethnic partnerships in New Zealand from 1981 to 2006 using census data. It applied statistical and social simulation methodologies to the census data to answer the following research questions:

1. What changes can be seen in inter-ethnic cohabitation patterns in the period 1981 to 2006?
2. What factors and/or social processes influence patterns of ethnic partnership formation?

This research is unique in a number of ways, through its content and methodology. Although some research has previously been conducted on New Zealand marriage/partnership data, it has tended to focus on a single point in time, and rely on descriptive statistics (see Section 2.1.3: Research in New Zealand). This research takes data from six full census data sets, and extends the analysis from simple proportions to a variety of log-linear models, in order to better describe the patterns observed in the data.

It also demonstrates the first social simulation model of the New Zealand marriage market. The simulation is populated with unit-level census data, and then has an evolutionary algorithm applied to it via the Auckland BeSTGRID cluster. The patterns of ethnic partnership are examined during the matching process and are compared and contrasted with the previous statistical models.

8.1. Statistical Analyses

The initial investigation of the data, in Section 5.1, examined the proportion of homogamous couples. The focus is on the 18 to 30 age group in order to establish patterns of emergence rather than prevalence. This analysis was extended in Section 5.2, with log-linear models used to re-examine the patterns independently of the relative sizes of the ethnic groups. The models included the quasi-independence model, which examined the “diagonal dominance” of the frequency tables, the crossing parameter model, which extended the quasi-independence model by looking at partial matches of ethnicity, and the quasi-symmetry model, which examined the relationship of the off-diagonal cells. A logistic regression model (Section 5.3) was also used to examine the effect of age, education and location on the odds of a given couple having a homogamous partnership. The proportions and log-linear models combine to answer the first research question, by describing the patterns of ethnic partnering in New Zealand between 1981 and 2006. The logistic regression takes the first steps towards answering the second, by looking at some of the factors influencing the patterns of ethnicity.

The proportions showed that most Europeans had a European partner, whilst the smaller ethnic groups had lower proportions of homogamous partnerships. The main disadvantage of using proportions to measure homogamy is that the proportions are a function of the size of the groups they relate to. It should not be surprising that there is a high proportion of Europeans with a European partner because there are a large number of European people. The solution to this was to re-examine the tables using log-linear models, which examine the patterns and changes after accounting for the relative group sizes.

The log-linear models showed some patterns that were not obvious in the proportions. The quasi-independence model (Sections 5.2.1-5.2.4) used parameters to measure the level of homogamy for different groups by modelling the diagonal dominance of the frequency tables. The European Only group had a much lower rate of homogamy than was reflected in the proportions. With the log-linear models controlling for the size of the ethnic groups, the rate of homogamy for the European Only group was very similar to that of the Maori Only group. Both groups saw little change in the rate of homogamous partnership in the 18 to 30 year-old cohort over the 1981 to 2006 period. By contrast, the Pacific Only group saw an increasing rate of homogamy over time, with approximately twice the rate of homogamous partnerships in 2006 as there were in 1981 amongst New Zealand-born couples. The Asian Only group had the largest decrease in homogamy, moving from the most ethnically homogamous group, with a rate of homogamy that was about ten times greater than the other ethnicities in 1981, to approximately two-to-three times greater than the European Only and Maori Only groups in 2006. The dual ethnic groups had much lower rates of homogamy than the single ethnicity groups, although the quasi-independence model did not account for partial matches of ethnicity, such as a Maori Only individual with a European/Maori partner.

Crossing parameter models (Section 5.2.5) extended the quasi-independence model to measure partial, as well as full matches, of homogamy. These models showed that people were more likely to have a partial match of ethnicity than no match. For example, a European/Maori person is more likely to have a partner who is European Only or Maori Only than a completely different ethnicity. However, individuals were still more likely to have an identical ethnic match than a partial one.

The quasi-symmetry model (Section 5.2.6) examined the relationship of the off-diagonal cells of the frequency tables. It showed statistical evidence of asymmetry in the tables. Examining the frequency tables, the two most significant patterns in the off-diagonal cells are the increasing numbers of European men with an Asian partner (relative to European women with an Asian partner) and the number of Pacific women with a European partner (relative to Pacific men with a European partner). The logistic regression model showed a decreasing likelihood over time of an Auckland couple has a homogamous relationship. It showed limited power of age and education to predict the odds of a relationship being homogamous, but the low R^2 values suggested that there could be other factors – beyond those available in the census – with predictive power in this relationship.

8.2. Simulation Modelling

The simulation modelling followed on from the logistic regression in answering the second research question. It explored different partnership matching methods and scenarios, using abstract (Chapter 6) and empirical simulation (Chapter 7) to examine the patterns that were created by matching the single people in the population (who, by definition, use were eligible to form couples - rather than analysing the couples who had already formed, as in the previous statistical analysis).

The initial abstract simulation in Chapter 6 was conducted on a small artificial set of agents to examine how different matching methods and parameters affect partnering patterns. Since the data set was artificial, it examined patterns in a generic hierarchical trait, rather than ethnicity. It compared a random matching model with one where agents were attracted to a mate with the highest level of the trait, and one where agents were attracted to a mate with the most similar level of the trait. It found that the method of matching had an effect on the degree of homogamy. However, this was based on an ordinal characteristic, rather than a nominal one like ethnicity. The number of possible mates also had an effect, with a smaller pool of possible mates leading to less homogamy due to fewer options for the agents to choose from.

In Chapter 7 an empirical simulation model was applied to unit-level census data to investigate the second research question of what factors and social processes were involved in the partnership matching process. The second research question was broken down into three sub-goals for the empirical simulation model:

1. To examine individually the effect of each of the scoring variables (age, education, macro, random).
2. To find the weighted combination of the scoring variables that produced the most similar set of inter-ethnic cohabitation patterns to those that actually occurred.
3. To investigate the possibility of a micro-macro relationship within ethnic partnering patterns.

When the simulation was run with each of the single scoring variables (Section 7.4.2), the frequencies for the same ethnicity partnerships tended to be under-estimated, whilst the frequencies for the mixed ethnicity partnerships were generally over-estimated. This suggests that there was some level of same ethnic preference that was not captured in the simulation model. It also corresponds with the quasi-independence model findings of ethnic preferences, as shown by the diagonal dominance parameters on the frequency tables. Of the four variables, the macro variable had the most interesting results. It shifted from promoting homogamous partnerships in the earlier census periods to promoting heterogamous ones in the later census periods. This would imply that not only does macro-level "social pressure" have an impact on micro-level partnering decisions, but it also mimicked the shift towards more mixed-ethnicity couples as the regions became more diverse.

The evolutionary algorithm (Section 7.4.3) found that the weights were generally dominated by the age and education variables, reinforcing previous studies that had demonstrated that age similarity and educational similarity are both important variables for partnership matching. However, there was some weight on the macro variable, and at times the random variable, providing evidence of a possible micro-macro link, and also random variation in the matching process that was not adequately described by the other scoring variables.

8.3. Future Research Possibilities

There are a number of future research possibilities that could be pursued as a follow on to this piece of research, in both content and methodology. One extension of the research would be to supplement the data from the census with other quantitative and/or qualitative data. Using data-matching, quantitative data from other sources could be added to the statistical and simulation models. The statistical and simulation methodologies could also be extended. As well as incorporating other data, the statistical analysis could be extended by breaking down the ethnic categories into smaller groups. In particular, the sub-groups of the Asian Only and Pacific Only groups could be analysed to see whether the within-ethnicity patterns are changing in the same way as the between-ethnicity ones.

9. References

Abercrombie, N., Hill, S., & Turner, B. S. (2000). *The Penguin Dictionary of Sociology*. London: Penguin.

Agresti, A. (2002). *Categorical Data Analysis* (second ed.). New Jersey: John Wiley & Sons.

Aguirre, B. E., Saenz, R., & Hwang, S.-S. (1995). Remarriage and intermarriage of Asians in the United States of America. *Journal of Comparative Family Studies, 26*(2), 207-215.

Alam, S. J., & Meyer, R. (2008). *Comparing two sexual mixing schemes for modelling the spread of HIV/AIDS.* Paper presented at the World Congress on Social Simulation, George Mason University, Washington DC.

Alba, R. D., & Golden, R. M. (1986). Patterns of ethnic marriage in the United States. *Social Forces, 65*(1), 202-223.

Anderson, A. (1991). *Race against time: The early Maori-Pakeha families and the development of the mixed-race population in southern New Zealand*. Dunedin: University of Otago.

Ashlock, D. (2006). *Evolutionary Computation for Modeling and Optimization*. New York: Springer.

Åström, M., & Vencatasawmy, C. P. (2001). Incorporating artificial intelligence in microsimulation. *Geografiska Annaler Series B: Human Geography, 83*(2), 53-65.

Bacon, B., & Pennec, S. (2007). APPSIM - Modelling Family Formation and Dissolution. Working Paper No. 4, from http://www.canberra.edu.au/centres/natsem/publications

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy, 81*(4), 813-846.

Becker, G. S. (1974). A theory of marriage: Part II. *Journal of Political Economy, 82*(2), 11-26.

Best, S. (2005). *Understanding Social Divisions*. London: Sage Publications.

Blackwell, D. L. (1998). Marital homogamy in the United States: The influence of individual and paternal education. *Social Science Research, 27*(2), 159-188.

Blackwell, D. L., & Lichter, D. T. (2000). Mate selection among married and cohabiting couples. *Journal of Family Issues, 21*(3), 275-302.

Blackwell, D. L., & Lichter, D. T. (2004). Homogamy among dating, cohabiting and married couples. *Sociological Quarterly, 45*(4), 719-737.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*. New York: Free Press.

Bouffard, N., Easther, R., Johnson, T., Morrison, R. J., & Vink, J. (2001). Matchmaker, matchmaker, make me a match. *Brazilian Electronic Journal of Economics, 4*(2).

Bratter, J., & Zuberi, T. (2001). The demography of difference: shifting trends of racial diversity and interracial marriage 1960-1990. *Race and Society, 4*(2), 133-148.

Bull, H. H. (2005). Deciding whom to marry in a rural two-class society: Social homogamy and constraints in the marriage market in Rendalen, Norway, 1750-1900. *International Review of Social History, 50*, 43-63.

Callister, P. (1998). The "meet" market: Education and assortative mating patterns in New Zealand. *New Zealand Population Review, 24*, 43-70.

Callister, P. (2003). Maori/Non-Maori ethnic intermarriage. *New Zealand Population Review, 29*(2), 89-105.

Callister, P. (2004). Ethnicity measures, intermarriage and social policy. *Social Policy Journal of New Zealand, 23*, 109-140.

Callister, P. (2009). Measuring ethnicity: An introduction to the theme. *Social Policy Journal of New Zealand, 36*, 1.

Callister, P., & Blakely, T. (2004). Ethnic classification, intermarriage and mortality: Some methodological issues in relation to ethnic comparisons in Aotearoa/New Zealand. *Working Paper*.

Callister, P., Didham, R., & Potter, D. (2005). Ethnic intermarriage in New Zealand. *Statistics New Zealand Working Paper*.

Callister, P., Didham, R., & Potter, D. (2007). Ethnic intermarriage in New Zealand. *Official Statistics Research Series*, Vol 1, from http://www.statisphere.govt.nz/official-statistics-research/series/volume-1-2007.aspx

Carter, K., Hayward, M., Blakely, T., & Shaw, C. (2009). How much and for whom does self-identified ethnicity change over time in New Zealand? Results from a longitudinal study. *Social Policy Journal of New Zealand, 36*, 32-45.

Cheesbrough, S., & Scott, A. (2003). Simulating demographic events in the SAGE model, *Technical Note 4*. London: ESRC-Sage Research Group.

Chen, C.-Y. (2005). *Picking and choosing: The simulation of sequential mate selection process.* Paper presented at the Population Association of American 2006 Annual Meeting, Los Angeles, California.

Coleman, J. S. (1990). *Foundations of Social Theory*. Cambridge, MA: The Belknap Press of Harvard University Press.

Conte, R., Paolucci, M., & Di Tosto, G. (2006). Vampire bats and the micro-macro link. In F. C. Billari, T. Fent, A. Prskawetz & J. Scheffran (Eds.), *Agent-Based Computation Modelling: Applications in Demography, Social, Economic and Environmental Sciences* (pp. 173-194). Berlin: Physica-Verlag.

Crowder, K. D., & Tolnay, S. E. (2000). A new marriage squeeze for black women: The role of racial intermarriage by black men. *Journal of Marriage and the Family, 62*(3), 792.

Dempsey, K., & De Vaus, D. (2004). Who cohabits in 2001?: The significance of age, gender, religion and ethnicity. *Journal Of Sociology, 40*(2), 157-178.

Dharmalignam, A. (2002). Marriage, cohabitation and marital dissolution in New Zealand. *Journal of Population Research, September Special Issue*, 121-136.

Duncan, O. D. (1979). How destination depends on origin in the occupational mobility table. *American Journal of Sociology, 84*(4), 793-803.

Epstein, J. M. (2006). *Generative social science: Studies in agent-based computational modeling*. New Jersey: Princeton University Press.

Favreault, M., & Smith, K. (2004). A primer on the dynamic simulation of income model (DYNASIM3). from http://www.urban.org/UploadedPDF/410961_Dynasim3Primer.pdf

Ferguson, T. S. (1989). Who solved the secretary problem? *Statistical Science, 4*(3), 282-289.

Freeman, P. R. (1983). The Secretary Problem and its extensions: A review. *International Statistical Review / Revue Internationale de Statistique, 51*(2), 189-206.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly, 69*(1), 9-15.

Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. Maidenhead: Open University Press.

Giorgas, D., & Jones, F. L. (2002). Intermarriage patterns and social cohesion among first, second and later generation Australians. *Journal of Population Research, 19*(1), 47-64.

Goodman, L. A. (2007). Statistical magic and/or statistical serendipity: An age of progress in the analysis of categorical data. *Annual Review of Sociology, 33*(1), 1-19.

Gusfield, D., & Irving, R. W. (1989). *The Stable Marriage Problem*. Cambridge, MA.: The MIT Press.
Halpin, B. (1999). Simulation in sociology. *American Behavioral Scientist, 42*(10), 1488-1508.
Harding, A. (2007). *APPSIM: The Australian dynamic population and policy microsimulation model*. Paper presented at the Technical Workshops of the 1st General Conference of the International Microsimulation Association.
Harris, D. R., & Ono, H. (2005). How many interracial marriages would there be if all groups were of equal size in all places? A new look at national estimates of interracial marriage. *Social Science Research, 34*(1), 236-251.
Hayes, B. C. (1991). Religious identification and marriage patterns in Australia. *Journal for the Scientific Study of Religion, 30*(4), 469-478.
Hazelrigg, L. E., & Garnier, M. A. (1976). Occupational mobility in industrial societies: A comparative analysis of differential access to occupational ranks in seventeen countries. *American Sociological Review, 41*, 498-511.
Heard, G., Birrell, B., & Khoo, S.-E. (2009). Intermarriage between indigenous and non-indigenous Australians. *People and Place, 17*(1), 1-14.
Hills, T., & Todd, P. M. (2008). Population heterogeneity and individual differences in an assortative agent-based marriage and divorce model (MADAM) using search with relaxing expectations. *Journal of Artificial Societies and Social Simulation, 11*(4).
Hoffmann, A. O. I., Jager, W., & von Eije, J. H. (2006). Combining Empirical Data with Multi Agent Social Simulation: Investigating Micro-Macro Links in Stock Markets. Available at SSRN: http://ssrn.com/abstract=898826
Howard, S., & Didham, R. (2007). Ethnic intermarriage and ethnic transference amongst the Maori population: implications for the measurement and definition of ethnicity. *Official Statistics Research Series*, Vol 1, from http://www.statisphere.govt.nz/official-statistics-research/series/volume-1-2007.aspx
Hwang, S.-S., Saenz, R., & Aguirre, B. E. (1997). Structural and assimilationist explanations of Asian American intermarriage. *Journal of Marriage and the Family, 59*(3), 758-772.
Jacob, C. (2003). Stochastic search methods. In M. Berthold & D. J. Hand (Eds.), *Intelligent Data Analysis*. Berlin: Springer-Verlag.

