THE DETERMINANTS OF THE GEOGRAPHIC CONCENTRATION AMONG IMMIGRANTS: APPLICATION TO AUSTRALIA[1]

Barry R. Chiswick

Department of Economics, University of Illinois at Chicago, 601 S. Morgan Street, Chicago, IL 60607-7121, USA.

Yew Liang Lee

Department of Economics, University of Western Australia, Crawley, WA 6009, Australia.

Paul W. Miller

Department of Economics, University of Western Australia, Crawley, WA 6009, Australia.

ABSTRACT　　This paper develops a theoretical framework for the study of the tendency for immigrant groups to be geographically concentrated. Testing the model for Australia shows that the extent of geographic concentration of immigrant groups is negatively related to age at migration, duration of residence in Australia and the proportion of the birthplace group that is fluent in English. The extent of geographic concentration is also affected by the availability of ethnic media and the distance between the country of origin and the place of residence in Australia.

1. INTRODUCTION

A common characteristic of immigrants in various destinations and in various time periods is that they tend to be geographically concentrated, that is, they tend to live in immigrant concentrations or enclaves. Immigrants from a particular origin tend to live in areas where others from the same origin live, rather than distributing themselves across the regions of the destination in the same proportion as the native-born population. Moreover, these concentrations differ across immigrants from different countries of origin. Immigrant concentrations in the liberal democracies of Western Europe, North America, and Oceania are seldom a consequence of government policies to concentrate immigrants. Indeed, they arise even when public policy prefers to disperse immigrants across the host country.

If these geographic concentrations were without consequences they would be of little interest. Yet, they do have consequences. (See, for example, Case and

---

Katz, 1991; Chiswick and Miller, 1995, 2000; Goddard, Sparkes and Haydon, 1985; Hugo, 1995; Le, 1999, and Veltman, 1983.) The geographic concentration appears to have adverse effects on immigrants acquiring destination language skills, but they may have favourable effects on immigrant groups maintaining and passing on to their children their mother tongue and ethnic culture. Enclaves may facilitate immigrant entrepreneurship, although they appear to depress the nominal earnings of immigrants. Enclaves have an effect on the demand for "ethnic goods", as well as on the demand for publicly provided goods and services[2]. Moreover, enclaves may affect the political strength of immigrant groups at local and national levels.

What has been subject to less study is why immigrant groups tend to form concentrations or enclaves. The purpose of this paper is to present some facts (descriptive statistics) on immigrant concentrations in Australia, to develop a theoretical framework for the analysis of the determinants of these concentrations, and to use multivariate statistical analysis to test the empirical relevance of the theoretical model.

Section 2 reviews past research on the geographic concentration of immigrants and presents the geographic concentration index (G) employed in this study. The index is used in Section 3 for data from the 1991 and 1996 censuses of Australia to present descriptive statistics on the concentration of immigrants for the States and Territories of Australia by country of origin. A theoretical model for the determinants of the degree of immigrant concentration across States/Territories and birthplaces is developed in Section 4. The empirical estimation of the model for data on the concentration index by postal code from the 1991 Census is reported in Section 5. Section 6 contains a summary and conclusion.

2. MEASURING GEOGRAPHIC CONCENTRATION

The distribution of immigrants across regions in Australia differs from that for the Australia born. This has been the focus for considerable research. Hugo and Maher (1995) provide details on much of this research, and an overview of the pattern of settlement of migrants in 1991.

Table 1 illustrates the pattern of geographic concentration of immigrants in Australia. The information in this table is on the distribution of the major birthplace groups across States and Territories in 1996.

It is apparent from Table 1 that New South Wales and Victoria are the states most heavily populated by both the Australia born and the overseas born. Relatively few Australia-born persons and even fewer of the overseas born live in Tasmania, the Australian Capital Territory and the Northern Territory. There

---

[2] "Ethnic goods" are market and non-market goods and services consumed by members of an immigrant or ethnic group that are not consumed, or not consumed to the same extent, by members of other groups. These include ethnic churches and marriage markets, as well as food, clothing, and festivals specific to immigrant ethnic groups. For an application of this concept, see Chiswick and Miller (2000).

is a greater proportional representation of immigrants (by 2 to 3 percentage points) in New South Wales, Victoria and Western Australia than is the case of the Australia born, and a smaller representation in the other regions, particularly in Queensland. There are differences, however, in the regions of residence of immigrants from English-speaking and non-English speaking countries[3]. Thus, immigrants from English-speaking countries have a smaller relative proportional representation in New South Wales and Victoria than the Australia born, and a much higher representation in Western Australia. There are much higher proportional representations of immigrants from non-English speaking countries in New South Wales and Victoria, and a much smaller proportional representation in Queensland.

The seven most numerous groups of immigrants in the 1996 Census of Population and Housing in Australia, each with at least 100,000 immigrants, are from China, Germany, Greece, Italy, New Zealand, the United Kingdom and Vietnam, and data on the geographic distributions for these birthplace groups are also provided in Table 1. Immigrants from Germany have a distribution across States and Territories that is very similar to that of the Australia born. But there are marked differences between the geographic distributions of the remaining birthplace groups in Table 1 and the geographic distribution of the Australia born. The majority of Greek immigrants, for example, have settled in Victoria and New South Wales (48.8 percent and 32.5 percent, respectively), and relatively few have settled in Queensland and Western Australia (3.5 and 2.7 percent, respectively). The main feature of the geographic distribution of Italian immigrants is the high representation in Victoria, where 41.3 percent of the group reside, in contrast to only 24.0 percent for those born in Australia and 26.6 percent for immigrants from English-speaking countries.

Most immigrants from New Zealand have located in states just across the Tasman Sea, Queensland (34.9 percent), followed by New South Wales (30.2 percent). In the case of immigrants born in the United Kingdom, 27 percent are in New South Wales, and about 20 percent each in Victoria and Western Australia.

Immigrants from Vietnam and China are more highly concentrated than the other birthplace groups included in Table 1. Fully 59 percent of immigrants from China reside in New South Wales, and most Vietnamese immigrants live in New South Wales and Victoria (40.5 percent and 36.5 percent, respectively).

Hence the immigrant groups in Table 1 are distributed across States and Territories in ways that are quite different from the geographic distribution of the Australia-born population. The geographic concentrations differ across immigrant groups: not only is the English-speaking/non-English speaking distinction important, but within each of these two major categories there are considerable differences across individual birthplaces. To illustrate further the

---

[3] As close to 90 percent of immigrants from South Africa speak English only, it is categorised as a predominately English-speaking country in this study.

