

Logistic Population Growth in the World's Largest Cities

Gordon F. Mulligan

Department of Geography and Regional Development, University of Arizona, Tucson, AZ

This article demonstrates that recent population growth in the world's largest cities has conformed to the general parameters of the logistic process. Using data recently provided by the United Nations, logistic population growth for 485 million-person cities is analyzed at 5-year intervals during 1950–2010, with the UN projections for 2015 adopted as upper limits. A series of ordinary least-squares regression models of increasing complexity are estimated on the pooled data. In one class of models, the logarithms of population proportions are specified to be linear in time, which is the standard approach, but in a second class of models those proportions are specified as being quadratic. The most complex models control logistic growth estimates for (i) city-specific effects (e.g., initial population), (ii) nation-specific effects (e.g., economic development, age distribution of population), and (iii) global coordinates (for unobserved effects). Moreover, the results are segregated according to each city's membership in four different growth clubs, which was an important finding of previous research.

Introduction

Much research has been devoted to understanding employment growth and population change in large cities. Three perspectives now dominate the literature. One group of observers has used *conceptual* models to chronicle growth in the world's largest urban agglomerations (Hall 1984; Friedmann 1986, 1995; Sassen 1991, 2000, 2002). These world cities continuously reorganize—along economic, political, and social lines—to cope with the various stresses and strains that accompany accelerating globalization (Knox and Taylor 1995; Scott 2001). Moreover, global cities perpetually compete for strategic positions in transnational exchange systems. Debate in this research stream is largely *exclusive*, in part because world cities are believed to be qualitatively different from smaller urban places; some notable exceptions include Smith and Timberlake (1995) and Taylor (1997). However, a

Correspondence: Gordon F. Mulligan, Department of Geography and Regional Development, University of Arizona, Harvill Building, Box #2, Tucson, AZ 85721
e-mail: mulligan@email.arizona.edu

Submitted: April 19, 2005. Revised version accepted: February 3, 2006.

second group of observers has adopted a more *inclusive* perspective, one that generalizes about employment growth and population change in cities of virtually all population sizes (Borchert 1967; Berry and Horton 1970; Gaile and Hanink 1985; Glaeser et al. 1992; Cheshire 1995; Henderson and Wang 2005). World cities are now seen as being only quantitatively larger variants of their smaller relatives. Of course, larger places are fully recognized to be more complex than smaller places, but the growth prospects of all cities are believed to be determined by the same general mechanisms. In this research stream, *empirical* models are created to clarify how initial conditions and other contextual factors differentially influence urban growth. Yet a third group of observers has pioneered the urban analysis of the New Economic Geography (Henderson 1988; Arthur 1994; Fujita, Krugman, and Venables 1999; Brakman, Garretsen, and van Marrewijk 2001). Here, *mathematical* models, often solved by numerical simulations, are designed to capture dynamic interactions between growth factors like comparative advantage, increasing returns, innovation, and circular causation. While this research stream is certainly the most analytical of the three, the overall perspective on city growth and change is again largely inclusive (Stelder 2005).

Adopting the second perspective, Mulligan and Crampton (2005) recently summarized various facets of late 20th-century and early 21st-century population growth in the world's largest urban centers. Extensive use was made of a cross-national population data set, recently updated by the United Nations (2004), which summarizes the growth trajectories of all cities whose population exceeded 750,000 in the year 2000. Census-derived figures are provided by the UN study at 5-year intervals, stretching back to the year 1950, and most likely growth scenarios are used to project urban populations out to the year 2015. After deciding to omit the final 5-year interval, Mulligan and Crampton analyzed population growth in the 485 so-called *million-person cities*—those being all urban agglomerations whose populations exceeded one million ($\text{pop} > 1\text{m}$) sometime during 1950–2010. They found that the “average” million-person city would have grown by a remarkable 953% at the end of this 60-year study period, which equates to a non-compounded yearly rate of 15.9%. In other words, the “typical” million-person city will be 8.5 times as populous in 2010 as it was in 1950.

More importantly, Mulligan and Crampton (2005) showed that these million-person cities exhibited remarkably different growth patterns over time. Some places started small and other places started large; some places grew early and others grew late; and some places grew moderately fast while others grew very fast. So the authors identified clusters of cities that shared similar population-growth trajectories. For the purposes of that earlier article, 10 different growth types were identified by cluster analysis, and the members of each type were mapped out so that broad geographic patterns could be discerned. However, for the purposes of the present article, less classificatory detail is needed and these 10 clusters have been consolidated into four fairly homogeneous *clubs*. The acronyms, numbers, and mean 1950–2010 population growth figures for each of these four city clubs are as

follows: (1) HYP, which includes 117 hypergrowth (HYP) places averaging 2717% growth; (2) FAS, 125 fast-growth (FAS) places, 749%; (3) MOD, 113 moderate-growth (MOD) places, 343%; and (4) SLO, 130 slow-growth (SLO) places, 91%. Later in the article, extensive use is made of this relatively simple city-growth typology.

The main intent of this article is to test for logistic population growth across the sample of 485 million-person cities during the same 60-year time span. Logistic growth commonly occurs in a variety of natural and human systems when entities in those systems must cope with resource constraints or environmental limits. Common sense dictates that *individual* cities—much like larger regions and entire nations—should exhibit some sort of an S-shaped pattern in their population growth (Berry 1973; Keyfitz 1980). Here, the main underlying supposition is that intracity constraints and intercity competition, acting together, necessarily lead to logistic (or near logistic) growth patterns within individual cities. The results of this study clearly bear out this claim; moreover, somewhat different logistic growth patterns are established for the four growth clubs of cities mentioned above. The secondary intent of this study is to explore how different types of city-specific and nation-specific initial conditions have influenced population growth in these very large cities. Here, the findings are somewhat preliminary, if only because the demographic and economic conditions chosen as control variables do not represent all of the factors that have influenced recent growth in these large places. Nevertheless, the results are very promising and clearly suggest that this line of research is worthy of more attention.

Logistic growth

The logistic function is approximately characterized by a bell-shaped or normal curve as its probability density function and by a sigmoid curve as its cumulative density function. The function has been used very widely for a long time, and some applications apparently even precede the groundbreaking theoretical work of Verhulst during the 1840s. Well-known applications can be found in business organization (Hannan and Freeman 1989), cultural transmission (Cavalli-Sforza and Feldman 1981), ecology (Rose 1987), epidemiology (Cliff et al. 1981), knowledge growth (van Duijn 1983), migration or mobility (Morrill 1965), social communication (Hagerstrand 1965), and spatial diffusion (Hudson 1972; Casetti and Semple 1969). Useful pedagogical overviews are available for geographers in Abler, Adams, and Gould (1971) and Thomas and Huggett (1980). Two reasons seem to dictate the wide application of the logistic function. First, there are a host of natural and human processes, including urbanization in general, that appear to follow an S-shaped growth pattern. And second, the parameters of the logistic process can be estimated very easily—something that cannot be said for many other (even similar) processes that are formally described as differential equations. For this reason, the logistic function is widely recognized and adopted as the simplest representation of

density-dependent population regulation. Nevertheless, many observers question the usefulness of the logistic function for making population forecasts, usually because individuals having different attributes (e.g., age, income) and motivations are all lumped together in the estimation (Keyfitz and Caswell 2005). But these same skeptics likely would agree that the logistic approach continues to work reasonably well as a means for summarizing past population trends, which is the sole intent of this article.

In any case, the function is denoted by the following formula:

$$dP(t)/dt = rP(t)[U - P(t)]/U$$

where r is the known growth rate, $P(t)$ is the time-specific population count of individuals, and U —which typically exceeds all $P(t)$ levels—is some upper limit (or long-run equilibrium) to that population count. This equation indicates how the instantaneous change $dP(t)/dt$ in population is related to intrinsic exponential growth $rP(t)$ that is tempered by some inhibiting factor $[U - P(t)]/U$ that becomes stronger over time. Integration leads to

$$P(t) = U/(1 + e^{rt})$$

which is usually presented as

$$P(t) = U/[1 + e^{(a-bt)}]$$

where $P(t)$ is the proportion of the upper limit U that has been reached at some time t . Time usually indicates the interval as the growth process started but it sometimes only represents the interval since measurement of the process commenced. The value U is often visualized as the carrying capacity for natural processes or the saturation level for human processes; in any case, it normally represents the maximum sustainable value for $P(t)$. Because logistic processes are often interpreted in ratio or percentage terms, U is commonly seen as being equal to unity or to 100. The term e is the base of the natural logarithms (so $\ln = \log_e$), the intercept term a ($a > 0$) determines the value of the process when time $t = 0$, and the slope term b ($b > 0$) represents the specific rate at which $P(t)$ changes with the passage of time. These last two parameters are typically estimated by ordinary least squares (OLS) regression, where the growth rate r is seen to decline over time from its highest level a at time $t = 0$. Note that

$$\ln\{[U - P(t)]/P(t)\} = a - bt$$

When the process is half-complete and the logistic curve has reached its inflection point, the left-hand side (LHS) of the equation can be set equal to zero (as $\ln 1 = 0$). This indicates that a major change in the nature of the process occurs at time $t^* = a/b$. Time t^* is frequently used to make claims about the so-called *half-life* of the logistic process.