Jaffe, K. (1999). On the adaptive value of some mate selection strategies. *Acta Biotheoretica, 47*(1), 29-40.

Johnson, R. A. (1980). *Religious Assortative Marriage in the United States*. New York: Academic Press.

Jones, F. L. (1991). Ethnic intermarriage in Australia, 1950-52 to 1980-82: Models or indices? *Population Studies, 45,* 27-42.

Kalick, S. M., & Hamilton, T. E. (1986). The matching hypothesis reexamined. *Journal of Personality and Social Psychology, 51*(4), 673-682.

Kalick, S. M., & Hamilton, T. E. (1988). Closer look at a matching simulation: Reply to Aron. *Journal of Personality and Social Psychology, 54*(3), 447-451.

Kalmijn, M. (1991). Shifting boundaries: Trends in religious and educational homogamy. *American Sociological Review, 56*(6), 786-800.

Kalmijn, M. (1993). Trends in black/white intermarriage. *Social Forces, 72*(1), 119-146.

Kalmijn, M. (1994). Assortative mating by cultural and economic occupational status. *American Journal of Sociology, 100*(2), 422-452.

Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, and trends. *Annual Review of Sociology, 24*(1), 395-421.

Kalmijn, M., & Vermunt, J. K. (2007). Homogeneity of social networks by age and marital status: A multilevel analysis of ego-centered networks. *Social Networks, 29*(1), 25-43.

Kelly Raley, R. (1996). A shortage of marriageable men? A note on the role of cohabitation in black-white differences in... *American Sociological Review* (Vol. 61, pp. 973): American Sociological Association.

Khoo, S.-E., Birrell, B., & Heard, G. (2009). Intermarriage by birthplace and ancestry in Australia. *People and Place, 17*(1), 15-28.

Klein, T. (2001). Intermarriages between Germans and foreigners in Germany. *Journal of Comparative Family Studies, 32*(3), 325-346.

Knuth, D. E. (1997). *Stable Marriage and its Relation to Other Combinatorial Problems: An Introduction to the Mathematical Analysis of Algorithms*. Providence, Rhode Island: American Mathematical Society.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research, 33*(2), 188-229.

Levi-Strauss, C. (1969). *The Elementary Structures of Kinship*. Boston: Beacon Press.

Logan, J. A., Hoff, P. D., & 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), 559-569.
Luna, F., Alba, E., Nebro, A. J., Mauroy, P., & Pedraza, S. (2008). Applying evolutionary algorithms to solve the automatic frequency planning problem. In A. Nayak & I. Stojmenovic (Eds.), *Handbook of Applied Algorithms* (pp. 271-286). New Jersey: John Wiley & Sons.
Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology, 28*(1), 143-166.
Mamet, R., Jacobson, C. K., & Heaton, T. B. (2005). Ethnic intermarriage in Beijing and Xinjiang, China, 1990. *Journal of Comparative Family Studies, 36*(2), 187-204.
Manning, W. D., & Smock, P. J. (1995). Why marry? Race and the transition to marriage among cohabitors. *Demography, 32*(4), 509-520.
Mare, R. D. (2001). Five decades of educational assortative mating. *American Sociological Review, 56*(1), 15-32.
McCaa, R., & Schwartz, S. B. (1983). Measuring marriage patterns: Percentages, Cohen's Kappa, and log-linear models. *Comparative Studies in Society and History, 25*(4), 711-724.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415-444.
Menard, S. (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician, 54*(1), 17-24.
Merton, R. K. (1941). Intermarriage and the social structure: Fact and theory. *Psychiatry, 4*, 361-374.
Miller, G. F., & Todd, P. M. (1998). Mate choice turns cognitive. *Trends in Cognitive Sciences, 2*(5), 190-198.
Milligan, S., Fabian, A., Coope, P., & Errington, C. (2006). *Family Wellbeing Indicators from the 1981-2001 New Zealand Censuses*: Statistics New Zealand.
Ministry of Education. (2007). Education counts: Participation rates in tertiary education. Retrieved 30 April 2008, from http://www.educationcounts.govt.nz/indicators/student_participation/tertiary_education/participation_rates_in_tertiary_education
Muttarak, R. (2003). *Who Intermarries in Britain? Ethnic Intermarriage.* Unpublished Masters Thesis, University of Oxford.

Nabrzyski, J., Schopf, J. M., & Weglarz, J. (Eds.). (2004). *Grid Resource Management: State of the Art and Future Trends*. Norwell, MA: Kluwer Academic Publishers.

Nayak, A., & Stojmenovic, I. (Eds.). (2008). *Handbook of Applied Algorithms*. New Jersey: John Wiley & Sons.

O'Donoghue, C., Lennon, J., & Hynes, S. (2009). The Life-Cycle Income Analysis Model (LIAM): A study of a flexible dynamic microsimulation modelling computing framework. *International Journal of Microsimulation, 2*(1), 16-31.

O'Leary, R. (2001). Modernization and religious intermarriage in the Republic of Ireland. *The British Journal of Sociology, 52*(4), 647-665.

Orcutt, G. H. (1957). A new type of socio-economic system. *The Review of Economic Statistics, 39*(2), 116-123.

Orcutt, G. H., Caldwell, S., & Wertheimer II, R. (1976). *Policy Exploration Through Microanalytic Simulation*. Washington DC: Urban Institute Press.

Parimal, R., & Hamilton, I. (2000). Intermarriage among Italians: Some regional variations in Australia. *Journal of Comparative Family Studies, 31*(1), 63.

Parry, H. R., Evans, A. J., & Heppenstall, A. J. (2006). *Millions of agents: Parallel simulations with the RePast agent-based toolkit.* Paper presented at the Austrian Society for Cybernetic Studies, Vienna.

Parsons, T. (1954). *Essays in Sociological Theory*. New York: Free Press.

Pennec, S., & Bacon, B. (2007). APPSIM - Modelling fertility and mortality. *National Centre for Social and Economic Modelling (NATSEM) Working Paper*, No 7., from http://www.canberra.edu.au/centres/natsem/publications

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.

Pujol, J. M., Flache, A., Delgado, J., & Sanguesa, R. (2005). How can social networks ever become complex? Modelling the emergence of complex networks from local social exchanges. *Journal of Artificial Societies and Social Simulation, 8*(4).

Qian, Z., & Lichter, D. T. (2001). Measuring marital assimilation: Intermarriage among natives and immigrants,. *Social Science Research, 30*(2), 289-312.

Qian, Z., & Lichter, D. T. (2007). Social boundaries and marital assimilation: Interpreting trends in racial and ethnic intermarriage. *American Sociological Review, 72*(1), 68-94.

Regis, R. G., & Shoemaker, C. A. (2004). Local function approximation in evolutionary algorithms for the optimization of costly functions. *IEEE Transactions on Evolutionary Computation, 8*(5), 490-505.

Rephann, T. J., & Holm, E. (2004). Economic-demographic effects of immigration: Results from a dynamic spatial microsimulation model. *International Regional Science Review, 27*(4), 379-410.

Richmond, P. (2003). *Never the twain shall meet?: Causal factors in Fijian-Indian intermarriage.* Unpublished Masters Thesis, University of Hawaii.

Riddell, K. (2000). Improving the Maori: Counting the ideology of intermarriage. *New Zealand Journal of History, 34*(1), 80-97.

Robbins, M. C. (1981). A measure of social homogamy. *American Anthropologist*, *83*(2), 379-385.

Romney, A. (1971). Measuring Endogamy. In P. Kay (Ed.), *Explorations in Mathematical Anthropology* (pp. 191-213). Cambridge: MIT Press.

Rosenfeld, M. J. (2005). A critique of exchange theory in mate selection. *American Journal of Sociology, 110*(5), 1284-1325.

Roy, P., & Hamilton, I. (1997). Interethnic marriage: Identifying the second generation in Australia. *The International Migration Review, 31*(1), 128-142.

Ryan, M., Boddington, B., & Dunstan, K. (2005). Age differences between partners at the time of marriage: 1963, 1983 and 2003. In *Demographic Trends 2004* (pp. 3-8). Wellington: Statistics New Zealand.

Rytina, S., Blau, P. M., Blum, T., & Schwartz, J. (1988). Inequality and intermarriage: A paradox of motive and constraint. *Social Forces, 66*(3), 645-675.

Saam, N. J. (1999). Simulating the micro-macro link: New approaches to an old problem and an application to military coups. *Sociological Methodology, 29,* 43-79.

Sawyer, R. K. (2000). Simulating emergence and downward causation in small groups. In S. Moss & P. Davidsson (Eds.), *Multi Agent Based Simulation, Lecture Notes in Artificial Intelligence* (Vol. 1979, pp. 49-67). Berlin: Springer-Verlag.

Sawyer, R. K. (2003). Artificial societies: Multiagent systems and the micro-macro link in sociological theory. *Sociological Methods and Research, 31*(3), 325-363.

Sawyer, R. K. (2005). *Social emergence: Societies as complex systems*. New York: Cambridge.

Schoen, R., & Weinick, R. M. (1993). Partner choice in marriages and cohabitations. *Journal of Marriage and the Family, 55*(2), 408-414.
Schwartz, C. R., & Mare, R. D. (2005). Trends in educational assortative marriage from 1940 to 2003. *Demography, 42*(4), 621-646.
Schwefel, H.-P. (1981). *Numerical Optimization of Computer Models*. New York: John Wiley & Sons.
Schwefel, H.-P. (2000). Advantages (and disadvantages) of evolutionary computation over other approaches. In T. Back, D. B. Fogel & T. Michalewicz (Eds.), *Evolutionary Computation 1: Basic Algorithms and Operators* (pp. 20-22). Bristol: Institute of Physics Publishing.
Sherkat, D. E. (2004). Religious intermarriage in the United States: trends, patterns, and predictors. *Social Science Research, 33*(4), 606-625.
Silverman, E., & Bryden, J. (2007). From artificial societies to new social science theory. In F. Almeida e Costa, L. M. Rocha, E. Costa, I. Harvey & A. Coutinho (Eds.), *European Conference on Artificial Life, Lecture Notes in Artificial Intelligence* (Vol. 4648, pp. 565-574). Berlin: Springer-Verlag.
Simão, J., & Todd, P. M. (2001). A Model of Human Mate Choice with Courtship That Predicts Population Patterns. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim & J. Ziegler (Eds.), *Advances in Artificial Life* (pp. 377-380). Dortmund: Springer.
Simão, J., & Todd, P. M. (2002). Modeling mate choice in monogamous mating systems with courtship. *Adaptive Behavior, 10*(2), 113-136.
Simão, J., & Todd, P. M. (2003). Emergent patterns of mate choice in human populations. *Artificial Life, 9*(4), 403-417.
Simonoff, J. S. (2003). *Analysing Categorical Data*. New York: Springer.
Smits, J. (2003). Social closure among the higher educated: trends in educational homogamy in 55 countries. *Social Science Research, 32*(2), 251-277.
Smits, J., Ultee, W., & Lammers, J. (1999). Occupational homogamy in eight countries of the European Union, 1975-89. *ACTA Sociologica, 42*, 55-68.
Spielauer, M., & Vencatasawmy, C. P. (2001). FAMSIM: Dynamic microsimulation of life course interactions between education, work, partnership formation and birth in Austria, Belgium, Italy, Spain and Sweden. *Brazilian Electronic Journal of Economics, 4*(2).
Statistics New Zealand. (2004). Report of the review of the measurement of ethnicity. from http://www2.stats.govt.nz/domino/external/web/aboutsnz.nsf/htmldocs/Review+of+the+Measurement+of+Ethnicity#Report

Statistics New Zealand. (2005). Statistical Standard for Ethnicity. Retrieved 17/11/2006, from http://www.stats.govt.nz/reports/analytical-reports/review-measurement-of-ethnicity.aspx
Statistics New Zealand. (2006a). QuickStats about culture and identity. Retrieved 4 December 2008, from http://www.stats.govt.nz/census/2006censushomepage/quickstats/quickstats-about-a-subject/culture-and-identity.aspx
Statistics New Zealand. (2006b, 4 May 2006). Statistics Act 1975. Retrieved 1/12/2007, from http://www.stats.govt.nz/about-us/who-we-are/statistics-act-1975.htm
Statistics New Zealand. (2007a). Guidelines for using ethnicity data: 2006 census. from http://www.stats.govt.nz/reports/analytical-reports/review-measurement-of-ethnicity/papers.aspx
Statistics New Zealand. (2007b). Labour market statistics (Publication. Retrieved 30 April 2008: http://www.stats.govt.nz/publications/workknowledgeandskills/labour-market-statistics-2007.aspx
Statistics New Zealand. (2007c). Profile of New Zealander responses, ethnicity question: 2006 census. from http://www.stats.govt.nz/census/about-2006-census/profile-of-nzer-responses-ethnicity-question-2006-census.aspx
Statistics New Zealand. (2008a). Statistical standard for partnership status in current relationship. Retrieved 7 October, 2009, from http://www.stats.govt.nz/~/media/Statistics/class-stnd/partnership-status-in-current-relationship/Partnershipstatus-statistical-standard.ashx
Statistics New Zealand. (2008b). Statistical standard for relationship between members in a private dwelling. Retrieved 7 October, 2009, from http://www.stats.govt.nz/~/media/Statistics/class-stnd/relationship/RELAT08-statistical-standard.ashx
Statistics New Zealand. (2009). Draft report of a review of the official ethnicity statistical standard: Proposals to address the 'New Zealander' response issue. from http://www.stats.govt.nz/~/media/Statistics/Methods%20and%20Services/Review%20ethnicity/draft-reportof-reviewo-official-ethnicity-standard.ashx
Stephan, C. W., & Stephan, W. G. (1989). After intermarriage: Ethnic identity among mixed-heritage Japanese-Americans and Hispanics. *Journal of Marriage and the Family, 51*(2), 507-519.

Stolk, H. J., Hanan, J., & Zalucki, M. P. (2007). Subpopulation Agents Emerge from Individual Agents in Metapopulations. In L. Oxley & D. Kulasiri (Eds.), *MODSIM 2007 International Congress on Modelling and Simulation* (pp. 46-52): ISBN : 978-0-09758400-4-7.

Thomas, A., & Sawhill, I. (2002). For richer or for poorer: Marriage as an antipoverty strategy. *Journal of Policy Analysis and Management, 21*(4), 587-599.

Todd, P. M. (1997). Searching for the Next Best Mate. In R. Conte, R. Hegselmann & P. Terna (Eds.), *Simulating Social Phenomena* (pp. 419-436). Berlin: Springer.

Todd, P. M., & Billari, F. C. (2003). Population-wide marriage patterns produced by indivdual mate-search heuristics. In F. C. Billari & A. Prskawetz (Eds.), *Agent-Based Computational Demography: Using Simulation to Improve Our Understanding of Demographic Behaviour* (pp. 117-137). New York: Physica-Verlag.

Todd, P. M., Billari, F. C., & Simão, J. (2005). Aggregate age at marriage patterns from individual mate search heuristics. *Demography, 42*(3), 559-574.

Todd, P. M., & Miller, G. F. (1999). From pride and prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer, P. Todd & ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 287-308). New York: Oxford University Press.

Todd, P. M., & Miller, G. F. (2002). From pride and prejudice to persuasion: Satisficing in mate search. In G. Gigerenzer & P. Todd (Eds.), *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Troitzsch, K. G. (1996). *Social Science Microsimulation*. Berlin: Springer.

