Pergamon

S0264-2751(96)00003-0

Cities, Vol. 13, No. 3, pp. 165–174, 1996

0264-2751/96 $15.00 + 0.00

The demographic profiles of the world's largest cities

A baseline analysis and policy implications

Xiangming Chen

Department of Sociology (M/C 312), University of Illinois at Chicago, 1007 West Harrison Street, Chicago, IL 60607-7140, USA

This paper examines the demographic profiles of 74 of the world's largest cities through cluster analysis, which classifies the cities into distinctive groups based on population size, net population change, rate of natural increase, sex ratio, age composition, and dependency ratios. Discussions of the results focus on: the strikingly different demographic profiles of the large cities in more developed countries (MDCs) and less developed countries (LDCs); the simultaneous similarities and differences among MDC cities and socialist and former socialist cities; and the particular demographic characteristics of certain MDC and LDC cities. The paper also draws research implications for the data and method used and urban policy implications for the demographic and socioeconomic conditions in very large MDC and LDC cities. Copyright © 1996 Elsevier Science Ltd

Demographic profiles of cities reflect their basic socioeconomic conditions and past and future growth tendencies. Population size is a basic indicator of a city's status and functional influence in an urban hierarchy. A city in economic decline through the loss of jobs usually has slow or even net negative population growth, as some older industrial central cities in the USA have experienced over the past two or three decades. A city with a disproportionately large young (or old) population may shoulder a burden in terms of labour supply and social support. Thus a focused analysis of cites' demographic profiles provides a basic insight into their more systemic features and carries implications for urban policy and planning.

This paper fits loosely with world city research which has emerged from the general comparative urban literature into a distinctive body of theoretical and empirical scholarship. The theoretical works (Friedmann and Wolff, 1982; Friedmann, 1986; Sassen, 1991) suggested that world cities are a small number of centres at the top of the global urban hierarchy that play crucial roles in organizing and controlling the world economy. The empirical studies are split between comparative case studies of a handful of the world's largest and economically influential cities (Dogan and Kasarda, 1988a; Hall, 1984; Markusen and Gwiasda, 1994; Sassen, 1991), and cross-national quantitative analyses of some attributes of the world's major cities (Frey, 1988; Sufian, 1993) and of the air travel linkages among a score of world cities (Smith and Timberlake, 1995). Instead of testing the world city hypothesis, which is impossible with the data and cluster analysis here, this study takes an inductive approach to generating broad and systematic comparative demographic evidence for theorizing about world cities. It makes an empirical contribution similar to Sufian's (1993) focused analysis of quality of life in 98 of the world's largest metropolitan areas.

The premise of this study is that the demographic profiles of cities in developed and developing countries differ considerably. In developed countries, large central cities generally experience slow or no growth, or even decline as a result of a low birth rate and rapid suburbanization of population and economic activities, whereas central cities, especially very large or primate cities in developing countries, continue to grow, many rapidly, because of a relatively high birth rate, large in-migration and little deconcentration of economic and residential activities. The general demographic differences between developed and developing countries, however, may vary further within either group of countries. First, the demographic profiles of cities in developed countries may differ across regions, such as the sunbelt vs the snowbelt in the USA (see Frey, 1990; Sawyer and Tabb, 1984). Between country differences in Europe produced varied patterns of city growth within regions (Cheshire, 1990; Hall and Hay, 1980). Second, demographic profiles of cities may vary across countries of contrasting political and economic systems (capitalist vs socialist). Third, developing countries experience uneven industrialization, which may be associated with varied demographic profiles of their cities.

The above assumptions are addressed by showing, through cluster analysis, how the demographic profiles of the world's largest cities vary across countries and regions, and how certain variables differentiate these cities' demographic profiles. The next section introduces the data, sample, indicators and cluster analysis. The third section presents the results, which are discussed in the fourth section. The last section draws research and policy implications.

Data, sample, indicators and cluster analysis

Data source

The data were derived from *Book of World City Ranking* (Marlin *et al*, 1986), which compiled fairly complete demographic indicators on 105 major cities in 50 countries around 1980. Although the data are a little dated, they provide an adequate retrospective and prospective baseline for studying the demographic profiles of these cities. The data not only reflect the cumulative yet varied demographic consequences of continued suburbanization and metropolitanization in developed countries and of the rapid growth of large cities in developing countries prior to 1980 (see Dogan and Kasarda, 1988b) but also serve as a reference point for further comparative analysis based on more recent data.

The unit of analysis and sample

Any cross-national comparative study of cities must deal with the problem of what constitutes a city in different countries. Marlin *et al* (1986) differentiated their data according to three different urban units – central city, metropolitan area and region. For the majority of cities, data are available for central cities governed by a mayor or its legal equivalent. For a small number of cities, data refer to either the metropolitan area or the region.[1] The lack of standard cross-national definition and data comparability requires a careful decision on the unit to be cluster analysed.