The general thinking behind the logistic process is simple. When the process first begins, there are few or no resource constraints for individuals to contend with

and growth essentially assumes an exponential form. Eventually, though, these constraints are reached and all future growth of the population is inhibited by an ever-decreasing stock of resources. In ecological systems, predators find that prey become scarcer and scarcer; in postfrontier agricultural systems farmers find less and less arable land to cultivate; and in epidemiological or diffusion processes there are fewer and fewer susceptible individuals remaining to adopt the spreading disease or innovation. So eventually the process must approach some upper limit. And, in theory, even if those population numbers overshoot this upper limit and then collapse, the numbers should eventually climb back to that maximum sustainable value as a long-run equilibrium.

However, any logistic process can be influenced by a variety of initial or ongoing conditions. From the geographer's viewpoint, one of these contextual parameters must always be location, as measured in either absolute or relative terms. But many other factors might affect the growth process of interest. It is known, for example, that birth rates and death rates can quickly (and interdependently) change in both biological and demographic systems and these changes will impact the time-specific growth rate. In this article, I focus on the population growth of large cities, so I can fully expect that the process will be impacted or constrained by a host of demographic, economic, and political parameters operating at a variety of spatial scales.

In this article, the logarithm of proportions is estimated on the LHS in two somewhat different ways. In one class of models, only the intercept term a on the right-hand side (RHS) is extended where

$$a = a_0 + \mathbf{cV}$$

and $\mathbf{c}$ represents estimates of different contextual variables in the vector $\mathbf{V}$. As interest centers only on conditions occurring near the beginning of the study period, these city-specific and nation-specific attributes constitute the initial conditions for subsequent logistic population growth. Moreover, the variable YEAR is now used to measure time since 1950, which is the beginning of the study period. So the following function is estimated in this first class of models:

$$\ln\{[U - P(t)]/P(t)\} = a_0 + \mathbf{cV} - b \times \text{YEAR}$$

But in a second class of models, the slope term b on the RHS of the standard logistic function is modified where

$$b = b_0 + \mathbf{dW}$$

and $\mathbf{d}$ represents estimates of different contextual variables in the vector $\mathbf{W}$. Now the process may no longer be purely logistic, in the strict sense that growth is linear in time, but the model certainly qualifies as a member of the extended logistic family (Abler, Adams, and Gould 1971). In the discussion that follows, the time variable now affects the logistic process in two other ways: first, as an extra-quadratic term and then as a term interacting with the initial population of the city. So

the following function is estimated in the second class of models

$$\ln\{[U - P(t)]/P(t)\} = a_0 + \mathbf{cV} - (b_0 + \mathbf{dW}) \times \text{YEAR}$$

Special attention is given to the signs of the estimates in vectors $\mathbf{V}$ and $\mathbf{W}$. In those cases where the entire group of 485 million-person cities is examined, the timing of population growth is of the greatest interest. A positive (negative) sign on a variable in $\mathbf{V}$ indicates that this variable induces a late (an early) start to the logistic growth process. Estimates for $\mathbf{c}$ are in fact additive shift parameters that complement the intercept term a_0 . But when the four clubs of cities are examined separately, interest focuses on interclub differences in both the timing and rapidity of growth. A positive (negative) sign on a variable in $\mathbf{W}$ means that this variable induces a steeper (shallower) slope in the logistic growth process. The expectation is that cities in some clubs experienced delayed but much more rapid population growth than cities in other clubs. The estimates for $\mathbf{d}$ are shift parameters that complement the intercept term b_0 , and indicate the rapidity of this population growth.

The logistic model seems especially suited for understanding both the timing and the year-specific rate of urban population growth across the city systems of different nations. Cities compete with one another for scarce regional, national, and international resources and, eventually, their growth is tempered by different sorts of resource constraints (Kasarda and Crenshaw 1991). These resources are of both the private and public variety. Factors like diminishing returns, congestion and high transportation costs, high land values and exorbitant real estate taxes, high crime rates, and rising pollution will eventually inhibit *absolute* growth in nearly every city. But diminished *relative* growth occurs even when cities have enjoyed past success in adapting to environmental circumstances. Less well appreciated is how public fiscal decisions (involving conflicts between efficiency and equity) are also responsible for the geography of urban growth (Tolley and Thomas 1987; Anand and Ravallion 1993; Canaleta, Arzoz, and Gárate 2004). Moreover, city growth can be viewed from the wider perspective of regional, national, and international demographic circumstances, where the population growth rates of specific cities are ultimately constrained by such factors as regional and national fertility rates, interregional migration rates, and national policies with regard to immigration and emigration. In fact, a whole litany of reasons—arising from the twin perspectives of labor demand and labor supply—suggest that a general S-shaped pattern in population growth should occur among all interacting and competing cities. So the logistic model seems a very appropriate one to adopt for understanding city population growth in different places at different times.

Examples

For illustrative purposes, the standard logistic model was first applied to four cities in the UN data set. These particular cities were chosen to represent just a few of the

many urban growth scenarios that were summarized by Mulligan and Crampton (2005). Two of these places exhibited truly remarkable population growth—Brasilia (selected from HYP), which grew from 0.04 to 2.50 m, and Delhi (selected from FAS), which grew from 1.39 to 18.21 m. The other two places—Vancouver, which increased from 0.56 to 2.20 m, and Tokyo, which increased from 6.92 to 27.09 m—were both drawn from MOD but these are two cities having very different initial population sizes. OLS regression procedures were used to estimate $L = \ln[(U - P)/P]$ in each place. As indicated above, U is the highest stated population level and it can be interpreted as being 100 because all year-specific estimates are shown as percentages. Recall that P represents the time-specific ratio of current (or cumulative) population to that maximum figure of U . The variable YEAR varies continuously from time 0 (which is 1950) to time 60 (which is 2010). Note that

$$L(\text{Brasilia}) = 3.825 - 0.1060 \times \text{YEAR}$$
$$L(\text{Delhi}) = 2.890 - 0.0687 \times \text{YEAR}$$
$$L(\text{Vancouver}) = 1.707 - 0.0658 \times \text{YEAR}$$
$$L(\text{Tokyo}) = 1.497 - 0.1050 \times \text{YEAR}$$

$$\text{Ad. } R^2 = 0.98 \text{ SEE} = 0.30$$
$$\text{Ad. } R^2 = 0.96 \text{ SEE} = 0.29$$
$$\text{Ad. } R^2 = 0.92 \text{ SEE} = 0.38$$
$$\text{Ad. } R^2 = 0.97 \text{ SEE} = 0.35$$

(SEE, standard errors of the estimate; Ad. R^2 , R^2 value adjusted for degrees of freedom.)

The proportions $(U - P)/U$ were calculated for each city at 13 separate points in time, beginning in 1950 and ending in 2010. The UN population projections for 2015 were used to establish the upper limit U in each case, but the proportions were only calculated up to the year 2010. In this way, the estimates for the parameters a and b would not be overweighted by the large values for the proportions that occur at the very end of the logistic process. Values for both the observed and expected proportions of the four cities are shown in Table 1.

Table 1 Estimates of Logistic Population Proportions: Four Representative Cities

YEAR	Brasilia		Delhi		Vancouver		Tokyo	
	Ob	Ex	Ob	Ex	Ob	Ex	Ob	Ex
0	2.67	2.14	6.66	5.26	22.12	15.36	25.45	18.29
10	2.85	5.92	10.93	9.95	24.67	25.94	40.37	39.01
20	10.95	15.38	16.91	17.80	41.58	40.35	60.68	64.63
30	30.97	34.42	26.62	30.37	49.82	56.64	80.38	83.93
40	50.64	60.23	39.30	46.43	62.04	71.61	92.24	93.72
50	75.59	81.38	59.57	63.26	81.54	82.96	97.26	97.71
60	93.85	92.66	87.22	77.38	95.68	99.06	99.64	99.19

NOTE: All estimates are year-specific cumulative population percentages standardized by each city's upper limit of $U = 100\%$. YEAR = 0 indicates 1950, YEAR = 60 indicates 2010, and U is established by using the UN's projection for 2015 (YEAR = 65). In each year Ob indicates the actual proportion and Ex indicates the predicted proportion based on the standard (linear) logistic model.

