Modeling the spatial distribution of the immigrant population in Greece

Stelios Zimeras[1] and Cleon Tsimbos[2]

[1]Department of Statistics and Actuarial-Financial Mathematics, University of the Aegean, Samos, Greece.
[2] Department of Statistics and Insurance Science, University of Piraeus, Greece.

Abstract

Since 1990 Greece has been transformed from a traditional sending to a receiving country. According to the 2001 census data, foreign citizens represent 7.3% of the total population of the country, while 97% of the population change, which occurred during the intercensal, period 1991-2001 is attributed to net immigration. The regional distribution of immigrants has several demographic and socio-economic implications. These effects can be studied using spatial models and techniques of visualization and exploration.

This study investigates the spatial distribution of immigrants using data from the last population census of Greece (2001). The statistical analysis involves the study of various socio-economic and demographic variables for the 51 administrative departments of the country. Multivariate techniques are used to identify clusters of low/high indicators and neighborhood structures of the regions. Spatial statistical analysis methods, like spatial autocorrelation and Moran coefficients, are employed for the final presentation and interpretation of the results.

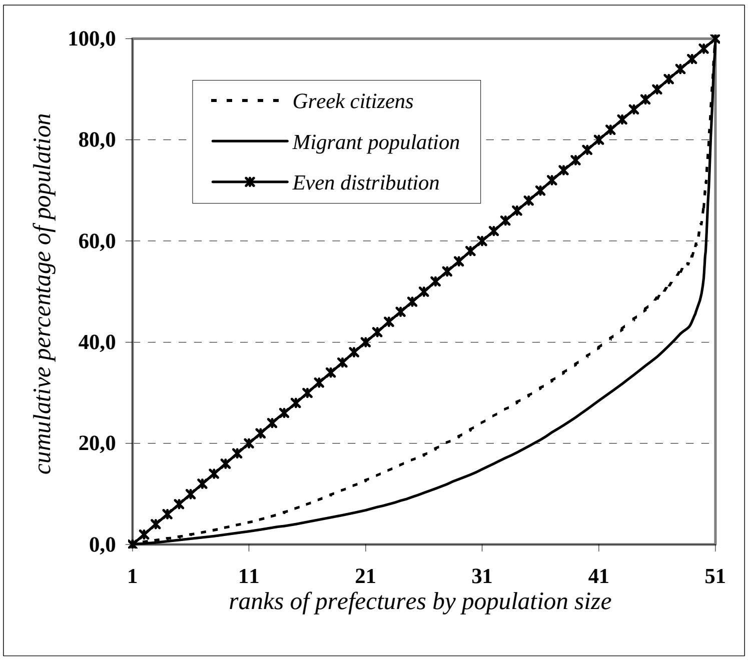
INTRODUCTION

Migration has been always an important population component for Greece. Each historic period is characterized by different types of movements and pathways but at all times it has been associated with the prevailing national and international economic and political conditions (Psimmenos and Georgoulas, 2001). The period of intense Greek emigration (1945-1973) – mainly to W. Germany but also to the USA, Canada and Australia – was followed by a period of relatively intense repatriation of Greek refugees as well as economic migrants (1974-1985). In the former period unemployment, poor economic conditions and political instability operated as strong push factors for outflows; in the latter, it was the stable political situation and the economic development of the country together with the new labor market conditions in the Western European countries that played a significant role (Fakiolas, 2000, King, 2000). However, since 1974 not only return migration but also the first immigration flows of foreign workforce have been observed. The entry of Greece to the EEC made possible the implementation of ambitious plans for economic, social and political harmonization and liberalization that encouraged geographical mobility (Psimmenos, 1997). The existence of a large informal Greek economy and a large service sector run according to the internal demands for cheap and seasonal-flexible labour (Iosifidis and King, 1998); these structural changes resulted in the first migration intakes from the Maghreb countries in the 1970s and from the Far East (mainly Fillippines), Balkan countries and the Eastern European regions in the 1980s.