Table 1. Population Distribution – Birthplace by States/Territories, 1996

Birthplace	NSW	VIC	QLD	SA	WA	TAS	ACT	NT	Total	% of Aust. Population
All Persons	33.7	24.6	18.7	8.0	9.7	2.6	1.7	1.1	100.0	100.0
Australia Born	33.2	24.0	20.0	8.1	8.9	3.0	1.7	1.1	100.0	76.7
Total Immigrants	35.5	26.6	14.3	7.7	12.2	1.2	1.7	0.8	100.0	23.3
Total English-Speaking Countries[a]	28.8	18.6	20.6	9.7	17.8	1.9	1.7	0.9	100.0	9.2
Total Non-English Speaking Countries	39.8	31.7	10.1	6.5	8.7	0.8	1.7	0.7	100.0	14.1
New Zealand[b]	30.2	14.6	34.9	3.3	13.4	1.2	1.2	1.2	100.0	1.7
United Kingdom	26.9	20.1	17.0	12.3	19.1	2.1	1.7	0.8	100.0	6.4
China	59.0	25.3	6.7	2.7	4.1	0.3	0.3	1.5	100.0	0.7
Germany	29.4	26.9	17.5	12.0	9.1	1.8	1.0	2.3	100.0	0.7
Greece	32.5	48.8	3.5	10.0	2.7	0.5	1.1	0.9	100.0	0.8
Italy	27.7	41.3	7.2	11.4	10.6	0.5	1.1	0.3	100.0	1.4
Vietnam	40.5	36.5	7.3	7.1	6.7	0.1	1.5	0.3	100.0	0.9

[a] The English-speaking countries are Canada, Ireland, New Zealand, South Africa, the United Kingdom, and the United States of America.

[b] In order of numerical importance the birthplaces are United Kingdom (1.16 million), New Zealand (315,100), Italy (259,100), Vietnam (164,200), Greece (141,800), China (121,100) and Germany (120,800).

Source: 1996 Census of Population and Housing for state data, and Tables 38 and 41 of ABS (1998) for the data in the final column. ABS (1998) adjusts the census data for under-enumeration and for Australian residents overseas at the time of the Census, and provides a more accurate measure of the birthplace composition of the Australian population.

extent of some of these differences, it is noted that around 20 percent of immigrants from Vietnam would have to shift State/Territory for their distribution across States/Territories to be the same as that of the Australia born. For immigrants from China, 28 percent would need to move to achieve a distribution across States/Territories the same as that of the Australia born, whereas for immigrants from Germany, only 8 percent would have to move.

If smaller geographic areas than States are used as the basis for defining a region, even greater levels of concentration of immigrant groups would be expected because of varying degrees of geographic concentration within states. Indices presented by Burnley, Murphy and Fagan (1997), for example, show that when the unit of analysis is the Local Government Area (LGA) the discrepancies between the geographic distributions of the Australia born and immigrant groups are much larger than those evident in Table 1. For example, considering the 43 LGAs in the Sydney metropolitan area in the 1991 Census, 67 percent of those born in Vietnam would have to shift their residence to have the same distribution across LGAs in Sydney as the Australia born. The comparable figures for those born in Greece and China are 50 and 44 percent, respectively, while for the UK and Ireland, only 12 percent would need to shift across LGAs for the geographic distributions of this birthplace category and the Australia born to be congruent.

This pattern is typical of that described in the literature: immigrants from English-speaking countries have a distribution across geographic units that is much more like that of the Australia born than is the case for the immigrants from non-English speaking countries. Immigrants from Vietnam, predominately refugees and relatively recent arrivals in Australia, have the distribution across geographic units that is most dissimilar to that of the Australia-born population.

In this study the degree of geographic concentration is measured by the coefficient of geographic association (G) (see Haggett, Cliff and Frey, 1977; Bartel, 1989)[4]. It is computed as

$$G = \sum_{i=1}^{n} \text{Max}([g_{io} - g_{ip}], 0),$$

where g_{io} is the percentage representation of the specific birthplace group in the ith geographic area, g_{ip} is the percentage representation of the total population in the ith geographic area and n is the number of geographic areas. Note that the benchmark is the total population, and not those born in Australia. Where the overseas born group has a distribution across regions the same as the total population, G will equal 0. Where the overseas born group is completely segregated, the upper bound of G will equal (100-GS), where GS is the percent of the total population accounted for by the specific birthplace group. If a group is 1 percent of the population, the maximum value of the index G is 99; if it is 5 percent the maximum value is 95. This means that comparisons across birthplace

---

[4] There are several excellent review articles on the measurement of segregation. See in particular Massey and Denton (1988) and Watts (1998). Bartel (1989, p.375) notes that standard errors cannot be computed for the G index.

groups can be affected by group size. The only birthplace group with a large enough share of the population to cause the upper bound to diverge appreciably from 100 is the UK, where the upper bound would be 94. The birthplace dummy variables included in the statistical analyses reported below will capture, in part, this group size effect.

The size of the *G* index is not unique and varies with the number of the areas used in the calculation. The larger the size of the region (*i.e.*, the smaller the number of areas) the more likely it is that differences in geographic distributions at a local level will be averaged out. So the index is expected to be smaller the larger the size of the region—it will be smaller if the states of Australia are the units of observation than if Local Government Areas are used[5]. Appendix B contains further discussion of the indices used in the Australian literature.

There is no set way of defining the area to use in the calculation: this will depend on the purpose of the study. The current analyses are based on postcode data, which is an intermediate level of geographic disaggregation compared to the Local Government Area and Collectors Districts mainly used in the literature. The average population count in the 2,387 postcode areas in 1996 is 7,164 persons.

3. CONCENTRATION OF IMMIGRANTS: 1991 AND 1996

Table 2 presents the indices of geographic concentration *G* of immigrants using the 1991 and 1996 Censuses. In this analysis the geographic unit is defined with respect to postcode of residence. An alternative would be to compute the index based on place of work. However, the Australian Census of Population and Housing does not provide the necessary information. The indices can be interpreted as the percentage of immigrants from the specific birthplace group that, along with a similar number of others (native born and other immigrants), would need to shift across postcode areas in order for the distribution of that birthplace group across postcode areas to be the same as the distribution for the total population[6]. For example, the figure in Table 2 for 1991 shows that approximately 21 percent of immigrants from English-speaking countries, and an equal number of other people, would have to shift across postcode areas if the geographic distribution across postal codes of persons from English-speaking

---

[5] To illustrate this point, the index for immigrants from non-English speaking countries at the postcode level in 1996 is 30.6 (see Table 2). A total of 2,387 postcode areas are used in this computation. When the eight States/Territories are used as the unit of analysis, the index for immigrants from non-English speaking countries is only 13.2.

[6] Note that the requirement for other members of the population to shift to replace the members of the specific group ensures the geographic spread of the total population is maintained. This is a sensible condition to impose in terms of making use of the existing infrastructure (housing, services etc.). This requirement will generally be implicit in the discussion that follows.

countries was to mirror that of the total population[7].

Table 2. Indices of Geographic Concentration of Immigrants, 1991 and 1996[a] [b]

	% of the Australian Population, 1996	1991	1996
Canada	0.15	31.8	29.4
Ireland	0.30	26.2	25.1
New Zealand	1.72	27.6	26.9
South Africa	0.34	38.3	38.1
United Kingdom	6.42	22.1	22.0
United States of America	0.30	33.9	31.5
Total English-speaking Countries	9.23	20.6	20.4
China	0.66	52.6	54.3
Germany	0.66	19.6	18.4
Greece	0.77	57.3	56.9
Hong Kong	0.43	51.9	54.6
India	0.46	39.8	40.6
Italy	1.42	45.8	45.5
Lebanon	0.42	67.3	67.4
Malaysia	0.46	45.0	45.1
Malta	0.30	56.5	54.4
Netherlands	0.52	24.6	23.5
Philippines	056	39.9	38.9
Poland	0.39	39.5	37.6
Vietnam	0.90	68.8	68.3
Total Non-English-speaking Countries[c]	14.07	33.5	30.6
Total Immigrants	23.26	21.8	22.1

[a]   Birthplace groups in which data are available in both the 1991 and 1996 Census.