The intercept and slope terms were found to be highly significant in each case. The coefficients of determination were very high for each run and the standard errors of the estimate (SEE) were remarkably similar. As stated above, the intercept term provides evidence regarding the beginning time of the logistic process. Evidently, growth started earlier in Tokyo and Vancouver than it did in Delhi and Brasilia. In fact, the first two cities experienced considerable growth before 1950. But the slope term provides other evidence about the rapidity of the logistic process. Evidently, growth was fastest for Brasilia and Tokyo. But the population proportions in Table 1 indicate that these two cities experienced their rapid growth over two very different time periods: Tokyo from 1950 to 1980 and Brasilia from 1970 to 2000. The half-life estimate for each city sheds light on this issue as it indicates when population growth was the steepest in each city. The inflection point for Brasilia's logistic curve occurred in 1986, some 36.1 (i.e., $3.825/0.1060$) years after 1950. The inflection points for population growth in the other three cities were Delhi (1992, 42.07 years after), Vancouver (1976, 25.9 years after), and Tokyo (1964, 14.3 years after). These estimates substantiate that the periods of most rapid growth for Delhi and Brasilia lagged considerably after the periods of most rapid growth for Tokyo and Vancouver.

The pooled data

As stated earlier, a main intent of this study is to generate year-specific estimates of the logistic function across all of the cities in the UN data set. So a pooled data set was constructed having a total of 6305 observations (i.e., 485 cities at 13 points in time) for both the city-specific proportion ($(U - P)/P$) and the year since 1950. Pooling was adopted to generate more efficient estimates for the logistic coefficients a and b (Pindyck and Rubinfeld 1991; Greene 2000). It seemed a very natural thing to do because the cities appeared to be heterogeneous units and they were believed to represent, when combined in space and time, all possible stages of urban logistic growth. As above, city population in 2015 was used to set the upper limit U on the various population proportions. A handful of cities—including Budapest, Milan, and London—actually dipped under this projected figure and in those cases P was set below 100 at one or more points in time.

The variables

This study estimates logistic population growth recognizing each million-person city's individual circumstances as well as its wider contextual conditions (mean $L = -0.3701$). The variables chosen to capture these attributes are for the most part initial conditions. Later research can extend or update the list of variables to allow covariation or even adjustment between these various attributes and logistic growth. In all, this study uses six city-specific initial conditions and seven orthogonal national-level factors as variables.

Global coordinates are also introduced to capture unobserved effects, which might include a litany of local circumstances.

City-specific variables

POP50: the “initial” city population in millions in 1950 (minimum = 0.01, maximum = 12.34; mean = 0.6929). Because this is a pooled data set, the cities vary considerably in their population sizes. Cities that were large in 1950 would have started the growth process earlier than their smaller counterparts, a property that is analogous to having a leftward shift in the intercept term a in the logistic function. To compensate, the slope term b on time for smaller cities in 1950 might be steeper. In any case, *POP50* should be negatively related to logistic growth (United Nations 2004).

CAPITAL: a binary variable indicating whether or not the city served as a national capital sometime during the study period (minimum = 0, maximum = 1, mean = 0.22). In many nations, capital cities serve to concentrate both economic and political power, so analysts have speculated that such urban centralization often precedes rapid population growth (Carroll 1982). The expected sign on *CAPITAL* is negative (Hoffman 1988; United Nations 2004).

MAJPORT: a binary variable indicating whether or not the city served as a major port sometime during the study period (minimum = 0, maximum = 1, mean = 0.23). Many analysts have noted that port cities enjoy a decided location advantage and this induces employment and subsequent population growth in most cases (Fujita and Mori 1996). Ports normally induce several ancillary service activities like insurance (Porter 1990), benefit from break-of-bulk manufacturing (Lloyd and Dicken 1977), and often enjoy circular and cumulative growth. The expected sign on *MAJPORT* is also negative (Hoffman 1988; American Association of Port Authorities 2005).

DISTAN3: sum of the Great Circle distances (Wikipedia 2005), in thousands of kilometers, to all three world cities of London, New York, and Tokyo (minimum = 23.534, maximum = 35.035, mean = 30.115), an index capturing each city’s accessibility, both to the world’s major markets and to the most important capital-goods-supplying regions (Gallup, Sachs, and Mellinger 1999). Nearness should promote early population growth, so the expected sign on *DISTAN3* is positive; that is, a low (i.e., favorable) accessibility index should shift the logistic intercept term to the left.

LONGIT: longitude in degrees west (negative) or east (positive) of London with a reversal of sign at the International Date Line (minimum = -123.2, maximum = 174.7, mean = 35.9). Given the late urbanization of the Middle and Far East, the expected sign on *LONGIT* is negative (Showers 1989).

LATIT: latitude in degrees north (positive) or south (negative) of the equator (minimum = -37.8, maximum = 59.9, mean = 35.9). Given the early urbanization of the Northern Hemisphere, the expected sign on *LATIT* is negative (Showers 1989).

National-level variables and factors

In the tradition of Berry (1961a, b), data were collected for 21 different national-level indicators believed to capture many of the contextual conditions for urban growth. Special attempts were made to capture features of rural-to-urban migration (Harris and Todaro 1970) and to measure national levels of trade and infrastructure (Tolley and Thomas 1987; Ray 1998). Taken together, the entire array was designed to capture the nation's general level of economic development, its position in the process of demographic transition, its economic openness to neighbors, the density of its economic activities, and the degree of primacy that was exhibited in its urban hierarchy. Short captions and descriptive statistics for this array of variables are made available in Appendix A. As demographic, economic, and other information is much more widespread and reliable for 1960 than for 1950, most of the data pertain to that later year.

This data matrix was next treated by factor analysis in order to reduce the number of variables down to a small set of representative dimensions. A total of seven orthogonal factors were identified that, together, explained nearly 90% of the variance in the 21 original variables (see Appendix A). Nation-specific factor scores, using a Varimax rotation, were then generated and allocated as "new" variables to the 6305 city-specific observations in the pooled data set. Each rotated factor that was retained accounted for at least 5% of the variance in the original data matrix. In comparison with the city-specific variables, it is sometimes difficult to specify anticipated signs for the factors in part because they capture the composite effects of numerous variables. So the expected signs of factors in the logistic regression equations are specified in only three of the seven cases.

FACTOR1 captures the overall level of *economic development* and corresponds to the technological scale addressed by Berry (1961a). Core nations have positive scores and peripheral nations have negative scores. Among the variables that load highly on this dimension are Gross Domestic Product per capita, level of urbanization, and the percentage of the labor force either in manufacturing or in services. Nations scoring high in economic development (e.g., Canada, United States) were often the first to urbanize while nations scoring low in economic development (e.g., India, Indonesia) were just beginning to urbanize in 1960. This factor should have an effect on city growth that is analogous to that of the initial city population, so the expected sign in the logistic regression is negative.

FACTOR2 differentiates nations according to their main *population characteristics*. This factor largely corresponds to Berry's (1961a) demographic dimension. Table A2 indicates that national population loads highly negative while both crude birth and death rates load moderately positive. But this factor also differentiates the more export-oriented and free-market economies from their more self-sufficient and centrally planned neighbors. Countries like Argentina, Japan, and the Netherlands have high positive scores while others like China, Pakistan, and Russia have high negative scores. In general, nations with high levels of population momentum

will urbanize later, and this will shift the starting time for urban population growth. So the expected sign in the logistic regression equation is positive.

FACTOR3 addresses the varying *population densities* of the different nations. Countries like Bangladesh and the Netherlands have high positive scores while countries like Argentina and Canada have high negative scores. Evidently, these high densities were not related to the nation's overall level of economic development, but countries having high population densities did tend to share slightly higher loadings on both infrastructure and urban primacy.

FACTOR4 largely recognizes national differences in the degree of *urban primacy*. High factor scores are especially prevalent throughout Latin America, including Argentina and Mexico, but are also found in more developed nations like France and the United Kingdom where world cities like Paris and London continue to dominate (Moomaw and Alwosabi 2004). Negative factor scores are commonplace in nations with either balanced city-size distributions (e.g., United States, Nigeria) or those where several cities compete for national prominence (e.g., Italy).

