

Urbanization and city growth: The role of institutions [☆]

J. Vernon Henderson ^{*}, Hyoung Gun Wang

Brown University, USA

Received in revised form 12 October 2006; accepted 6 November 2006

Available online 10 January 2007

Abstract

This paper examines how urbanization is accommodated by increases in numbers and in sizes of cities. Political institutions play a key role. Estimation uses a worldwide data set on all metro areas over 100,000 from 1960 to 2000. The degree of democratization and technological advances strongly affect growth in both city numbers and individual city sizes. Effects on city sizes are heterogeneous. Technology improvements help bigger cities relative to smaller ones. Increasing democratization levels the playing field across the urban hierarchy, helping smaller cities. Given these opposing effects, the overall relative size distribution of cities worldwide is unchanged over the time period.

© 2006 Elsevier B.V. All rights reserved.

JEL classification: H7; O3; O18; R

Keywords: Urban growth; City size distribution; Democratization; Urbanization

Countries urbanize as they develop, with urbanization transforming the national landscape. Urbanization has three inter-related dimensions that the literature studies: changes in the size distribution of cities (Eaton and Eckstein, 1997; Dobkins and Ioannides, 2001), growth in individual city population sizes (Glaeser et al., 1995; Black and Henderson, 2003), and growth in city numbers (Dobkins and Ioannides, 2001; Black and Henderson, 2003). We examine all three dimensions: each offers a particular perspective and relates to a particular literature and the three dimensions complement each other. Although we try to ground the hypotheses in the theoretical literature, this is an empirical paper that identifies the key forces driving urbanization outcomes.

[☆] This research was supported by a grant from the Research Committee of the World Bank. Rupa Ranganathan of the World Bank prepared the original data and documentation. The work has benefits from comments by participants in presentations at LSE, the Wider Institute, and the Regional Science Association International meetings.

^{*} Corresponding author. Department of Economics, Brown University, Providence, RI, 02912, USA. Tel.: +1 401 863 2886; fax: +1 401 863 1970.

E-mail address: J_Henderson@Brown.edu (J.V. Henderson).

First is technological progress, which in our data and in modeling drives the expansion of the urban relative to rural sector (Henderson and Wang, 2005). Technological progress also fosters growth in individual city sizes, because knowledge accumulation either interacts with and enhances urban scale economies or improves the ability to manage cities through, for example, innovations in commuting technology. Both lead to larger equilibrium and efficient city sizes. Intuition might suggest some innovations are more important for bigger than smaller cities. Larger cities have higher commuting costs and congestion and pose greater managerial challenge. Their more business service and high tech oriented production may benefit relatively more from knowledge spillovers, than the standardized manufacturing found in smaller and medium size cities.

Second, institutions affect the size distribution of cities and growth of small relative to large cities. In particular, we argue that an increase in the degree of democracy leads to faster city formation, less of the national population growth being accommodated in bigger cities, and a reduction in the degree of spatial inequality, as defined later by spatial Gini's. Democratization implies the election of regional representatives to a national assembly which leads to increased regional representation. It in turn is associated with increasing fiscal decentralization and greater local autonomy. These changes help level the playing field upon which cities compete for national resources, limiting the ability of a national ruling class to favor the national capital and other key cities in which they live (Ades and Glaeser, 1995).

While technological growth and democratization are the key factors we focus on, globalization and geography are also factors to consider. Cities are heterogeneous in terms of geography having better or worse access to national and international markets in particular. Given this, there is a debate (e.g., Krugman and Elizondo, 1996) about whether increased openness of countries favors coastal cities versus hinterland cities. One hypothesis in the new economic geography literature is that increased openness fostered by tariff and transport cost reductions, opens a greater array of markets for hinterland cities, enhancing their growth relative to coastal cities. But there is a more traditional view that coastal cities, which are the gateways to trade and FDI, distinctly benefit more from openness than their hinterland competitors.

To study urbanization, we assemble a worldwide data set for metropolitan areas over 100,000 from 1960 to 2000, with a base of 142 countries. In terms of world urbanization, prior work has focused mostly on the largest metro areas, such as national capitals or primate cities, examining the effects on urban primacy of institutions and growth and the consequences of excessive or deficient concentration of resources in the largest cities (e.g., Ades and Glaeser, 1995; Henderson, 2003). This paper makes several contributions. First, it looks at the entire size distribution of metro areas in countries, not just the largest cities, which we hope will influence some of the debates about urbanization processes. The popular press (e.g. Economist, Time, Newsweek, National Geographic) and international agencies tend to presume that urban growth is dominated by mega-city development, with an increasingly skewed distribution of cities. International agencies warn about the dangers of mega-city growth and the perils associated with the development of "unbalanced urban hierarchies", and advocate the development of medium size cities (UN, 1993). The *World Development Report (2000, Chapter 7)* emphasizes the grim side of life in mega-cities. While there are certainly issues with certain mega-cities, one underlying premise of these reports is wrong. Urbanization is not concentrated in mega-cities and urban hierarchies are not becoming increasingly unbalanced with urbanization. First, the worldwide relative size distribution of cities has been rock stable over the last 40 years. Second, much of urbanization occurs through the development of new cities and growth of smaller metro areas and little of the world's urban population lives in mega-cities. In summary, one contribution of this

paper is to provide a set of facts about urbanization based on looking at the entire size distribution of worldwide metro areas, rather than just one or two cities in a country.

Second, the paper examines key forces governing the urbanization process, showing the effects of technological progress and democratization on city formation and city size growth. As such we will explain why the relative size distribution of cities has remained invariant over the last 40 years: the forces of technological progress and increased democratization have offset each other over this time period, the former acting to increase relative urban concentration and the latter to reduce it. Demonstrating the effect of democratization is difficult because institutions change slowly and typically country coverage is limited. Our large sample of countries and cities over a 40 year time span makes this identification possible. Finally, we show that many effects are heterogeneous across cities, in ways which may accord with intuition. For example, increased openness benefits coastal cities more than hinterland ones; and cities, given their place in space, have very different levels and changes in access to national markets over time.

1. A conceptual framework and issues in the literature

The systems of cities literature predicts that national urban population growth induced by technological change will be accommodated by growth in potentially both city population sizes and the number of cities in a country (Black and Henderson, 1999; Duranton and Puga, 2001; Henderson and Wang, 2005). In terms of city sizes, one starts by modeling equilibrium city sizes at a point in time. In urban models, real income per worker in a city is an inverted-U shaped function of city size (population). The inverted-U represents a trade-off between the initial scale benefits of increasing city size versus the commuting and other organizational diseconomies that set in as a city gets larger. In a world in which the national population is perfectly mobile, there is an effectively unlimited supply of equal quality sites upon which cities form, and city governments (public or private as in “large” developers) have the autonomy to form and invest in new cities and to limit populations through zoning and land use regulation (Henderson and Becker, 2000). Cities operate at the sizes where real income per worker peaks. There is of course not just one type of city with one equilibrium size, but a system of cities specialized in different products where each type has its own equilibrium size.