The central city was chosen as the unit of analysis for the following reasons. First, the existing literature has a heavy emphasis on metropolitan areas in developed countries where suburbanization is far advanced and central cities and suburban rings are closely connected. Central cities in developing countries, on the other hand, are much less integated with their suburban peripheries. Focusing on central cities maximizes the contrast between the two groups of countries. Second, although the definition of central city varies from country to country, it has greater between-case comparability than the more varied and ambiguous definition of the broader entity of metropolitan area across countries. Third, since the majority of the indicators in the data source pertain to the central city, using it as the unit of analysis maximizes the number of cases to be studied and the potential for broader generalizations.

Selection of the final set of cities for cluster analysis involved the following procedures. First of all, the cities with missing values on the demographic indicators were excluded. Then those cities whose demographic information was collected for either metropolitan area or region (other than central city) were removed. This double exclusion trimmed the 105 cities down to 74, which are listed in Table 1 by country, region, and continent. The large share of USA, West European and Japanese cities weights the sample toward developed countries. But the presence of cities in developing, socialist. and former socialist countries in the sample allows a broad comparison across countries of varied development levels and political systems. Although not randomly selected from the global urban system, this 'convenient' sample provides a rare opportunity to examine the demographic profiles of the world's largest cities fully.

Demographic indicators

Eleven demographic indicators are used as input measures for cluster analysis:

(1) *POP*: population of central cities;
(2) *APOPC*: annual net population change (%);
(3) *BIRTH*: births per 100 population (%);
(4) *DEATH*: deaths per 100 population (%);
(5) *RNI*: rate of natural increase (birth minus death, %);
(6) *SEXR*: sex ratio (number of males per 100 females);
(7) *YGPOP*: % younger population (ages 0–19[2]);

[1]The central city represents the legal city boundaries. In the case of London, for example, the central city refers to Inner London, including the downtown London boroughs. The metropolitan area refers to the cental city plus its suburban ring, the principal urbanized area of a city. An example is Greater Athens. The region refers to the metropolitan area plus the exurban ring within its sphere of economic influence, such as London's southeast region (Marlin *et al*, 1986, p 6).

[2]Marlin *et al* (1986) defined the younger segment of the population as those aged 0–19. The most common age group s 0–14, although an alternative range is ages 0–18 (Shryock *et al*, 1976, p 133). Using ages 0–19 slightly overestimates the dependency of the younger population on the working population, as a certain proportion of the 15–19 age group in MDCs and LDCs, especially the latter, is in the labour force.

(8) *ODPOP*: % older population (ages 65 and over[3]);
(9) *ADPENR*: age dependency ratio $((P_{0-19} + P_{65+})/P_{20+64} \times 100)$;
(10) *CDPENR*: child dependency ratio $(P_{0-19}/P_{20-64} \times 100)$;
(11) *ODPENR*: old-age dependency ratio $(P_{65+}/P_{20-64} \times 100$[4]).

The populations of these cities ranged from Mexico City's 9.7 million to Tel Aviv's 0.3 million, with a mean of 2.3 million. By these statistics, the cities constitute a sample of the world's largest cities, which enhances comparability among the cases but limits generalization based on size alone. *APOPC* refers to the net annual population change in these cities for a minimum of one year around 1980, with data on the overwhelming majority of the cities being bounded between 1978 and 1982. While *APOPC* measures the cities' population growth or decline including through the effect of in- and out-migration, indicators (3) to (5) gauge the portion of that growth or decline accounted for by natural increase and its birth and death components. Indicators (6) to (8) are self-explanatory. *ADPENR* measures the ratio of the population of dependent ages to the working population, while the last two indicators differentiate the age dependency ratio into its child and old-age components. The pairing of indicators (7) and (10) and (8) and (11) reflect the types of dependent populations of the cities in relation to their total dependent populations.

Cluster analysis

Cluster analysis refers to statistical procedures for classifying data points into distinct groups in such a way that the distance among observations within each group is minimized and that among groups maximized. In hierarchical cluster analysis, each observation begins in a cluster by itself. The two closest clusters are merged to form a new cluster replacing the two old clusters. This series of successive fusions or merging of two closest clusters is repeated at multiple stages until all the observations in a data set constitute one big cluster (Everitt, 1974; SAS Institute, 1985).

The primary issue concerning cluster analysis is to determine the 'ideal' and optimal number of clusters. Milligan and Cooper (1985) evaluated 30 procedures (or the so-called stopping rules) for determining the number of clusters in a data set. A common approach is to examine the values and plots of what are known as the cubic clustering criterion (CCC), pseudo F and t^2 statistics.[5] These criteria, however, require the assumption that the clusters in the data are compact and multivariate normal. Since this assumption may not hold, clustering methods based on non-parametric probability density estimates should be considered.