FACTOR5 clearly differentiates nations according to their *transportation and communications infrastructure*. Small nations like Denmark and the United Kingdom are given high positive scores and large nations like Brazil and Australia are given large negative scores. Not surprisingly, there is some evidence that these infrastructure-rich countries are also wealthier.

FACTOR6 captures the relative *youthfulness* of the national population. Like FACTOR2 exports are prominent but, in contrast to that other factor, the nation's physical size is now a more important variable. Demographically dynamic countries like Brazil and the United States, both of which were more self-sufficient in 1960, are given high positive scores. But most aging and well-developed European countries—such as Germany, Italy, and the United Kingdom—are given high negative scores. This factor should have an effect on city growth that is similar to that of the second factor, so the expected sign in the logistic regression is positive.

FACTOR7 is related to an extreme degree of *openness* in a small number of nations. There is also some evidence that this factor is related to high degrees of urban primacy. Factor scores are highly positive for countries like Ghana and Singapore, which had very high amounts of overall trade (either in raw commodities or manufactured goods) but scores are very low for societies like Japan and South Korea, which were relatively closed (and still recovering from war) in 1960.

Results

Table 2 shows pooled estimates of $L = \ln[(U - P) / P]$ for the very simplest models. In Model 1, we see first that the intercept estimate is $a = 1.650$, which suggests that the low time-series estimates for Vancouver and Tokyo are more in line with that for all million-person cities than are the high estimates for Brasilia and Delhi. On the other hand, the slope estimate is $b = 0.06733$, which indicates that the low time-series estimates for Delhi and Vancouver are closer to that of the entire UN data set

Table 2 Estimates of Logistic Population Proportions: The Basic Models

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	1.650 (48.96)	2.032 (63.32)	2.281 (52.11)	-1.309 (-8.87)	-0.900 (-6.44)
YEAR	-6.733E-02 (-70.64)	-6.733E-02 (-78.18)	-4.945E-02 (-16.30)	-6.733E-02 (-81.69)	-4.945E-02 (-17.13)
YEARSQ			-3.766E-04 (-7.84)		-3.766E-04 (-8.23)
POP50		-0.552 (-37.68)	-1.460 (-42.14)	-0.526 (-35.97)	-1.436 (-42.96)
POP50SQ			0.103 (30.17)		0.102 (31.30)
YRPOP50			6.804E-03 (9.40)		6.803E-03 (9.87)
CAPITAL				0.379 (9.78)	0.412 (11.57)
MAJPORT				0.299 (7.86)	0.344 (9.83)
DISTAN3				0.105 (22.70)	9.945E-02 (23.28)
SEE	1.4160	1.2793	1.1839	1.2245	1.1266
Ad. R ²	0.442	0.544	0.610	0.583	0.647

NOTE: All *t* scores are shown in parentheses. SEE, standard error of the estimate and Ad. R², R² value adjusted for degrees of freedom. N = 6305.

than are the high estimates for Brasilia and Tokyo. The entire regression equation for Model 1 is

$$L(485 \text{ cities}) = 1.650 - 0.06733 \times \text{YEAR}$$

The estimate for the inflection point of the logistic curve, which indicates the half-life of the growth process, is $t^* = a/b = 1.650/0.06733 = 24.5$ years. So, on average, population growth was the fastest—and the logistic curve was the steepest—across the entire group of 485 large cities sometime during the mid 1970s.

Model 2 discloses how the typical city's initial population size affected its subsequent growth. The inclusion of this one extra variable considerably reduces the standard error and permits another full 10% of the variance to be accounted for. This encouraging result suggests that the model can be improved upon even more by controlling city growth for other initial conditions. The estimate for POP50 is -0.552 , which is significant. The sign indicates that the proportion of the maximum population that was attained during any year was always lower for those cities that were initially small in 1950. The estimates (at the means) indicate, for example, that a city of 1 m initial population had reached 40.0% of its upper limit in 1970 but a smaller city of 0.5 m initial population had only reached 33.6% of its upper limit at that date. But by 2000 these saturation proportions were 86.8% and 83.3%, respectively, indicating that the growth gap separating large and small centers had narrowed considerably 10 years out from the end of the study period.

Model 3 provides the alternative to the standard logistic function, as discussed above. Quadratic terms are introduced for both time and initial population size, and an interaction term between time and size is also adopted. The standard error is reduced to 1.1839 and a remarkable 61% of the variance in logistic proportions is now accounted for. The results suggest that population growth in the world's largest cities did not exactly conform to the standard logistic function. In fact, the better-fitting quadratic model indicates that the S-shaped growth curve was initially flatter, but later steeper, than that depicted in the usual linear model. And, for the statistically average city in the data set, the half-life of the growth process increased by 3.25 years from $t^* = 24.5$ years in Model 2 to $t^* = 27.75$ years in Model 3. This is further evidence that the population growth of the million-person cities was initially somewhat slower than that depicted by the standard (linear) logistic process.

Models 4 and 5 are straightforward extensions of Models 2 and 3, respectively. But now three other city-specific initial conditions are introduced: dummies for both capital cities and major ports and distances from cities to the three major global markets. In both instances, all three of these new variables prove to be significant. Evidently, the logistic growth curve (both linear and quadratic versions) is shifted to the right when a city served as either a national capital or a major port (note the positive estimates), two results that were not expected. But nearness, both to world markets and capital goods supply regions, shifted the growth process to the left, as was anticipated.

Table 3 revisits several of these models for each of the four population-growth clubs identified earlier. Members of the SLO club constitute the null or excluded observations in each regression run, where dummies are used for those observations in the other three clubs: MOD, FAS, and HYP. While separate regressions could easily have been estimated for each growth club (giving exactly the same results), the dummy approach is superior for making summary comparisons with the models outlined above. Note that all significance tests for MOD, FAS, and HYP are based on comparisons with SLO; entirely separate regressions would have to be estimated to compare the estimates of MOD with those of FAS, for example.

Table 3 Estimates of Logistic Population Proportions: The Four Growth Clubs

Variable	Model 6	Model 7	Model 8
Intercept	− 0.360 (− 8.39)	− 0.182 (− 3.90)	0.340 (4.70)
YEAR	− 5.450E−02 (− 44.99)	− 5.450E−02 (− 45.82)	− 6.876E−02 (− 15.43)
YEARSQ			1.205E−04 (1.73)
POP50		− 0.121 (− 8.84)	− 0.531 (− 13.60)
POP50SQ			2.995E−02 (8.89)
YRPOP50			4.806E−03 (6.75)
MOD	1.907 (30.36)	1.867 (27.58)	0.853 (8.12)
MYEAR	− 5.301E−03 (− 2.98)	− 5.301E−03 (− 3.04)	5.778E−02 (8.87)
MYEARSQ			− 9.029E−04 (− 8.85)
MPOP50		− 6.879E−02 (− 2.29)	0.429 (4.79)
MPOP50SQ			− 3.193E−02 (− 2.56)
MYRPOP50			− 7.401E−03 (− 4.74)
FAS	2.655 (43.41)	2.555 (38.58)	1.815 (17.66)
FYEAR	− 1.741E−02 (− 10.06)	− 1.741E−02 (− 10.25)	2.839E−02 (4.47)
FYEARSQ			− 6.336E−04 (− 6.38)
FPOP50		− 6.567E−02 (− 1.34)	3.769E−02 (0.25)
FPOP50SQ			0.120 (2.46)
FYRPOP50			− 6.656E−03 (− 2.63)
HYP	3.651 (58.67)	3.774 (52.86)	3.556 (30.13)
HYEAR	− 2.949E−02 (− 16.75)	− 2.949E−02 (− 17.06)	2.823E−03 (0.43)
HYEARSQ			− 5.117E−04 (− 5.06)
HPOP50		− 2.593 (− 9.25)	− 9.847 (− 5.06)
HPOP50SQ			17.006 (8.01)
HYRPOP50			4.418E−02 (3.04)
SEE	0.9317	0.9148	0.8873
Ad. R^2	0.758	0.767	0.781

NOTE: All t scores are shown in parentheses. The excluded group (with no dummies) is the slow growth club SLO. Dummies are used with the moderate (MOD), fast (FAS), and hyper (HYP) growth clubs. These are the growth-club dummy extensions of Models 1, 2, and 3 shown in Table 2. SEE, standard error of the estimate; Ad. R^2 , R^2 value adjusted for degrees of freedom.