[b]   Data obtained from Tables 38 and 41 of ABS (1998) for column (i), CDATA91 for column (ii) and CDATA96 for column (iii).

[c]   Includes countries not listed.

---

[7] There are many other indices that could be computed (see Massey and Denton, 1988; Watts, 1998). To examine whether the findings were sensitive to the choice of measure, the comparisons in Table 2 were repeated using the Herfindahl index (see Bartel and Koch, 1991). Appendix A contains relevant results. The findings do not appear to be sensitive to the choice of index.

There are three main features of Table 2. First, the indices computed for each birthplace group are fairly consistent for 1991 and 1996. During this short interval there was little movement toward dispersion or concentration of immigrants by birthplace. Second, there is greater geographic concentration among immigrants from non-English speaking countries than among immigrants from English-speaking countries. The index for the total non-English-speaking countries is around 1.5 times that for the total immigrants from English-speaking countries (*i.e.,* around 30 compared to 20).

Third, within the English-speaking and non-English speaking aggregates, there is considerable variation in the index of geographic concentration. Immigrants from South Africa are the most highly concentrated group among the English-speaking birthplaces (index of 38). Among the non-English speaking countries, the greatest concentrations are recorded by immigrants from Vietnam and Lebanon (indices close to 70), while intermediate levels of concentration are recorded among immigrants from China, Greece, Hong Kong and Malta[8]. Immigrants from Germany and the Netherlands have relatively low levels of geographic concentration.

Table 3 shows the index of geographic concentration for each major immigrant group computed within each State/Territory in 1991. The number of postcode areas used in the calculations is 583 in New South Wales, 618 in Victoria, 397 in Queensland, 308 in South Australia, 296 in Western Australia, 106 in Tasmania, 23 in the ACT and 24 in the Northern Territory. As the number of postcode areas differs across the States/Territories, the indices are not strictly comparable across States, but they are comparable within States. While there are relatively few immigrants in the Northern Territory, Tasmania and the Australian Capital Territory, indices are presented for these groups for completeness. The discussion will not cover these areas, however, because of the greater sampling error. The main feature of Table 3 is that the relative indices vary from one State/Territory to another for each birthplace group.

Among immigrants from the English-speaking countries, South Australia appears to be a State with high levels of concentration, and Western Australia a State with low levels of concentration, relative to natives. For immigrants from the non-English speaking countries, New South Wales and Victoria are the areas which appear to be characterised by high concentrations of many birthplace groups relative to natives, and Queensland, South Australia and Western Australia have more diversified geographic distributions for most birthplace groups.

Among the English-speaking countries, South Africa is characterised by a

---

[8] Hugo (1995, p.9) suggests an Index of Dissimilarity of more than 40 can be categorised as "very high", an index of between 25 and 39 as "moderate" and an index of less than 25 as "low". Burnley *et al.* (1997) use a categorisation of more than 80 as "segregated to a large degree", between 50 and 80 as "strong concentration", 20-50 as "moderate concentration" and less than 20 as "lightly concentrated". The cut-offs will vary according to the size of the geographical unit used in the analysis.

relatively high degree of residential concentration in each State, the value of the *G* index being as high as 40. At the other end of the spectrum, the United Kingdom is characterised by consistently low values of the *G* index across the States, with values between 16 and 23.

Table 3. Indices of Geographic Concentration for Each State/Territory, 1991

	NSW	VIC	QLD	SA	WA	TAS	ACT	NT
Canada	33.6	30.1	25.6	34.3	27.3	39.2	20.3	20.3
Ireland	26.0	22.1	21.5	25.4	19.3	28.0	30.9	0.1
New Zealand	22.5	19.5	22.7	16.8	14.4	16.4	25.4	15.0
South Africa	40.5	39.4	28.1	35.8	28.2	34.5	27.7	19.7
United Kingdom	19.1	20.1	16.3	23.0	15.7	12.6	27.9	12.2
USA	37.1	32.7	27.6	36.7	26.8	34.4	30.6	42.4
Total English-speaking Countries	18.8	18.2	17.1	21.0	13.4	13.0	7.3	12.7
China	51.1	48.6	45.5	45.3	43.3	41.9	39.1	49.1
Germany	19.0	19.0	18.1	14.2	12.6	20.8	27.1	12.9
Greece	58.2	46.4	49.5	45.7	49.7	52.7	39.6	50.6
Hong Kong	51.2	48.6	47.0	52.7	42.2	47.8	35.1	47.2
India	37.8	36.6	31.4	30.8	31.4	43.6	30.6	30.5
Italy	43.3	40.5	43.7	44.0	39.5	41.3	33.5	23.4
Lebanon	61.7	54.2	67.3	43.9	42.2	60.9	40.6	52.3
Malaysia	39.5	45.6	33.4	43.2	40.8	47.2	30.7	35.3
Malta	49.8	58.5	38.7	35.4	37.8	55.1	35.0	28.2
Netherlands	22.2	30.1	18.1	19.6	16.1	26.2	32.0	12.8
Philippines	43.6	37.5	22.2	31.5	27.8	27.0	35.1	33.2
Poland	37.0	39.3	36.4	35.6	33.1	44.2	30.6	25.8
Vietnam	71.9	63.8	71.6	63.5	62.7	62.5	41.1	33.0
Total Non-English-speaking Countries[a]	35.2	30.9	22.8	25.6	22.9	21.5	25.4	28.9
Total Immigrants	24.8	20.3	17.4	15.0	12.9	14.2	5.0	16.8
Number of Postal Codes	583	618	397	308	296	106	23	24

[a] Include countries not listed.

4. SPECIFICATION OF THE MODEL

This section develops the theoretical model for the econometric analyses of the degree of concentration among immigrants. The considerable variation in the geographic concentration of immigrants across birthplace groups within States could arise for a range of reasons. The way the *G* index is constructed means that differences across States may be due in part to the different numbers of postcode areas in each State. The literature also suggests that the degree of immigrant segregation will decrease with a longer period of time that immigrants have been in Australia. For example, Hugo (1995, p.6) notes that "those who have been established in Australia longer have an increasing tendency to settle outside major cities". Veltman (1983) also draws attention to the fact that migrants may change residential location as their socioeconomic status rises. The same theme appears in recent research regarding immigrant (see Bartel and Koch, 1991) and racial segregation in the United States (see Cutler, Glaeser and Vigdor, 1999).