Of course these trends could not predict the influx of immigrants arriving in the country since the early 1990s; this new migration movement took place after the collapse of the former USSR, the liberalization of Eastern European countries and the shift in the political regime in Albania. Greece as part of the system of Globalization and the post-1989 New World Order received a considerable number of documented and undocumented immigrants (King, 2002) and underwent a transformation process from a traditional emigration to a new immigration country. Due to the geographical proximity, the morphology of the Greek borderlines and linguistic-religious agencies (Kassimati, 1998) the bulk of the immigrants originate from Albania (57.5%) and other Balkan countries (8.0%) as well the former USSR (9.2%).

According to the 2001 census, foreign citizens account for 797,091 persons representing 7.3% of the total population of Greece, a figure 3.8 times that of the respective share observed a decade earlier. The official demographic data indicate that since the early 1990s migration has been the most powerful driving force behind population growth in Greece, with 97.3% of the population change of the intercensal period 1991-2001 being attributed to net immigration; the age-sex composition of the migrant population (young age structure for both genders, relatively high proportion of males) carry important demographic effects (Tsimbos, 2006) and socio-economic implications for the national insurance system (Bagavos, 2003).

As international experience indicates, migrants tend to concentrate in a limited number of regions in the host country; the spatial distribution of immigrants seems to be rather uneven but different migrant groups usually exhibit different patterns of geography (Bucher, 1996, Coleman, 1994, Van der Gaag and Van Wissen, 2001, White, 1993). According to the 2001 census, 48.6% of the non-Greek citizens reported as usual place of residence the region of Greater Athens and another 8.8% the department of Thessalonica (the second largest city of Greece). As a general rule, migrants tend to prefer urbanized areas. Lianos (2003) analyzing data based on the 1998 Migrant Legalization Program and socio-economic statistics for the 51 prefectures of the country has found that the urbanization level, the relative volume of the primary sector of the economy and the per capita gross national product form significant pull-factors for migration inflows. Rovolis and Tragaki (2005) applying regression analysis procedures to cross-sectional census data for the administrative divisions of the country have resulted in similar findings and in addition they demonstrate the importance of the construction and tourism sectors. The differences in the regional distribution of the Greek and the migrant populations are presented in Figure 1 in which the ranks of the prefectures of the country (according to their population size) are plotted against the corresponding cumulative percentages of the population. Taking into account the population density of the 51 regional divisions of the country and following the procedure presented by Shryock et al. (1975) it is estimated that the Gini Concentration Ratio is 0.52 for the Greek population and 0.68 for the migrants (the corresponding Indices of Spatial Dissimilarity are 40.8% and 53.0% respectively).



Source: Elaboration based on the 2001 census of Greece

Figure 1: Lorenz curves comparing the regional distribution between Greek citizens and immigrants: 2001 census of Greece

DATA AND VARIABLES

In this paper we investigate the spatial distribution of immigrants using data from the last population census of Greece (2001). In our analysis, in order to distinguish the immigrant population we adopt the concept of citizenship; in spite of the potential complications due to the inclusion of naturalizations in such data (Van der Gaag and Van Wissen, 2002) the very nature of the available data dictates to the choice of such a definition.

For each prefecture of the country ($i = 1,2,...,51$) 78 indices were calculated expressing the demographic, socio-economic and geographical characteristics of the Greek and the migrant population. Following the procedure described in the next section, it was found that the following variables played an important role:

1. **Net migration rate (**net_i**):**

$$net_i = \frac{\left(P_i^{2001} - P_i^{1991}\right) - \left(B_i^{1991-01} - D_i^{1991-01}\right)}{\frac{1}{2} \cdot \left(P_i^{1991} + P_i^{2001}\right)} \cdot 100$$

where: P_i^{1991}, P_i^{2001} the resident census population of region i in 1991 and 2001 respectively, $B_i^{1991-01}$ the number of livebirths (classified by the usual place of residence of mother) registered in region i during the intercensal period 1991-2001, and $D_i^{1991-01}$ the number of deaths (classified by the usual place residence of the descendent) registered in region i during the intercensal period 1991-2001.

2. **National growth migration rate (**$ngmrr_i$**):**

$$ngmrr_i = \left[\frac{P_i^{2001} - P_i^{1991}}{P_i^{1991}} - \frac{P_T^{2001} - P_T^{1991}}{P_T^{1991}} \right] \cdot 100$$

where: P_T^{1991}, P_T^{2001} the total resident census population of the country in 1991 and 2001 respectively and P_i^{1991}, P_i^{2001} as above.