Integrating both non-parametric and parametric methods, this analysis proceeded in two stages, the first of which involved three steps. The first step was to standardize all 11 measures with a mean of 0 and standard deviation of 1 due to their variations in scale and metric. Such standardization prevented any variables from biasing the clustering process with sheer weights, and would facilitate comparative interpretation. Second, two non-parametric methods – *density linkage* and *two-stage density linkage* – were employed to find out the possible number of clusters. Both methods suggested 18, 8, 5, 3 and 2 clusters. Third, seven parametric methods (see Everitt, 1974; SAS Institute, 1985, pp 263–269) were applied consecutively to the data. Then the CCC, pseudo F and t^2 statistics generated by each method (in both numeric and graphic forms) were examined closely. Balancing all these criteria, 18, 13, 8, 5 and 3 clusters were determined as the final clustering solution.[6]

The second stage of the analysis involved the identification and determination of which variables are 'dominant' in structuring and characterizing the clusters. A 'dominant' variable is one that distinguishes a cluster by deviating from the averages of the variables for the population (all 74 cities). The statistical test[7] in Berlage and Terweduwe (1988) was used to identify the 'dominant' variables.

[3]For 44 (60%) of the 74 cities, the older or retired population refers to those aged 65+, while 60+ is the cut off for the other cities.

[4]Indicators 9, 10, 11 were constructed using the formulae in Shryock *et al* (1976, pp 130–142). With a value range between 0 and 100, they measure the number of dependents supported by 100 people of working age, generally ranged 20–64.

[5]The cubic clustering criterion (CCC) is a test statistic available in the SAS programming package. It functions as a stopping rule in that its maximum value across the hierarchy levels indicates the optimum number of clusters in the data. The CCC competes well against 29 other stopping rules, ranking 6th on determining the 'correct' number of clusters (Milligan and Cooper, 1985; Sarle, 1983). The maximum values on pseudo F and minimum values on t^2 suggest possible clusters.

[6]By setting the maximum number of clusters at 25, no method suggests possible number of clusters beyond 18. The Average Linkage method suggested 18, 7 and 3 clusters. Ward's Minimum Variance method suggested 12, 8, 6, and 3; the Centroid method 14, 9, 6, and 3; the Complete Linkage 13, 8, 5, and 3; the Flexible-Beta method 13, 9, 6, and 3; the McQuitty method 18, 10, 4, and 2, and the Median method 18, 10, 7 and 2. Considering the numbers of clusters suggested by the two non-parametric methods, there is almost consensus that the cities may be 'correctly' classified into 18, 8 and 3 clusters. As two methods suggested 13 clusters, and one method each 14 and 12 clusters respectively, the analysis should include a 13-cluster solution to avoid the big drop between the 18-cluster hierarchy and the subsequent single-digit clustering. Although there are competing suggestions for both a 5- and 6-cluster solution, the analysis went along with the 5-cluster hierarchy after comparing the descriptive statistics on both.

[7]Based on the distribution of z-scores, the test takes the following form: $z_i = (\bar{x}_i - \mu_i)/\sigma_i/\sqrt{n})$ where $\bar{x}_i$ = the mean with respect to each variable for the cluster, μ_i = the mean with respect to each variable for the population, σ_i = the standard deviation with respect to each variable for the population and n = the number of

Table 1 The name, country, region and continent of 74 of the world's largest cities

City name[a]	Country[b]	Region[c]	Continent[c]
Baltimore	United States (MDC)	North America	North America
Boston	United States (MDC)	North America	North America
Chicago	United States (MDC)	North America	North America
Dallas	United States (MDC)	North America	North America
Detroit	United States (MDC)	North America	North America
Houston	United States (MDC)	North America	North America
Los Angeles	United States (MDC)	North America	North America
New York	United States (MDC)	North America	North America
Philadelphia	United States (MDC)	North America	North America
Phoenix	United States (MDC)	North America	North America
San Diego	United States (MDC)	North America	North America
San Francisio	United States (MDC)	North America	North America
St Louis	United States (MDC)	North America	North America
Washington, DC	United States (MDC)	North America	North America
Calgary	Canada (MDC)	North America	North America
Winnipeg	Canada (MDC)	North America	North America
Birmingham	United Kingdom (MDC)		
Edinburgh	United Kingdom (MDC)		
Glasgow	United Kingdom (MDC)		
Leeds	United Kingdom (MDC)	Northern Europe	Europe
Liverpool	United Kingdom (MDC)	Northern Europe	Europe
Manchester	United Kingdom (MDC)	Northern Europe	Europe
Dublin	Ireland (MDC)	Northern Europe	Europe
Copenhagen	Denmark (MDC)	Northern Europe	Europe
Helskini	Finland (MDC)	Northern Europe	Europe
Milan	Italy (MDC)	Southern Europe	Europe
Naples	Italy (MDC)	Southern Europe	Europe
Rome	Italy (MDC)	Southern Europe	Europe
Madrid	Spain (MDC)	Southern Europe	Europe
Amsterdam	Netherlands (MDC)		Europe
The Hague	Netherlands (MDC)	Western Europe	Europe
Rotterdam	Netherlands (MDC)	Western Europe	Europe
Lyon	France (MDC)	Western Europe	Europe
Marseille	France (MDC)	Western Europe	Europe
Paris	France (MDC)	Western Europe	Europe
Düsseldorf	West Germany (MDC)	Western Europe	Europe
Frankfurt	West Germany (MDC)	Western Europe	Europe
Hamburg	West Germany (MDC)	Western Europe	Europe
Munich	West Germany (MDC)	Western Europe	Europe
West Berlin	West Germany (MDC)	Western Europe	Europe
Brussels	Belgium (MDC)	Western Europe	Europe
Vienna	Australia (MDC)	Western Europe	Europe
Bucharest	Romania (FSC)	East Central Europe	Europe
Budapest	Hungary (FSC)	East Central Europe	Europe
East Berlin	East Germany (FSC)	East Central Europe	Europe
Prague	Czecholovakia (FSC)	East Central Europe	
Warsaw	Poland (FSC)	East Central Europe	