With economic growth, cities grow in size as technological change shifts the inverted-U shape curves up and out. In a simple deterministic version, different types of cities grow in parallel, or at the same rate, so that the relative size distribution of cities is constant over time. But even in more complex models with technology shocks, city size distributions evolve to an invariant form as growth proceeds (Rossi-Hansberg and Wright, 2004; Duranton, 2004). But as suggested above, it may be more realistic to think of technological progress being biased towards enhancing the efficiency of larger compared to smaller cities. In particular per capita infrastructure requirements of cities appear to escalate with city size (Richardson, 1987), and so do commuting and congestion costs. Innovations that help improve efficiency of infrastructure investments, commuting technology, or urban management techniques may enhance growth of bigger cities compared to smaller ones.

The literature also examines the effect of democratization on urban concentration. Evidence suggests that in less democratic societies, national elites tend to favor the national capitals or the larger cities where they live (Ades and Glaeser, 1995; Davis and Henderson, 2003). Favoritism can involve provision of better public services such as better quality schooling and location of and admission to elite colleges. But it can be much more extensive. It can take the form of favored capital allocations to cities higher in a political hierarchy with China being a good example

(Jefferson and Singhe, 1999), the allocation of export or import licenses to favored cities (Henderson and Kuncoro, 1996), the allocation of government owned enterprises such as heavy industry in Brazil, or the spatial provision of transport and telecommunications infrastructure. In urban models (Henderson, 1988; Ades and Glaeser, 1995), migrants are drawn to the favored cities, both increasing the sizes of these cities and dissipating the benefits of favoritism. Elites in these cities can inhibit immigration through trying to restrict land use development, forcing migrants into unserved, illegal “squatter” settlements or through legal restrictions on migration as in China (Au and Henderson, 2006). Such restrictions raise the costs of in-migration, retard (but not eliminate) the extent to which favored cities expand, and help preserve some of the benefits of favoritism for initial residents.

Why might democratization help smaller cities? Although there are no urban political economy models that fully address this issue, there appear to be three sets of reasons. First, democratization may limit the ability of national elites to impose an agenda on the majority, which favors these elites. Second, as noted earlier democratization involves election of regional representatives to a national assembly. These representatives will call for better services for their regions, improved national road and telecommunication systems linking hinterland cities to markets, and better access to national capital markets and provision of import and export licenses and the like. With national governance by an elected majority in the legislature, versus a centrally based political elite, at least some hinterland regions are more likely to get a greater share of services. Moreover regional representation may lead to pressure for greater regional autonomy. Certainly as we will see below democratization and governmental decentralization are very strongly correlated in the data. Decentralization of national government powers to local and regional governments allows hinterland cities to compete on a level playing field with larger, traditionally more favored cities. With enhanced responsibilities and revenues, they can provide better services and compete for industry and migrants.

In addition, democratization may involve development of better and more transparent institutions. In the systems of cities model we outlined achieving efficient equilibrium city sizes requires certain powers for local governments. At a minimum local governments need the ability to set up new cities — in essence to finance infrastructure investments in towns expanding into cities and enable viable land use development in well functioning land markets (with clear title to land and transparent regulations). And the process of obtaining an efficient equilibrium is helped by the ability of existing cities to limit expansion through zoning and land use regulation. Absent such institutions cities tend to be too large and too few in number (Henderson and Becker, 2000). The institutions solve the coordination problem inherent in getting masses of people to move to new cities when it is efficient to do so. For a country starting with limited local government autonomy and limited land market functioning, improved institutions governing local land markets and the functioning of local governments make it easier to form new cities. That speeds city formation and reduces urban concentration as new cities attract resources away from old cities. This discussion begs the question of what leads to increased democratization. The empirical association is with economic growth, but we want to distinguish effects of determinants of economic growth from a pure democracy effect. Fortunately in identification countries have different histories of institutions and, apart from growth, different tendencies to democratization.

The new economic geography literature in examining the effects of transport costs and openness on hinterland versus core region development doesn't deal with issues of city size distributions (Krugman and Elizondo, 1996; Fujita et al., 1999), although there are attempts to integrate certain aspects of the two approaches in empirical (Au and Henderson, 2006) and theoretical contexts (see Abdel-Rahman and Anas, 2004 for a review). All the economic

geography literature argues that increases in market access, or market potential for a city will lead to increased demand for local goods and hence local labor, leading to increased city sizes through in-migration. We contribute to the literature by trying to quantify this effect. But the real debate as noted earlier is whether increased openness helps hinterland cities relative to coastal ones, or not; and we examine this issue.

2. Facts about cities and urbanization

The urbanization rate, defined as increases in the percent of the national population that is urbanized, is most rapid at low income levels, and then tails off as countries become “fully urbanized” (e.g., [World Development Report, 2000](#)). The definition of full urbanization and the rural/urban division under full urbanization vary across countries, so full urbanization usually ranges from 65 to 85% of the population being urbanized. This paper looks at the portion of the urban population in larger cities, or metro areas.

Our base sample is the metro areas of the world with populations over 100,000 every 10 years from 1960 to 2000. Data on metro area populations are from a variety of sources, cited in the Appendix B. The 100,000 cut-off is chosen for practical reasons — it is the cut-off employed by many countries. None have a higher cut-off and most do not provide consistent data over time on cities below 100,000. Even USA metro areas which in theory have a cut-off of 50,000, in practice only include comprehensively urban counties with over 85,000 urban residents ([Black and Henderson, 2003](#)). For the largest cities (typically metro areas over 750,000 in 1990) the UN World Urbanization Prospects data are utilized for 1960–2000. For smaller cities, populations for metro areas (especially post-1970) are available from the UN *Demographic Year Book* and country annual statistical yearbooks. Also utilized are www.citypopulation.de and www.World-Gazetteer.com for recent years. In most countries, urban population data are based on metro area definitions for bigger cities, and for all but the smallest cities in later years. However, in some countries, metro area definitions in earlier years (1960 and sometimes 1970) are based on municipality jurisdictions. As described in the documentation for this data (see Appendix B and web link), we draw different samples. The best data years are 2000 and 1960 (often pre- or near independence in many developing countries) for which we present a comprehensive picture of world urbanization in cities over 100,000 and how that has changed over the last 40 years.

When we move to decade-by-decade growth analysis, such as examining growth in the numbers of cities in a country, a few countries have missing urban data for particular decades as detailed later. For data on individual city growth analysis, apart from arrival of new cities, the city-country panel is unbalanced because some individual cities have 1 or 2 years of “bad data” (e.g., a 1970 number that is 1/4 or fourfold the 1960 number and 1975 or 1980 numbers that are 30% higher than the 1960 number). If a country has over 25% of its cities with at least one bad observation in a given decade, the whole country is dropped for that decade; otherwise data are utilized for the sample of cities with sequences of good data.¹ All decade figures and growth rates are consistent so if data are from 1962 to 1970, 1962 is extrapolated back to 1960 based on the annual 1962–1970 growth rate.²

¹ As will be discussed below we utilize a measure of market potential, which describes a city’s access to other cities’ populations within a country. This variable is calculated on as broad a base as possible including city populations under 100,000 when available and cities with possibly flawed numbers (where bad numbers are replaced by interpolations) that are dropped otherwise from the sample.