There is a wide array of other variables that could affect the extent to which immigrants locate among concentrations of their compatriots. These factors could be like duration of residence in that they are part of the immigrant adjustment phenomenon. Other factors may be representative of the attitudes of immigrants or features of the immigrant groups that impact on the settlement pattern. Finally, the socioeconomic environment in the host country may influence where immigrants settle initially and subsequently.

The conceptual model proposed in this analysis seeks to account for the variations in indices of the type presented in Table 3 as follows:

$$G_j = f(\text{Adjustment}_j, \text{Attitudes}_j, \text{Environment}_j).$$

English language skill is a factor which has been interpreted in the literature as an indicator of immigrant adjustment (see, for example, McManus, Gould and Welsh, 1983; Chiswick and Miller, 1995). Higher levels of adjustment are argued to be associated with both greater proficiency in English and with a greater shift to English as the language spoken at home. Immigrants proficient in English presumably have a much wider range of job and consumption opportunities than their counterparts with limited English skills, and they may therefore have less incentive to live among others from the same country of origin. The indices presented in Burnley *et al.* (1997, p.43), which were disaggregated by English language skills, show that the degree of geographic concentration is less among those who are fluent in English. For example, among immigrants from China living in Sydney in 1991, the index of dissimilarity was 47 for those who did not speak English well, and 40 for those who spoke English well. These data suggest that immigrants' choice of residence among compatriots may be influenced, in part, by the protection this offers against their limited English skills (see Hugo, 1995, p.19).

Another variable which has been highlighted in the immigrant adjustment literature that may impact on geographic concentration is educational attainment. Educational attainment is a key variable in understanding immigrant economic adjustment, and the correlates with economic adjustment, such as dominant

language fluency. The better educated will have higher levels of earnings than their less-well educated counterparts, and yet may have the same fixed costs of internal migration. This results in a tendency for there to be a national labour market for highly educated workers, but much more of a tendency towards local labour markets for those with little education. For this reason, the better educated are expected to be more likely to move from a place of initial settlement to regions where the dominant attraction is employment opportunities rather than concentrations of compatriots.

A final factor that has obvious links to immigrant adjustment is the likelihood of return migration. It has been argued that the probability of return migration is lower (and thus the incentive for immigrant adjustment more intense) when the costs associated with it are higher (Chiswick and Miller, 2001). This will be the case the greater the physical distance between the origin and destination countries. The greater distance also increases the favourable selectivity in the initial migration stream, and immigrants more favourably selected are likely to have a lower return migration rate. Hence, extending the arguments above, it would be expected that the degree of geographic concentration will be higher in a community that expects to be transient (that is, it has a relatively high expected propensity for return migration). This will tend to occur where the distances between the countries of origin and Australia are shorter.

Age at migration, marital status and gender are characteristics of immigrants that may impact on the settlement pattern. Chiswick and Miller (1995) have established that the younger the age at migration, the greater the adaptability of an immigrant to the host country. In particular, the younger the age at migration, the greater the likelihood that an immigrant will learn English. Presumably other indices of the process of adjustment are similarly affected, including adjusting to the local culture. Accordingly, it might be expected that those who migrate at an early age will have a reduced tendency toward geographic concentration. However, a young age at migration may simply mean that the person migrated with his or her parents, and these family ties may have been maintained and thus contribute to their having a relatively high degree of geographic concentration. The impact of age at migration on geographic concentration is therefore an empirical issue.

Similarly, as marriage among immigrants is likely to be to a person of the same birthplace group, there might be a positive relationship between the proportion of the immigrant group married and the degree of geographic concentration for purely mechanical reasons. Working against this is the possibility that the support of a partner may substitute for the support that other compatriots might offer.

The proportion of an immigrant group that is female may also influence the degree of geographic concentration if the strength of kinship varies by gender.

There are various ways that the environment can influence the settlement pattern. Many ethnic groups have a well-established "ethnic infrastructure", comprising ethnic media, ethnic organisations, places of worship and ethnic shops. The existence of such an infrastructure can be taken as a general indicator of clustering among the particular ethnic group. Hence, the more well-developed

the ethnic infrastructure, the more intense the expected degree of geographic concentration. However, an ethnic infrastructure could also substitute for geographic concentration. For example, if there are many newspapers printed in Italian that are easily available throughout Australia, then this could, *ceteris paribus*, replace the value of settling among compatriots for immigrants from Italy. It is expected, however, that one would need a regional concentration in the first place to establish the newspaper.

The empirical counterpart of the conceptual equation presented above is:

$$G_{ji} = f \text{ (Birthplace}_j, \text{ State}_i, \text{ Limited English}_{ji}, \text{ Duration}_{ji}, \text{ Education}_{ji},$$
$$\text{Age at Migration}_{ji}, \text{ Married}_{ji}, \text{ Female}_{ji}, \text{ Distance}_{ji}, \text{ Ethnic Media}_{ji})$$

where "Birthplace$_j$" is the jth country of origin for which the particular indices are computed, "State$_i$" is the ith State or Territory of current residence, and "Limited English$_{ji}$" is the proportion of immigrants in the jth birthplace group in the ith state with limited English skills. "Duration$_{ji}$", "Education$_{ji}$", "Age at Migration$_{ji}$", "Married$_{ji}$", and "Female$_{ji}$" are, respectively, the mean duration of residence in Australia, the mean educational attainment, the mean age at migration, proportion married and proportion female of members of the jth birthplace group from the ith State. "Distance$_{ji}$" is the distance (in thousands of kilometers) between the capital city of the immigrant's country of origin and the capital city of the State or Territory of residence, while "Ethnic Media$_{ji}$" is the number of newspapers in the main language of the immigrant's country of origin that are available in the State or Territory of residence. In preliminary estimations, the number of ethnic organisations in the State or Territory of residence was included in the model. This was not significant, however. Hence only the Ethnic media variable is used to represent the influence of ethnic groups. Appendix B contains definitions of these variables, and means and standard deviations.

Some modifications to this schema are, however, necessary in the empirical application. First, to provide maximum flexibility in the construction of the measures of limited English skills, duration of residence, educational attainment, age at migration, marital status and gender, these have been computed from the 1991 Census of Population and Housing Household Sample File. One potential advantage when using a unit record file to compute the means is that appropriate variables to test for non-linearities can be introduced into the estimating equation[9]. The 1991 Census Household Sample File enables data for New South Wales, Victoria and Queensland to be obtained. However, for South Australia, Western Australia, Tasmania and the Territories, the regions are categorised as: "Adelaide", "Perth" and "Tasmania & ACT" and "Remainder of South Australia, Western Australia and the Northern Territory". Data for non-metropolitan areas in Western Australia and South Australia, and data for the Northern Territory,

---

[9] If the explanatory variable is $\overline{X}$, then the appropriate squared term is $\sum X_i^2 \big/ n$ rather than $\left(\overline{x}\right)^2$. The two are related as $\sum X_i^2 \big/ n = \left(\overline{x}\right)^2 + VAR(X)$.

are therefore not included in the analysis. As few immigrants reside in these areas this is not a major limitation. The indices reported in Table 3 have been re-computed using these six geographic areas.