Table 1—continued opposite

Results

The boxes in Figure 1 represent clusters of cities at the five stages of 18, 13, 8, 5 and 3 clusters. The arrows indicate the aggregation or joining of new and old clusters. As the hierarchical clustering moves through the first three stages, clusters of cities with similar demographic profiles join to form new clusters. Cluster 1b absorbs 1 to 4 and 1a to 3a, whereas clusters 7, 8 and 9 merge into 6a and 3b, which also captures 6 and 5a. Clusters 13 to 16 merge into 6b via 9a, 10a and 11a. The other clusters remain distinctive. At the last two stages, there is a very large cluster (1c and 1d) and a small cluster (4c and 2d). Clusters 2c and 3c remain separate at stage 5 but merge into 1d at the last stage. Manila remains in a one-city cluster throughout.

Figure 2 shows the 'dominant' variables in the clusters at the five stages. Cities in cluster 1, for example, are distinguished by a lower death rate, more females and a smaller older population than all cities. New York and Tokyo cluster by themselves initially and as a pair later because of size. Bangkok,

[7]*continued*
cases (cities) in each cluster. If a particular variable in a cluster has a large and statistically significant *z*-value, it distinguishes the variable as 'dominant' in that it sets the cities in that cluster apart from the averages (means) of all the cities in the analysis.

Table 1—Continued

City name[a]	Country[b]		Region[c]		Continent[c]
Brisbane Melbourne Perth Sydney	Australia (MDC)	→	Oceania	→	Oceania
Kobe Kyoto Nagoya Osaka Sapporo Tokyo Yokohama	Japan (MDC)	→	East Asia		Asia
Beijing Shanghai Tianjin	China (LDC/SC)	→	East Asia		Asia
Pusan Seoul	South Korea (LDC)	→	East Asia		Asia
Bangkok	Thailand (LDC)		Southeast Asia		Asia
Jakarta	Indonesia (LDC)		Southeast Asia		Asia
Manila	Phillipines (LDC)		Southeast Asia		Asia
Singapore	Singapore (LDC)		Southeast Asia		Asia
Istanbul	Turkey (LDC)		Western Asia		Asia
Tel Aviv	Israel (MDC)		Western Asia		Asia
Havana	Cuba (LDC/SC)	→	Caribbean		Latin America
Mexico City	Mexico (LDC)	→	Central America		Latin America
Buenos Aires	Argentina (LDC)		South America		Latin America
São Paulo	Brazil (LDC)		South America		Latin America
Cairo	Egypt (LDC)	→	Northern Africa	→	Africa

[a] West Berlin and East Berlin were two separate cities and did not become one city until West and East Germany were reunified in 1990.
[b] MDC = more developed country: LDC = less developed country; SC = socialist country; FSC = former socialist country.
These categories are loosely defined according to the United Nation's geographical and World Bank's per capita GNP—and income-based classifications.
[c]Regional and continental classification according to *1990 World Population Data Sheet* (Population Reference Bureau, 1990).

Pusan, Seoul and São Paulo cluster together (in 13, 9a) not only for their size but also for a larger younger population and a higher rate of natural increase than the averages of all cities.

As the initial clusters merge subsequently, the relative influence and order of the 'dominant' variables shift. At the last stage, besides the extreme case of Manila, the cities belong to two distinctive clusters with contrasting demographic profiles. The much larger cluster, which contains predominantly cities in more developed countries (MDCs), have a smaller proportion of younger population and a lower rate of natural increase, whereas the eight cities (clusters 4c, 2d) in less developed countries (LDCs) differ sharply from the MDC cities on 10 of the 11 indicators. While the final stage produces the most general and 'neat' classifications, the intermediate three stages also contain interesting and revealing information for discussion.