Tucker, M. B., & Mitchell-Kernan, C. (1990). New trends in black American interracial marriage: The social structural context. *Journal of Marriage and the Family, 52*(1), 209-218.

Tzeng, J. M. (2000). Ethnically heterogamous marriages: The case of Asian Canadians. *Journal of Comparative Family Studies, 31*(3), 321-340.

Ultee, W., & Luijkx, R. (1990). Educational heterogamy and father-to-son occupational mobility in 23 industrial nations: General societal openness or compensatory strategies of reproduction. *European Sociological Review, 6*(2), 125-149.

Urdan, T. C. (2005). *Statistics in Plain English*. Mahwah, NJ: Lawrence Elbaum.

Uunk, W. J. G., Ganzeboom, H. B. G., & Róbert, P. (1996). Bivariate and multivariate scaled association models. An application to homogamy of social origin and education in Hungary between 1930 and 1979. *Quality and Quantity, 30*(3), 323-343.

van Imhoff, E., & Post, W. (1998). Microsimulation methods for population projection. *Population: An English Selection, 10*(1), 97-138.

Voas, D. (2003). Intermarriage and the demography of secularization. *The British Journal of Sociology, 54*(1), 83-108.

Waters, M. S., & Ressler, R. W. (1999). An economic model of cohabitation and divorce. *Journal of Economic Behavior & Organization, 40*(2), 195-206.

White, D. (1999). Controlled simulation of marriage systems. *Journal of Artificial Societies and Social Simulation, 3*(2).

White, D. (2004). Ring cohesion theory in marriage and social networks. *Mathématiques et Sciences Humaines, 168*(4), 5-28.

White, D., & Jorion, P. (1996). Kinship networks and discrete structure theory: Applications and implications. *Social Networks, 18*, 267-314.

Xie, Y., Raymo, J. M., Goyette, K., & Thornton, A. (2003). Economic potential and entry into marriage and cohabitation. *Demography, 40*(2), 351-367.

Yue, T. X., Wang, Y. A., Chen, S. P., Liu, J. Y., Qiu, D. S., Deng, X. Z., et al. (2003). Numerical simulation of population distribution in China. *Population and Environment, 25*(2), 141-163.

Zedlewski, S., R. (1990). The development of the dynamic simulation of income model (DYNASIM). In G. H. Lewis & R. C. Michel (Eds.), *Microsimulation Techniques for Tax and Transfer Analysis*. Washington, DC: Urban Institute Press.

Zhao, Z. W. (2000). Coresidential patterns in historical China: A simulation study. *Population and Development Review, 26*(2), 263-293.

A. Appendix A: Partnership Frequency Tables

Ethnicity of Male Partner	Ethnicity of Female Partner											Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Not elsewhere included	
European only	546294	5823	1917	1212	81	48	15504	48	1500	540	5142	578109
Maori only	4797	16869	246	21	0	3	3786	90	96	12	282	26199
Pacific only	1314	843	9030	3	0	0	447	36	303	3	366	12351
Asian Only	822	81	54	5133	0	0	78	0	12	27	60	6267
MELAA only	102	6	0	3	36	0	6	0	0	0	6	159
Other ethnicity only	39	3	0	3	0	42	3	0	3	0	0	96
Maori & European	14646	2865	150	45	0	3	9897	66	225	51	534	28485
Maori & Pacific	66	93	18	0	0	0	66	6	6	0	12	267
Pacific & European	1062	114	366	0	0	0	327	3	318	3	114	2307
Asian & European	441	18	9	21	0	0	39	0	9	60	18	615
Not elsewhere included	3459	327	393	45	3	0	537	9	117	12	2952	7854
Total	573045	27045	12177	6486	123	102	30690	258	2586	717	9486	662706

A.1 - 1981: All couples

Ethnicity of Male Partner	Ethnicity of Female Partner											Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Not elsewhere included	
European only	484533	5808	1323	867	63	33	15480	45	1233	405	4557	514344
Maori only	4794	16866	243	21	0	0	3783	87	96	15	279	26190
Pacific only	1146	840	522	0	0	0	447	33	90	0	69	3147
Asian Only	681	78	9	1539	0	0	78	3	3	18	21	2439
MELAA only	90	6	0	0	12	0	3	0	0	0	6	120
Other ethnicity only	36	3	0	0	0	3	3	0	0	0	0	51
Maori & European	14628	2865	150	45	3	3	9891	66	222	51	531	28452
Maori & Pacific	63	93	12	0	0	0	66	6	3	0	9	258
Pacific & European	978	114	36	0	0	0	327	3	66	0	42	1566
Asian & European	369	21	6	9	0	0	42	0	3	6	9	465
Not elsewhere included	3153	327	57	21	3	0	534	6	48	6	1926	6078
Total	510477	27018	2352	2505	81	42	30657	249	1767	504	7455	583104

A.2 - 1981: At least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner											Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Not elsewhere included	
European only	91266	1383	378	261	12	18	5358	21	450	150	708	100005
Maori only	1719	4677	132	6	0	0	1539	63	72	9	117	8327
Pacific only	465	339	252	0	0	0	243	21	63	0	39	1421
Asian Only	264	36	9	480	0	0	30	0	0	6	12	830
MELAA only	15	0	0	0	0	0	0	0	0	0	3	18
Other ethnicity only	18	0	0	3	0	0	0	0	0	0	0	21
Maori & European	5775	954	75	15	3	3	4389	42	129	24	288	11694
Maori & Pacific	48	54	9	0	0	0	54	3	3	0	9	187
Pacific & European	426	57	18	0	0	0	177	0	39	0	18	741
Asian & European	105	6	0	3	0	0	18	0	3	3	6	138
Not elsewhere included	663	120	33	3	0	0	258	6	24	6	252	1365
Total	100761	7626	903	768	15	21	12069	159	783	195	1449	124749

A.3 - 1981: Couples with male partner aged 18-30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner											Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Not elsewhere included	
European only	393267	4422	945	606	51	15	10122	24	780	252	3849	414339
Maori only	3075	12192	114	15	0	0	2244	27	24	6	159	17859
Pacific only	681	501	270	0	0	0	204	12	27	0	30	1725
Asian Only	420	45	0	1062	0	3	51	0	3	15	12	1608
MELAA only	75	6	0	0	12	0	6	0	0	0	3	102
Other ethnicity only	18	3	0	0	0	3	3	0	0	0	0	30
Maori & European	8856	1911	75	30	0	0	5502	24	96	27	243	16758
Maori & Pacific	18	36	3	0	0	0	12	0	0	0	0	72
Pacific & European	552	54	18	0	0	0	147	0	30	0	21	825
Asian & European	264	15	3	6	0	0	24	0	3	3	6	324
Not elsewhere included	2487	207	24	18	3	0	273	0	21	3	1677	4713
Total	409716	19395	1452	1737	66	18	18588	90	984	309	6006	458358

A.4 - 1981: Couples with male partner aged greater than 30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	**Maori only**	**Pacific only**	**Asian Only**	**MELAA only**	**Other ethnicity only**	**Maori & European**	**Maori and Pacific**	**Pacific and European**	**Asian and European**	**Two groups not elsewhere**	**Maori and Pacific and European**	**Not Elsewhere Included**	
European only	564498	13347	2829	2298	171	2949	7062	57	777	375	444	87	7950	602841
Maori only	12915	28380	573	84	3	207	1662	105	114	15	87	36	711	44889
Pacific only	2166	1518	12486	24	3	108	258	72	153	3	81	12	237	17127
Asian Only	1071	108	90	6756	0	45	36	0	3	9	27	3	87	8235
MELAA only	231	15	3	6	135	9	9	0	3	0	3	0	9	420
Other ethnicity only	2319	225	96	30	3	1263	57	3	6	0	9	0	45	4056
Maori & European	6480	1194	96	24	6	48	2160	27	114	18	54	39	189	10443
Maori & Pacific	84	105	27	0	0	0	27	30	3	0	0	0	9	282
Pacific & European	558	96	147	3	0	3	141	6	132	3	18	9	48	1158
Asian & European	327	18	6	12	0	0	24	0	6	33	3	0	9	438
Two groups not elsewhere	327	108	102	15	3	6	84	3	21	0	81	0	24	774
Maori & Pacific & European	96	21	6	0	0	0	33	3	6	0	3	3	6	180
Not Elsewhere Included	14520	1023	288	165	3	117	387	3	45	18	24	9	3771	20379
Total	605589	46158	16749	9417	333	4758	11934	306	1383	480	834	198	13089	711225

A.5 - 1986: All couples

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori and Pacific	Pacific and European	Asian and European	Two groups not elsewhere	Maori and Pacific and European	Not Elsewhere Included	
European only	504504	13329	1992	1662	132	2583	7053	57	681	294	360	87	6363	539103
Maori only	12906	28371	564	81	6	204	1662	105	111	15	87	36	711	44856
Pacific only	1896	1512	681	6	0	39	255	63	81	3	15	9	45	4602
Asian Only	891	111	15	1830	0	12	36	3	3	6	12	3	21	2940
MELAA only	198	12	3	0	15	6	6	0	0	0	3	0	0	249
Other ethnicity only	2058	228	36	15	0	843	54	3	6	0	3	0	24	3267
Maori & European	6471	1194	99	24	3	48	2154	27	111	18	51	39	189	10431
Maori & Pacific	81	102	18	0	0	0	27	9	6	0	0	0	6	249
Pacific & European	534	96	36	0	0	3	141	6	36	3	9	6	24	894
Asian & European	291	18	0	6	0	3	21	0	3	3	0	0	3	357
Two groups not elsewhere	303	108	9	9	0	6	81	3	6	3	18	3	15	564
Maori & Pacific & European	96	24	3	0	0	0	33	3	6	0	0	3	3	171
Not Elsewhere Included	11970	1020	27	42	0	81	384	3	21	9	12	9	27	13602
Total	542202	46122	3480	3675	159	3822	11919	279	1071	357	576	195	7434	621288

A.6 - 1986: At least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori and Pacific	Pacific and European	Asian and European	Two groups not elsewhere	Maori and Pacific and European	Not Elsewhere Included	
European only	84126	3477	438	363	27	309	2004	21	240	87	90	27	1233	92433
Maori only	4476	8475	297	18	3	51	729	66	75	6	33	24	255	14505
Pacific only	663	591	453	3	0	9	132	33	54	3	9	6	24	1983
Asian Only	222	36	3	462	0	3	9	0	0	3	3	0	9	750
MELAA only	45	9	0	3	0	0	3	0	0	0	0	0	0	60
Other ethnicity only	336	66	6	3	0	99	15	3	6	0	0	0	3	534
Maori & European	1977	399	42	9	3	18	795	15	69	6	18	18	72	3441
Maori & Pacific	51	63	12	0	0	0	18	6	3	0	0	0	6	162
Pacific & European	216	48	24	0	0	0	66	6	27	3	6	3	15	411
Asian & European	63	6	0	3	0	0	9	0	0	0	0	0	0	84
Two groups not elsewhere	93	27	6	3	0	0	27	3	3	0	3	0	3	174
Maori & Pacific & European	48	9	3	0	0	0	15	3	3	0	0	3	3	84
Not Elsewhere Included	81	18	3	3	0	0	33	3	3	0	3	3	12	162
Total	92400	13221	1284	867	30	492	3852	153	486	108	165	84	1638	114777

A.7 - 1986: Couples with male partner aged 18-30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori and Pacific	Pacific and European	Asian and European	Two groups not elsewhere	Maori and Pacific and European	Not Elsewhere Included	
European only	420381	9849	1557	1302	105	2274	5052	39	444	207	270	60	5130	446670
Maori only	8430	19896	267	63	3	150	933	39	36	9	57	15	453	30351
Pacific only	1230	921	228	3	3	30	123	30	27	0	6	3	18	2622
Asian Only	669	75	12	1365	0	9	30	0	3	6	6	3	15	2190
MELAA only	153	6	0	0	15	3	6	0	0	0	3	0	0	186
Other ethnicity only	1725	159	27	9	3	747	39	0	3	0	3	3	21	2733
Maori & European	4497	792	54	15	0	30	1359	12	42	12	33	18	117	6993
Maori & Pacific	30	42	6	0	0	0	6	3	0	0	0	0	3	90
Pacific & European	318	51	12	3	0	0	75	0	9	0	3	6	9	486
Asian & European	225	15	0	3	0	0	15	0	0	3	3	3	3	273
Two groups not elsewhere	207	78	3	9	3	6	54	0	0	0	15	0	9	390
Maori & Pacific & European	48	12	0	0	0	0	18	0	3	0	3	0	3	87
Not Elsewhere Included	11886	999	24	39	0	81	351	3	18	9	9	6	15	13440
Total	449802	32898	2196	2808	129	3330	8067	126	582	249	411	111	5799	506511

A.8 - 1986: Couples with male partner aged greater than 30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	566985	14694	3246	4263	399	1662	6735	84	600	405	435	9297	608802
Maori only	15207	26259	678	108	12	123	1668	102	153	60	102	891	45366
Pacific only	2553	1722	16224	51	3	66	321	66	177	24	66	339	21612
Asian Only	1389	147	222	14946	12	48	51	6	9	51	27	303	17205
MELAA only	510	21	15	39	504	3	9	0	3	0	9	21	1137
Other ethnicity only	1368	147	57	36	3	333	30	3	3	3	0	30	2010
Maori & European	5892	1005	123	63	3	30	1782	33	87	48	30	240	9342
Maori & Pacific	108	111	42	0	0	0	39	6	6	3	3	18	333
Pacific & European	486	111	99	3	0	0	87	6	66	6	12	36	912
Asian & European	315	45	9	60	0	3	39	3	6	48	6	18	549
Two groups not elsewhere	378	75	90	27	6	6	60	0	18	3	120	24	804
Not Elsewhere Included	15426	1302	420	654	24	30	432	18	42	24	33	4371	22773
Total	610620	45642	21231	20247	966	2301	11253	330	1170	669	837	15585	730845

A.9 - 1991: All couples

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	512010	14670	2370	3210	294	1383	6732	84	558	345	351	7599	549609
Maori only	15189	26244	663	108	12	123	1668	102	156	60	102	888	45312
Pacific only	2286	1713	1248	18	0	18	321	60	138	21	24	66	5916
Asian Only	1155	141	21	2163	0	6	51	6	6	30	9	33	3621
MELAA only	390	24	3	6	18	0	9	3	3	0	3	0	453
Other ethnicity only	1137	141	12	9	0	105	30	6	0	3	3	12	1455
Maori & European	5883	1005	123	63	6	30	1782	33	90	48	30	237	9324
Maori & Pacific	108	111	36	0	0	0	39	6	6	3	3	15	321
Pacific & European	465	111	54	0	0	0	90	6	33	6	6	24	798
Asian & European	288	48	6	24	0	0	39	0	6	9	6	15	438
Two groups not elsewhere	321	75	9	9	0	3	57	3	9	0	15	12	513
Not Elsewhere Included	12810	1296	69	51	0	18	423	15	27	21	12	84	14823
Total	552042	45579	4605	5661	333	1683	11238	318	1032	543	561	8985	632580

A.10 - 1991: At least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	74349	3207	432	555	72	186	1587	30	183	99	69	1206	81972
Maori only	4407	6579	300	24	3	30	606	48	90	36	33	210	12369
Pacific only	696	654	750	12	0	6	138	33	75	12	12	39	2427
Asian Only	282	45	6	390	0	3	18	3	3	9	3	6	759
MELAA only	102	6	0	0	0	0	3	3	3	0	0	0	117
Other ethnicity only	180	33	3	3	3	15	12	3	3	3	0	0	246
Maori & European	1539	306	51	21	0	9	603	18	39	24	9	66	2679
Maori & Pacific	57	51	21	0	0	0	18	3	3	0	3	9	162
Pacific & European	213	63	33	3	0	0	45	3	21	6	3	18	408
Asian & European	84	15	3	3	0	0	12	0	3	0	3	3	132
Two groups not elsewhere	78	24	6	3	0	0	21	3	3	0	3	6	144
Not Elsewhere Included	1938	345	30	15	0	6	135	6	12	6	6	33	2526
Total	83925	11319	1635	1038	78	249	3189	150	432	189	141	1596	103944