Second, there are certain limitations on the birthplace variables. The birthplaces of UK and Ireland are aggregated in the Household Sample File. Hence, the combined grouping of "UK and Ireland" will be used in the analysis. Canada is not separately identified in the Household Sample File. Hence, the indices for this birthplace cannot be included in the regression analysis. Thus, the analysis is based on indices constructed for 19 birthplace groups in each of six regions. However, a reliable measure of the mean age of arrival for immigrants from Lebanon living in Tasmania and the ACT could not be obtained from the Household Sample File. Hence this observation was excluded from the data set and the analysis is based on 113 observations[10].

Third, the Limited English skills, Duration of residence, Education, Age at migration, Married and Gender variables are the means of cells in a matrix of birthplace by region of residence. These means are related to the birthplace data, with the correlation between the birthplace dummy variables and each of the other explanatory variables (Limited English skills, Duration of residence etc.), being between .73 and .98[11]. Given the method of construction, it is not feasible to include the birthplace variables (other than a dichotomous non-English origin country variable) with the behavioural variables in an encompassing equation[12]. The State variables are included, however, in order to control for the different number of geographic units in the States used in the construction of the index of geographic concentration and to control for otherwise unmeasured State fixed effects. It is possible to replace the State variables by a continuous variable given as the number of postcode areas for each State (see Lewis, 1985 for a similar approach in the context of a study of occupational segregation on the basis of gender.) Experiments show that the degree of explanation achieved with this new variable is about one-fifth less than that achieved when the State variables are used, indicating that there are other factors that account for the variation in the *G* index across the States.

Fourth, each of the socioeconomic or behavioural variables is entered in the estimating equation in linear form only. Tests were conducted for non-linearities in the underlying behavioural relationships for these variables, but in each case the inclusion of the squared term in the estimating equation resulted in both the linear and squared terms being insignificant. This indicates that the relationship being modeled is approximately linear over the range of data represented in the

---

[10] Replacing this missing information by the mean for immigrants from Lebanon from the other States yields results similar to those reported.

[11] The correlations between these variables and the State dummy variables are only between .04 and .31.

[12] In practice this can be, and was, done. Due to the high degree of collinearity discussed, relatively few of the regressors are significant in the encompassing equation.

sample of mean values for birthplace/region cells that is employed here[13]. The relationships between the index of geographic concentration and the Distance and Ethnic media variables are, however, non-linear, and this is reflected in the preferred model.

Fifth, the primary methods of analysis are Ordinary Least Squares (OLS) and Weighted Least Squares (WLS), with the latter method being employed to take account of the different sizes of the birthplace/State groups. The model is also estimated using Least Absolute Deviations (LAD) in order to assess whether the results are influenced unduly by several observations for some smaller birthplace groups that appear to be extreme data points. The Breusch-Pagan (1979) test statistic indicates there is heteroscedasticity in the residuals, and all '*t*' statistics for the least squares estimators have therefore been computed using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

5. ECONOMETRIC ESTIMATION OF THE MODEL

This section reports the econometric analyses of the multivariate model developed in Section 4. The econometric results are presented in Table 4.

The specifications considered in columns (i) and (ii) contain only the State and birthplace information. It is apparent that these variables explain a large part of the variation in the *G* index of geographic concentration, the $\overline{R}^2$ in the first column being 0.87[14]. According to the coefficients on the State variables, Queensland, Adelaide and Perth are characterised by relatively low degrees of geographic concentration. As noted above, this could be due to the use of smaller numbers of postcode areas in the construction of the *G* index in these States.

The coefficients on the birthplace dummies can be grouped into three broad categories. First, there are the "low concentration groups", that is, birthplaces with concentrations that differ by little from that of the benchmark group of the UK and Ireland. Second, there are the "moderate concentration groups", that is, birthplace groups which have a level of geographic concentration between 15 and 30 percentage points higher than the approximately 20 percent value of the *G* index for immigrants from the UK and Ireland. Finally, there are the "high concentration groups", with a level of concentration over 30 percentage points higher than that of the UK and Ireland.

---

[13] Given the use of mean values, there is little range in the data so that $\overline{X}$ and $\left(\overline{X}\right)^2$ are highly collinear. There is little improvement if VAR(*X*) is added to the specification (see footnote 9).

[14] The R^2 reported for the WLS equation is a zero-mean R^2. It is not comparable with the conventional R^2 reported for the OLS equations, which looks at the explanation of variations around the mean value of the dependent variable.

Table 4. Models of the Determinants of Geographic Concentration Among Immigrants in Australia, Males and Females of all Ages: OLS, WLS and LAD Results[a]

	OLS (i)	WLS (ii)	OLS (iii)	WLS (iv)	LAD (v)
Constant	22.753 (12.34)	20.461 (41.46)	34.278 (2.97)	34.058 (1.52)	33.243 (2.87)
Education	[b]	[b]	0.031 (0.04)	-0.471 (0.30)	0.152 (0.21)
Age at Arrival	[b]	[b]	-0.565 (3.51)	-0.845 (2.76)	-0.560 (3.84)
Duration of Residence	[b]	[b]	-0.334 (1.94)	-0.628 (1.71)	-0.294 (1.47)
Lacks English Fluency	[b]	[b]	31.240 (3.31)	35.228 (2.28)	35.618 (4.18)
Non-English-speaking Origin	[b]	[b]	3.750 (1.43)	4.146 (1.18)	4.654 (1.89)
Proportion Married	[b]	[b]	-10.647 (1.35)	13.200 (0.80)	-10.822 (1.84)
Proportion Female	[b]	[b]	8.956 (1.20)	19.829 (0.99)	5.344 (0.78)
Distance (kms/1000)	[b]	[b]	4.109 (3.75)	3.885 (2.03)	3.544 (3.07)
Distance Squared	[b]	[b]	-0.189 (3.63)	-0.197 (1.88)	-0.165 (3.12)
Ethnic Media	[b]	[b]	1.164 (1.46)	2.076 (2.43)	1.743 (1.94)
Ethnic Media Squared	[b]	[b]	-0.033 (0.47)	-0.145 (2.26)	-0.089 (1.18)
State					
Victoria	-1.003 (0.84)	-1.069 (2.02)	-1.092 (0.49)	-2.092 (1.54)	-0.058 (0.03)
Queensland	-5.650 (4.26)	-3.789 (3.68)	-3.902 (1.62)	-4.235 (2.14)	-3.059 (1.37)
Adelaide	-10.772 (7.83)	-4.270 (2.26)	-10.174 (4.35)	-7.141 (3.02)	-8.983 (3.94)
Perth	-12.652 (9.23)	-9.116 (8.78)	-12.416 (4.59)	-14.035 (3.77)	-11.624 (4.99)
ACT-Tasmania	0.020 (0.02)	0.396 (0.23)	1.998 (0.86)	-0.448 (0.13)	2.377 (0.94)

Table 4 *(contd).* Models of the Determinants of Geographic Concentration Among Immigrants in Australia, Males and Females of all Ages: OLS, WLS and LAD Results[a]