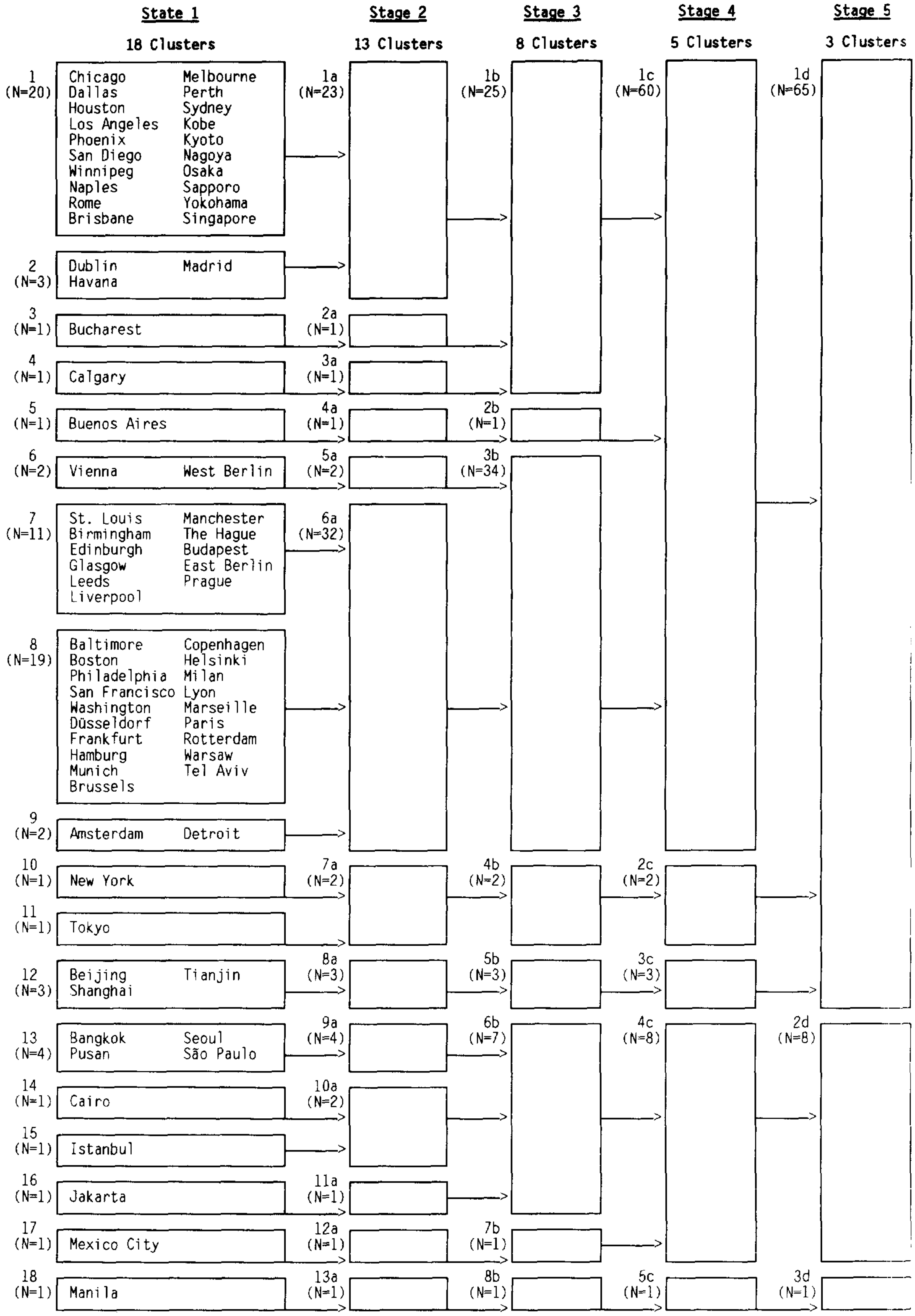
Discussions

Differences between the MDC and LDC cities

The cities in MDCs and LDCs are classified 'cleanly' into two groups expected of their general development levels, with few LDC cities crossing over into the earlier and later MDC clusters as somewhat special cases. Singapore is a city state, has had a very successful family planning programme, and was well on its way to 'graduate' from the developing world as a newly industrializing country (NIC) around 1980. Havana, a LDC/SC city, clusters initially with Dublin and Madrid, capital cities of two marginal MDCs (Ireland and Spain). With a relatively large older population and a high birth rate, Buenos Aires combines some demographic characteristics of both MDC and LDC cities, thus staying in a cluster by itself before joining the large MDC cluster at stage 4. Beijing, Shanghai and Tianjin—three LDC/SC cities in China—are not absorbed by the cluster of MDC cities until the last stage. A study of giant East Asia cities (Yeung, 1988) showed that around 1980 Singapore and Shanghai grew much more slowly than Bangkok, Jakarta and Seoul (in clusters 6b, 4c, 2d), and Manila. This combined evidence indicates that few very large LDC cities deviate from the general pattern for country-specific economic and political reasons.