² All that means in estimation is that estimates of the 10-year growth rate may be based on 8 or 9 (or 11 or 12) years.

Table 1
World city size distribution, 2000

	Count	Mean	Share ^a
$17,000,000 \leq n2000$	4	20,110,250	4.5
$12,000,000 \leq n2000 < 17,000,000$	7	13,412,714	5.2
$8,000,000 \leq n2000 < 12,000,000$	13	10,447,769	7.5
$4,000,000 \leq n2000 < 8,000,000$	29	5,514,207	8.8
$3,000,000 \leq n2000 < 4,000,000$	41	3,422,461	7.8
$2,000,000 \leq n2000 < 3,000,000$	75	2,429,450	10.1
$1,000,000 \leq n2000 < 2,000,000$	249	1,370,773	18.9
$500,000 \leq n2000 < 1,000,000$	355	703,043	13.8
$250,000 \leq n2000 < 500,000$	644	349,695	12.5
$100,000 \leq n2000 < 500,000$	1267	156,706	11.0
	2684	673,340	100.0

^a A ratio of total population in the group to total population of cities with $\geq 100,000$.

2.1. The 2000 worldwide size distribution of cities

Table 1 and Fig. 1 show the size distribution of cities over 100,000 in the year 2000 combining data on metro areas for 142 countries. The first observation is that although there is a wide spread in the size distribution, most cities are smaller. 84% of cities over 100,000 people are less than 1 m population and those cities account for 37% of the population of all cities over 100,000. In the medium size category, say 1–3 m, are 12% of cities accounting for 29% of the population of all cities over 100,000. After that we are into bigger cities — the 94 cities over 3 m. Of these, if we draw the line for mega-cities at 12 m, there are 11 such cities accounting for 9.6% of the population of cities over 100,000. If we draw the line at 10 m, there are 19 such cities accounting for 15% of this population.

Wherever the line is drawn, despite popular notions, most of the world's 1.8 b people, who live in significant size cities (over 100,000) represented in Table 1, live outside mega-cities. And if we put the world's total urban population in 2000 at 2.9 b (WDR, 2000), the 1.8 b in our sample of cities over 100,000 account for only 62% of the total urban population of the world. The rest are in even smaller cities. Mega-cities over 12 m only account for 6% of the total urban population of the

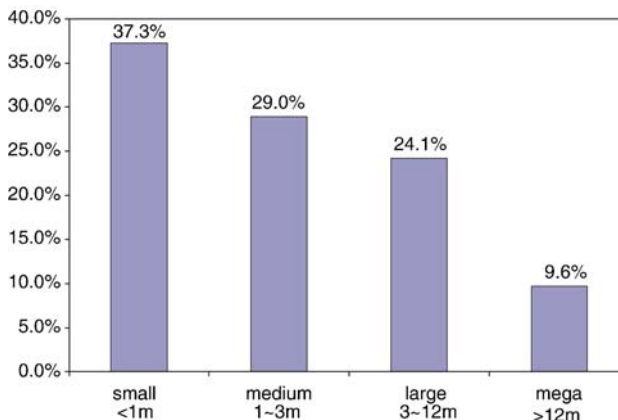


Fig. 1. Share by size category of world population in cities over 100,000 in the year 2000.

Table 2
World city size distribution, 1960

	Count	Mean	Share ^a
$17,000,000 \leq n_{1960}$	0	.	0.0
$12,000,000 \leq n_{1960} < 17,000,000$	1	14,164,000	2.3
$8,000,000 \leq n_{1960} < 12,000,000$	3	9,648,667	4.8
$4,000,000 \leq n_{1960} < 8,000,000$	14	5,763,286	13.4
$3,000,000 \leq n_{1960} < 4,000,000$	7	3,487,286	4.0
$2,000,000 \leq n_{1960} < 3,000,000$	25	2,351,720	9.7
$1,000,000 \leq n_{1960} < 2,000,000$	65	1,374,756	14.8
$500,000 \leq n_{1960} < 1,000,000$	161	705,902	18.8
$250,000 \leq n_{1960} < 500,000$	263	346,121	15.1
$100,000 \leq n_{1960} < 250,000$	681	151,224	17.0
	1220	495,101	100.0

^a The ratio of total population in the group to total population of cities with $\geq 100,000$.

world and less than 3% of its total population. The vast majority of the world's total urban population lives in smaller and medium size cities under 2 to 3 m. While the quality of life in mega-cities is an important issue, the agenda of international agencies could focus more on developing institutions and policies for the smaller metro areas that house most of the world's population.

2.2. Changes in the numbers, sizes, and size distribution of cities worldwide: 1960 versus 2000

The same size categories of cities used for cities in 2000 are given in Table 2 for 1960, which relative to the 2000 sample of 142 countries drops 5 small countries.³ By comparing Tables 1 and 2, two facts emerge, with details for regions of the world given in Table A1. First, the number of cities over 100,000 has increased by about 120% from 1960 to 2000. Second, the sizes of cities have increased. The average size has risen by about 36%. The number of cities over 12 m has increased from 1 to 11. The number of cities from 3 to 12 m has increased from 24 to 83. More critically what we consider “large” or “medium” has changed substantially. In 1960 a city of 1 m would be considered fairly large; today it would be at best a medium size city.

With development, there is increasing absolute spatial concentration of national populations as people urbanize and as cities grow in average size with technological change. Given that, we want to compare the shape of relative size distributions of cities over time to determine if there are changes in the degree of relative spatial concentration across metro areas — the proportion of people living in relatively smaller versus relatively larger cities. In the over time comparisons, there are two issues. First, absolute size distributions are shifting right; and second, our data refer just to the cities in the upper part of size distributions. In comparisons over time we want to try to look at the same portion of the size distribution. To do so, first we normalize city sizes by the average size of cities in the relevant sample in that time period (Eaton and Eckstein, 1997). Second, we alter the relevant sample in each period, raising the minimum size absolute cut-off point to keep the same relative size slice of the pie and the same relative standard to be a city. In models of urban growth, this standard maintains the same relative size cut of the pie if there is

³ 1960 data drop Macedonia, Moldova, Slovenia, Tajikistan, and Mali from the base sample of 142. The remaining actual 1960 data cover 114 countries, but the decline from 137 countries is accounted for by dropping small African and island countries for which we have city data in 1960, but no cities over 100,000.