3. **Percentage of male population in 2001 (** $males_i$ **):**

$$males_i = \frac{MP_i^{2001}}{P_i^{2001}} \cdot 100$$

where: MP_i^{2001} the male resident census population of region i in 2001 and P_i^{2001} as above.

4. **Percentage of male migrant population in 2001 (** $mgmales_i$ **):**

$$mgmales_i = \frac{m_i^{2001}}{M_i^{2001}} \cdot 100$$

where m_i^{2001}, M_i^{2001} the number of male migrants and the total male population of the region i in 2001 respectively.

5. **Migrants per 100 population in 2001 (** $mgrate_i$ **):**

$$mgrate_i = \frac{M_i^{2001}}{P_i^{2001}} \cdot 100$$

6. **Percentage of Albanian migrants in 2001 (** $albanians_i$ **):**

$$albanians_i = \frac{A_i^{2001}}{M_i^{2001}} \cdot 100$$

where: A_i^{2001} the number of Albanian migrants of region i in 2001 and M_i^{2001} as above.

7. **Percentage of population aged 15-64 in 2001 (** pwa_i **):**

$$pwa_i = \frac{P_i^{15-64}}{P_i^{2001}} \cdot 100$$

where: P_i^{15-64} the resident population aged 15-64 of region i in 2001 and P_i^{2001} as above.

8. **Percentage of population aged 65 and over in 2001 (** old_i **):**

$$old_i = \frac{P_i^{65+}}{P_i^{2001}} \cdot 100$$

where: P_i^{65+} is the resident population aged 65+ of region i in 2001 and P_i^{2001} as above.

MULTIVARIATE TECHNIQUES

At the first stage, 12 socio-economic and demographic variables were analysed. Due to the large number of variables, the use of statistical methodologies comprising multivariate analysis techniques (Mardia et. al., 1979) may elucidate the variables of overriding importance for the complex problems under investigation and depict correlations. Multivariate statistics facilitate the analysis of complex sets of data. Multivariate techniques are recommended when there are many independent and

possible dependent variables, which are correlated to each other to varying degrees. It reduces attribute space from a larger number of variables to a smaller number of factors. In many scientific fields, particularly behavioral and social sciences, variables cannot be measured directly. Such variables, called latent variables, can be measured by other 'quantifiable' variables, which reflect the underlying variables of interest. Factor analysis attempts to explain the association between variable values in terms of the underlying factors, which are not directly observable.
Principal components analysis (PCA) is a statistical technique applied to a single set of variables to discover which sets of indicators in the set form coherent subsets that are relatively independent of one another. PCA is used as a tool to reduce a large set of variables to a more meaningful structure. The main purpose of PCA is to reduce the dimensionality from p to d, where $d<p$, while at the same time accounting for as much as the variation in the original data as possible. With PCA, we transform the data to a new set of coordinates that are a linear combination of the original variables. The observations in the new principal component space are uncorrelated. The idea behind this procedure is that we can gain information and understand better the structure of the data in the new space.
The factor analysis model extracts only that proportion of variance, which is due to the common factors and shared by several items. The proportion of variance of a particular item that is due to common factors (shared with other items) is called communality. The proportion of variance that is unique to each item is then the respective item's total variance minus the communality.
Following the above methodology, eight variables are proposed that are related to the regional distribution of immigrants. For each one, the communalities are estimated (Table 1) suggesting that the these components describes the main characteristics of the immigrant population in Greece

Variables	**net**	**mgmrr**	**males**	**mgmales**	**mgrate**	**albanias**	**pwa**	**old**
Communalities	0.905	0.926	0.993	0.757	0.936	0.723	0.992	0.998

Table 1: Calculation of the communalities for each variable using PCA.

Thus, instead having 12 variables, based on the multivariate analysis we proposed 4 groups (factors F_i), which explain the 90.35% of the initial information; these factors with the appropriate explanations are:
1st factor F_1: **age structure** (variables: pwa, old)
2 d factor F_2: **population growth** (variables: net, mgmrr, mgrate)
3 d factor: F_3: **ethnic and migrant sex composition** (variables: albanians, mgmale)
4 th factor F_4: **sex composition of the population** (variable: male)

The sum of the squared factor loadings could be defined as the weight for the particular prefecture of the country ($i = 1,2,...,51$).