	OLS (i)	WLS (ii)	OLS (iii)	WLS (iv)	LAD (v)
Birthplace					
New Zealand	1.107 (0.56)	2.716 (3.03)	[b]	[b]	[b]
Italy	22.913 (9.44)	22.672 (34.36)	[b]	[b]	[b]
Yugoslavia	23.820 (8.76)	25.667 (16.00)	[b]	[b]	[b]
Greece	29.568 (13.71)	29.060 (21.10)	[b]	[b]	[b]
Vietnam	46.452 (17.07)	47.946 (25.91)	[b]	[b]	[b]
Germany	-0.410 (0.20)	-0.750 (0.60)	[b]	[b]	[b]
Netherlands	4.085 (1.95)	4.956 (2.25)	[b]	[b]	[b]
China	26.070 (13.94)	29.374 (19.96)	[b]	[b]	[b]
Philippines	14.395 (5.59)	18.620 (5.03)	[b]	[b]	[b]
Malaysia	21.287 (9.72)	22.188 (11.09)	[b]	[b]	[b]
Lebanon	29.961 (7.46)	39.170 (13.32)	[b]	[b]	[b]
Poland	17.370 (9.62)	17.755 (12.53)	[b]	[b]	[b]
India	14.370 (7.06)	15.787 (7.86)	[b]	[b]	[b]
Hong Kong	27.915 (12.64)	29.681 (17.04)	[b]	[b]	[b]
Malta	24.408 (6.84)	32.701 (6.32)	[b]	[b]	[b]
USA	13.502 (6.67)	14.052 (7.42)	[b]	[b]	[b]
South Africa	17.370 (5.67)	19.648 (5.44)	[b]	[b]	[b]

Table 4 *(contd).* Models of the Determinants of Geographic Concentration Among Immigrants in Australia, Males and Females of all Ages: OLS, WLS and LAD Results[a]

	OLS (i)	WLS (ii)	OLS (iii)	WLS (iv)	LAD (v)
USSR	19.800 (10.80)	20.107 (13.73)	[b]	[b]	[b]
$\overline{R}^2$	0.869	0.991[c]	0.657	0.973[c]	0.700
Breusch-Pagan Statistic	51.46	124.35	24.50	53.07	[d]
F-statistic	33.19	103.80	14.42	45.78	[d]
Sample size	113	113	113	113	113

[a] White's heteroscedasticity-adjusted '*t*' statistics in parentheses; New South Wales and UK-Ireland are the benchmark State and birthplace, respectively.
[b] Variables not entered.
[c] The R^2 is a zero-mean R^2.
[d] Not relevant.

Table 5. Categorisation of Birthplace Groups According to the *G* Index

Type of Concentration	Birthplace Groups
Low Concentration: *G* index about 20%	UK & Ireland, New Zealand, Germany, Netherlands
Moderate Concentration: *G* index around 35-50%.	Italy, Yugoslavia, Greece, China, Philippines, Malaysia, Poland, India, Hong Kong, Malta, USA, South Africa, USSR
High Concentration: *G* index over 50%	Vietnam, Lebanon

Source: Table 4.

The estimates presented in column (ii) weight each observation by the number of immigrants from the birthplace group in the particular State. These weights are scaled to the actual sample size of 113 to avoid artificial inflation of test statistics. The main impact of the weighting procedure is to reduce the State fixed effects. While the coefficients for Lebanon and Malta are increased by around 9 percentage points, all other birthplace effects are unaffected.

The estimates presented in the remaining columns in Table 4 attempt to find more fundamental reasons why there is this variation in the *G* index across the birthplaces. It does this through regressing the *G* index on a set of behavioural variables for Educational attainment, Age at arrival, Duration of residence, Limited English skills, a non-English speaking country dummy, proportion married, proportion females, the distance between the capital city of the origin

country and the capital city of the State or Territory of residence, and an Ethnic media variable. This approach is similar to that taken in Chiswick and Miller (2001), where it is shown the explanatory variables based on birthplace with behavioural interpretations provide almost as much explanatory power as a set of birthplace dummy variables in models for destination language skills.

The degree of explanation in columns (iii) and (iv) is lower than that achieved when birthplace dummies replace the behavioural variables in the model, columns (i) and (ii). However, in the equation estimated using OLS, almost two-thirds of the variation in the *G* index can be accounted for by using just the behavioural variables and State dummy variables. This suggests that underlying the birthplace dummy variables there are more fundamental socio-economic or behavioural factors that give rise to variations in the degree of geographic concentration of immigrant groups, and that these factors are amenable to formal modelling.

There are some differences between the results obtained when the various methods of estimation are applied to the model with the behavioural variables, but these are generally not statistically significant[15]. The discussion below will concentrate on the WLS results.

The results in column (iv) of Table 4 indicate that there is a strong, negative association between age at arrival and the degree of geographic concentration. As noted earlier, there are various explanations for the links between age at arrival and geographic concentration, including explanations founded on the phenomenon of immigrant adjustment, and arguments based on the person's family ties. The negative association suggests that those who arrive in Australia at an early age may have or develop ties to family and compatriots that are difficult to erode.

The negative relationship between the index of geographic concentration and duration of residence is consistent with the literature, and adds a further dimension to the process of immigrant adjustment documented by Chiswick and Miller (1995), among others. According to these results, the index of geographic concentration falls by almost two-thirds of a percentage point with each extra year of residence in Australia. In other words, a difference of about 13 percentage points is expected in the value of the *G* index for recent arrivals and the typical (average duration) immigrant in Australia.

The strongest relationship evident in this set of results is between the variable for Limited English skills and the *G* index. As the measure of "lack of English

---

[15] Weighted Least Squares (WLS) is used to account for differences in the population size of the unit of observation that may affect the reliability of the explanatory variables. The Least Absolute Deviations (LAD) estimator is used for similar reasons. In terms of differences in results, the effects of age at arrival and duration of residence are more pronounced when WLS is used in comparison with the OLS results. However, the effect of age at arrival in the LAD estimates is similar to that obtained using OLS, the effects of duration of residence is less in the LAD estimates than for either OLS or WLS, and the effects of limited English skills and a non-English speaking background are stronger in the LAD results than for OLS.

fluency" in practice ranges between 0 and 0.54 across the 113 birthplace/State units of observation, the results suggest that the *G* index would vary by about 19 percentage points between the extreme values actually represented in the data.

The coefficients on the variables for the distance between the origin country and the State/Territory of residence in Australia indicate that the degree of geographic concentration initially increases at a decreasing rate with this measure. Beyond approximately 10,000 kilometres (*e.g.*, New Delhi to Sydney), however, the degree of geographic concentration declines with distance. As the mean distance from the origin country is approximately 12,000 kilometres, it would appear that the distance variable is capturing effects other than those hypothesised. Inspection of Tables 2 and 4 suggests that it is reflecting the relatively higher value of the indices of geographic concentration among immigrants from the Asian region.

The relationship between the *G* index and the Ethnic media (number of newspapers) variable is also non-linear. As the number of newspapers increases, the *G* index initially increases. The relationship between the number of newspapers and the *G* index peaks at around 7 newspapers. The mean value of this measure is 4, with a standard deviation of 4. Hence, for most of the sample an increase in the extent of ethnic infrastructure, as measured by the number of newspapers in the origin language, is associated with a more intense concentration in the immigrant settlement pattern.