A.11 - 1991: Couples with male partner aged 18-30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	**Maori only**	**Pacific only**	**Asian Only**	**MELAA only**	**Other ethnicity only**	**Maori & European**	**Maori & Pacific**	**Pacific & European**	**Asian & European**	**Two groups not elsewhere**	**Not Elsewhere Included**	
European only	437664	11463	1938	2652	225	1197	5145	57	375	249	282	6393	467634
Maori only	10782	19665	363	84	6	90	1059	57	66	21	69	675	32943
Pacific only	1593	1059	498	6	0	12	186	27	63	9	9	24	3486
Asian Only	876	99	12	1770	0	6	33	3	3	21	6	30	2862
MELAA only	288	18	0	3	15	0	3	0	0	0	3	0	336
Other ethnicity only	957	108	9	6	0	90	18	3	0	0	0	9	1209
Maori & European	4344	699	72	42	3	21	1176	15	48	24	24	174	6642
Maori & Pacific	51	63	15	0	0	0	18	3	3	3	0	6	159
Pacific & European	255	48	24	0	0	0	42	3	12	0	3	6	390
Asian & European	201	30	0	18	0	0	30	0	3	9	3	9	309
Two groups not elsewhere	243	54	3	6	3	0	39	0	6	0	12	9	369
Not Elsewhere Included	10872	951	36	33	3	12	291	9	15	12	9	54	12294
Total	468120	34260	2967	4626	255	1434	8046	171	597	351	420	7389	528636

A.12 - 1991: Couples with male partner aged greater than 30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	546663	11163	2862	5712	495	4389	14058	150	1383	900	2643	15381	605802
Maori only	12531	20763	543	105	6	504	3264	255	231	141	216	1953	40509
Pacific only	2403	1413	15681	60	6	363	546	162	537	81	270	915	22434
Asian Only	1455	129	225	25476	15	237	75	3	18	81	159	702	28572
MELAA only	579	18	15	57	1329	18	15	0	6	0	48	66	2154
Other ethnicity only	4686	597	372	249	15	1428	267	15	42	12	63	270	8019
Maori & European	13857	2241	216	129	18	162	3417	69	198	153	126	1290	21876
Maori & Pacific	210	291	102	3	0	12	114	51	18	15	9	51	873
Pacific & European	1257	183	414	24	0	18	255	12	276	36	36	213	2730
Asian & European	783	99	45	105	3	21	129	15	27	147	12	144	1521
Two groups not elsewhere	2943	219	291	210	30	51	171	9	63	18	1425	378	5808
Not Elsewhere Included	23016	2070	867	1941	72	327	1533	60	228	153	378	6852	37503
Total	610383	39189	21633	34065	1986	7536	23844	801	3021	1743	5388	28206	777798

A.13 - 1996: All couples

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	492630	11094	2088	4248	339	3861	14007	147	1263	777	2388	13155	545991
Maori only	12489	20712	537	105	3	495	3261	252	228	141	216	1938	40371
Pacific only	2148	1389	1734	12	0	66	543	138	294	72	81	414	6891
Asian Only	1182	126	24	2319	0	30	72	6	9	36	39	105	3948
MELAA only	435	18	3	6	12	3	15	0	3	0	9	18	519
Other ethnicity only	4200	582	66	39	3	366	261	12	30	9	48	186	5808
Maori & European	13806	2238	213	126	18	162	3414	69	195	153	126	1281	21801
Maori & Pacific	210	294	66	3	0	9	114	33	15	15	9	48	810
Pacific & European	1200	183	153	18	0	15	255	12	78	33	21	150	2118
Asian & European	717	99	21	54	0	15	129	15	15	42	9	117	1230
Two groups not elsewhere	2748	219	57	45	6	42	171	6	36	18	1017	276	4638
Not Elsewhere Included	19518	2049	255	150	9	207	1521	60	171	120	279	2427	26769
Total	551280	39006	5208	7125	390	5259	23763	750	2337	1419	4242	20109	660891

A.14 - 1996: At least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	62013	1797	297	582	69	543	3141	33	315	198	366	2106	71463
Maori only	2718	4383	189	24	0	108	1083	99	108	72	57	468	9306
Pacific only	606	486	861	9	0	30	234	72	162	42	39	192	2730
Asian Only	231	33	9	282	0	6	21	0	6	12	9	27	633
MELAA only	120	0	3	0	0	3	6	0	3	0	3	6	141
Other ethnicity only	747	171	24	9	3	30	81	9	12	3	6	45	1143
Maori & European	3297	624	81	36	6	51	1167	33	87	57	36	408	5886
Maori & Pacific	69	96	36	3	0	3	54	12	12	6	6	21	321
Pacific & European	375	87	78	3	0	6	111	9	30	18	9	60	783
Asian & European	204	33	9	15	0	3	60	6	9	9	3	39	396
Two groups not elsewhere	549	54	27	9	0	9	54	0	18	6	123	54	900
Not Elsewhere Included	3063	519	96	27	0	45	402	18	66	39	48	513	4833
Total	73995	8283	1710	993	78	840	6417	291	822	465	705	3930	98529

A.15 – 1996: Couples with male partner aged 18-30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner												Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Not Elsewhere Included	
European only	430617	9294	1791	3666	270	3315	10866	114	948	579	2025	11046	474531
Maori only	9771	16329	345	84	3	387	2178	153	120	69	159	1470	31068
Pacific only	1542	906	873	6	0	36	306	63	132	33	42	225	4161
Asian Only	951	96	18	2037	0	24	54	3	6	27	33	78	3315
MELAA only	315	18	0	3	12	0	9	0	0	3	6	12	378
Other ethnicity only	3453	411	39	33	3	333	180	6	15	6	42	141	4665
Maori & European	10512	1614	129	90	12	111	2247	39	111	93	87	873	15918
Maori & Pacific	138	195	30	0	0	3	60	21	3	6	6	27	489
Pacific & European	822	96	75	12	0	9	141	3	51	18	15	90	1335
Asian & European	513	66	9	39	3	9	66	9	6	30	6	75	831
Two groups not elsewhere	2196	165	27	39	6	36	117	6	18	12	894	225	3738
Not Elsewhere Included	16455	1530	159	126	9	162	1119	42	108	84	231	1914	21936
Total	477282	30717	3498	6132	312	4422	17346	459	1518	960	3534	16179	562362

A.16 - 1996: Couples with male partner aged greater than 30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	549300	11214	3258	7821	696	5355	14835	144	1278	603	660	291	8322	603777
Maori only	13878	22272	783	153	27	516	3537	264	276	39	123	126	663	42654
Pacific only	3048	1866	19809	102	12	261	885	207	510	21	159	87	348	27312
Asian Only	1770	138	276	37107	30	270	90	6	24	84	63	3	612	40473
MELAA only	930	33	45	132	2388	39	45	0	9	3	30	0	60	3717
Other ethnicity only	6453	723	336	318	33	2073	369	9	39	9	30	12	138	10539
Maori & European	13146	1890	294	177	12	162	4128	81	297	54	69	114	396	20823
Maori & Pacific	213	261	90	9	0	9	141	21	24	0	12	12	12	801
Pacific & European	1260	204	255	36	6	15	327	15	183	9	15	30	60	2415
Asian & European	498	21	9	81	3	6	42	0	9	147	6	0	12	831
Two groups not elsewhere	708	135	159	78	24	30	99	3	30	6	243	9	36	1563
Maori & Pacific & European	330	87	33	15	0	9	129	9	15	0	9	27	24	687
Not Elsewhere Included	16197	1035	495	1641	120	207	882	24	105	42	54	33	3141	23979
Total	607731	39876	25845	47664	3348	8952	25512	783	2796	1023	1473	747	13824	779580

A.17 - 2001: All couples

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	495615	11169	2481	5904	459	4377	14802	141	1206	522	528	285	6714	544212
Maori only	13845	22224	771	147	24	498	3534	264	276	42	123	126	651	42528
Pacific only	2751	1851	3084	36	3	48	882	195	426	21	72	87	111	9567
Asian Only	1461	138	42	2655	3	27	87	6	15	45	24	3	39	4542
MELAA only	678	33	12	9	24	6	45	0	9	3	9	0	3	828
Other ethnicity only	5568	705	69	87	9	426	366	9	33	6	21	12	51	7356
Maori & European	13101	1890	294	171	12	159	4122	84	294	54	66	114	390	20751
Maori & Pacific	213	261	84	9	0	9	141	18	24	0	12	12	9	789
Pacific & European	1221	204	165	30	6	15	324	15	117	9	9	30	48	2190
Asian & European	444	21	3	45	3	6	42	0	9	39	3	0	9	624
Two groups not elsewhere	585	135	39	24	3	12	99	6	24	3	54	9	24	1017
Maori & Pacific & European	330	87	24	12	0	6	129	9	15	0	6	24	24	675
Not Elsewhere Included	13275	1023	102	87	3	48	873	24	84	27	27	33	105	15708
Total	549090	39738	7176	9210	546	5643	25446	768	2529	774	957	735	8178	650787

A.18 - 2001: At least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	50856	1341	288	666	81	363	3081	33	306	150	78	72	690	57999
Maori only	2349	4020	222	24	9	78	1131	111	126	12	27	51	108	8277
Pacific only	642	537	1494	18	3	18	336	99	210	9	36	39	54	3498
Asian Only	258	42	6	255	0	3	21	3	6	12	3	0	9	615
MELAA only	135	9	3	3	0	0	15	0	0	3	0	0	0	168
Other ethnicity only	654	147	18	15	3	45	99	3	12	0	3	0	9	1008
Maori & European	2940	525	93	45	3	39	1428	33	132	27	15	45	102	5427
Maori & Pacific	72	105	48	3	0	3	75	9	12	0	3	6	9	345
Pacific & European	396	72	81	15	0	6	153	6	45	3	3	12	18	810
Asian & European	141	9	3	15	3	0	18	0	6	3	0	0	6	198
Two groups not elsewhere	120	30	18	9	0	3	21	0	12	3	6	6	9	240
Maori & Pacific & European	111	48	12	9	0	0	60	3	6	0	3	9	9	279
Not Elsewhere Included	1440	195	42	12	0	3	207	9	27	9	3	9	33	1998
Total	60114	7080	2334	1083	102	564	6648	306	903	234	189	258	1053	80865

A.19 - 2001: Couples with male partner aged 18-30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	444759	9828	2196	5235	378	4014	11718	111	900	372	453	216	6027	486213
Maori only	11496	18204	549	123	18	420	2403	153	150	27	96	75	540	34251
Pacific only	2109	1314	1593	15	3	30	546	99	216	9	39	48	63	6072
Asian Only	1206	96	33	2400	3	24	66	3	9	33	18	0	33	3924
MELAA only	543	21	9	9	27	6	27	3	6	3	9	0	3	660
Other ethnicity only	4914	558	51	69	3	384	264	6	21	6	18	12	42	6348
Maori & European	10161	1365	198	129	6	120	2694	51	162	27	51	69	288	15321
Maori & Pacific	141	156	39	3	0	9	66	9	9	0	9	3	3	444
Pacific & European	822	135	84	15	3	12	171	9	72	3	6	21	30	1380
Asian & European	303	12	3	33	0	3	24	0	3	36	3	0	6	426
Two groups not elsewhere	465	105	24	18	3	12	75	3	12	0	48	0	15	780
Maori & Pacific & European	219	39	12	3	0	6	72	6	9	0	3	15	15	393
Not Elsewhere Included	11835	825	60	75	3	45	666	15	57	21	24	21	69	13713
Total	488976	32658	4842	8124	441	5079	18801	465	1629	537	768	480	7125	569922

A.20 - 2001: Couples with male partner aged greater than 30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	571806	11763	3426	11613	1191	3960	17277	183	1440	741	7029	417	9666	640512
Maori only	15072	23184	900	228	24	528	4224	339	333	48	351	153	786	46167
Pacific only	3567	2130	23232	225	18	351	1086	324	609	27	246	150	474	32442
Asian Only	2292	171	381	62067	45	396	135	6	33	78	237	6	927	66774
MELAA only	1089	48	78	210	3912	45	54	3	9	6	60	0	69	5583
Other ethnicity only	5325	759	378	507	54	2187	387	24	42	12	105	21	177	9978
Maori & European	14463	2211	384	321	33	171	4797	105	348	75	273	174	510	23862
Maori & Pacific	270	351	126	3	0	6	201	42	24	3	18	15	21	1083
Pacific & European	1359	201	282	60	12	18	396	21	180	12	48	42	81	2712
Asian & European	453	21	6	108	6	9	51	0	12	69	18	0	15	768
Two groups not elsewhere	6867	315	222	360	51	66	366	6	57	12	2766	18	420	11523
Maori & Pacific & European	435	120	39	15	0	9	219	9	24	3	18	27	33	954
Not Elsewhere Included	16125	1077	582	2055	126	165	1023	36	105	33	495	45	3402	25272
Total	639117	42354	30030	77775	5475	7908	30216	1089	3216	1116	11667	1068	16584	867618

A.21 - 2006: All couples

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	504654	11697	2625	8589	774	3234	17205	183	1371	633	6585	411	7788	565755
Maori only	15003	23106	885	219	24	522	4215	333	333	45	348	153	780	45972
Pacific only	3231	2103	4164	75	9	78	1074	309	534	24	180	144	234	12153
Asian Only	1761	168	60	2766	3	15	132	6	21	21	123	6	54	5145
MELAA only	711	48	27	9	24	3	54	3	6	3	24	3	3	912
Other ethnicity only	4491	735	90	69	3	312	378	24	39	9	90	21	69	6327
Maori & European	14409	2208	375	309	36	165	4791	105	345	75	267	174	504	23766
Maori & Pacific	267	351	120	0	0	9	198	39	24	3	18	15	21	1068
Pacific & European	1320	201	204	51	12	18	396	21	129	9	42	42	69	2508
Asian & European	405	21	3	51	3	6	54	0	9	9	18	0	12	588
Two groups not elsewhere	6513	312	126	240	24	54	363	6	48	9	2262	21	378	10353
Maori & Pacific & European	429	117	33	12	0	6	222	9	24	3	18	24	30	930
Not Elsewhere Included	13101	1047	177	102	3	60	1008	36	87	21	438	45	228	16350
Total	566295	42120	8892	12495	915	4473	30090	1074	2973	864	10410	1056	10167	691827

A.22 - 2006: At least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	47328	1317	231	1119	147	315	3345	39	294	222	594	114	756	55824
Maori only	2403	3864	225	42	6	78	1191	135	120	21	60	48	132	8325
Pacific only	690	537	1647	39	3	27	372	144	240	9	69	69	84	3933
Asian Only	402	63	24	276	3	6	57	3	12	9	21	3	15	891
MELAA only	135	15	9	3	6	0	18	0	0	0	6	0	0	192
Other ethnicity only	528	132	30	6	0	21	99	9	15	3	6	9	9	867
Maori & European	3093	567	90	72	9	39	1542	42	117	42	57	72	126	5868
Maori & Pacific	99	147	60	3	0	3	108	21	15	3	6	12	15	489
Pacific & European	399	63	87	21	3	9	150	12	54	3	15	24	24	867
Asian & European	126	6	3	24	0	0	24	0	3	3	9	0	6	204
Two groups not elsewhere	696	57	30	39	9	3	84	0	12	3	177	9	57	1176
Maori & Pacific & European	132	42	15	6	0	3	111	6	9	3	9	6	9	351
Not Elsewhere Included	1194	189	54	24	0	6	210	12	30	9	63	21	48	1866
Total	57225	6993	2508	1677	189	510	7311	432	924	327	1095	384	1275	80850