Model 6 is the club-specific extension of Model 1. By introducing dummies on the intercepts and on the single variable YEAR, the standard error is reduced from 1.2793 to 0.9317 (a reduction of 27.2%) and the coefficient of determination is improved from 0.544 to 0.758 (a remarkable increase of 39.3%). These results alone suggest that logistic growth processes unfolded very differently across the four clubs of million-person cities.

Model 7 indicates that the overall results for Model 2 can be ascribed to the various clubs as follows:

$$L(\text{SLO}) = -0.182 - 0.05450 \times \text{YEAR} - 0.121 \times \text{POP50}$$

$$L(\text{MOD}) = 1.685 - 0.0598E \times \text{YEAR} - 0.190 \times \text{POP50}$$

$$L(\text{FAS}) = 2.373 - 0.07191 \times \text{YEAR} - 0.187 \times \text{POP50}$$

$$L(\text{HYP}) = 3.592 - 0.08399 \times \text{YEAR} - 2.715 \times \text{POP50}$$

Among other things, these estimates substantiate that logistic growth occurred earlier (later) in the slowest (fastest) growing cities of the data set. And when growth did take off, the process was much steeper in the HYP club than in the other three clubs. These estimates indicate that (at the group means) inflection points for the logistic process occurred at the following times: SLO: -6.6 years; MOD: 25.8 years; FAS: 31.9 years; and HYP: 39.2 years. So the half-life parameter of the SLO club preceded the overall figure (i.e., $t^* = 24.5$) by some 31 years while the half-life parameter of the HYP club lagged some 15 years behind the overall figure. Evidently, the HYP cities experienced their most rapid growth just before 1990 but the SLO cities experienced their most rapid growth sometime near the end of World War II, which was some half-decade before the first year of the study period. Not surprisingly, estimates (not shown) from other regression runs indicate that the poorest statistical fit is in the SLO club, evidently because many of those cities had already surpassed the inflection point of their logistic growth curve by 1950.

Model 8 provides club-specific estimates for the alternative, quadratic version of logistic growth. The standard error of its predecessor, Model 5, is reduced from 1.1839 to 0.8873, and the amount of the variance now explained in the population proportions jumps dramatically from 58.3% to 78.1%. The alternative club-specific estimates are

$$L(\text{SLO}) = -0.340 - 0.0688 \times \text{YEAR} + 0.0001 \times \text{YEARSQ} - 0.5310 \times \text{POP50} \\ + 0.0299 \times \text{POP50SQ} + 0.0048 \times \text{YRPOP50}$$

$$L(\text{MOD}) = 1.193 - 0.0110 \times \text{YEAR} - 0.0008 \times \text{YEARSQ} - 0.1020 \times \text{POP50} \\ - 0.0020 \times \text{POP50SQ} - 0.0026 \times \text{YRPOP50}$$

$$L(\text{FAS}) = 2.156 - 0.0404 \times \text{YEAR} - 0.0005 \times \text{YEARSQ} - 0.4930 \times \text{POP50} \\ + 0.1500 \times \text{POP50SQ} - 0.0018 \times \text{YRPOP50}$$

$$L(\text{HYP}) = 3.896 - 0.0659 \times \text{YEAR} - 0.0004 \times \text{YEARSQ} - 10.3780 \times \text{POP50} \\ + 17.0360 \times \text{POP50SQ} + 0.0490 \times \text{YRPOP50}$$

Even with the use of only two base variables, YEAR and POP50, it is difficult to make generalizations about the importance of the regression coefficients once the different transformations are introduced. However, as before, the estimates do indicate that the inflection points for the logistic process occurred at different times for the various growth clubs. In fact, the quadratic estimates of the half-life parameter t^* (for the average city in each club) can be shown to be as follows: SLO: -4.7 years; MOD: 30.5 years; FAS: 34.3 years; and HYP: 40.3 years. So now the alternative half-life parameter of the SLO club preceded the appropriate overall figure (i.e., $t^* = 27.75$) by nearly 32.5 years, while the half-life parameter of the HYP club lagged some 12.5 years behind the new overall figure. Moreover, these club-specific results endorse the finding, noted above, that the standard logistic model underestimates that point in time when population growth was the most rapid in the million-person cities.

Models 9 and 10 are extensions of Models 4 and 5 incorporating the influence of the seven orthogonal factors that were discussed earlier (Table 4). All of the factors except factor 4 prove to be significant, and the signs on factors 1, 2, and 6 are consistent with prior expectations. Evidently, the timing for logistic growth in the world's million-person cities was heavily impacted by both a nation's initial level of economic development and by its specific demographic characteristics. Moreover, the inclusion of these factors diminishes the significance that was attributed earlier to a city being a capital or a port. Note that the quadratic version of the logistic model improves the adjusted R^2 to 0.770 from 0.647, and reduces the SEE to 0.9098 from 1.1266. In other words, introducing the seven factors as supplemental explanatory variables improves the goodness-of-fit statistics of the logistic model by approximately 20%.

Models 11 and 12 next extend these two models by introducing city-specific coordinates, which control for unobserved variables occurring throughout the study period. But latitude was first transformed by using the standard Mercator projection in order to make the two axes orthogonal and reduce any collinearity between latitude and longitude. For the most part, this operation created only slight changes in the estimates, although nontrivial shifts (greater than 2%) were experienced by both LATIT and DISTAN3 (Weisstein 2005). Compared with earlier models, the only noticeable difference that occurs among the factors is a reversal in the significance of factors 4 and 5, both of which are relatively marginal anyway. That is, primacy becomes significant but infrastructure becomes insignificant with the introduction of the location coordinates. As expected, the signs on both longitude and

Table 4 Estimates of Logistic Population Proportions: The Extended Models

Variable	Model 9	Model 10	Model 11	Model 12
Intercept	-0.363 (-1.86)	-0.259 (-1.37)	-1.608 (-6.05)	-1.469 (-5.75)
YEAR	-6.727E-02 (-105.24)	-4.914E-02 (-21.08)	-6.728E-02 (-106.52)	-4.919E-02 (-21.40)
YEARSQ		-3.807E-04 (-10.30)		-3.800E-04 (-10.43)
POP50	-0.204 (-16.30)	-0.779 (-26.42)	-0.203 (-16.45)	-0.788 (-27.07)
POP50SQ		4.901E-02 (17.58)		5.017E-02 (18.23)
YRPOP50		6.788E-03 (12.20)		6.790E-03 (12.38)
CAPITAL	-2.082E-02 (-0.59)	3.552E-02 (1.05)	-7.930E-03 (-0.23)	4.990E-02 (1.49)
MAJPORT	1.202E-02 (0.39)	5.731E-02 (1.95)	2.410E-02 (0.79)	6.896E-02 (2.36)
DISTAN3	7.152E-02 (11.16)	7.085E-02 (11.54)	0.121 (13.64)	0.119 (14.13)
FACTOR1	-0.554 (-30.38)	-0.508 (-28.73)	-0.575 (-31.25)	-0.527 (-29.63)
FACTOR2	0.375 (21.48)	0.345 (20.51)	0.306 (16.11)	0.269 (14.74)
FACTOR3	-7.350E-02 (-5.78)	-5.506E-02 (-4.50)	-3.895E-02 (-3.02)	-1.840E-02 (-1.49)
FACTOR4	-5.537E-02 (-0.40)	-1.282E-02 (-0.97)	-3.686E-02 (-2.65)	-4.632E-02 (-3.47)
FACTOR5	-4.527E-02 (-3.60)	-3.264E-02 (-2.70)	-6.202E-03 (-0.48)	8.208E-03 (0.68)
FACTOR6	0.539 (43.00)	0.486 (39.15)	0.408 (24.98)	0.348 (21.81)
FACTOR7	0.102 (8.10)	8.725E-02 (7.24)	0.112 (8.79)	9.588E-02 (7.88)
LONGIT			-3.662E-03 (-10.16)	-3.701E-03 (-10.74)
LATIT			-3.782E-03 (-5.75)	-4.250E-02 (-6.75)
SEE	0.9495	0.9098	0.9382	0.8970
Ad. R ²	0.749	0.770	0.755	0.776

SEE, standard error of the estimate; Ad. R², R² value adjusted for degrees of freedom.

latitude are negative. The addition of these two variables only slightly improves the two goodness-of-fit statistics.

Finally, Models 13 and 14 estimate these last two models within each of the four growth clubs. Model 13 represents the linear version (Table 5), and is somewhat easier to interpret. Model 14, on the other hand, represents the quadratic version (Table 6) of the logistic function.