According to the estimates reported in column (iv), educational attainment, marital status, gender and non-English speaking origin are not significant determinants of variations in the degree of geographic concentration.

6. CONCLUSION

Among many immigrant groups, there is a tendency for recent arrivals to settle in areas where others from their country of origin live. It is well documented, for example, that there is a high degree of geographic concentration among Vietnamese immigrants in Australia, and a low degree of geographic concentration among immigrants from the UK and Ireland.

This study attempts to model variations in the index of geographic concentration (*G*) across birthplaces and regions using information on the mean characteristics of the birthplace groups for whom the indices have been calculated. The behavioural or socioeconomic variables considered in the analysis are for Limited English skills, Educational attainment, Duration of residence in Australia, Age at migration, Martial status, Gender, non-English-speaking origin, the distance between the country of origin and the State/Territory of current residence, and Ethnic media. Up to two-thirds of the variation in the *G* index around its mean value can be explained using these variables.

It is shown that the extent of geographic concentration of immigrant groups is negatively related to age at migration, duration of residence in Australia, and the percentage of the birthplace group that is fluent in English. There are non-linear relationships between the extent of geographic concentration and both the availability of ethnic media and the distance between the country of origin and

the place of residence in Australia. The pattern of effects discerned for the distance variable (where *G* first increases to about 10,000 kilometres and then decreases with the distance from the country of origin) suggests that it is reflecting the relatively higher value of the indices of geographic concentration among immigrants from the Asian region. For most of the sample there is a positive relationship between the degree of concentration in the settlement pattern and the number of ethnic newspapers. While the *G* index initially increases as the number of newspapers increases, it does so at a decreasing rate. This suggests the presence of diminishing returns in terms of the intensity of the ethnic infrastructure represented by the larger number of newspapers.

The results suggest that there are likely to be behavioural patterns in the immigrant settlement process that transcend the birthplace and regional categories that have been the focus of research to date. Further study of these behavioural patterns has the potential to considerably enhance understanding of the immigrant settlement process[16].

REFERENCES

Australia (1992) *Directory of Ethnic Community Organisations in Australia.* Department of Immigration and Ethnic Affairs/Australian Government Publishing Service: Canberra.

Australian Bureau of Statistics (1993) *CDATA91 with Mapinfo: 1991 Census of Population and Housing (Australia)*, Greenwich, New South Wales: Peripheral Systems. Australian Bureau of Statistics: Canberra.

Australian Bureau of Statistics (1994) *1991 Census of Population and Housing Household Sample File* [computer file]. Australian Bureau of Statistics: Canberra.

Australian Bureau of Statistics (1997) *CDATA96 with Mapinfo Professional: 1996 Census.* Troy, N.Y.: Mapinfo Corp. Australian Bureau of Statistics: Canberra.

Australian Bureau of Statistics (1998) *Migration*, Catalogue No. 3412, Australian Bureau of Statistics: Canberra.

Bartel, A.P. (1989) Where do the new U.S. immigrants live? *Journal of Labor Economics*, 7(4), pp. 371-391.

Bartel, A.P. and Koch, M.J. (1991) Internal migration of US immigrants. In: J. M. Abowd and R. B. Freeman (eds), *Immigration, Trade and the Labor Market.* The University of Chicago Press: Chicago, pp. 121-134.

Breusch, T.S. and Pagan, A.R. (1979) A simple test of heteroscedasticity and random coefficient variation. *Econometrica*, 47, pp. 1287-1294.

Burnley, I.H. (1976) *The Social Environment: A Population and Social Geography of Australia.* McGraw Hill: Sydney.

---

[16] In a companion paper in progress an attempt is made to ascertain if the empirical relationships established in this paper for the aggregates formed for the various State/birthplace cells carry over to the analysis of information on individual immigrants.

Burnley, I. (1996) *Atlas of the Australian People - 1991 Census: New South Wales*. Australian Government Publishing Service: Canberra.

Burnley, I., Murphy, P. and Fagan, R. (1997) *Immigration and Australian Cities.* The Federation Press: Sydney.

Case, A.C. and Katz, L.F. (1991) The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths. *Harvard Institute of Economic Research*, Discussion Paper Number 1555.

Chiswick, B.R. and Miller, P.W. (1995) The endogeneity between language and earnings: an international analysis. *Journal of Labor Economics*, 13(2), pp. 246-288.

Chiswick, B.R. and Miller, P.W. (2000) Do enclaves matter in immigrant adjustment? Mimeo. Department of Economics, The University of Western Australia.

Chiswick, B.R. and Miller, P.W. (2001) A model of destination language acquisition: application to male immigrants in Canada. *Demography,* 38(3), pp. 391-409.

Cutler, D.M., Glaeser, E.L. and Vigdor, J.L. (1999) The rise and decline of the American ghetto. *Journal of Political Economy*, 107(3), pp. 455-506.

Duncan, O.D. and Duncan, B. (1955a) Methodological analysis of segregation indexes. *American Sociological Review*, 20(2), pp. 210-217.

Duncan, O.D. and Duncan, B. (1955b) Residential distribution and occupational stratification. *The American Journal of Sociology*, 60(5), pp. 493-503.

Fitzpatrick, G.L. and Modlin, M.J. (1986) *Direct-Line Distances: International Edition.* The Scarecrow Press Inc.: Metuchen, N.J.

Goddard, R.F., Sparkes, L.H. and Haydon, J.A. (1985) Demographic consequences of immigration. In: Neville R. Norman and Kathryn F. Meikle (eds), *The Economic Effects of Immigration on Australia.* Committee for Economic Development of Australia: Melbourne, pp. 47-161.

Haggett, P., Cliff, A.D and Frey, A. (1977) *Locational Analysis in Human Geography*, 2nd Edition, Vol. 2, Edward Arnold: London.

Hugo, G. (1995) *Understanding Where Immigrants Live*, Australian Government Publishing Service: Canberra.

Hugo, G. and Maher, C. (1995) *Atlas of the Australian People – 1991 Census: National Overview.* Australian Government Publishing Service: Canberra.

Le, A.T. (1999) Empirical studies of self employment. *Journal of Economic Surveys*, 13(4), pp. 381-416.

Lewis, D.E. (1985) The sources of changes in the occupational segregation of Australian women. *Economic Record*, 61(175), pp. 719-736.

McManus, W., Gould, W. and Welsh, F. (1983) Earnings of hispanic men: the role of English language proficiency. *Journal of Labor Economics*, 1(2), pp.101-130.

Massey, D.S. and Denton, N.A. (1988) The dimensions of residential segregation. *Social Forces*, 67(2), pp. 281-315.

Moir, H. and Selby-Smith, J. (1979) Industrial segregation in the Australian labour market. *Journal of Industrial Relations*, 21, pp. 281-291.

Veltman, C. (1983) *Language Shift in the United States.* Moulton Publishers: Berlin.