Table 3
Relative size distribution comparisons: 1960 versus 2000

a) Counts and sizes of cities							
Cutoff ^a	Variable	Count	Mean	Median	SD	Min	Max
Absolute	<i>n</i> 1960	1220	495,101	210,789	969,391	100,000	14,164,000
	<i>n</i> 2000	2684	673,340	264,950	1,469,465	100,009	26,444,000
Relative	<i>n</i> 1960	1220	495,101	210,789	969,391	100,000	14,164,000
	<i>n</i> 2000	1644	1,009,682	472,734	1,798,187	204,366	26,444,000

b) Decomposition: share of new cities since 1960 in 1960–2000 growth of total population in metro areas				
Cut-off	World	Developed countries	Soviet bloc	All other countries
Absolute	.32	.19	.43	.35
Relative	.26	.11	.29	.28

^a Absolute cutoff=100,000; relative cutoff: minimum/mean size=.202, given a 1960 minimum size of 100,000.

parallel (same rate) growth of different types of cities in a country. More generally it is consistent with any model or outcome in which the relative size distribution of cities is constant over time, as will be the case in our data. We take the ratio of the minimum (100,000) to mean (495,101) size for 1960 and apply that ratio (.202) to 2000 (see [Black and Henderson, 2003](#)). That is, we define the cut-off point to qualify as a city in the sample for a particular year as the first *s* cities (ordered by size) such that the *s* + 1 city would be below that relative cut-off size; or we choose *s* such that in time *t*

$$\min \left[s(t); N_{s+1}(t) / \sum_{i=1}^{s+1} N_i(t) / (s + 1) \leq .202 \right] \quad (1)$$

where $N_i(t)$ is the population of city *i* in time *t*. For the year 2000, out of the possible 2684 cities in the world over 100,000, this gives us 1644 cities with an average size of 1,009,682 and a minimum absolute size city of 204,366.⁴

[Table 3a](#) compares 1960 and 2000 urban patterns, based on metro areas in 137 countries. Given the relative cut-off points, our data are for 1220 cities over 100,000 in 1960 and 1644 cities over 204,366 in 2000. Numbers for the absolute cut-off point of 100,000 in both years are also given. In [Table 3a](#), the mean city sizes after relative cut-offs increase by 104% from 1960 to 2000 and the number of cities increase by 35%. Urbanization is accommodated by both increasing numbers and sizes of relatively large cities. In [Table 3b](#), we decompose the world (and region) growth in total population in metropolitan areas, into the share of that increase in total population found in new cities — ones present in 2000, but not in 1960. For the world, about 26% of the increase is accommodated in new cities and the rest in growth of existing 1960 cities. Note that for developed or more fully urbanized countries, the share of new cities is much smaller; while for developing and especially former Soviet bloc countries, the role of creation of new cities in accommodating urban population growth is more important.

We do three main comparisons of city size distributions in 1960 with 2000. First, we compare plots of size distributions, then spatial Gini coefficients, and then coefficients from estimation of

⁴ Note we can't draw the cut-off line at the 100,009/673,340=.1485 where the mean for 2000 is 673,340, since that would require a relative cut-off point in 1960 of around 56,000; and we don't have data for cities under 100,000 in 1960 on a consistent basis.

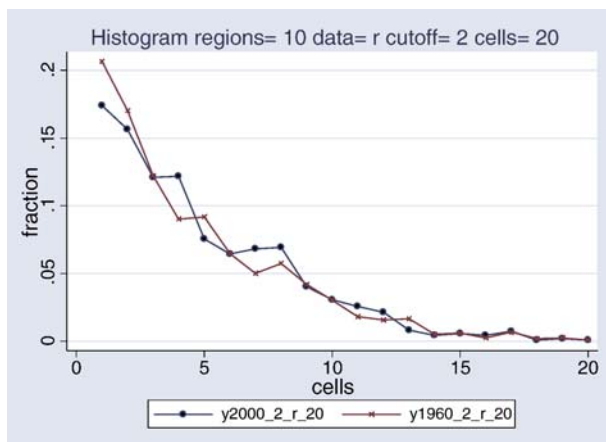


Fig. 2. Relative size distribution for cities in all countries.

Pareto distributions. The basic results emerge from the simple plots, comparing the 1960 and 2000 relative size distributions (city sizes relative to world mean city size). In Figs. 2–4, for 20 cells, we plot the share of number of world cities in each cell. Cells divide the line of length $\ln(100,000/\text{mean city size } 1960)$ to $\ln(\text{max city size } 1960/\text{mean city size } 1960)$ equally, so that there is an equal percent change in city size as we move up the size distribution.

Fig. 2 shows the world comparison. The 1960 and 2000 relative size distributions almost perfectly overlap. In Fig. 3 for developed countries and in Fig. 4 for other countries, the strong overlap also applies. Over the last 40 years there has been almost no change in the relative size distribution of cities: we have the same proportions of relatively small, medium and large size cities. Cities are not converging to some common size. The spread of relative city sizes remains constant over time, suggesting on-going roles for cities of all relative sizes. As noted above, this result is consistent with systems of cities models, where there are different types of cities specializing in different types of products, with each having different equilibrium sizes. A second

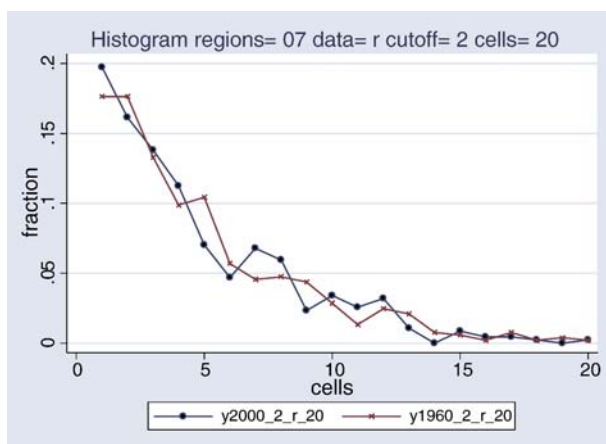


Fig. 3. Relative size distribution of cities in developed countries.

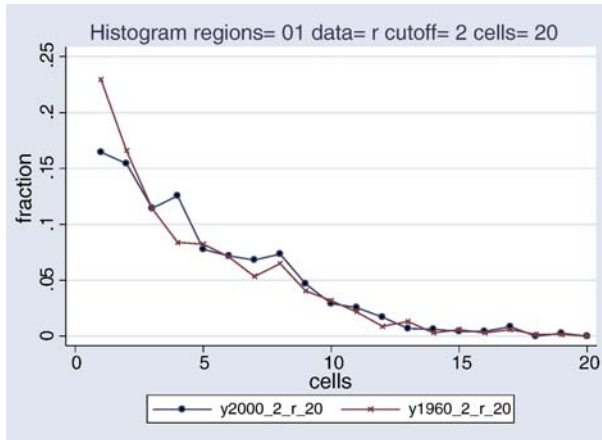


Fig. 4. Relative size distribution of cities in developing and transition countries.

basic point is that urban hierarchies are not becoming increasingly unbalanced, with a greater role for mega-cities. The hierarchy is rock stable in terms of the relative size distribution.