Model 13 was first estimated for all 485 cities by introducing appropriate dummies as was discussed above. This increased the adjusted R^2 of Model 11 from 0.755 to 0.827, and reduced the SEE from 0.9390 to 0.7874. Model 14 was then estimated and, compared with Model 12, the coefficient of determination was raised from 0.776 to 0.838 and the standard error was reduced from 0.8970 to 0.7635. The improvement in the second statistic over earlier models was approximately 20% again! In other words, even with the inclusion of the seven factors, which controlled for national differences in economic development and demography, each of the four clubs exhibited remarkably different logistic growth patterns during the study period. The club-specific regressions shown in Tables 5 and 6 are very revealing with regard to these differences. One very interesting finding is that the orthogonal factors did not always influence growth across the four clubs in a uniform manner (e.g., see factor 4), although the signs on factors 1, 2, and 6 are still largely consistent with expectations. Moreover, being a national capital now has an important effect on the timing of logistic growth, moving the starting time up for most of the slower-growing cities (as anticipated) but moving the starting time back for the fastest-growing cities. And distance to world markets is no longer significant for all city clubs, although the sign is correct in all four cases. But the most revealing results are the goodness-of-fit statistics, which indicate that the estimates of logistic population growth become increasingly better as the city clubs exhibit faster growth. By far the worst fit is for the SLO group, for many of these cities had already passed the inflection point in their growth curve before 1950.

A few other remarks should be made about the sizes of the β coefficients in Model 13 (which is easier to interpret), as these provide evidence regarding the relative importance of the different regression variables in the logistic growth process. As the raw estimates suggest, the β coefficient (absolute value) on YEAR does become increasingly larger as the clubs comprise faster-growing cities: for instance, in SLO the estimate is -0.590 but in HYP the estimate is -0.907 . Initial population shows a similar trend in its impact, increasing from -0.049 in SLO to -0.126 in HYP, but distance to world markets exhibits the opposite trend, declining from 0.403 in SLO to 0.032 in HYP. Six of the seven orthogonal factors have large β coefficients in the slowest-growing club (e.g., -0.231 for factor 1 and -0.202 for factor 3) but only two factors have coefficients exceeding 0.100 in the fastest-growing club (-0.109 for factor 1 and 0.134 for factor 2). Across the four regressions, FACTOR1 has the most ubiquitous effect, outside of YEAR, and clearly dominates initial population as an influence on growth in three of the four city clubs.

Table 5 Club-Specific Estimates: Model 13 (Linear Version)

Variable	Slow	Moderate	Fast	Hyper
Intercept	-4.529 (-5.06)	1.450 (4.32)	0.684 (1.72)	2.865 (6.78)
YEAR	-5.446E-02 (-34.85)	-5.979E-02 (-78.42)	-7.190E-02 (-93.51)	-8.395E-02 (-109.07)
POP50	-5.256E-02 (-2.55)	-8.843E-02 (-4.51)	-0.164 (-4.88)	-2.604 (-14.21)
CAPITAL	-0.241 (-2.38)	-0.246 (-4.51)	-6.257E-02 (-1.43)	0.228 (5.06)
MAJPORT	0.121 (1.45)	-5.538E-02 (-1.33)	-2.551E-02 (-0.65)	-8.074E-02 (-2.02)
DISTAN3	0.179 (5.87)	4.400E-03 (0.40)	5.051E-02 (3.93)	1.918E-02 (1.42)
FACTOR1	-0.450 (-4.04)	-0.235 (-10.71)	-0.465 (-13.25)	-0.226 (-6.61)
FACTOR2	0.384 (2.28)	0.101 (3.42)	0.102 (4.37)	0.243 (9.58)
FACTOR3	-1.024 (-5.78)	-3.569E-02 (-3.33)	-0.512 (-9.00)	8.135E-02 (1.48)
FACTOR4	-0.255 (-5.47)	0.183 (7.86)	-0.153 (-8.36)	-6.332E-02 (-3.12)
FACTOR5	0.190 (5.67)	0.178 (4.84)	4.028E-03 (0.14)	6.160E-02 (1.99)
FACTOR6	0.146 (2.38)	0.156 (4.61)	8.699E-02 (2.28)	-4.002E-02 (-0.95)
FACTOR7	-0.205 (-1.67)	3.004E-02 (1.39)	3.816E-02 (2.56)	-9.224E-03 (-0.53)
LONGIT	-3.898E-03 (-3.52)	9.337E-04 (2.12)	-2.639E-03 (-4.08)	-6.836E-04 (-1.30)
LATIT	-4.505E-03 (-1.88)	2.071E-03 (2.32)	2.507E-04 (0.25)	-3.393E-03 (-3.11)
N	1690	1469	1625	1521
SEE	1.2019	0.5468	0.5799	0.5616
Adj. R ²	0.515	0.821	0.853	0.895

NOTE: The combined model, using all N = 6305 observations and three club-specific dummies, has a standard error of the estimate (SEE) = 0.7874 and an R² value adjusted for degrees of freedom (Adj. R²) = 0.827.

Table 6 Club-Specific Estimates: Model 14 (Quadratic Version)

Variable	Slow	Moderate	Fast	Hyper
Intercept	– 4.047 (– 4.55)	1.007 (3.34)	0.394 (1.03)	3.011 (7.46)
YEAR	– 6.855E–02 (– 11.48)	– 1.095E–02 (– 4.19)	– 4.034E–02 (– 14.23)	– 6.579E–02 (– 22.62)
YEARSQ	1.180E–04 (1.27)	– 7.828E–04 (– 19.11)	– 5.134E–04 (– 11.56)	– 3.933E–04 (– 8.92)
POP50	– 0.341 (– 6.18)	– 9.438E–02 (– 1.95)	– 0.416 (– 3.89)	– 7.735 (– 13.33)
POP50SQ	1.590E–02 (3.40)	1.565E–02 (2.15)	0.118 (3.38)	10.149 (7.50)
YRPOP50	4.794E–03 (5.02)	– 2.598E–03 (– 3.41)	– 1.852E–03 (– 1.22)	4.908E–02 (5.63)
CAPITAL	– 0.211 (– 2.10)	– 0.247 (– 5.07)	– 6.304E–03 (– 0.14)	0.212 (4.95)
MAJPORT	0.123 (1.48)	– 3.856E–02 (– 1.02)	4.386E–03 (0.13)	– 6.057E–02 (– 1.59)
DISTAN3	0.176 (5.78)	4.369E–03 (0.45)	5.132E–02 (4.27)	1.907E–02 (1.49)
FACTOR1	– 0.435 (– 3.94)	– 0.246 (– 12.18)	– 0.461 (– 13.66)	– 0.221 (– 6.82)
FACTOR2	0.359 (2.15)	9.231E–02 (3.47)	0.109 (4.87)	0.230 (9.54)
FACTOR3	– 0.961 (– 5.46)	– 3.321E–02 (– 3.46)	– 0.526 (– 9.64)	5.516E–02 (1.05)
FACTOR4	– 0.252 (– 5.46)	0.182 (8.74)	– 0.159 (– 9.03)	– 5.467E–02 (– 2.83)
FACTOR5	0.191 (5.77)	0.174 (5.27)	4.879E–03 (0.18)	7.389E–02 (2.51)
FACTOR6	0.130 (2.14)	0.163 (5.38)	8.925E–02 (2.46)	– 4.634E–02 (– 1.16)
FACTOR7	– 0.201 (– 1.66)	3.113E–02 (1.61)	2.982E–02 (2.06)	– 6.579E–03 (– 0.38)
LONGIT	– 4.071E–03 (– 3.72)	8.137E–04 (2.05)	– 2.457E–03 (– 3.96)	– 7.057E–04 (– 1.41)
LATIT	– 5.405E–03 (– 2.27)	1.8521E–03 (2.31)	7.914E–04 (0.81)	– 2.349E–03 (– 2.25)
N	1690	1469	1625	1521
SEE	1.1894	0.4870	0.5557	0.5332
Ad. R ²	0.525	0.858	0.865	0.905

NOTE: The combined model, using all N = 6305 observations and three club-specific dummies, has a standard error of the estimate (SEE) = 0.7635 and an R² value adjusted for degrees of freedom (Ad. R²) = 0.838.