A.23 - 2006: Couples with male partner aged 18-30, at least one partner born in New Zealand

Ethnicity of Male Partner	Ethnicity of Female Partner													Total
	European only	Maori only	Pacific only	Asian Only	MELAA only	Other ethnicity only	Maori & European	Maori & Pacific	Pacific & European	Asian & European	Two groups not elsewhere	Maori & Pacific & European	Not Elsewhere Included	
European only	457326	10380	2391	7470	627	2919	13863	141	1080	411	5988	300	7029	509928
Maori only	12600	19242	660	174	18	441	3027	198	213	24	291	105	651	37644
Pacific only	2541	1569	2520	36	3	48	702	165	291	12	111	72	150	8223
Asian Only	1359	108	36	2490	3	9	75	3	12	12	102	6	39	4251
MELAA only	579	33	18	6	21	0	33	0	6	0	18	0	3	720
Other ethnicity only	3963	603	60	66	0	288	279	12	24	3	81	12	60	5460
Maori & European	11313	1644	285	237	27	126	3252	60	231	33	213	102	381	17898
Maori & Pacific	171	204	60	0	3	3	90	15	9	0	12	3	9	582
Pacific & European	921	138	114	30	6	9	243	9	78	6	27	18	45	1647
Asian & European	279	15	0	27	3	6	30	0	6	6	6	0	3	384
Two groups not elsewhere	5817	258	96	198	15	51	279	6	33	6	2082	9	321	9174
Maori & Pacific & European	300	75	18	6	0	6	111	3	15	0	9	18	18	582
Not Elsewhere Included	11904	861	123	75	0	54	798	21	57	9	375	27	180	14487
Total	509073	35124	6387	10821	729	3966	22779	639	2046	534	9315	672	8892	610977

A.24 - 2006: Couples with male partner aged greater than 30, at least one partner born in New Zealand

B. Appendix B: Statistics Computer Code

B.1. SAS Code

```
******* marcos ***********;
%include 'F:/FWWPmacros.sas';
%include 'F:/additional_formats.sas';
%include 'F:/MoSC Data Sets/MOSC_2006_Formats.sas';
%each_session(action=A);
libname lyndon 'f:\Lyndon_data';

**** get each dataset *******************;
data mosc06;
	set ro.mosc_2006_final_dep;

	* adults living in residence;
	where recode_family_role_code ne '41' and recode_family_role_code ne '42'
	and  dwell_rec_type_code = '1' and recode_family_grp_code ne '55'
	and individual_rec_type_code in ('1','3','5');

	* mixed ethnicity indicator;
	format mixedeth $mixedeth.;
	if ethnic_rand6_grp2_code > "90000" then mixedeth = '0';
	else mixedeth = '1';

	*ethnicity as per stats nz protocol;
	length ethnicity $2;
	*1 = euro only;
	if substr(ethnic_rand6_grp1_code,1,1) = '1' and mixedeth='0' then ethnicity = '01';
	else if substr(ethnic_rand6_grp1_code,1,1) = '1' and
substr(ethnic_rand6_grp2_code,1,1) = '1' and ethnic_rand6_grp3_code > '90000'     and
ethnic_rand6_grp4_code > '90000' and ethnic_rand6_grp5_code > '90000' and
ethnic_rand6_grp6_code > '90000' then ethnicity='01';
	else if substr(ethnic_rand6_grp1_code,1,1) = '1' and
substr(ethnic_rand6_grp2_code,1,1) = '1' and substr(ethnic_rand6_grp3_code,1,1) and
ethnic_rand6_grp4_code > '90000' and ethnic_rand6_grp5_code > '90000' and
ethnic_rand6_grp6_code > '90000' then ethnicity='01';
	else if substr(ethnic_rand6_grp1_code,1,1) = '1' and
substr(ethnic_rand6_grp2_code,1,1) = '1' and substr(ethnic_rand6_grp3_code,1,1)='1' and
substr(ethnic_rand6_grp4_code,1,1)='1' and ethnic_rand6_grp5_code > '90000' and
ethnic_rand6_grp6_code > '90000' then ethnicity='01';
	else if substr(ethnic_rand6_grp1_code,1,1) = '1' and
substr(ethnic_rand6_grp2_code,1,1) = '1' and substr(ethnic_rand6_grp3_code,1,1)='1' and
substr(ethnic_rand6_grp4_code,1,1)='1' and substr(ethnic_rand6_grp5_code,1,1)='1' and
ethnic_rand6_grp6_code > '90000' then ethnicity='01';
	else if substr(ethnic_rand6_grp1_code,1,1) = '1' and
substr(ethnic_rand6_grp2_code,1,1) = '1' and substr(ethnic_rand6_grp3_code,1,1)='1' and
substr(ethnic_rand6_grp4_code,1,1)='1' and substr(ethnic_rand6_grp5_code,1,1)='1' and
substr(ethnic_rand6_grp6_code,1,1)='1' then ethnicity='01';
	else if ethnic_rand6_grp1_code = '61118' then ethnicity = '01';

	* 2 = maori only;
	else if ethnic_rand6_grp1_code = '21111' and mixedeth='0' then ethnicity = '02';
	else if ethnic_rand6_grp1_code = '21111' and ethnic_rand6_grp2_code = '21111' and
ethnic_rand6_grp3_code > '90000' and ethnic_rand6_grp4_code > '90000' and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000' then ethnicity='02';
	else if ethnic_rand6_grp1_code = '21111' and ethnic_rand6_grp2_code = '21111' and
ethnic_rand6_grp3_code = '21111' and ethnic_rand6_grp4_code > '90000' and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000' then ethnicity='02';
	else if ethnic_rand6_grp1_code = '21111' and ethnic_rand6_grp2_code = '21111' and
ethnic_rand6_grp3_code = '21111' and ethnic_rand6_grp4_code = '21111' and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000' then ethnicity='02';
	else if ethnic_rand6_grp1_code = '21111' and ethnic_rand6_grp2_code = '21111' and
ethnic_rand6_grp3_code = '21111' and ethnic_rand6_grp4_code = '21111' and
ethnic_rand6_grp5_code = '21111' and ethnic_rand6_grp6_code > '90000' then ethnicity='02';
```

```
        else if ethnic_rand6_grp1_code = '21111' and ethnic_rand6_grp2_code = '21111' and
ethnic_rand6_grp3_code = '21111' and ethnic_rand6_grp4_code = '21111' and
ethnic_rand6_grp5_code = '21111' and ethnic_rand6_grp6_code = '21111' then ethnicity='02';

        * 3 = pacific only;
        else if ethnic_rand6_grp1_code >= '30000' and ethnic_rand6_grp1_code < '40000' and
mixedeth='0' then ethnicity = '03';
        else if (ethnic_rand6_grp1_code >= '30000' and ethnic_rand6_grp1_code < '40000') and
(ethnic_rand6_grp2_code >= '30000' and ethnic_rand6_grp2_code < '40000') and
ethnic_rand6_grp3_code > '90000' and ethnic_rand6_grp4_code > '90000' and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000'  then ethnicity =
'03';
        else if (ethnic_rand6_grp1_code >= '30000' and ethnic_rand6_grp1_code < '40000') and
(ethnic_rand6_grp2_code >= '30000' and ethnic_rand6_grp2_code < '40000') and
(ethnic_rand6_grp3_code >= '30000' and ethnic_rand6_grp3_code < '40000') and
ethnic_rand6_grp4_code > '90000' and ethnic_rand6_grp5_code > '90000' and
ethnic_rand6_grp6_code > '90000'  then ethnicity = '03';
        else if (ethnic_rand6_grp1_code >= '30000' and ethnic_rand6_grp1_code < '40000') and
(ethnic_rand6_grp2_code >= '30000' and ethnic_rand6_grp2_code < '40000') and
(ethnic_rand6_grp3_code >= '30000' and ethnic_rand6_grp3_code < '40000') and
(ethnic_rand6_grp4_code >= '30000' and ethnic_rand6_grp4_code < '40000') and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000'  then ethnicity =
'03';
        else if (ethnic_rand6_grp1_code >= '30000' and ethnic_rand6_grp1_code < '40000') and
(ethnic_rand6_grp2_code >= '30000' and ethnic_rand6_grp2_code < '40000') and
(ethnic_rand6_grp3_code >= '30000' and ethnic_rand6_grp3_code < '40000') and
(ethnic_rand6_grp4_code >= '30000' and ethnic_rand6_grp4_code < '40000') and
(ethnic_rand6_grp5_code >= '30000' and ethnic_rand6_grp5_code < '40000') and
ethnic_rand6_grp6_code > '90000'  then ethnicity = '03';
        else if (ethnic_rand6_grp1_code >= '30000' and ethnic_rand6_grp1_code < '40000') and
(ethnic_rand6_grp2_code >= '30000' and ethnic_rand6_grp2_code < '40000') and
(ethnic_rand6_grp3_code >= '30000' and ethnic_rand6_grp3_code < '40000') and
(ethnic_rand6_grp4_code >= '30000' and ethnic_rand6_grp4_code < '40000') and
(ethnic_rand6_grp5_code >= '30000' and ethnic_rand6_grp5_code < '40000') and
(ethnic_rand6_grp6_code >= '30000' and ethnic_rand6_grp6_code < '40000') then ethnicity =
'03';

        * 4 = Asian only;
        else if ethnic_rand6_grp1_code >= '40000' and ethnic_rand6_grp1_code < '50000' and
mixedeth='0' then ethnicity = '04';
        else if (ethnic_rand6_grp1_code >= '40000' and ethnic_rand6_grp1_code < '50000') and
(ethnic_rand6_grp2_code >= '40000' and ethnic_rand6_grp2_code < '50000') and
ethnic_rand6_grp3_code > '90000' and ethnic_rand6_grp4_code > '90000' and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000'  then ethnicity =
'04';
        else if (ethnic_rand6_grp1_code >= '40000' and ethnic_rand6_grp1_code < '50000') and
(ethnic_rand6_grp2_code >= '40000' and ethnic_rand6_grp2_code < '50000') and
(ethnic_rand6_grp3_code >= '40000' and ethnic_rand6_grp3_code < '50000') and
ethnic_rand6_grp4_code > '90000' and ethnic_rand6_grp5_code > '90000' and
ethnic_rand6_grp6_code > '90000'  then ethnicity = '04';
        else if (ethnic_rand6_grp1_code >= '40000' and ethnic_rand6_grp1_code < '50000') and
(ethnic_rand6_grp2_code >= '40000' and ethnic_rand6_grp2_code < '50000') and
(ethnic_rand6_grp3_code >= '40000' and ethnic_rand6_grp3_code < '50000') and
(ethnic_rand6_grp4_code >= '40000' and ethnic_rand6_grp4_code < '50000') and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000'  then ethnicity =
'04';
        else if (ethnic_rand6_grp1_code >= '40000' and ethnic_rand6_grp1_code < '50000') and
(ethnic_rand6_grp2_code >= '40000' and ethnic_rand6_grp2_code < '50000') and
(ethnic_rand6_grp3_code >= '40000' and ethnic_rand6_grp3_code < '50000') and
(ethnic_rand6_grp4_code >= '40000' and ethnic_rand6_grp4_code < '50000') and
(ethnic_rand6_grp5_code >= '40000' and ethnic_rand6_grp5_code < '50000') and
ethnic_rand6_grp6_code > '90000'  then ethnicity = '04';
        else if (ethnic_rand6_grp1_code >= '40000' and ethnic_rand6_grp1_code < '50000') and
(ethnic_rand6_grp2_code >= '40000' and ethnic_rand6_grp2_code < '50000') and
(ethnic_rand6_grp3_code >= '40000' and ethnic_rand6_grp3_code < '50000') and
(ethnic_rand6_grp4_code >= '40000' and ethnic_rand6_grp4_code < '50000') and
(ethnic_rand6_grp5_code >= '40000' and ethnic_rand6_grp5_code < '50000') and
(ethnic_rand6_grp6_code >= '40000' and ethnic_rand6_grp6_code < '50000') then ethnicity =
'04';

        * 5 = MELAA only ;
```

```
	else if ethnic_rand6_grp1_code >= '50000' and ethnic_rand6_grp1_code < '60000' and
mixedeth='0' then ethnicity = '05';
	else if (ethnic_rand6_grp1_code >= '50000' and ethnic_rand6_grp1_code < '60000') and
(ethnic_rand6_grp2_code >= '50000' and ethnic_rand6_grp2_code < '60000') and
ethnic_rand6_grp3_code > '90000' and ethnic_rand6_grp4_code > '90000' and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000'  then ethnicity =
'05';
	else if (ethnic_rand6_grp1_code >= '50000' and ethnic_rand6_grp1_code < '60000') and
(ethnic_rand6_grp2_code >= '50000' and ethnic_rand6_grp2_code < '60000') and
(ethnic_rand6_grp3_code >= '50000' and ethnic_rand6_grp3_code < '60000') and
ethnic_rand6_grp4_code > '90000' and ethnic_rand6_grp5_code > '90000' and
ethnic_rand6_grp6_code > '90000'  then ethnicity = '05';
	else if (ethnic_rand6_grp1_code >= '50000' and ethnic_rand6_grp1_code < '60000') and
(ethnic_rand6_grp2_code >= '50000' and ethnic_rand6_grp2_code < '60000') and
(ethnic_rand6_grp3_code >= '50000' and ethnic_rand6_grp3_code < '60000') and
(ethnic_rand6_grp4_code >= '50000' and ethnic_rand6_grp4_code < '60000') and
ethnic_rand6_grp5_code > '90000' and ethnic_rand6_grp6_code > '90000'  then ethnicity =
'05';
	else if (ethnic_rand6_grp1_code >= '50000' and ethnic_rand6_grp1_code < '60000') and
(ethnic_rand6_grp2_code >= '50000' and ethnic_rand6_grp2_code < '60000') and
(ethnic_rand6_grp3_code >= '50000' and ethnic_rand6_grp3_code < '60000') and
(ethnic_rand6_grp4_code >= '50000' and ethnic_rand6_grp4_code < '60000') and
(ethnic_rand6_grp5_code >= '50000' and ethnic_rand6_grp5_code < '60000') and
ethnic_rand6_grp6_code > '90000'  then ethnicity = '05';
	else if (ethnic_rand6_grp1_code >= '50000' and ethnic_rand6_grp1_code < '60000') and
(ethnic_rand6_grp2_code >= '50000' and ethnic_rand6_grp2_code < '60000') and
(ethnic_rand6_grp3_code >= '50000' and ethnic_rand6_grp3_code < '60000') and
(ethnic_rand6_grp4_code >= '50000' and ethnic_rand6_grp4_code < '60000') and
(ethnic_rand6_grp5_code >= '50000' and ethnic_rand6_grp5_code < '60000') and
(ethnic_rand6_grp6_code >= '50000' and ethnic_rand6_grp6_code < '60000') then ethnicity =
'05';

	* 6 = other single ethnicity;
	else if mixedeth = '0' then ethnicity = '06';
	*'12' = 'Maori and Pacific and European';
	else if maori_ethnic_ind_code = '2' and european_ethnic_ind_code = '2' and
pacific_island_ethnic_ind_code = '2' then ethnicity = '12';
	*'13' = '3 or more groups';
	else if ethnic_rand6_grp3_code ^= '99999' then ethnicity = '13';
	* 7 = Maori & European;
	else if maori_ethnic_ind_code = '2' and european_ethnic_ind_code = '2' then
ethnicity = '07';
	* '8' = 'Maori and Pacific';
	else if maori_ethnic_ind_code = '2' and pacific_island_ethnic_ind_code = '2' then
ethnicity = '08';
	*'9' = 'Pacific and European';
	else if pacific_island_ethnic_ind_code = '2' and european_ethnic_ind_code = '2' then
ethnicity = '09';
	*'10' = 'Asian and European';
	else if asian_ethnic_ind_code = '2' and european_ethnic_ind_code = '2' then
ethnicity = '10';
	*'11' = 'Two groups not elsewhere';
	else if ethnic_rand6_grp3_code > '90000' then ethnicity = '11';

	*'14" = other;
	else ethnicity = '14';
	format ethnicity $ethnicity.;

	age=age_code;
	*age codes;
	if age_code > '017' and age_code < '031' then age1830 = '1';
	else age1830 = '0';
	if age_code > '019' and age_code < '035' then age2034 = '1';
	else age2034 = '0';
	if age_code > '022' and age_code < '036' then age2335 = '1';
	else age2335 = '0';

	* immigrant for consistency using birthplace rather than year in nz;
	if birth_country_code2d = 12 then immigrant = '0';
	else immigrant = '1';
```

```
        * north island;
        if URRegC06 < '10' then north = '1';
        else if URRegC06 > '10' and URRegC06 < '20' then north = '0';

        * auckland indicator;
        if URRegC06 = '02' then auckland_ind = '1';
        else auckland_ind = '0';

        * nzsco99 level 1;
        length nzsco991 $2;
        if nzsco99 < '200' then nzsco991 = '01';
        else if nzsco99 < '300' then nzsco991 = '02';
        else if nzsco99 < '400' then nzsco991 = '03';
        else if nzsco99 < '500' then nzsco991 = '04';
        else if nzsco99 < '600' then nzsco991 = '05';
        else if nzsco99 < '700' then nzsco991 = '06';
        else if nzsco99 < '800' then nzsco991 = '07';
        else if nzsco99 < '900' then nzsco991 = '08';
        else if nzsco99 < '950' then nzsco991 = '09';
        else nzsco991 = '10';
        format nzsco991 $nzsco.;

        * pacific indicators;
        if ethnic_rand6_grp1_code = '31111' or ethnic_rand6_grp2_code = '31111' or
ethnic_rand6_grp3_code = '31111'
        or ethnic_rand6_grp4_code = '31111' or ethnic_rand6_grp5_code = '31111' or
ethnic_rand6_grp6_code = '31111' then samoan_ind_code = '1';
        else samoan_ind_code = '0';
        if ethnic_rand6_grp1_code = '32100' or ethnic_rand6_grp2_code = '32100' or
ethnic_rand6_grp3_code = '32100'
        or ethnic_rand6_grp4_code = '32100' or ethnic_rand6_grp5_code = '32100' or
ethnic_rand6_grp6_code = '32100' then cook_ind_code = '1';
        else cook_ind_code = '0';
        if ethnic_rand6_grp1_code = '33111' or ethnic_rand6_grp2_code = '33111' or
ethnic_rand6_grp3_code = '33111'
        or ethnic_rand6_grp4_code = '33111' or ethnic_rand6_grp5_code = '33111' or
ethnic_rand6_grp6_code = '33111' then tongan_ind_code = '1';
        else tongan_ind_code = '0';

        *highest qualification;
        length highestqual $1;
        if highest_qual_code = '00' then highestqual = '1';
        else if highest_qual_code in ('01', '02', '03', '04') then highestqual = '2';
        else if highest_qual_code in ('05','06', '07', '08', '09','10') then highestqual =
'3';
        else if highest_qual_code in ('11', '12', '13','14') then highestqual = '4';
        else highestqual='5';
        format highestqual $highestqual.;
run;

******************* match 2006 couples ******************************************;
proc sort data=lyndon.mosc06 out=couples06;
by id_family;
run;

data couples06_A;
        set couples06;
        by id_family;

        *delete those under 16;
        if age_code < '016' then delete;

        *adults who live at the residence;
        where dwell_rec_type_code = '1' and recode_family_grp_code ne '55'
        and recode_family_grp_code ne '54'
        and individual_rec_type_code in ('1','3','5')
        and recode_family_role_code in ('11','31');

        * delete singles;
        if (first.id_family=1 and last.id_family=1) then delete;
run;
```

```
*rename variables for female dataset;
data fem_couples06;
        set couples06_A;
        where sex_code='2';

        fprioreth = prioreth;  fmixedeth = mixedeth;
        fethnicity = ethnicity; format fethnicity $ethnicity.;
        fage1830 = age1830; fage2034 = age2034;fage2335 = age2335; fimmigrant = immigrant;
        fnorth = north; fnzsco991 = nzsco99;fauckland_ind = auckland_ind;   fpct_col =
        pct_col; fhighestqual = highestqual;  format fhighestqual $highestqual.;
fCNRegC06=CNRegC06; fCNTA06=CNTA06; fId_Dwell=Id_Dwell; fId_ExtFamily=Id_ExtFamily;
fId_Family=Id_Family;fId_Person=Id_Person; fId_UR_Area_Unit=Id_UR_Area_Unit;
fNZDep2006=NZDep2006; fNZDep_score_2006=NZDep_score_2006;
fSQKM=SQKM; fURRegC06=URRegC06; fURTA06=URTA06; fage_code=age_code;
fage_monitor_code=age_monitor_code; fanzsco2006=anzsco2006; fanzsic96=anzsic96;
fanzsic2006=anzsic2006; fasian_ethnic_ind_code=asian_ethnic_ind_code;
fbedroom_count_code=bedroom_count_code; fbirth_country_code2d=birth_country_code2d;
fcensus_year=census_year; fchild_depend_code=child_depend_code;
fchild_depend_family_type_code=child_depend_family_type_code;
fdwell_rec_type_code=dwell_rec_type_code; fdwell_substitute_code=dwell_substitute_code;
fdwell_type_code=dwell_type_code; femp_status_code=emp_status_code;
fethnic_rand6_grp1_code=ethnic_rand6_grp1_code;
fethnic_rand6_grp2_code=ethnic_rand6_grp2_code;
fethnic_rand6_grp3_code=ethnic_rand6_grp3_code;
fethnic_rand6_grp4_code=ethnic_rand6_grp4_code;
fethnic_rand6_grp5_code=ethnic_rand6_grp5_code;
fethnic_rand6_grp6_code=ethnic_rand6_grp6_code;
feuropean_ethnic_ind_code=european_ethnic_ind_code;
fextended_family_type_code=extended_family_type_code; ffamily_grp_code=family_grp_code;
ffamily_role_code=family_role_code; ffamily_type_code=family_type_code;
ffertility_code=fertility_code; fheat_fuel=heat_fuel; fhhld_composn_code=hhld_composn_code;
fhighest_qual_code=highest_qual_code; fincome_srce10_code=income_srce10_code;
fincome_srce10_ext_family_code=income_srce10_ext_family_code;
fincome_srce10_family_code=income_srce10_family_code;
fincome_srce10_hhld_code=income_srce10_hhld_code; fincome_srce11_code=income_srce11_code;

drop highestqual pct_col nzsco991 auckland_ind samoan_ind_code cook_ind_code
tongan_ind_code north ethnicity immigrant age1830 age2034 mixedeth prioreth  age2335
CNRegC06 CNTA06 Id_Dwell Id_ExtFamily Id_Person Id_UR_Area_Unit NZDep2006 NZDep_score_2006
SQKM URRegC06 URTA06 age_code age_monitor_code anzsco2006 anzsic96 anzsic2006
asian_ethnic_ind_code bedroom_count_code birth_country_code2d census_year child_depend_code
child_depend_family_type_code dwell_rec_type_code dwell_substitute_code dwell_type_code
emp_status_code ethnic_rand6_grp1_code ethnic_rand6_grp2_code ethnic_rand6_grp3_code
ethnic_rand6_grp4_code ethnic_rand6_grp5_code ethnic_rand6_grp6_code
european_ethnic_ind_code extended_family_type_code family_grp_code family_role_code
family_type_code fertility_code heat_fuel hhld_composn_code highest_qual_code
income_srce10_code income_srce10_ext_family_code income_srce10_family_code
income_srce10_hhld_code income_srce11_code income_srce11_ext_family_code
income_srce11_family_code income_srce11_hhld_code income_srce12_code
income_srce12_ext_family_code income_srce12_family_code income_srce12_hhld_code
income_srce13_code income_srce13_ext_family_code income_srce13_family_code
income_srce13_hhld_code income_srce14_code income_srce1_code income_srce1_ext_family_code
income_srce1_family_code income_srce1_hhld_code income_srce2_code
income_srce2_ext_family_code income_srce2_family_code income_srce2_hhld_code
income_srce3_code income_srce3_ext_family_code income_srce3_family_code
income_srce3_hhld_code income_srce4_code income_srce4_ext_family_code
income_srce4_family_code income_srce4_hhld_code income_srce5_code
income_srce5_ext_family_code income_srce5_family_code income_srce5_hhld_code
income_srce6_code income_srce6_ext_family_code income_srce6_family_code
income_srce6_hhld_code income_srce7_code income_srce7_ext_family_code
income_srce7_family_code income_srce7_hhld_code income_srce8_code
income_srce8_ext_family_code income_srce8_family_code income_srce8_hhld_code
income_srce9_code income_srce9_ext_family_code income_srce9_family_code
income_srce9_hhld_code income_srce_count_code income_srce_family_count_code
income_srce_hhld_count_code income_support_count_code income_support_family_code
income_support_hhld_count_code individual_rec_type_code individual_substitute_code
job_ind_code landlord_code language1_code language2_code language3_code language4_code
language5_code language6_code languages_count_code legal_marital_status_recode
maori_ethnic_ind_code melaa_ethnic_ind_code motor_vehicle_count_code nzsco99 nzsei96
official_language_code other_ethnic_ind_code pacific_island_ethnic_ind_code
```

```
post_school_qual_code2d prior_eth recode_family_grp_code recode_family_role_code
recode_maori_descent_code related_family_grp_code related_family_role_code religion1_code
religion2_code religion3_code religion4_code scndry_school_qual_code sex_code
sex_monitor_code smoke_ever_code smoke_regular_code smoking_status_code
social_marital_status_recode std_highest_qual_code telecomm1_code telecomm2_code
telecomm3_code telecomm4_code telecomm5_code tenure_code tenure_holder_code
total_income_code total_income_family_code total_income_hhld_code total_work_hrs_code
usual_resdnt_code usual_resdnt_count_code usual_resdnt_monitor_code visitor_only_dwell_code
weekly_rent_code wklfs_code wklfs_monitor_code years_at_addr_code years_in_nz_code ;

run;
data male_couples06;
	set couples06_A;
	where sex_code='1';
	run;

proc sort data=male_couples06;
	by id_family;
run;
proc sort data=fem_couples06;
	by id_family;
run;
data lyndon.mergedcouples06;
	merge male_couples06 fem_couples06;
	by id_family;

	* partners of same ethnicity;
	if ethnicity = fethnicity then sameeth = '1';
	else sameeth = '0';

	*mixed couple indicator;
	mixed = 1 - sameeth;

	*delete same sex couples;
	if sex_code='1' and fsex_code='2' then het='1';
	else delete;

	* both immigrants;
	if immigrant =1 and fimmigrant = 1 then immcoup =1;
	else immcoup=0;

	*age difference;
	length agediffcat $1;
	agediff = age_code - fage_code;

	if	asian_ethnic_ind_code = 1 then lasian_ethnic_ind_code =1;
	else lasian_ethnic_ind_code =0;
	if	fasian_ethnic_ind_code = 1 then lfasian_ethnic_ind_code =1;
	else lfasian_ethnic_ind_code =0;
	if european_ethnic_ind_code   = 1 then leuropean_ethnic_ind_code = 1;
	else leuropean_ethnic_ind_code = 0;
	if feuropean_ethnic_ind_code  = 1 then lfeuropean_ethnic_ind_code = 1;
	else lfeuropean_ethnic_ind_code = 0;
	if maori_ethnic_ind_code = 1 then lmaori_ethnic_ind_code = 1;
	else lmaori_ethnic_ind_code=0;
	if fmaori_ethnic_ind_code = 1 then lfmaori_ethnic_ind_code = 1;
	else lfmaori_ethnic_ind_code=0;
	if pacific_island_ethnic_ind_code = 1 then lpacific_island_ethnic_ind_code=1;
	else lpacific_island_ethnic_ind_code=0;
	if fpacific_island_ethnic_ind_code = 1 then lfpacific_island_ethnic_ind_code=1;
	else lfpacific_island_ethnic_ind_code=0;

	*married;
	if legal_marital_status_recode = '2' or legal_marital_status_recode='3' then married
= '1';
	else married='0';

run;
```