The second way to examine spatial concentration is to calculate spatial Gini coefficients, which give an overall measure of spatial inequality for the entire distribution. To calculate the Gini for a country, we rank all cities from smallest to largest on the x -axis and on the y -axis we calculate their Lorenz curve — the cumulative share of the total sample population. If all cities were of (almost) equal size, the plotted line would be (approximately) the 45° line. The Gini is the share of the area between the 45° line and the plotted curve, relative to the area below the 45° line. The greater the area, the “less equal” is the size distribution, since smaller cities account for a smaller (cumulated) share of the sample population.⁵

In Table 4, we give Gini’s for 1960 versus 2000 for the world, developed countries, (former) Soviet bloc countries, and the rest (developing countries). Table 4 reinforces Figs. 2–4. Gini coefficients for the world, developed countries, and less developed countries, are very similar in 1960 and 2000. In Table 4, if anything, spatial inequality is slightly higher in the developed world, than in the less developed world. Soviet bloc countries, as is commonly perceived, have lower Gini’s reflecting a more even spread of population across cities; and these Gini’s have fallen over time. Policy in former planned economies was to spread the urban population out and limit the growth of the biggest cities, perhaps for reasons of political stability. The table also shows that with rapid world economic growth, relative urban concentration, or spatial inequality has not increased over time.⁶ This is illustrated as well for 7 large countries, where their Gini’s are similar

⁵ If cities are ranked 0 to s where s is the largest size city, the x -axis goes from 0 to s and the y -axis from 0 to $\sum_{i=1}^s N_i(\bar{N}s)^{-1}$ ($=1$) where \bar{N} is average size. The area under the 45° line is $s/2$. The area between the 45° line and the Lorenz curve is $(s+1)/2 - (s\bar{N})^{-1} \sum_{i=1}^s (s-i+1)N_i$ and the Gini is this area divided by $s/2$.

⁶ Table 4 also shows that while the number of cities above the minimum relative size has increased by 95% in the developing world, they have declined by 11% in the developed world. Having relatively stagnant city sizes in the developed world means some cities in 1960 between 100,000 and 200,000 did not grow fast enough to meet the 204,366 cut-off in 2000. Using relative cut-off points is also critical to the calculations. If the 2000 cut-off were 100,000, that would add many relatively small city sizes with tiny shares of world city population, increasing spatial inequality and the Gini. For example for 2000, the world Gini for cities over 100,000 is .626, compared to .562 when we use a relative cut-off point.

Table 4
Spatial inequality in 1960 versus 2000

	1960		2000	
	Count	Gini coefficient	Count	Gini coefficient
World	1220	.5848	1644	.5619
Developed countries	527	.6125	471	.5791
Soviet bloc	193	.5117	198	.4451
All other countries	500	.5659	975	.5605
Brazil	26	.6618	63	.6536
China	108	.4719	222	.4234
India	95	.5561	138	.5821
Indonesia	22	.5239	29	.6140
Japan	95	.6161	77	.6597
USA	167	.5768	195	.5385
Russia	90	.5301	89	.4598

in 1960 versus 2000. Note Russia and China have distinctly lower Gini's than other countries, and Brazil and Japan higher.

A third way used to assess spatial inequality is to impose a Pareto distribution under Zipf's Law and estimate rank size coefficients, by regressing $\ln(\text{rank})$ on $\ln(\text{city size})$ (where largest is rank 1). Lower slope coefficients, or flatter lines (for rank on the y -axis) imply greater inequality: a given change in rank corresponds to a larger change in city sizes. Indeed for 15 countries in 1960 and 2000 for which we calculated Gini's and rank size coefficients, the two are strongly negatively correlated ($R^2 = .90$). We footnote the 1960 and 2000 slope coefficients for these countries and note that the absolute values of the slope coefficients for the world in 1960 and 2000 are respectively 1.086 and 1.107, showing almost no change.⁷

We are next going to examine factors correlated with changes in spatial inequality between 1960 and 2000. For this we are going to examine differences in Gini's rather than in slope coefficients under Zipf's Law for several reasons. First, Gini's don't impose a specific size distribution (Pareto for rank size coefficients). Recent detailed work by Duranton (2004) and Rossi-Hansberg and Wright (2004) suggest significant deviations of actual distributions from Zipf's Law. Finally, in modeling, the emergence of Zipf's Law is based on Gibrat's Law for the upper tail of the city size distribution (such as we are looking at), which says that city growth rates are a random walk and independent of base period size (Gabaix and Ioannides, 2004).⁸ Our data reject Gibrat's Law and a unit root process, showing significant mean reversion in the raw data. While we more fully develop the error structure for city growth equations in Section 4, here we simply report on standard tests for unit roots under the hypotheses that $\beta = 0$ in a model where $\Delta \ln \text{city pop}_i(t) = d_i + \beta \ln \text{city pop}_i(t-1) + \theta_i + \varepsilon_{it}$ for city i (Im et al., 2003; Levin et al., 2002). We strongly reject the hypothesis $\beta = 0$ for this specification, as well as one where data is demeaned at the individual country level. We find a β of $-.037$ ($t = -26.1$) without city fixed effects, a β of $-.225$ ($t = -79.9$) with city fixed effects, and a β of $-.543$ ($t = -223.8$) with city

⁷ The 2000 (and 1960) absolute values of slope coefficients by country are Brazil .86 (.73), China 1.33 (1.10), India 1.07 (1.14), Indonesia .90 (.94), Mexico 1.05 (.96), Nigeria .98 (1.53), France .97 (.93), Germany .73 (.74), Italy .76 (.83), Japan 1.04 (1.14), Spain .96 (1.01), UK .83 (.82), USA 1.12 (1.08), Russia 1.34 (1.18), and Ukraine 1.31 (1.05).

⁸ In fact, from Gabaix (1999), it is necessary to place lower bounds on how small cities can get. Otherwise a log normal distribution arises, rather than a Pareto.

fixed effects and a time trend. These negative β 's imply that, on average, smaller cities grow faster than bigger ones in accommodating urbanization, especially once the overall trend in growth of all cities is accounted for.

2.3. *Why do size distributions differ across countries?*

We have shown that there are differences in the size distribution of cities across countries and that some countries have had changes in their size distributions. We would like to explain these differences and changes. But we have also shown that most size distributions remained remarkably stable over 40 years. Why is there this stability? One can appeal to the theory in [Gabaix \(1999\)](#) or [Rossi-Hansberg and Wright \(2004\)](#) that underlying stochastic processes drive city size distributions towards a limit Pareto distribution such as Zipf's Law. But it seems doubtful that the needed requirements to obtain such a distribution are met in practice, as already noted for the case of Gibrat's Law. Moreover with rapid advances in technology as it affects cities, the advances in globalization, and the enormous political–institutional changes that affect city governance over the last 40 years, it would not be surprising to see shifts in the relative size distribution of cities (and even if in the limit, distributions approach Pareto ones, changes in fundamentals could lead to changes in the Pareto coefficient). We have already previewed our basic result. While there are forces acting to alter size distributions, so far the two main forces of democratization and technological advances oppose each other and have been offsetting in their effects. To do a first exploration of this theme, before turning to detailed econometric models of city growth, we look at the determinants of the size distribution in 2000 and changes between 1960 and 2000.