EXPLORATORY SPATIAL DATA ANALYSIS

Various techniques have been suggested to investigate the nature and extent of spatial correlation between demographic variables. These techniques constitute what is now called exploratory spatial data analysis (ESDA) by reference to exploratory data analysis (EDA) proposed by Tukey in the late 1970s. These techniques include visual and quantitative methods to summarize the spatial properties of a variable, to describe its specific patterns in space, spot extreme values or outliers, and to identify specific geographical subsets. The availability of data in a GIS (geographic information system) format allows the systematic spatial exploration of the data.

The general idea behind these techniques is the examination of the nature of spatial variation (referring as spatial auto-correlation) between values of the same variable at different spatial locations (referring as neighborhood observations). Once the concept of “neighboring observations” is defined, the correlation between neighbors may be compared to the general variance of the sample in the same way as in ordinary correlation analysis. The resulting measure of spatial autocorrelation is a first indication of the spatialized nature of the phenomenon studied: this correlation may be non-existent, low or strong according to the variables used.

Moran introduced in 1950 the first measure of spatial autocorrelation in order to study stochastic phenomena, which are distributed in space in two or more dimensions. Moran's index has been subsequently used in almost all studies employing spatial autocorrelation. Moran’s I is used to estimate the strength of this correlation between observations as a function of the distance separating them (correlograms). Like a correlation coefficient the values of Moran's I range from +1 (meaning strong positive spatial autocorrelation) to 0 (meaning a random pattern) and to –1 (indicating strong negative spatial autocorrelation); negative autocorrelation is extremely unusual in social sciences. Values near +1 indicate similar values tend to cluster; values near -1 indicate dissimilar values tend to cluster; values near -1/(n-1) (which goes to 0 as n gets large) indicate values tend to be randomly scattered. Since spatial data are easily mapped, it is thus only natural that techniques have been developed for generating and mapping local counterparts to many global measurements.

The definition of Moran’s I (Anselin, 1995) for a spatial proximity matrix w_{ij} for a variable y_i at location *i* is defined below as

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)} = \frac{\sum_{i\neq j} w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{n\sigma^2(y)}$$

where $\sigma^2(y)$ is the sample variance. Usually, the proximity matrix w_{ij} is everywhere 0 except for contiguous locations *i* and *j* where it takes the value 1. Based on the proximity matrix computation for different distances could be take place. This provides a complete correlogram of spatial autocorrelation by distance class and the impact of distance on the strength of spatial autocorrelation for each variable can be examined. Note that values at neighboring lags of a correlogram are highly correlated, since the correlation at larger lags is in part a function of correlations at smaller lags.

The basic geostatistical description of chosen variables used here is shown in the following figures. For mapping the basic indices we have used a three-group classification of colours (low, middle and high), which are based on the percentiles of these variables. Figure 2 illustrates the spatial distribution of national growth migration rate ($ngmrr_i$) in Greece. It is clear that most regions of central and northern Greece exhibit negative or low net migration rates as a result of very low fertility as well as high internal migration rates. The global Moran’s I statistic is 0.203 indicating positive spatial correlation (not very strong clustering).

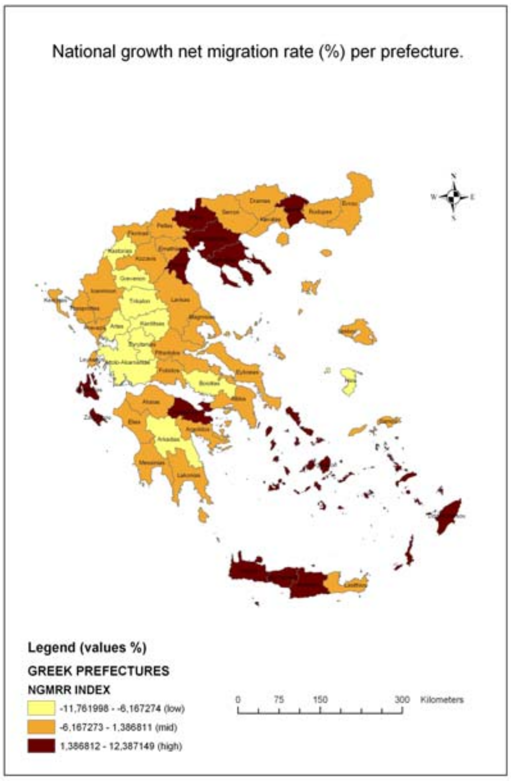


Figure 2: Spatial distribution of national growth migration rate ($ngmrr_i$)