B.2. R Code

```
loglinear.fit = function(list.table, interaction.fit=FALSE, qi.fit=FALSE, qsim.fit=FALSE,
qcp.fit=FALSE, time=FALSE){

# table must be a square matrix of counts or list of tables (time)
# fits diagonal dominance model (quasi independence) or quasi symmetry + time variable
# Uses: library(catspec)

if (!is.list(list.table)) list.table=list(list.table)

for (i in 1:length(list.table)) {
	if(!is.matrix(list.table[[i]]))stop("Input must be a matrix")
	if(dim(list.table[[i]])[1] != dim(list.table[[i]])[2]) stop("Input must be a square
matrix")
	if(all(round(list.table[[i]]) != list.table[[i]])) stop("Input must be a square
matrix of integers")
	if(!all(list.table[[i]]>= 0)) stop("Input must be a square matrix of non-negative
integers")
}

if (!time) {
	i=1
	p=dim(list.table[[i]])[1]
	row.var = as.factor(rep(dimnames(list.table[[i]])[[2]],p))
	col.var = as.factor(rep(dimnames(list.table[[i]])[[2]],rep(p,p)))
	counts = as.vector(list.table[[i]])

	model<-counts~row.var+col.var

	# add interaction term to model
	if (interaction.fit==TRUE){
		model<-update(model, .~. + row.var:col.var)}
	# add quasi-independence term to model
	if (qi.fit==TRUE){
		model<-update(model, .~. + mob.qi(row.var,col.var))}
	# add quasi-symmetry term to model
	if (qsim.fit==TRUE){
		model<-update(model, .~. + mob.symint(row.var,col.var))}
}

else if (time==TRUE) {
	row.var = NULL
	col.var = NULL
	time.var = NULL
	counts =NULL

	for (i in 1:length(list.table)) {
		p=dim(list.table[[i]])[1]

		row.var = as.factor(c(row.var,rep(dimnames(list.table[[i]])[[2]],p)))
		col.var = as.factor(c(col.var,rep(dimnames(list.table[[i]])[[2]],rep(p,p))))
		counts = as.vector(c(counts,(list.table[[i]])))
		time.var = as.factor(c(time.var, rep(i,p*p)))
	}

	model<-counts~
row.var:col.var+row.var:time.var+col.var:time.var+time.var:mob.qi(row.var,col.var)
}
glm(model, family=poisson)
}

# example
data.df.red.06<-read.table(file.choose(),header=T, sep="\t")
my.table.red.06 = as.matrix(data.df.red.06[,-1])
library(catspec)
fit.red.06.qi = loglinear.fit(my.table.red.06, qi.fit=TRUE)
summary(fit.red.06.qi)

exp(cbind(fit.red.81.qi$coefficients,fit.red.86.qi$coefficients,fit.red.91.qi$coefficients,
fit.red.96.qi$coefficients, fit.red.01.qi$coefficients,fit.red.06.qi$coefficients))

# deviance residuals
round(matrix(residuals(fit.red.06.qi),ncol=9), digits=2)
```