[Table 5](#) provides a first look at these determinants. First in column (1), we regress Gini's for 2000 on a set of base period covariates, using 1970 variables as the base (rather than 1960 which limits the sample due to missing values on covariates). Column (1) describes the base period environment under which cities developed. In column (2), we examine 2000 Gini's as a function of the same covariates, but using contemporaneous values; then in column (3) we instrument for 2000 covariates with 1970 variables, to remove the simultaneity effects of contemporaneous shocks. Finally in column (4) we look at the change in Gini's from 1970 to 2000. We focus on results in columns (1) and (4). Instrumental variable [IV] estimates of coefficients in column (3) are very similar to those in column (1), but the standard errors are much larger under the IV process.

In [Table 5](#), covariates are national urban population, income per capita defined as purchasing power parity real GDP per capita, income per capita squared and then three political variables. A national land area variable, once national urban population size is controlled for, has a zero coefficient and is dropped. National urban population is a measure of the scope of urbanization, where bigger and more advanced economies may support a wider range of products and types of cities, with a wider city size distribution and hence greater spatial "inequality". Income per capita is the measure of the level of technology, as well as the wealth needed to invest in the infrastructure for urbanization. Later we will use educational attainment measures to capture growth in technology; but for the level of technology we use income per capita. While we have an extensive discussion of political variables later, for now we use the Polity 4 measure of democracy which takes values from -10 to 10 ([Jaggars and Garr, 1995](#)), a dummy variable for whether a country was a planned economy (Soviet bloc), and a dummy for whether a country was defined as federalist by [Gastil \(1978\)](#). Note the drop in sample size from prior tables. We have about 75 countries for which we have complete data on covariates; dropped are those new countries for which we can't infer data historically and countries, particularly African ones, which have poor

Table 5
Determinants of spatial Gini's^a (standard errors in parentheses)

	Gini 2000 OLS-1970 covariates (1)	Gini 2000 OLS-2000 covariates (2)	Gini 2000 2SLS-2000 covariates (3)	Change in Gini 1970–2000 (4)
ln(nat. urban pop.) 2000 [1970]	.0660** (.00884)	.0703** (.0114)	.0573** (.0217)	
ln(GDP p.c.) 2000 [1970]	.595** (.202)	.175 (.182)	.609* (.364)	
ln(GDP p.c.) sq. 2000 [1970]	-.0361** (.0126)	-.0104 (.0107)	-.0295 (.0237)	
Planned economy	-.208** (.0572)	-.129** (.0571)	-.229* (.122)	
Federal structure 1990 [1970]	-.0792** (.0378)	-.106** (.0444)	-.0894 (.0719)	
Democracy index 2000 [1970]	-.00387** (.00193)	.00527 (.00359)	-.0261 (.0416)	
Change ln(nat. urb. Pop) 1970–2000				.155** (.0201)
Change in democracy index 1970–2000				-.00414** (.00156)
Change in ln(GDP p.c.) 1970–2000				-.0318 (.0258)
Constant	-2.89 (.803)	-1.35 (.759)	-3.26 (1.49)	-.0856 (.0268)
<i>N</i>	63	65	59	46
Adj. <i>R</i> ²	.559	.382		.601
Sargan test [<i>p</i> -value]			1.51[.471]	

** Significant at 5% level; * Significant at 10% level.

^a Instruments are 1970 values of ln(urban population), ln(GDP p.c.), ln(GDP p.c.) sq., democracy index, dummy for Gastil federal country, dummy for a planned economy, ln(land area), and the ratio of labor in agriculture to total labor. Note the first and last two variables instrument for 2000 urban population, where errors in how that population and the Gini are measured may be correlated and sources of measurement error may persist over time. For example a liberal measure for what is urban may result in more small cities being counted and a higher Gini.

panel data on economic and institutional characteristics. In addition for Gini comparisons, we drop 25 countries with just one city.

What stands out in Table 5 is that political variables have huge impacts. Being a planned economy lowers the Gini by .21, where the mean and standard deviation of the Gini are respectively .50 and .14. Being a federal country lowers it by .08. The additional impact of the democracy index is more muted, where a one standard deviation increase in democracy lowers the index by .027. The democracy variable is the only one with time variation, to use in cross time comparisons. We graph the partial effect of democracy on Gini levels in Fig. 5a and b. Fig. 5a is the partial effect of democracy controlling for just the urban population size and GDP per capita variables, whereas 5b also controls for the federalism and planned economy dummy variables.⁹

⁹ That is, we are plotting the residuals of the Gini from a regression on covariates against the residuals from a regression of democracy on the same covariates.

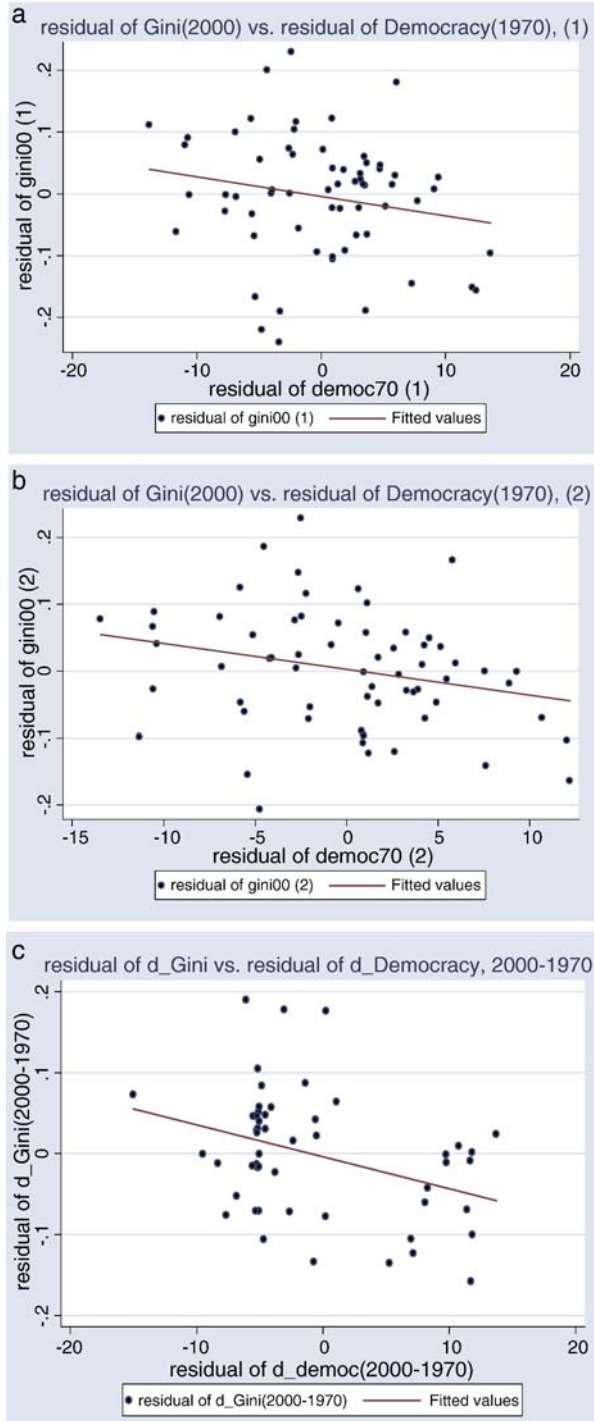


Fig. 5. a. Effect of democracy on the Gini (no control for federalism and planned). b. Effect of democracy on the Gini (full controls). c. Effect of change in democracy on change in Gini.