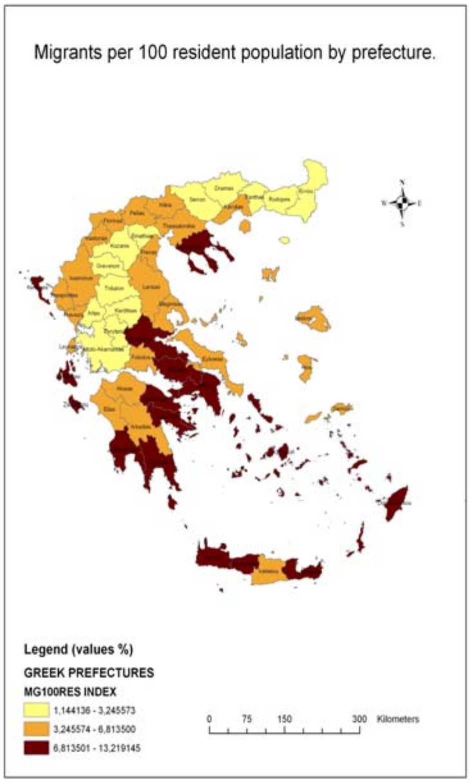


Figure 3: Spatial distribution of migrants per 100 populations in 2001 ($mgrate_i$)

Figure 3 illustrates the spatial distribution of migrants per 100 population in 2001 ($mgrate_i$) in Greece. It is clear that most regions of central and northern Greece exhibit low migration rates. The global Moran's I statistic is 0.2951 indicating positive spatial correlation (not very strong clustering).

Figure 4 illustrates the percentage spatial distribution of Albanian migrants in 2001 ($albanians_i$) in Greece. Very high percentages of Albanian migrants (71.8% or over) are observed in the central and northern Greece as well as in the Ionian Islands; these are regions which are rather close to Albania or regions with high economic activity in the primary sector (agriculture) or the construction sector. On the other hand, rather low percentages are observed in Crete and other islands as well as southern Greece, i.e. areas that have great distance from Albania. The global Moran's I statistic is 0.464 indicating a strong positive spatial correlation (clustering of like values).

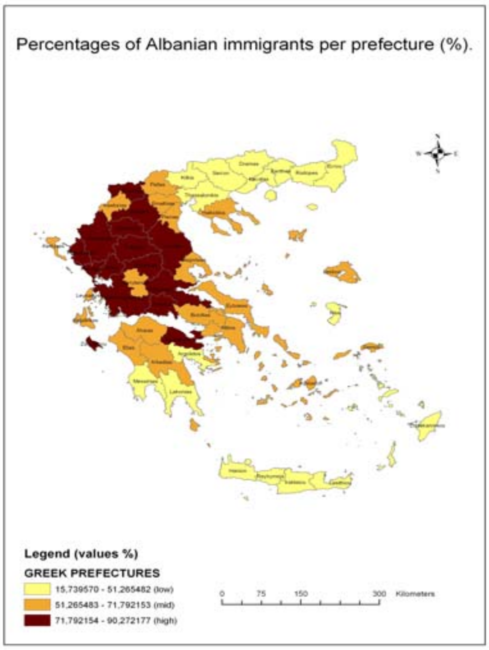


Figure 4: Spatial distribution of percentage of Albanian migrants in 2001 $(albanians_i)$

CONCLUSIONS

In this work we focus on the recent migration intakes to Greece. The regional distribution of immigrants has several demographic and socio-economic implications. These effects can be studied using spatial models and techniques of visualization and exploration. We investigate the spatial distribution of immigrants using data from the last population census of Greece (2001). In our analysis, in order to study the regional migrant population patterns, various demographic and socio-economic variables are considered. Multivariate techniques are used to identify clusters of low/high indicators and neighborhood structures of the regions. Spatial statistical analysis methods, such as spatial autocorrelation and Moran coefficients, are employed for the final presentation and interpretation of the results. Results are then used to provide a comparative analysis of spatial patterns of demographic data: As the measurement of spatial autocorrelation of demographic indices is still uncommon, comparative analysis is crucial to understand the spatial structure of different demographic factors including the national growth migration rate, net migration rate, the proportion of male population and indices of age composition.