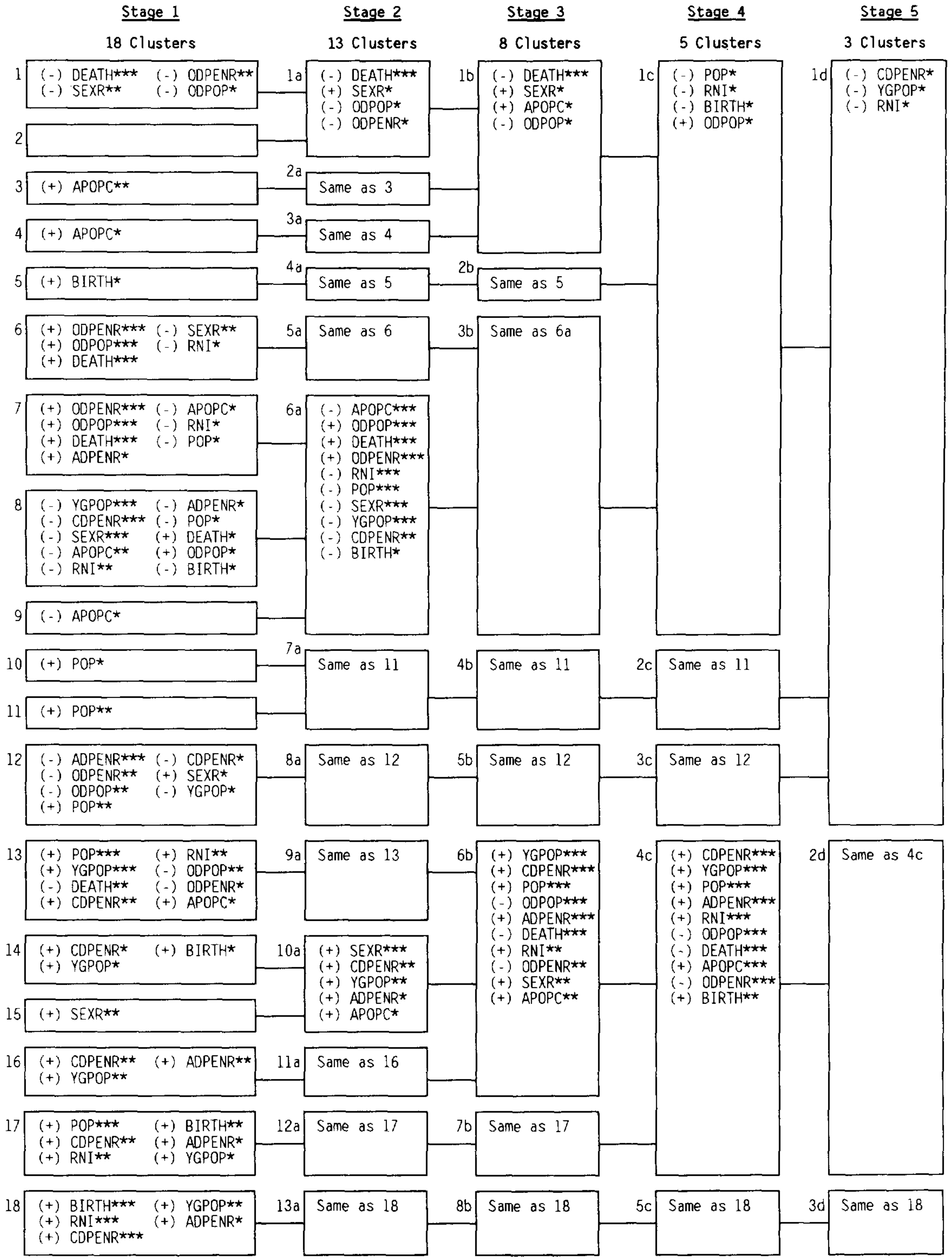
Similarities and differences among the MDC cities

The similar demographic profile of the MDC cities is reflected in all Australian and Japanese cities (except Tokyo) clustering with some US and West European cities (1, 1a). They together constitute



Note: [a]Cities in each cluster at stage 1 are grouped by the countries to which they belong to facilitate cross-national comparison within and across clusters.

Figure 1 Classification of 74 of the world's largest cities[a] by population size, net change, natural increase, sex ratio, age composition and dependency ratios, *ca* 1980



Notes: [a]For the definition and measurements of the variables. see the data section of and footnotes 2, 3, and 4 to the paper. The dominant variables for each cluster are listed in the descending magnitude of their *z*-values.

The (+) and (-) refer to a positive and negative *z*-value respectively.

* = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$ (two-tailed tests).

Figure 2 "Dominant" variables[a] for the clusters of 74 cities what may be called Type I MDC cities, which are characterized by a low death rate and less old population.

Of the 54 MDC 34 (63%) cities experienced negative annual population growth around 1980. And most of these are US and West European cities (in clusters 7, 8, 6a and 3b) that also have an older (than Type I) population, and can be labelled Type II MDC cities. Their shared central demographic feature of negative growth ties in with Hall and Hay's (1980) finding that the cores of European metropolises, like their US counterparts, lost population through the 1970s.

There are also major differences among the MDC cities between and within countries. That the cities in Australia, Japan, the United Kingdom, West Germany and France cluster together indicates the effect of country-specific characteristics in differentiating the cities. Within-country differences are highlighted by the so-called 'sunbelt' and 'snowbelt' US cities clustering into two distinctive intranational regional groups, which would be perfectly verified if Chicago and San Francisco had traded places between clusters 1 and 8, 1a and 6a, and 1b and 3b.

Similarities and differences among the socialist and former socialist cities

The former socialist cities had a similar demographic profile, as four of the five East Central European cities cluster together (7, 8, 6a, 3b). The most striking similarity between the former socialist cities is their negative population growth, which is consistent with Friedrichs' (1988, p 133) observation that the inner districts of Budapest, Prague and Warsaw had population losses of −2.3%, −61.9%, and −24.8% during 1970–80.