For the change formulation in column (4), the change in democracy has significant effects as illustrated in Fig. 5c.¹⁰ Moving from the lowest (−10) to highest (10) value for democracy decreases the Gini by .083, about .6 of a standard deviation of the Gini. The implication is that greater regional representation under democratization and greater local autonomy significantly reduce the degree of urban concentration. The paper will investigate this phenomenon in detail, looking at city growth formulations and dealing with econometric identification issues.

Turning to income effects, increases in per capita income at first lead to increases in the Gini, where effects peak at 8.24, or a GDP p.c. of \$3800, before declining. The 95% confidence interval for this peak calculated by the delta method is 2460 to 5919. The idea of a peak mirrors older results in the literature which suggest urban concentration as measured by either primacy (the ratio of the population of the largest city to national urban population) or a Hirschman–Herfindahl index has an inverted-U relationship with per capita income (Wheaton and Shishido, 1981; Junius, 1999; Davis and Henderson, 2003). The idea is that at low levels of development, initial urbanization is spatially concentrated because resources for urban infrastructure and inter-city transport infrastructure are limited; skilled urban workers are in short supply, and knowledge is limited and spatially concentrated perhaps at points of entry to international markets. As the economy develops, it garners the ability to disperse and the economy diversifies (Williamson, 1965). But we note a GDP of \$3800 in 1970 is pretty high and urban concentration increases with income growth for most countries, suggesting that there may be more than just a resource story here. In particular as noted earlier, the income per capita measure is meant to also capture technological progress and that progress may differentially favor the biggest cities, enhancing urban concentration. Given these potentially conflicting forces, summary effects are fairly modest. Moving from the peak at a GDP of \$3800 up one standard deviation in $\ln(\text{GDP p.c.})$ to an income of \$10,600 lowers the Gini by .037. And in the change formulation in column (4), quadratic effects (not shown) are not apparent and a linear term is also insignificant.

For remaining variables, in all columns in Table 5, increases in the total urban population in a country increase the Gini, perhaps an economic diversity effect of greater national urban scope. In column (1), a one standard deviation increase in urban population (1.4) increases the Gini by .09. The change effect in column (4) is even larger and may suggest over the last 30 years of rapid urbanization that rapid growth spurs spatial inequality. We will return to this notion later when looking at determinants of city size growth. Finally, not shown in table, we note that openness in 1970 has no significant effect on spatial Gini's per se, as they emerge in 2000. Later we will argue that openness is important for cities in certain locations but not for the overall size distribution.

2.4. Decade by decade comparisons of the size distribution of cities

So far this section has focused on comparing 2000 with either 1960 or 1970. Decade by decade comparisons allow us to look further at the urbanization process. The characteristics of the sample we use for decade by decade analysis are given in Table 6 (with a relative cut-off point over time for what defines a metro area¹¹). Relative to the base sample of 142 countries, 24 countries are

¹⁰ For concerns about contemporaneous correlation, we also ran the equation with changes in the Gini from 1980–2000 as a function of covariates changes from 1970–1990. Results are very similar.

¹¹ For this sample the relative minimum to average size cut-off is .185. Minimum absolute sizes by construction grow at the same rate as average sizes, so for this sample in 2000, the minimum size is 177,300.

Table 6
Total numbers of cities and sizes

	1960	1970	1980	1990	2000
Number of cities	1020	1183	1414	1551	1552
Mean size	540,424	625,550	678,264	794,930	957,942
Median size	243,497	269,829	300,169	363,984	443,430
Minimum size	100,082	115,195	125,528	147,294	177,300

dropped; in constructing the sample, we require countries to have city data for all decades.¹² In Table 6, the number of cities grows between decades by 16%, 20%, 9.7%, and 0.06% for 1960–70, 1970–1980, 1980–1990, and 1990–2000, while average sizes increase by 16%, 8.4%, 17% and 21%. In early years, the numbers grow at a faster rate than sizes, while in the 1980–2000 time period the reverse is the case, which will be consistent with empirical and theoretical results below. Growth through city numbers is greater when countries are less urbanized, as is the case in 1960, compared to 2000. Overall in this new sample, for the period 1960–2000, the numbers of cities rise by 52% and average sizes by 77%. Again clearly both dimensions to urban growth are critical.

We also looked at the evolution of the size distribution by decade, city transitions through the size distribution, and emergence of any steady-state distributions, based on the sample in Table 6. We did transition analysis, following Dobkins and Ioannides (2001), Eaton and Eckstein (1997), and Black and Henderson (2003), dividing the size distribution into 5 cells in 1960 containing approximately 35%, 30%, 15%, 10% and 10% of all cities starting from the bottom, with fixed relative cell cut-off points.¹³ We calculated entry rates of new cities and show two typical patterns. With entry of new cities in the bottom cells, many existing cities get “pushed up” into higher cells (so transition probabilities of moving up are high relative to moving down). Second, cities in the top cell have extremely low probabilities of moving down. The relatively big, with their enormous long-lived public capital stocks and scale externalities, stay big. Although worldwide transition matrices are not stationary (unlike individual large country ones) and applying transition analysis worldwide, as opposed to by country is suspect, we did calculate the predicted 2000 distribution and the steady-state one.¹⁴ Starting from the actual 1960 distribution by cell of .35, .30, .15, and .10, the 2000 actual distribution is .33, .30, .15, .11, and .11 while the 2000 predicated and steady-state distributions are respectively .31, .32, .15, .11, and .11 and .30, .31, .15, .12, and .12. This rock stability just mirrors Figs. 2–4.

3. Numbers of cities

In this section we first examine the determinants of city numbers, a country aggregate varying across countries and time. In practice identifying growth in numbers of cities is difficult and the data are noisy. We of course are not looking at development of entirely new cities, but calculating the growth in numbers of cities that pass some cut-off point in terms of size. As explained in the previous section we need a relative cut-off point, so that in essence we are capturing the same

¹² In any decade the number of countries represented may be less than 118 (=142–24) because some countries have no cities above the minimum cut-off for that decade.

¹³ Upper cut-off points (relative to mean) are .32, .66, 1.18, 1.95, end open.

¹⁴ For M the transition matrix, i entry rates overall of cities and Z the vector distribution of entrants across cells, under stationarity the steady state distribution is $[I - (1 - i)M']^{-1}iZ$.

right tail of the relative size distribution of cities at each point in time. We define numbers of cities in each decade as in Section 2.4. In essence, for a stable relative size distribution of cities worldwide where the numbers of cities covered in that distribution is growing over time, we try to explain the decade by decade variation across countries in the numbers of cities that qualify to be a metro area, by passing the same relative size standard. While the exercise has issues given the assumptions we impose to define the dependent variable, the results are compelling.