Watts, M. (1998) Occupational gender segregation: index measurement and econometric modelling. *Demography*, 35(4), pp. 489-496.

White, H. (1980) A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, pp. 817-838.

APPENDIX A. The Herfindahl Index

Table A1. Herfindahl Indices, 1991 and 1996[a]

	1991	**1996**
Canada	0.017	0.015
Ireland	0.018	0.017
New Zealand	0.018	0.018
South Africa	0.035	0.033
United Kingdom	0.016	0.015
United States of America	0.022	0.018
Total English-speaking Countries	0.015	0.014
China	0.053	0.053
Germany	0.013	0.013
Greece	0.048	0.046
Hong Kong	0.042	0.052
India	0.031	0.031
Italy	0.037	0.035
Lebanon	0.136	0.140
Malaysia	0.055	0.039
Malta	0.131	0.113
Netherlands	0.015	0.014
Philippines	0.041	0.048
Poland	0.031	0.029
Vietnam	0.160	0.168
Total Non-English-speaking Countries[b]	0.024	0.021
Total Immigrants	0.016	0.015
Born in Australia	0.011	0.011

[a] Birthplace groups in which data are available in both the 1991 and 1996 Census; the indices listed have been multiplied by 10 for expositional reasons.

[b] Include countries not listed.

The higher the Herfindahl index the more regionally segregated the birthplace group. It is perhaps easiest to interpret the data with reference to the value of 0.011 for Australia. Birthplaces with similarly low values of the Herfindahl index are Ireland, New Zealand, United Kingdom, United States of America, Germany and the Netherlands. At the other extreme, birthplaces with high values of the Herfindahl index, indicating a more segregated settlement pattern, include South Africa, China, Greece, Hong Kong, the Philippines and, especially, Lebanon, Malta and Vietnam.

APPENDIX B. Description of Data

Population

Males and females born in Australia or overseas. The age group differs across variables. The regional data used to compute *G* are available only for the total population. For many other variables the 15 and over age bracket can be used.

Dependent Variable

Index of Geographic Concentration

This is computed as $G = \sum_{i=1}^{n} Max([g_{io} - g_{ip}], 0)$, where g_{io} is the percentage representation of the specific overseas born group in the ith geographic area, g_{ip} is the percentage representation of the total population in the ith geographic area, and n is the number of regions used in the computations. Where the overseas born group has a distribution across regions the same as the total population, *G* will equal 0. Where the overseas born group is completely segregated, the upper bound of *G* will equal (100-*GS*), where *GS* is percent of the total population accounted for by the group.

G is interpreted as the percentage of the specific overseas born group that would need to shift, together with a similar number of members of other groups in the population, to achieve the same geographic distribution as the total population. Allowing members of the other groups to replace members of the specific birthplace group ensures the overall population distribution is unaltered.

The *G* index is similar in construction to the Index of Dissimilarity (Duncan and Duncan, 1955a) that has been used in much of the Australian literature on the geographic concentration of immigrants (see, for example, Burnley *et al.,* 1997; Hugo, 1995). This is defined as $D = 0.5 \times \sum_{i=1}^{n} |g_{io} - g_{ia}|$, where g_{io} is the percentage representation of the specific overseas born birthplace group in the ith geographic area and g_{io} is the percentage representation of the Australia born in the same geographic area. This index can be interpreted as the percentage of a birthplace group (either the Australia born or the specific overseas born group) that would have to shift across regions for the geographic distributions of the two birthplace groups to be the same. Replacement of the group that moves by members of other birthplace groups is not required with this measure, which means the geographic distribution of the total population would change.

It is not always clear how the index of dissimilarity has been computed in the literature. For example, Burnley (1996) described the index as

$$D = \frac{0.5 \times \sum_{i=1}^{n} \left| g_{io} - g_{ia} \right|}{1 - \dfrac{G_o}{G_a}},$$

where G_o is the total number of the specific overseas born group, and G_o is the total number of the Australia born. This index will have a minimum of 0 where there is no dissimilarity, but the maximum could be greater than 100. As G_o is small relative to G_o (see Table 1), in practice the upper limit will be close to 100, enabling an interpretation the same as the conventional Index of Dissimilarity (Duncan and Duncan, 1955a). See also Burnley (1976, Appendix V). The adjustment in the denominator in the above index is generally required only where the benchmark group is defined as the total population (Australia born plus the specific overseas born group). See Duncan and Duncan (1955b, p.494).

The coefficient of geographic association (G) is equivalent to the modified index of dissimilarity ($0.5 \times \sum_{i=1}^{n} \left| g_{io} - g_{ip} \right|$) discussed by Moir and Selby-Smith (1979).

Independent Variables

Educational Attainment: The mean educational attainment of immigrants of specific birthplace groups aged 15 or more in each State. The age restriction is used to permit a focus on completed levels of schooling.

Duration of Residence: This variable is the mean duration of residence for immigrants of specific birthplace groups in each State. It is calculated for immigrants aged 15 or more years, these being the groups who are likely to be "decision makers" that will impact on the G index.

Age at Arrival: This variable is created from data on age and duration of residence of the overseas born. The mean is calculated within each State for immigrants aged 15 or more.

Lacks English Fluency: Individuals who speak a language other than English at home and who report their English proficiency as either "not well" or "not at all" are classified as having limited English skills. The proportion of each birthplace group in each State aged 15 or more years with limited English skills is the measure used in the analysis.

Married: This variable records, for each state, the proportion of each birthplace group aged 15 or more years who are married.

Female: The proportion of the members of the specific birthplace group that resides in each State that is female. The population for the construction of this variable is consistent with other variables, being those aged 15 or more.

Ethnic Press: This is defined as the number of newspapers associated with the particular birthplace group in each State. Where newspapers were indicated as being available in the capital city, it was assumed that they could be obtained across the entire State.

Postcodes in State: This variable is the number of postcode areas in the State used in the construction of the *G* index.

Distance: This is a measure of the distance between the capital city of the birthplace and the capital city of the State (Canberra for the ACT-Tasmania grouping).

Sources: The *G* index and number of postcodes is each state are computed from CDATA91, the Educational attainment, Duration of residence, Age at arrival, Lacks English fluency, Married, and Female variables are obtained from the 1991 Australian Census of Population and Housing, Household Sample File (Australian Bureau of Statistics 1994), the Ethnic press variable is constructed from information in Australia (1992) and the physical distance measure is obtained from Fitzpatrick and Modlin (1986).

Table B1. Means and Standard Deviations of Selected Variables[a]

Variable	Mean	Standard Deviation
Education (years)	11.842	1.63
Age at Arrival (years)	25.877	5.02
Duration of Residence (years)	19.181	7.74
Lacks English Fluency (%)	13.7	15.4
Married (%)	66.9	13.7
Female (%)	49.8	11.0
Non-English Speaking Origin (%)	78.8	41.1
Ethnic Press (Number)	4.19	3.99
Distance from Origin Country (kms)	11,813	4,352

Note: Unweighted data used in calculations.

(a) These are means and standard deviations of variables computed by state and birthplace for foreign born persons aged 15 and over.

Source: Various, see Appendix B.