ACKNOWLEDGE

The authors would like to thank Mr. Vaggelis Kontzamanoglou from National Statistical Service of Greece, for his useful involvement for the mapping of the data using ArcGis 9.

REFERENCES

Anselin, L., 1995, “ Local indicators of spatial association - LISA”, *Geographical Analysis*, Vol. 27, n°2.

Bagavos, Ch. (2003) *Demographic Shifts, Labour Market and Pension in Greece and European Union*, Athens: Gutenberg.

Bailey, T. C., and Gatrell, A. C., 1995, *Interactive Spatial Data Analysis*, Longman, Harlow.

Bucher, H. (1996) "Migration and the Impact on the Internal Distribution and Redistribution of Population", in M. Frey and U. Mammey (eds.): *Impact of Migration in the Receiving Countries*, Germany, pp. 45-47, Geneva: International Organization for Migration.

Coleman, D. (1994) "International Migrants in Europe: Adjustment and Integration Process and Policies", in United Nations Economic Commission for Europe: *International Migration: Regional Process and Responses*, pp. 41-76, New York and Geneva: United Nations.

Fakiolas. R. (2000) "Migration and Unregistered Labour in the Greek Economy". in R. King et al.. eds.: *Eldorado or Fortress? Migration in Southern Europe*. pp. 57-79. London: Macmillan.

Iosifidis. Th. and R. King (1998) "Socio-spatial Dynamics and Exclusion of three Immigrant Groups in Athens Conurbation". South European Society and Politics. 3(2): 205-227.

Kassimati. K. (1998) "Pontioi Metanastes stin Ellada kai Koinonikos Apokleismos". in K. Kassimati. ed. *Koinonikos Apokleismos: I elliniki empeiria*. pp. 275-307 (in Greek). Athens: Gutenberg..

King. R. (2000) "Southern Europe in the Changing Map of Migration" in R. King et al.. eds.: *Eldorado or Fortress? Migration in Southern Europe*. pp. 1-27. London: Macmillan.

King. R. (2002) "Towards a New Map of European Migration". *International Journal of Population Geography*. 8: 89-106.

Lianos, T. (2003) *Contemporary Migration to Greece: Economic Aspects,* in Greek, in Greek, Athens: Center of Economic Studies and Planning.

Mardia K.V., Kent J. I. και Bibby M. (1979): Multivariate Analysis, *1st Edition, Academic Press.*

Psimmenos. I. (1997) *Globalisation and Employee Participation*. Aldershot: Avebury Press.

Psimmenos. I. and Georgoulas S. (2001) Migration Pathways: A Historic. Demographic and Policy Review of the Greek Case. IAPASIS Project "*Does implementation matters? Informal administration practices and shifting immigrant strategies in four member states*". Brussels: European Commission.

Rovolis, A, and A. Tragaki (2005) “The Regional Dimension of Migration in Greece: Spatial Patterns and Causal Factors”, Working Paper, 45th Congress of the European Regional Science Association, “Land Use and Water Management in a Sustainable Network Society”, Vrije Universiteit Amsterdam.

Shryock, H. S., et al. (1975) *The Methods and Materials in Demography*, Vol. I, Chapter 6, U.S. Bureau of the Census, Washington: U.S. Government Printing Office.

Tsimbos. C. (2006) “The Impact of Migration on Growth and Ageing of the Population in a New Receiving Country: the Case of Greece”. *International Migration* (forthcoming).
Van der Gaag, N. and L. Van Wissen (2001) “Determinants of the Subnational Distribution of Immigration”, *Journal of Economic and Social Geography* 92(1): 27-41.

Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley Publishing.

Van der Gaag. N. and L. Van Wissen (2002) “Modeling Regional Immigration: Using Stocks to Predict Flows”. *European Journal of Population* 18: 387-409.

White, P. (1993) “Immigrants and the Social Geography of European Cities”, in R. King (ed.): *Mass Migration in Europe, the Legacy and the Future*, pp. 47-66, London: Belhaven Press.