The differences between the socialist and former socialist cities are more noticeable, however. A lower age dependency ratio and a younger population, coupled with large size, set the three Chinese cities apart from their former counterparts in East Central Europe. Bucharest and Havana are absorbed into the Type I MDC cities (1a, 1b) and into the general MDC city group (1c, 1d), whereas Budapest, East Berlin, Prague and Warsaw cluster with the Type II MDC cities (7, 8, 6a, 3b) earlier, and join the combined MDC group subsequently. So do the three Chinese cities eventually. This broad similarity between the socialist and former socialist and MDC cities overshadows some of the specific differences between the former. It suggests that the demographic profiles of socialist and former socialist cities and capitalist MDC cities are less differentiated than expected from their contrasting political systems and uneven economic development.

Cities with particular demographic profiles

Finally, the cluster analysis has identified few cities with somewhat singular demographic profiles, which are defined as cities that remain in very small, especially one-city clusters at stage 3 and beyond. New York and Tokyo are two MDC examples, for which population size is the dominant discriminating factor (see Figure 2). Given their status as global financial centres (Sassen, 1991), New York and Tokyo would most probably stand apart from other major MDC cities if economic indicators were also used in the classification. Beijing, Shanghai and Tianjin are three special cases with the demographic features of both MDC and LDC/SC cities. Although they belong to the general MDC group because of a lower child dependency ratio and a smaller younger population, they resemble the LDC city group (4c, 2d) in terms of size and a smaller old population. However, a low natural increase rate and slow net population growth, the result of China's rigorous family planning programme and tight control on migration to large cities, differentiate the three Chinese cities sharply from their LDC counterparts.

Manila is an extreme outlier relative to both the MDC and LDC cities primarily because its birth rate of 6.4% (the highest of all 74 cities in the analysis) is almost three times the high average (2.4%) of the LDC city group in clusters 4c and 2d.

Research and policy implications

The research implications of this study concern the data and method used, while the substantive results have urban policy implications.

Cross-national comparative studies of cities are very challenging in that they require systematic, comprehensive and cross-nationally comparable data for multiple time points. The data used here are valuable because they meet some of these requirements. First, the data are systematic in the area of demography. Second, they are comprehensive in covering cities in diverse countries and regions and comparable across these countries. Third and most important, they provide an important baseline for examining change and continuity in the world's largest cities with more up-to-date data. As that baseline, the 1980 data are very timely because they marked the beginning of the decade when world city research was established. In this sense, the 1980 data serve as a reference point against which the results from subsequent analysis can be compared. The classified group of LDC cities based on these data was independently confirmed as belonging together by Sufian (1993) who, using the late 1980s data and discriminant analysis, found that the same cities (except the two South Korean cities of Seoul and Pusan) belong to one group with a similar living standard index. It is unfortunate that the cross-national quantitative studies of cities thus far, including the most recent network analysis of world city linkages (see Smith and Timberlake, 1995), have not gone beyond cross-sectional data. This study highlights the urgency in establishing longitudinal data for world city research in particular, and cross-national comparative analysis of cities more generally.

This paper has demonstrated the strengths of cluster analysis in integrating the advantages of both variable and case based methods. Capable of handling a large number of variables on a large number of cases, cluster analysis allows us to study the multiple attributes of many cities simultaneously. Cluster analysis based on demographic indicators here can be easily extended to accommodate other input variables to produce new categories of the same cities. Given its exploratory nature, cluster analysis can generate rich inductive typological evidence for developing testable hypotheses in world city research. Finally, cluster analysis is effective in isolating outliers. The few cities identified as outliers in this paper suggest themselves as candidates for more in-depth case studies.

The demographic profiles of large cities in MDC and LDC countries pose different challenges and require different policy responses from municipal governments and urban planners. The population loss in large MDC cities erodes their tax base and exerts financial pressure on the municipal governments. A large and continually growing older population in large MDC cities, coupled with a small and shrinking younger population due to a low birth rate and natural increase rate, further weakens the generally weak family-based support, and strains the governmental medical care and social services system. To accommodate this type of population in large MDC cities, the municipal governments and urban planners may need to modify or redesign the social and physical infrastructure, which requires large yet only minimally available resources during the current urban fiscal crisis.

For large LDC cities, a much younger population fuelled by a high birth rate and a natural increase rate imposes great pressure on child care facilities and primary and secondary schools. In addition, this age structure creates an annual oversupply of youths entering the labour market. The rapid growth of these already very large cities due to the combined effect of high natural increase and large in-migration pressures the local economies to grow fast to absorb the large number of new labour force entrants. These demographic forces contribute to housing shortages, congested streets, crowded public transport, lack of electricity and strained social services, which remain major problems in large LDC cities today. And the resources to deal with these problems are very limited at both the local and national levels of developing countries. Although these problems may be ameliorated by population and economic deconcentration from central cities, it certainly did not occur around 1980, as shown by the earlier analysis.