Conclusions

Using recently released UN data, this article has examined the logistic population growth experienced by nearly 500 of the world's largest cities. The results, although preliminary in the sense that not all relevant initial conditions were controlled for, are very encouraging. Logistic growth was shown to vary according to city-specific attributes, country-specific factors, and location. Those places that were large in 1950, the initial year for the study, were shown to begin logistic growth before those cities that were smaller in 1950. Cities in nations having low levels of economic development in 1960 were shown to have their logistic growth lag behind cities in nations having high levels of economic development. Those cities in countries with youthful populations in 1960 also lagged behind in their logistic growth. And places close to world markets generally began logistic growth sometime before those places found in more peripheral regions. Moreover, these estimates were shown to vary across four different population-growth clubs, a result that was based on the earlier work of Mulligan and Crampton (2005).

The research was also exploratory in that the behavioral mechanisms for generating urban logistic growth were only partially fleshed out. In fact, initial conditions of the sort adopted in this study can be expected to drive subsequent city growth for only a few decades at most. After that, the relative levels for the different contextual variables would inexorably shift, suggesting that the entire data set should be updated every decade or so. This indicates that a lagged model, with clearer behavioral relationships articulated among the variables (including the nation's level of urbanization), should be designed in an attempt to improve the results shown above. Moreover, efforts should be made to acquire earlier data for national birth and death rates, to find accurate and comprehensive measures of internal migration within nations (Cook 1991), to introduce suitable measures of political organization and openness, and to differentiate cities according to their roles in national and international hierarchies. Perhaps a return to the research of Gallup, Sachs, and Mellinger (1999) might be instructive, in that they were able to develop various empirical linkages between a nation's geographic attributes and its economic growth. And the work of the Loughborough group would be very helpful in trying to measure how the quality and intensity of intercity relationships vary across the world's largest places (Taylor 1997). For the million-person cities found in the more developed world, data might even indicate how logistic growth is related to land and housing values (Kelley and Williamson 1984), industrial and occupational composition of the labor force (Markusen and Schrock 2003), and natural and human-made amenities (Blomquist, Berger, and Hoehn 1988). Finally, a more geographic line of inquiry, based on geographically weighted regression, might disclose that the sorts of growth factors discussed above actually have substantially different impacts at different locations on Earth's surface (Fotheringham, Brunsdon, and Charlton 2000).

In a recent article, David (1999) challenged social scientists to analyze data on resource allocations in space and time in order to obtain a less stylized perspective on the economic geography of development. In line with this challenge, the present study has developed a preliminary foundation for understanding logistic population growth in the world’s largest cities. Implicit to this study’s results is that urban population is path dependent but the parameters of each path are somewhat different, depending upon both location and contextual factors. Urbanization occurs in similar ways at different places but the timing and magnitude of this urbanization varies from one place to another. Theoretical and applied problems still abound in this rich research area, including specification of the appropriate upper limit for measuring urban logistic growth, incorporation of feedback between employment growth and population change, and exploration of how different national policies (e.g., immigration, economic union) channel subsequent urban growth.

Acknowledgements

I thank the editor and three referees for their various helpful comments on an earlier version. Former colleagues John Kupfer and Richard Reeves also provided useful commentary on certain technical issues.

Appendix A

Table A1 National Indicators c. 1960

Variable	Descriptive statistics			
	Minimum	Maximum	Mean	Source
AREA	1.0	16,888.0	4778.2	4
BRATE	15.0	54.0	35.8	4
CITYDN	0.4	1000.0	16.4	4, 5
CNRATE	1.0	54.0	17.5	4
DRATE	6.0	31.0	14.3	4
EXPORT	1.0	99.0	52.9	2
GDPPC	75.0	10,707.0	2349.0	3
MAIL	0.0	19.5	1.3	1
MANUF	1.0	52.0	19.4	4
NUMBR	1.0	110.0	37.8	5
OPEN	4.5	341.8	28.0	2
POPDN	0.9	2950.0	80.3	3, 4
PRIME1	32.1	100.0	51.2	1
PRIME2	6.0	100.0	18.4	4
ROADS	0.1	302.0	19.3	1

Table A1 Continued

Variable	Descriptive statistics			
	Minimum	Maximum	Mean	Source
SCHOOL	1.0	90.0	27.8	1
SERVS	3.0	69.0	24.9	4
TOTPOP	0.1	692.7	240.9	4
URBAN	3.0	100.0	36.5	4
URBCON	0.1	9.1	2.0	4, 5
YOUTH	13.5	36.1	21.9	1

NOTES: Some of the late-1950s data from the *Atlas of Economic Development* pertained to colonial areas (especially in Africa) and subjective allocations were made to emerging nations. Likewise, other data from various sources pertained to national units (e.g., USSR, Yugoslavia) falling out of existence during the study period and again subjective allocations were made to the new nations. And in a few (mostly Eastern Bloc) cases the data for openness were based on the first date at which estimates were available. $N = 485$ million-person cities times 13 time periods = 6305 observations. All means are weighted according to the number of million-person cities in each nation. AREA, land area in thousands of square kilometers; BRATE, crude birth rate per thousand; CITYDN, density of million-person cities (per million square kilometers); CNRATE, children aged 1–4 years death rate per thousand; DRATE, crude death rate per thousand; EXPORT, % exports to industrial market economies; GDPPC, per capita gross domestic product (GDP) in 1997 US\$; MAIL, international mail outflow per 1000 people; OPEN, exports plus imports divided by GDP (adjusted for currencies); NUMBR, number of million-person cities; MANUF, % labor force in manufacturing; POPDN, density of total population (per square kilometer); PRIME1, population of largest city divided by population of four largest cities; PRIME2, % urban population residing in largest city; ROADS, road kilometers per 100 square kilometers; SCHOOL, % children aged 5–14 in primary school; SERVS, % labor force in services; TOTPOP, total population in millions; URBAN, % total population residing in cities; URBCON, urban population (in millions) divided by number of million-person cities; YOUTH, % total population aged 5–14. Sources: (1) Ginsburg, N. (1961) *Atlas of Economic Development*; (2) Heston, A. et al. (2002) *Penn World Table* (Version 6.1); (3) United Nations (1997) *Human Development Report 1997*; (4) World Bank (1981) *World Development Report 1981*; (5) United Nations (2004) *World Urbanization Prospects: The 2003 Revision*.

Table A2 Major Factor Loadings: Varimax Rotation

Variable	Factors						
	1	2	3	4	5	6	7
AREA	0.25	−0.75			−0.25	0.27	
BRATE	−0.78	0.35				0.42	
CITYDN			0.95				
CNRATE	−0.91						
DRATE	−0.72	0.46			0.30		

Table A2 Continued

Variable	Factors						
	1	2	3	4	5	6	7
EXPORT		0.83				0.35	
GDPPC	0.85				0.30		
MAIL	0.56		0.28	0.30			
MANUF	0.89						
NUMBR		− 0.90					
OPEN		0.26					0.91
POPDN			0.97				
PRIME1		0.43		0.69			0.30
PRIME2				0.91			
ROADS	0.61				0.65	− 0.25	
SCHOOL	0.89						
SERVS	0.88	0.36			0.34	0.35	
TOTPOP	− 0.30	− 0.86		− 0.25			
URBAN	0.90	0.29					
URBCON	0.55					− 0.68	
YOUTH	− 0.61	0.44			− 0.38	− 0.26	
% Variance	35.79	18.85	9.91	8.03	5.85	5.78	5.45

NOTE: Only factor loadings whose absolute values exceeded 0.250 are shown. The various factors are interpreted by the sizes of their respective loadings. factor 1, economic development; factor 2, population characteristics; factor 3, population density; factor 4, urban primacy; factor 5, infrastructure; factor 6, youthfulness; factor 7, openness (see the text for details and examples).

References

Abler, R., J. S. Adams, and P. Gould. (1971). *Spatial Organization: The Geographer's View of the World*. Englewood Cliffs, NJ: Prentice-Hall.

American Association of Port Authorities. (2005). Port Industry Statistics. www.aapa-ports.org/industryinfo

Anand, S., and M. Ravallion. (1993). "Human Development in Poor Countries: On the Role of Private Incomes and Public Services." *Journal of Economic Perspectives* 7, 133–150.

Arthur, W. B. (1994). *Increasing Returns and Path Dependence in the Economy*. Ann Arbor: University of Michigan Press.

Berry, B. J. L. (1961a). "A Statistical Analysis." In *Atlas of Economic Development*, 110–19, edited by N. Ginsburg. Chicago: University of Chicago Press.

Berry, B. J. L. (1961b). "City Size Distributions and Economic Development." *Economic Development and Cultural Change* 9, 573–88.