C. Appendix C: Simulation Code

C.1. Netlogo Abstract Simulation Code

```
globals [
rank_ave_m ; average rank of partnered men
rank_ave_w  ; average rank of partnered women
rank_t_m  ; sum of all rankings for partnered men
rank_t_w  ; sum of all rankings for partnered women
num_pair ; number of matches
list_of_men ; list of ids for men
list_of_women  ; list of ids for women
num_proposers ; number of men proposing to women in a given round
ave_diff ; average difference in education
]

turtles-own [
education ; years
list_f ; list of opposite sex potential partners in social network
list_p ; list of proposer (who has proposed to this turtle?)
available?
partner
rank ; rank of partner in turtle's list
rank_propose ; rank of turtle in partner's list
]

to setup
    ca
    ask patches [ set pcolor scale-color grey 5 0 10 ]
    crt numbermen + numberwomen ; create equal number of men and women


    ask turtles [

      ifelse population = "Random Uniform" [
        set education random levels ; random number between 0 and levels
      ] []

      ifelse population = "Normal" [
        set education round random-normal ((levels - 1) / 2)(((levels - 1) / 2) / 3);
      ] []

      ifelse population = "Right Skew" [
        set education round random-exponential ( levels / 6 )
      ][]

      ifelse population = "Left Skew" [
        set education round ( levels - random-exponential (levels / 6 ))
      ][]

      ifelse who < numbermen ; scale color so that men are blue, women are red, darker = 
higher education
        [set shape "person"]
        [set shape "woman" ]
      ifelse who < numbermen ; scale color so that men are blue, women are red, darker = 
higher education
        [ set color scale-color blue education levels 0 ]
        [set color scale-color red education levels 0 ]
    ]

    set list_of_men turtles with [who < numbermen]
    set list_of_women turtles with [ who >= numbermen ]
    ask turtles [
      setxy random-int-or-float world-width random-int-or-float world-height
      set available? true
      set xcor pxcor
```

```
      set ycor pycor
    ]
    ask turtles [ while [(count turtles-here) > 1] [ fd 1 set xcor pxcor set ycor pycor ]]
    ask turtles [setup-partner-list]
    set num_pair 0
    set rank_t_m 0
    set rank_t_w 0
end

to setup-partner-list
    locals [w temp_set similar_list diff pos t]
    ifelse neighborhood [
    ifelse (who < numbermen)
      [set temp_set (list_of_women in-radius-nowrap radius)]
      [set temp_set (list_of_men in-radius-nowrap radius)]
    ]
    [
    ifelse (who < numbermen)
        [set temp_set list_of_women ] ; men
        [set temp_set list_of_men] ; women
    ]
    set list_f []

    if ranking = "random"[
          repeat (count temp_set)
          [
            set w one-of temp_set  ; man or woman with highest education
            set list_f lput w list_f ; put w on list of possible parters
            set temp_set temp_set with [who !=  value-from w [who] ] ; remove w from
temp_set
          ]
    ]

    if ranking = "highest edu" [
      repeat (count temp_set)
      [
      set w max-one-of temp_set [ education ] ; man or woman with highest education
      set list_f lput w list_f ; put w on list of possible parters
      set temp_set temp_set with [who !=  value-from w [who] ] ; remove w from temp_set
      ]
   ]

   if ranking = "most similar" [
   set similar_list []
   set pos 0
   set temp_set values-from temp_set [self] ; turn agentset into a list
     repeat (length temp_set) ; compute difference between education of self and agents of
opposite sex
     [
       set w item pos temp_set ; take a person in list
       set diff abs((value-from w [education]) - (value-from self [education])) ; compute
absolute value of difference in education
       set similar_list lput diff similar_list
       set pos (pos + 1)
     ]
     repeat (length temp_set)
     [
       set diff min similar_list ; lowest value from list of differences
       set pos position diff similar_list ; position of lowest value
       set w item pos temp_set ; get person with most similar education
       set list_f lput w list_f ; put w on list of possible partners
       set temp_set remove w temp_set ; remove w from list of men or women
       set similar_list remove-item pos similar_list ; remove this difference in educations
from list of differences
     ]
   ]
end

to go
    locals [min_rank select_m select_w num w rank_m rank_w diff tot_diff]
    set num_proposers 0
```

```
    ask turtles with [available? = true] [set list_p []]
    ask turtles with [available? = true and length list_f > 0] [propose]
    if (num_proposers = 0) [stop]
    ask turtles with [available? = true and length list_p > 0] [evaluate]
    set min_rank (numbermen + numberwomen)

    ask turtles
    [
        if (available? and length list_p > 0)
        [
            set w rank + rank_propose ; sum of my rank of my partner and partner's rank of
me
            if (w < min_rank)
            [
                set min_rank w ; our joint rank is the new best rank
                ifelse who < numbermen
                    [set select_m self   ; info if turtle is male
                    set select_w partner
                    set rank_m rank
                    set rank_w rank_propose]

                    [set select_w self   ; info if turtle is female
                    set select_m partner
                    set rank_w rank
                    set rank_m rank_propose]
            ]
        ]
    ]


    if (min_rank < (numbermen + numberwomen))
    [

        ask select_m
        [
            set available? false
            if heading != (value-from select_w [heading]) [ set heading towards select_w ]
            repeat ((distance select_w))
            [fd 1 wait 0.02] ; slow movement down so user can see what's happening
            while [distance select_w > .2] [ fd .2] ; couple wants to be on same patch but
            while [distance select_w < .15] [fd .1] ; not on top of one another
            set pcolor 48
            set partner select_w
            output-print "Male" + select_m + " Edu " + value-from self [education] + "
Female" + select_w + " Edu " + value-from partner [education]

            ifelse file_output [
              file-open "c:/documents and settings/lwal036/desktop/results.txt"
              file-print "Male" + select_m + " Edu " + value-from self [education] + "
Female" + select_w + " Edu " + value-from partner [education]
              file-close ] [ ]  ;writes turtles education level and partner to results.txt
file on desktop
        ]
        ask select_w ;
        [
            set available? false
            set partner select_m
            set pcolor 48
        ]
        set num_pair num_pair + 1
        set rank_t_m rank_t_m + rank_m
        set rank_t_w rank_t_w + rank_w

    ]

    set tot_diff 0
    ask list_of_men with [available? = false] [
    set diff abs((value-from partner [education]) - (value-from self [education]))
    set tot_diff (tot_diff + diff)
    ]
```

```
;     ask turtles-with available? false
    if num_pair != 0 [set rank_ave_m rank_t_m / num_pair
                      set rank_ave_w rank_t_w / num_pair
                      set ave_diff tot_diff / num_pair]
  ;  type num_pair type "   " type (precision rank_ave_m 1) type "   " print (precision
rank_ave_w 1)
  ; print (precision ave_diff 3) ; average difference in education of partners

    graph

end

to propose
    foreach list_f
    ; loop over list of people proposing to
    [
         if(length list_f = 0)[stop]
         if (available?-of ?) = true
         ; if that person is available
         [
             set (list_p-of ?) lput self (list_p-of ?)
             ; put myself on list of proposers
             set num_proposers num_proposers + 1 ; increase global list num_proposers by 1
             stop ; only one proposal made by each turtle in each round
         ]
    ]
end

to evaluate
    locals [min_rank pos]
    set min_rank numbermen; worst possible ranking
    foreach list_p ; for each person on list of proposers
    [
        set pos position ? list_f ; what is the rank of this proposer?
        if (pos < min_rank) [set min_rank pos set partner ?] ; best ranked proposer becomes
partner
    ]
    set rank min_rank ; best proposer's rank in my list
    set rank_propose position self list_f-of partner  ; my rank on my best proposer's list
end

to graph
   ; set-current-plot "Average Ranking"
   ; set-current-plot-pen "men"
   ; plot rank_ave_m
   ; set-current-plot-pen "women"
   ; plot rank_ave_w
    set-current-plot "Average Educational Difference in Partnership"
    plot ave_diff

end
```

C.2. Java Code

```
//package simulation;

import java.io.*;
import java.util.*;

/***************
 *  * @author lwal036
 * simulate marriage partnerships
 * Usage: Simulation2 input.csv actual.txt weight1 weight2 weight3 weight4
 */

public class Simulation2
{
       // Global variables
       int timesteps = 5;
       final int startpool = 50;
       final int addpool = 10;

       // set lists for each variable
       ArrayList<Integer> sexlist;
       ArrayList<Integer> maleagelist;
       ArrayList<Integer> maleethlist;
       ArrayList<Integer> maleedulist;
       ArrayList<Integer> femaleagelist;
       ArrayList<Integer> femaleethlist;
       ArrayList<Integer> femaleedulist;
       ArrayList<Double> actual;

       // number of single men threads currently running
       static Simulation2 sim1;
       int [] malespartner;
       int [] femalespartner;
       int [][] singlepools;
       double [][] scores;
       double [] weights = new double[4];
       int samecouple = 0;
       int newcouples = 0;
       double couplesperperiod=0;

 /********************************************************************/
       // read in actual couple proportions

   public void readactual(String args)
   {
      actual = new ArrayList<Double>();
      try{
         FileReader input = new FileReader(args);
         BufferedReader bufRead = new BufferedReader(input);

         String line;
         line = bufRead.readLine();

         while (line != null){
            actual.add(Double.parseDouble(line));
            line=bufRead.readLine();
         }
         bufRead.close();
         }
         catch (ArrayIndexOutOfBoundsException e){
            System.err.println("Usage: java ReadFile filename\n");
         }
         catch(IOException e){
                 e.printStackTrace();
         }
   }

/**********************************************************************/
```

```
// count number of couples formed
   public void countnumberofcouples(String actuals)
   {
          double numberofcouples = 0;
          if(actuals.equals("auck86.txt"))  numberofcouples=17805;
          else if(actuals.equals("auck91.txt"))  numberofcouples=23982;
          else if(actuals.equals("auck96.txt"))  numberofcouples=26391;
          else if(actuals.equals("auck01.txt"))  numberofcouples=20088;
          else if(actuals.equals("auck06.txt"))  numberofcouples=29937;
          else if(actuals.equals("well86.txt"))  numberofcouples=7458;
          else if(actuals.equals("well91.txt"))  numberofcouples=9864;
          else if(actuals.equals("well96.txt"))  numberofcouples=8388;
          else if(actuals.equals("well01.txt"))  numberofcouples=6759;
          else if(actuals.equals("well06.txt"))  numberofcouples=9186;
          else if(actuals.equals("cant86.txt"))  numberofcouples=8493;
          else if(actuals.equals("cant91.txt"))  numberofcouples=9054;
          else if(actuals.equals("cant96.txt"))  numberofcouples=10332;
          else if(actuals.equals("cant01.txt"))  numberofcouples=6999;
          else if(actuals.equals("cant06.txt"))  numberofcouples=10611;
          sim1.couplesperperiod=Math.round(numberofcouples/sim1.timesteps);

   }


/****************************************************************************/
// read in census csv file
    public int [] readfile(String args)
    {
       sexlist = new ArrayList<Integer>();
       maleagelist = new ArrayList<Integer>();
       maleethlist = new ArrayList<Integer>();
       maleedulist = new ArrayList<Integer>();
       femaleagelist = new ArrayList<Integer>();
       femaleethlist = new ArrayList<Integer>();
       femaleedulist = new ArrayList<Integer>();
        int malecount = 0;
        int femalecount = 0;
       int [] counts = new int [2];

    // read in csv file (no header, order sex age eth edu)
       try {
              FileReader input = new FileReader(args);
              BufferedReader bufRead = new BufferedReader(input);
               String line;   // String that holds current file line

              // Read first line
               line = bufRead.readLine();

              // Read through file one line at time. Print line # and line
              while (line != null){
                     String[] tokens = line.split(",");
                     sexlist.add(Integer.parseInt(tokens[ 0 ]));
                     if (Integer.parseInt(tokens[0]) == 1){
                            maleagelist.add(Integer.parseInt(tokens[ 1 ]));
                            maleethlist.add(Integer.parseInt(tokens[ 2 ]));
                            maleedulist.add(Integer.parseInt(tokens[ 3 ]));
                     }
                     else {
                            femaleagelist.add(Integer.parseInt(tokens[ 1 ]));
                            femaleethlist.add(Integer.parseInt(tokens[ 2 ]));
                            femaleedulist.add(Integer.parseInt(tokens[ 3 ]));
                     }
                     line = bufRead.readLine();
                     if (Integer.parseInt(tokens[0]) == 1) malecount++;
                     else femalecount++;
               }
              bufRead.close();
             }

       catch (ArrayIndexOutOfBoundsException e){
```

```
            /* If no file was passed on the command line, this exception is generated. A
        message indicating how to the class should be called is displayed */
                System.out.println("Usage: java ReadFile filename\n");
                }
        catch (IOException e){
                // If another exception is generated, print a stack trace
            e.printStackTrace();
        }
        counts[0]= malecount;
        counts[1]= femalecount;
        return counts;
    }


/***************************************************************/
        // checks if item val is in testarray
    public boolean isin(int [] testarray, int val)
    {
        boolean in = false;
        int i = 0;
        while (i < testarray.length)
        {
                if (testarray[i]==val)
                {
                        in = true;
                        i = testarray.length;
                }
                i++;
        }
        return in;
    }

/***************************************************************/
        // initialises partner arrays with -1 for single
    public int [] partnersinit(int count)
    {
                int [] partner = new int [count];
        for (int i=0;i < count;i++)
                        partner[i]=-1;
                return partner;
    }


/***************************************************************/
        // returns an array of the positions of the single women
        public ArrayList<Integer> singlefemales(int [] partners, int [] singlepools)
        {
                ArrayList<Integer> singles = new ArrayList<Integer>();
                        for (int i=0;i < partners.length; i++)
                        if (partners[i]== -1 & !isin(singlepools, i))
                                singles.add(i);
                return singles;
        }


/***************************************************************/
        // takes a sample of size m from all single females
        public int [] sample1(int M, int[] singlepools)
        {
                ArrayList<Integer> singles = singlefemales(femalespartner, singlepools);
                int N = singles.size();
                if (N<M)
                        M=N;
                int [] perm = new int[N];

                for (int j = 0; j < N; j++)
                        perm[j]=j;

                for (int i=0; i < M; i++)
                {
                        int r = i + (int) (Math.random() * ((N-1)-i));
```

```
                        int t = perm[r];
                        perm[r] = perm[i];
                        perm[i] = t;
                }

                int samp [] = new int[M];
                for (int i=0;i<M;i++)
                        samp[i]=singles.get(perm[i]);

                return samp;
        }


/*****************************************************************/
// match partners by heighest attraction score
    public void pairs(double[][] a, int [][] pool) {
        // a is scores, pool is single pool

        double max = -1;
        int partnertemp [] = new int[2];
        double max1 = 1;
        boolean quitloop = false;
        sim1.samecouple=0;
        sim1.newcouples=0;

        while (sim1.newcouples < sim1.couplesperperiod & !quitloop)
        {
                for (int r=0; r < a.length; r++)
                {
                        if (malespartner[r] == -1 )
                        {
                                for (int c=0; c < a[r].length; c++)
                                {
                                        if (femalespartner[pool[r][c]] == -1 )
                                        {
                                                if (a[r][c] > max)
                                                {
                                                        partnertemp[0]=r; // best match (male)
                                                        partnertemp[1]=c; // best match
(female)
                                                        max = a[r][c];    // how good a match?
                                                }
                                        }
                                }
                        }
                 }

                if (max != -1)
                {
                malespartner[partnertemp[0]] = pool[partnertemp[0]][partnertemp[1]];
                femalespartner[pool[partnertemp[0]][partnertemp[1]]] = partnertemp[0];
                sim1.newcouples++;


                if(sim1.maleethlist.get(partnertemp[0])==sim1.femaleethlist.get(pool[partner
                temp[0]][partnertemp[1]]))
                        sim1.samecouple++;
                }

                if (max == -1)
                        quitloop = true;

                max1 = a[partnertemp[0]][partnertemp[1]];
                max = -1;
                a[partnertemp[0]][partnertemp[1]]=-1;
        }

    }


/*****************************************************************/
```

```
// allocate partners lists to array
       public void network(int time, int singleman)
       {
               if (time==1)   {
                       int [] newpeople = sim1.sample1(sim1.startpool,
sim1.singlepools[singleman]);
                       for (int i=0;i<newpeople.length;i++)
                               sim1.singlepools [singleman][i] = newpeople[i];
               }
               else {
                       int [] newpeople = sim1.sample1(sim1.addpool,
sim1.singlepools[singleman]);
                       for (int i=0;i<newpeople.length;i++)
                               sim1.singlepools
[singleman][sim1.startpool+sim1.addpool*(time-2)+i] = newpeople[i];
               }
       }



 /*****************************************************************/
 // allocate partners lists to array
    public int [][] partnerallocate(ArrayList<Integer> males, ArrayList<Integer> females)
    {
        int [] [] partners = new int [males.size()][2];
       int idx = 0;
        int idy = 0;

       for(Integer v : males)
               partners[idx++][0] = v;
       for (Integer x: females)
               partners[idy++][1] = x;

       return partners;
    }

/*****************************************************/
/* count number of different categories in a list */
       public double scores(int mindex, int findex)
       {
        double mage = sim1.maleagelist.get(mindex);
        double fage = sim1.femaleagelist.get(findex);
        double medu = sim1.maleedulist.get(mindex);
        double fedu = sim1.femaleedulist.get(findex);
        double macropressure = 0;

        // macro samecouple/newcouples or 1-samecouple/newcouples except on first iteration
        if (sim1.newcouples==0 &&
sim1.maleethlist.get(mindex)!=sim1.femaleethlist.get(findex))
               macropressure = 5;
        else if (sim1.newcouples==0 &&
sim1.maleethlist.get(mindex)==sim1.femaleethlist.get(findex))
               macropressure = 5;
        else if (sim1.maleethlist.get(mindex)!=sim1.femaleethlist.get(findex))
            macropressure = 10*(1- (double )sim1.samecouple / (double) sim1.newcouples);
        else macropressure = 10*( (double) sim1.samecouple / (double)newcouples);

        // random atraction component
        double attraction = 9*Math.random();

        return Math.exp(-Math.sqrt((80 - mage - fage) + sim1.weights[0]*((mage-fage)*(mage-
fage))+sim1.weights[1]*((medu-fedu)*(medu-fedu))
                        - sim1.weights[2]*macropressure + sim1.weights[3]*attraction ));
       }


/*****************************************************/
       /* count number of different categories in a list */

       public int numcats(ArrayList<Integer> list)
       {
```

```
		int num = 0;
		for (int i=0; i<Collections.max(list)+1; i++)
			if (list.contains(i))
				num++;

		return num;
	}

/******************************************************/
	/* make and print crosstabs */

	public int [][] makecrosstab(int [] malespartner)
	{
		int [][] ethstab = new int
[Collections.max(maleethlist)][Collections.max(femaleethlist)];

		for (int i=0; i<malespartner.length; i++)
			if (malespartner[i] != -1)
				ethstab[maleethlist.get(i)-
1][femaleethlist.get(malespartner[i])-1]++;
		return ethstab;
	}

	public void printcrosstab(int [][] ethstab)
	{
		for (int i=0;i<ethstab.length;i++){
			for (int j=0;j<ethstab[i].length;j++)
				System.out.print(ethstab[i][j] + " ");
			System.out.print(" | ");}
	}

/******************************************************/
	/* calculate sum of squared deviations */
	public double sqdev(int [][] ethstab)
	{

		ArrayList<Double> simprops = new ArrayList<Double>();
		 double freqsum = 0;

		for (int i=0;i<ethstab.length;i++)
			for (int j=0;j<ethstab[i].length;j++)
				freqsum=freqsum+ethstab[i][j];

		for (int i=0;i<ethstab.length;i++)
			for (int j=0;j<ethstab[i].length;j++)
				{
				Double prop = ethstab[i][j]/freqsum;
				simprops.add(prop);
				}

		double sqdevs=0;
		for (int i=0; i<Math.min(simprops.size(), sim1.actual.size());i++)
		{
			sqdevs=sqdevs + (simprops.get(i)-
sim1.actual.get(i))*(simprops.get(i)-sim1.actual.get(i));
		}

		return sqdevs;
	}


/*******************************************************************/
	public static void main(String [] args)
	{

		// initialise simulation
		sim1 = new Simulation2();

		// filenames for inputs
		String filename = args[0];
	 String actuals = args[1];
```

```
        sim1.countnumberofcouples(actuals);

                sim1.weights[0] = Double.parseDouble(args[2]);
        sim1.weights[1] = Double.parseDouble(args[3]);
        sim1.weights[2] = Double.parseDouble(args[4]);
        sim1.weights[3] = Double.parseDouble(args[5]);

                // read in actual values and singles csv
        sim1.readactual(actuals);
        int counts [] = sim1.readfile(filename);
                int malecount = counts[0];
                int femalecount = counts[1];

                // initialise partners arrays
                sim1.malespartner = sim1.partnersinit(malecount);
                sim1.femalespartner = sim1.partnersinit(femalecount);

                // initialise social network and score arrays
                sim1.singlepools = new
int[malecount][sim1.startpool+sim1.addpool*(sim1.timesteps-1)];
                sim1.scores = new
double[malecount][sim1.startpool+sim1.addpool*(sim1.timesteps-1)];

                // simulation
                for (int time=1; time < sim1.timesteps + 1; time++)
                {
                        for (int singleman=0; singleman < malecount; singleman++)
                        {
                                if (sim1.malespartner[singleman]==-1)
                                {
                                                // (initialise or) increase pool of available
partners
                                                sim1.network(time, singleman);

                                                // allocate scores to females that this guy
knows
                                                for (int i=0; i < (sim1.startpool +
sim1.addpool*(time-1)); i++)
                                                        sim1.scores[singleman][i] =
sim1.scores(singleman, sim1.singlepools[singleman][i]);
                                }
                        }

                        // compare scores, match from highest
                        sim1.pairs(sim1.scores, sim1.singlepools);
                }

                int [][] ethstab = sim1.makecrosstab(sim1.malespartner);
                double sqdev = sim1.sqdev(ethstab);
                System.out.print(sim1.weights[0]  + " " + sim1.weights[1]  + " " +
sim1.weights[2]  + " " + sim1.weights[3] + " | ");
                System.out.print(sqdev + " | ");
                sim1.printcrosstab(ethstab);
                System.out.println();

        }
}
```