The very large cities in China did not experience many of the problems mentioned above around 1980 because their demographic profiles were shaped largely by a strong government population control policy through almost draconian family planning and rigid migration restrictions. But the 1990s see the very large Chinese cities take on new demographic characteristics. First, their populations have been ageing, although not as rapidly as large MDC cities. Nevertheless, as a recent case study of Shanghai (Di and Rosenbaum, 1994) indicated, this ageing population structure, coupled with the small younger cohorts due to the one-child family policy, have already weakened China's traditional family based system of caregiving for the elderly. Second, large Chinese cities have grown faster than through the mid-1980s by drawing large numbers of temporary migrants (who have a higher birth rate than local residents) as household registration and migration controls have been loosened. In these and other aspects, large Chinese cities have become more like both MDC and LDC cities as China has become less socialist and more market-oriented (see Chen and Parish, 1995).

This final comparative point punctuates the need for designing urban policies to tackle the more common problems facing large MDC and LDC cities as their demographic profiles continue to shift.

Acknowledgements

The research and writing for this paper was supported in part by a postdoctoral fellowship from the American Council of Learned Societies and the Social Science Research Council during 1993–94. Richard Campbell, Kenneth Land, Tim Futing Liao, Angela O'Rand, Anthony Orum, Joe Lee Rogers, Moshe Semyonov and Joel Smith provided comments on earlier versions and drafts. You-Bei Lou provided computing assistance. I am solely responsible for the content of the paper.

References

Berlage, L and Terweduwe, D (1988) 'The classification of countries by cluster and by factor analysis' *World Development* **16** (12) 1527–1545

Chen, X M and Parish, W L (1996) 'Urbanization in China: reassessing an evolving model', in Gugler, J (ed) *The Urban Transformation of the Developing World: Regional Trajectories* Oxford University Press, London

Cheshire, P (1990) 'Explaining the recent performance of European Community's major urban regions' *Urban Studies* **27** (3) 311–333

Di, J X and Rosenbaum, E (1994) 'Caregiving system in transition: an illustration from Shanghai, China' *Population Research and Policy Review* **13** (1) 101–112

Dogan, M and Kasarda, J D (eds) (1988a) *The Metropolis Era: Mega-Cities* Vol 2, Sage Publications, Newbury Park, CA

Dogan, M and Kasarda, J D (eds) (1988b) 'Introduction: comparing giant cities', in Dogan, M and Kasarda, J D (eds) *The Metropolis Era: Mega-Cities* Sage Publications, Newbury Park, CA

Everitt, B (1974) *Cluster Analysis* John Wiley, New York

Frey, W H (1988) 'Migration and metropolitan decline in developed countries' *Population and Development Review* **14** (4) 595–628

Frey, W H (1990) *Metropolitan America: Beyond the Transition* Population Reference Bureau, Washington, DC

Friedmann, J (1986) 'The world city hypothesis' *Development and Change* **17** 69–83.

Friedmann, J and Wolff, G (1982) 'World city formation: an agenda for research and action' *International Journal of Urban and Regional Research* **6** (3) 309–344.

Friedrichs, J (1988) 'Large cities in Eastern Europe' in Dogan, M and Kasarda, J D (eds) *The Metropolis Era: A World of Giant Cities* Sage Publications, Newbury Park, CA

Hall, P (1984) *The World Cities* 3rd edn, Weidenfeld and Nicolson, London

Hall, P and Hay, D (1980) *Growth Centres in the European Urban System* University of California Press, Berkeley, CA

Markusen, A and Gwiasda, V (1984) 'Multipolarity and the layering of functions in world cities: New York City's struggle to stay on top' *International Journal of Urban and Regional Research* **18** (2) 167–193

Marlin, J T, Ness, I and Collins, S T (1986) *Book of World City Ranking* The Free Press, New York

Milligan, G W and Cooper, M C (1985) 'An examination of procedures for determining the number of clusters in a data set' *Psychometrika* **50** (2) 159–179

Population Reference Bureau (1990) *1990 World Population Data Sheet* Population Reference Bureau, Washington, DC

Sarle, W S (1983) *Cubic Clustering Criterion* SAS Technical Report A-108 SAS Institute, Cary, NC

SAS Institute (1985) *SAS User's Guide: Statistics* SAS Institute, Cary, NC

Sassen, S (1991) *The Global City: New York, London, Tokyo* Princeton University Press, Princeton, NJ

Sawyer, L and Tabb, W K (eds) (1984) *Sunbelt/Snowbelt: Urban Development and Regional Restructuring* Oxford University Press, New York

Shryock, H S and Siegel, J S and Associates (1976) *The Methods and Materials of Democracy* Academic Press, New York

Smith, D A and Timberlake, M (1995) 'Conceptualizing and mapping the structure of the world system's city system' *Urban Studies* **32** (2) 287–302

Sufian, A J M (1993) 'A multivariate analysis of the determinants of urban quality of life in the world's largest metropolitan areas', *Urban Studies* **30** (8) 1319–1329

Yeung, Y M (1988) 'Great cities of Eastern Asia' in Dogan, M and Kasarda, J D (eds) *The Metropolitan Era: A World of Giant Cities* Sage Publications, Newbury, Park, CA