Berry, B. J. L. (1973). *The Human Consequences of Urbanization*. New York: St. Martin's Press.

Berry, B. J. L., and F. E. Horton. (1970). *Geographic Perspectives on Urban Systems*. Englewood Cliffs, NJ: Prentice-Hall.

- Blomquist, G. C., M. C. Berger, and J. P. Hoehn. (1988). "New Estimates of Quality of Life in Urban Areas." *American Economic Review* 78, 89–107.
- Borchert, J. R. (1967). "American Metropolitan Evolution." *Geographical Review* 57, 301–32.
- Brakman, S., H. Garretsen, and C. van Marrewijk. (2001). *An Introduction to Geographical Economics*. Cambridge, UK: Cambridge University Press.
- Canaleta, C. G., P. P. Arzo, and M. R. Gárate. (2004). "Regional Economic Disparities and Decentralisation." *Urban Studies* 41, 71–94.
- Carroll, G. R. (1982). "National City-Size Distributions: What do We Know after 67 Years of Research?" *Progress in Human Geography* 6, 1–43.
- Casetti, E., and R. K. Semple. (1969). "Concerning the Testing of Spatial Diffusion Hypotheses." *Geographical Analysis* 3, 254–59.
- Cavalli-Sforza, L. L., and M. W. Feldman. (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.
- Cheshire, P. (1995). "A New Phase of Urban Development in Western Europe? The Evidence for the 1980s." *Urban Studies* 32, 1045–63.
- Cliff, A. D., P. Haggett, J. K. Ord, and G. R. Vesey. (1981). *Spatial Diffusion: An Historical Geography of Epidemics in an Island Community*. Cambridge, UK: Cambridge University Press.
- Cook, C. J. (1991). "Rural-Urban Migration Rates and Development: A Quantitative Note." *Review of Urban and Regional Development Studies* 11, 63–75.
- David, P. A. (1999). "Krugman's Economic Geography of Development: NEG, POG, and Naked Models in Space." *International Regional Science Review* 22, 162–72.
- Fotheringham, A. S., C. Brunsdon, and M. Charlton. (2000). *Quantitative Geography: Perspectives on Spatial Data Analysis*. London: Sage.
- Friedmann, J. (1986). "The World City Hypothesis." *Development and Change* 17, 69–83.
- Friedmann, J. (1995). "Where We Stand: A Decade of World City Research." In *World Cities in a World-System*, 21–47, edited by P. L. Knox and P. J. Taylor. Cambridge, UK: Cambridge University Press.
- Fujita, M., P. Krugman, and A. J. Venables. (1999). *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: MIT Press.
- Fujita, M., and T. Mori. (1996). "The Role of Ports in the Making of Major Cities: Self-Agglomeration and Hub Effect." *Journal of Development Economics* 49, 93–120.
- Gaile, G. L., and D. M. Hanink. (1985). "Relative Stability in American Metropolitan Growth." *Geographical Analysis* 17, 241–48.
- Gallup, J. L., J. D. Sachs, and A. D. Mellinger. (1999). "Geography and Economic Development." *International Regional Science Review* 22, 179–232.
- Ginsburg, N. (1961). *Atlas of Economic Development*. Chicago: University of Chicago Press.
- Glaeser, E. L., H. D. Kallal, J. A. Scheinkman, and A. Shleifer. (1992). "Growth in Cities." *Journal of Political Economy* 100, 1126–52.
- Greene, W. H. (2000). *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice-Hall.
- Hagerstrand, T. (1965). "A Monte Carlo Approach to Diffusion." *European Journal of Sociology* 6, 43–67.
- Hall, P. (1984). *The World Cities*, 3rd ed. New York: St. Martin's Press.
- Hannan, M. T., and J. Freeman. (1989). *Organizational Ecology*. Cambridge, MA: Harvard University Press.

- Harris, J., and M. Todaro. (1970). "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 40, 126–42.
- Henderson, J. V. (1988). *Urban Development: Theory, Fact, and Illusion*. Oxford, UK: Oxford University Press.
- Henderson, J. V., and H. G. Wang. (2005). Urbanization and City Growth: The Role of Institutions. www.econ.brown.edu/faculty/henderson/papers
- Heston, A., R. Summers, and B. Allen. (2002). *Penn World Table (Version 6.1)*. Center for International Comparisons. Philadelphia: University of Pennsylvania.
- Hoffman, M. S. (1988). *The World Almanac and Book of Facts 1989*. New York: Pharos Books.
- Hudson, J. C. (1972). *Geographical Diffusion Theory*. Department of Geography, Studies in Geography No. 19. Evanston, IL: Northwestern University.
- Kasarda, J. D., and E. M. Crenshaw. (1991). "Third World Urbanization: Dimensions, Theories, and Determinants." *Annual Review of Sociology* 17, 467–501.
- Kelley, A. C., and J. G. Williamson. (1984). *What Drives Third World City Growth?* Princeton, NJ: Princeton University Press.
- Keyfitz, N. (1980). "Do Cities Grow by Natural Increase or by Migration?" *Geographical Analysis* 12, 142–56.
- Keyfitz, N., and H. Caswell. (2005). *Applied Mathematical Demography*, 3rd ed. New York: Springer.
- Knox, P. L., and P. J. Taylor, eds. (1995). *World Cities in a World-System*. Cambridge, UK: Cambridge University Press.
- Lloyd, P. E., and P. Dicken. (1977). *Location in Space: A Theoretical Approach to Economic Geography*, 2nd ed. New York: Harper and Row.
- Markusen, A., and G. Schrock. (2003). "Cities as Hierarchists or Specialists? Evidence from Occupational Profiles." Working Paper #258, Humphrey Institute of Public Affairs. University of Minnesota, Minneapolis, MN.
- Moomaw, R. L., and M. Alwosabi. (2004). "An Empirical Analysis of Competing Explanations of Urban Primacy Evidence from Asia and the Americas." *The Annals of Regional Science* 38, 149–71.
- Morrill, R. (1965). "The Negro Ghetto: Problems and Alternatives." *Geographical Review* 55, 339–61.
- Mulligan, G. F., and J. P. Crampton. (2005). "Population Growth in the World's Largest Cities." *Cities* 22, 365–80.
- Pindyck, R. S., and D. L. Rubinfeld. (1991). *Econometric Models and Economic Forecasts*. New York: McGraw-Hill.
- Porter, M. (1990). *The Competitive Advantage of Nations*. New York: The Free Press.
- Ray, D. (1998). *Development Economics*. Princeton, NJ: Princeton University Press.
- Rose, M. R. (1987). *Quantitative Ecological Theory: An Introduction to Basic Models*. Baltimore: Johns Hopkins University Press.
- Sassen, S. (1991). *The Global City: New York, London, Tokyo*. Princeton, NJ: Princeton University Press.
- Sassen, S. (2000). *Cities in a World Economy*, 2nd ed. Thousand Oaks, CA: Pine Forge Press.
- Sassen, S., ed. (2002). *Global Networks: Linked Cities*. New York: Routledge.
- Scott, A. J., ed. (2001). *Global City Regions: Trends, Theory, Policy*. Oxford, UK: Oxford University Press.

- Showers, V. (1989). *World Facts and Figures*, 3rd ed. New York: Wiley.
- Smith, D. A., and M. Timberlake. (1995). "Conceptualising and Mapping the Structure of the World System's City System." *Urban Studies* 32, 287–302.
- Stelder, D. (2005). "Where do Cities Form? A Geographical Agglomeration Model for Europe." *Journal of Regional Science* 45, 657–79.
- Taylor, P. (1997). "Hierarchical Tendencies amongst World Cities: A Global Research Proposal." *Cities* 14, 323–32.
- Tolley, G. S., and V. Thomas, eds. (1987). *The Economics of Urbanization and Urban Policies in Developing Countries*. Washington, DC: The World Bank.
- Thomas, R. W., and R. J. Huggett. (1980). *Modelling in Geography: A Mathematical Approach*. Totowa, NJ: Barnes and Noble.
- United Nations. (1997). *Human Development Report 1997*. New York: Oxford University Press.
- United Nations. (2004). *World Urbanization Prospects: The 2003 Revision*. New York: Author.
- van Duijn, J. J. (1983). *The Long Wave in Economic Life*. London: George Allen and Unwin.
- Weisstein, E. W. (2005). <http://mathworld.wolfram.com/MercatorProjection.html>
- Wikipedia. (2005). http://en.wikipedia.org/wiki/Great_circle_distance
- World Bank. (1981). *World Development Report 1981*. New York: Oxford University Press.