C.3. Grid Code

C.3.1. Optimisation Code

```
# this script will execute Lyndon's simulation code
# with multiple parameters based on inputs file,
# select best parameters for the next itteration, and
#  then execute it again.

GATEWAY=ng2.auckland.ac.nz
DIR=/home/grid-lyndon
RSL_FILE=multi.rsl
CLASSNAME=Simulation2
# file contains weight deltas during each iteration.
INPUT_FILE=inputs
CSV=1986C1830.csv
EXPECTED=cant91.txt
NUMBER_OF_ITTERATIONS=$1
W1=$2
W2=$3
W3=$4
W4=$5

# if different inputs file is used, this variable points to suffix
INPUTS_SUFFIX=
# number of itterations  to change to different inputs file
ITTER=8
CURRENT_ITTER=0

OUTPUT_FILE=outputs

# generate job submission for single job and insert it into $RSL_FILE
# accepts four parameters for weights
function generateRSL(){

# don't do anything if one of the weights is 0 or less
if test 0 -eq $(echo "$1 >= 0"|bc)
then return
fi
if test 0 -eq $(echo "$2 >= 0"|bc)
then return
fi
if test 0 -eq $(echo "$3 >= 0"|bc)
then return
fi
if test 0 -eq $(echo "$4 >= 0"|bc)
then return
fi

if test 0 -eq $(echo "$1 <= 1"|bc)
then return
fi
if test 0 -eq $(echo "$2 <= 1"|bc)
then return
fi
if test 0 -eq $(echo "$3 <= 1"|bc)
then return
fi
if test 0 -eq $(echo "$4 <= 1"|bc)
then return
fi


    cat >> $RSL_FILE <<EOF
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
```

```
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://${GATEWAY}:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>$</directory>
 <argument>${CLASSNAME}</argument>
 <argument>${CSV}</argument>
 <argument>${EXPECTED}</argument>
 <argument>$1</argument>
 <argument>$2</argument>
 <argument>$3</argument>
 <argument>$4</argument>
 <jobType>single</jobType>
</job>
EOF
}


rm $OUTPUT_FILE

for i in `seq 1 $1`
do
CURRENT_ITTER=$(echo "$CURRENT_ITTER+1"|bc)
if $(test $CURRENT_ITTER -eq $ITTER)
then INPUTS_SUFFIX="2"
fi

rm $RSL_FILE

  cat >> $RSL_FILE <<EOF
<multiJob>
EOF

  cat inputs${INPUTS_SUFFIX}| while read I1 I2 I3 I4;
  do
  W1TEMP=`echo "$W1+$I1"|bc`
  W2TEMP=`echo "$W2+$I2"|bc`
  W3TEMP=`echo "$W3+$I3"|bc`
  W4TEMP=`echo "$W4+$I4"|bc`
  generateRSL $W1TEMP $W2TEMP $W3TEMP $W4TEMP
  done

  cat >> $RSL_FILE <<EOF
</multiJob>
EOF

# run multiJob, sort the results based on deviation and select the "best" weights
touch $OUTPUT_FILE
(cat $OUTPUT_FILE; globusrun-ws -submit -s   -J -S -f $RSL_FILE -F $GATEWAY) | sort -n -t
"|" -k 2,2 | head -n 1|
awk  -v output=$OUTPUT_FILE --  '{print $0 >> output}'

# increment weights variables
eval `tail -n 1 $OUTPUT_FILE| awk '{printf("W1=%s;W2=%s;W3=%s;W4=%s",$1,$2,$3,$4)}'`

echo $W1 $W2 $W3 $W4

done
```

C.3.2. Parallel Processing Code

```
<multiJob>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.10</argument>
 <argument>.15</argument>
 <argument>.40</argument>
 <argument>.35</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.10</argument>
 <argument>.15</argument>
 <argument>.50</argument>
 <argument>.25</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.10</argument>
 <argument>.25</argument>
 <argument>.40</argument>
 <argument>.25</argument>
 <jobType>single</jobType>
</job>
```

```
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.20</argument>
 <argument>.15</argument>
 <argument>.40</argument>
 <argument>.25</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.10</argument>
 <argument>.25</argument>
 <argument>.50</argument>
 <argument>.15</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.20</argument>
 <argument>.15</argument>
 <argument>.50</argument>
 <argument>.15</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
```

```
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.20</argument>
 <argument>.25</argument>
 <argument>.40</argument>
 <argument>.15</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.20</argument>
 <argument>.25</argument>
 <argument>.50</argument>
 <argument>.05</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.30</argument>
 <argument>.2</argument>
 <argument>.45</argument>
 <argument>.05</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
```

```
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.15</argument>
 <argument>.35</argument>
 <argument>.45</argument>
 <argument>.05</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.15</argument>
 <argument>.2</argument>
 <argument>.60</argument>
 <argument>.05</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>.30</argument>
 <argument>.05</argument>
 <argument>.45</argument>
 <argument>.2</argument>
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
```

```
 &lt;argument&gt;testtest.txt&lt;/argument&gt;
 &lt;argument&gt;.30&lt;/argument&gt;
 &lt;argument&gt;.2&lt;/argument&gt;
 &lt;argument&gt;.30&lt;/argument&gt;
 &lt;argument&gt;.2&lt;/argument&gt;
 &lt;jobType&gt;single&lt;/jobType&gt;
&lt;/job&gt;
&lt;job&gt;
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        &lt;wsa:Address&gt;
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        &lt;/wsa:Address&gt;
        &lt;wsa:ReferenceProperties&gt;
            &lt;gram:ResourceID&gt;PBS&lt;/gram:ResourceID&gt;
        &lt;/wsa:ReferenceProperties&gt;
    &lt;/factoryEndpoint&gt;
 &lt;executable&gt;java&lt;/executable&gt;
 &lt;directory&gt;/home/grid-lyndon&lt;/directory&gt;
 &lt;argument&gt;Simulation2&lt;/argument&gt;
 &lt;argument&gt;test2.csv&lt;/argument&gt;
 &lt;argument&gt;testtest.txt&lt;/argument&gt;
 &lt;argument&gt;.15&lt;/argument&gt;
 &lt;argument&gt;.35&lt;/argument&gt;
 &lt;argument&gt;.30&lt;/argument&gt;
 &lt;argument&gt;.2&lt;/argument&gt;
 &lt;jobType&gt;single&lt;/jobType&gt;
&lt;/job&gt;
&lt;job&gt;
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        &lt;wsa:Address&gt;
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        &lt;/wsa:Address&gt;
        &lt;wsa:ReferenceProperties&gt;
            &lt;gram:ResourceID&gt;PBS&lt;/gram:ResourceID&gt;
        &lt;/wsa:ReferenceProperties&gt;
    &lt;/factoryEndpoint&gt;
 &lt;executable&gt;java&lt;/executable&gt;
 &lt;directory&gt;/home/grid-lyndon&lt;/directory&gt;
 &lt;argument&gt;Simulation2&lt;/argument&gt;
 &lt;argument&gt;test2.csv&lt;/argument&gt;
 &lt;argument&gt;testtest.txt&lt;/argument&gt;
 &lt;argument&gt;.15&lt;/argument&gt;
 &lt;argument&gt;.05&lt;/argument&gt;
 &lt;argument&gt;.60&lt;/argument&gt;
 &lt;argument&gt;.2&lt;/argument&gt;
 &lt;jobType&gt;single&lt;/jobType&gt;
&lt;/job&gt;
&lt;job&gt;
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        &lt;wsa:Address&gt;
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        &lt;/wsa:Address&gt;
        &lt;wsa:ReferenceProperties&gt;
            &lt;gram:ResourceID&gt;PBS&lt;/gram:ResourceID&gt;
        &lt;/wsa:ReferenceProperties&gt;
    &lt;/factoryEndpoint&gt;
 &lt;executable&gt;java&lt;/executable&gt;
 &lt;directory&gt;/home/grid-lyndon&lt;/directory&gt;
 &lt;argument&gt;Simulation2&lt;/argument&gt;
 &lt;argument&gt;test2.csv&lt;/argument&gt;
 &lt;argument&gt;testtest.txt&lt;/argument&gt;
 &lt;argument&gt;0&lt;/argument&gt;
 &lt;argument&gt;.35&lt;/argument&gt;
 &lt;argument&gt;.45&lt;/argument&gt;
 &lt;argument&gt;.2&lt;/argument&gt;
```

```
 <jobType>single</jobType>
</job>
<job>
    <factoryEndpoint
            xmlns:gram="http://www.globus.org/namespaces/2004/10/gram/job"
            xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/03/addressing">
        <wsa:Address>
            https://ng2.auckland.ac.nz:8443/wsrf/services/ManagedJobFactoryService
        </wsa:Address>
        <wsa:ReferenceProperties>
            <gram:ResourceID>PBS</gram:ResourceID>
        </wsa:ReferenceProperties>
    </factoryEndpoint>
 <executable>java</executable>
 <directory>/home/grid-lyndon</directory>
 <argument>Simulation2</argument>
 <argument>test2.csv</argument>
 <argument>testtest.txt</argument>
 <argument>0</argument>
 <argument>.2</argument>
 <argument>.60</argument>
 <argument>.2</argument>
 <jobType>single</jobType>
</job>
</multiJob>
```