Before turning to institutions, we think of changes in the numbers of cities as being influenced by two factors: growth in the urban population which should lead to increases in numbers of cities and growth in technology which, for the same urban population, should lead to decreases in numbers of cities. Why the latter? As technology improves, as we argued in the introduction, equilibrium and efficient sizes increase. That occurs because knowledge accumulation may enhance urban scale economies on the production side, and will lower urban diseconomies by improving urban commuting and management methods. In analyzing the data, we need to specify how to measure technological progress. As in [Grossman and Helpman \(1991\)](#) and [Benhabib and Spiegel \(1994\)](#), we make technological progress a function of knowledge accumulation in the country, with the idea that greater human capital levels spur technology development and adoption, leading to larger size cities and more rural–urban migration. Our measure of knowledge accumulation is the percent of adults completing high school from [Barro and Lee \(1996\)](#).

In the raw data, a simple OLS regression of the percent change in number of cities by decades for 1970–2000 on the change in urban population and on education levels yields¹⁵

$$d\ln(\text{no. of cities}) = \underset{(.120)}{.402} d\ln(\text{total urban population}) - \underset{(.00140)}{.00309} (\% \text{ adults with high school}) + \text{time dummies.}$$

As expected, the number of cities rises with urban population growth and declines with knowledge accumulation. Time dummies are always insignificant in this equation and the ones to follow. The problem with this specification is that urban population itself is endogenous, also being driven by technological change. [Henderson and Wang \(2005\)](#) have a closed economy model of growth in city numbers under exogenous national population growth and endogenous technological change driven by endogenous human capital accumulation. Technology has two opposing effects — to reduce city numbers as efficient sizes of cities increase with improved technology and to increase city numbers as technology improvements move people from the agricultural to the urban sector. Rural to urban migration occurs in a context where technology improves with economic growth in both the urban and rural sectors, but the demand for agricultural products is income inelastic so the urban employment sector expands at the expense of rural employment through rural–urban migration. In net the overall effect of the two forces depends on the relative size of the rural sector. The bigger the rural sector, with a technological advance the more people absolutely are released from agriculture and the more likely this force is to outweigh the effect of city size increases, so the numbers of cities rises. A direct specification from [Henderson and Wang \(2005\)](#) is

$$\Delta \ln m_{jt} \equiv \ln m_{jt} - \ln m_{jt-1} = g_{jt} - b_0 h_{jt-1} + b_1 \left[\left(\frac{n_a}{n_u} \right)_{jt-1} h_{jt-1} \right] + \varepsilon_{jt} \tag{2}$$

¹⁵ If we look at the level number of cities in, say, 2000 that is driven entirely by total urban population. A regression of $\ln(\text{number of cities})$ on covariates in [Table 5](#) or variations thereof has $\ln(\text{total urban population})$ as the only significant variable (coefficient [standard error] of .891 [.0367]) with an R^2 of .93.

In Eq. (2), m_{jt} is the number of cities in country j at time t ; g_{jt} is the national rate of population growth which in the Henderson and Wang (2005) model has a coefficient of 1; h_{jt-1} is the base period level of knowledge accumulation that drives technological improvements; and n_a/n_u is the national ratio of rural to urban population which we measure by the ratio of national population not in our city sample to the total population in our cities. Details on data and variables are in Appendix B.

To this formulation, we experiment with adding measures reflecting the effects of democratization and openness. Greater effective federalism is associated with increases in local autonomy, reflecting both increased control of local governments over local public goods and the ability of hinterland cities to undertake growth enhancing investments in infrastructure and to compete on a level playing field with the national capital and other large cities. A problem is that we don't have measures of effective federalism for a large sample of countries. There is the Gastil index of federalism which is based on formal institutions and doesn't change over time. Arzaghi and Henderson (2005) construct an index of effective federalism for every 5 years from 1960 to 1995, where the index has six components which look at the ability of local and state governments to impose taxes, not be overruled by the center, and make independent expenditure decisions.¹⁶ Unfortunately that index only covers 37 countries in our estimating sample. What we do have for a large sample of countries is the commonly used measure of the extent of democracy, the Polity 4 index of democracy available through the University of Maryland website, which following Jaggars and Garr (1995) is the index of democracy minus the index of autocracy. It has values from -10 to $+10$, for which we have measures from 1960 on. Inherent in democracy is regional representation and increased state and local autonomy. The Polity 4 index and our index of federalism for 37 countries for which they overlap has a simple correlation coefficient of .71 in 1970 levels and of .53 in change form for 1970–1995. In Arzaghi and Henderson (2005) we also point out the high correlation between the federalism index and objective measures of fiscal decentralization such as the share of state and local governments in total government consumption. We will use the democracy index, since we are looking for the effects of changes in institutions. Openness is measured by the percent share of imports in GDP. We use the World Development Indicators data from the World Bank.

A key issue in estimation of Eq. (2) is how to identify causal effects beyond correlations. A common way to proceed is to assume $\varepsilon_{jt} = u_j + d_t + \tilde{\varepsilon}_{jt}$, where u_j is a fixed effect representing time invariant country level unmeasured cultural–political, geographic, and institutional variables affecting city formation and growth in city numbers. These items could also affect covariates. d_t is a “world” technology level, and $\tilde{\varepsilon}_{jt}$ represents a contemporaneous shock affecting growth in city numbers. However in our data, tests on serial correlation reject serial correlation of the composite error term ε_{jt} . Correspondingly, fixed or random effect specifications are rejected in favor of OLS by standard Breusch–Pagan and F -tests. For this equation, it appears that the simple differencing of numbers of cities removes all persistence.

The remaining concerns in estimation are endogeneity and measurement error. While we might think of national population growth and base human capital levels as exogenous to shocks to the numbers of cities, shocks to $\ln m_t - \ln m_{t-1}$ could be correlated with $(n_a/n_u)_{t-1}$ and that term may be measured with error. The formulation that best satisfied specification tests is one where $(n_a/n_u)_{t-1}$ is affected by shocks to $\ln m_t - \ln m_{t-1}$, while the rate of change of population growth

¹⁶ This index is available at <http://www.econ.brown.edu/faculty/henderson/papers.html>.

Table 7
Growth rate in city numbers^a (standard errors in parentheses)

	(1) IV	(2) OLS
Growth rate of national population (<i>t</i>)	1.02** (.134)	.944** (.125)
% adults with secondary education (<i>t</i> −1)	−.00326** (.00113)	−.00265** (.00116)
% adults with secondary education * (rural/urban pop.) (<i>t</i> −1)	.000647** (.000285)	.000102 (.000166)
Time dummies	yes	yes
<i>N</i> [countries]	215 [74]	215 [74]
<i>R</i> ²		.253
Sargan test stat. [<i>p</i> -value]	9.99 [.763]	

** Significant at 5% level. * Significant at 10% level. OLS errors are calculated allowing for country clustering.

^a The mean and standard deviation of the national population growth rate, % adults with secondary education, and rural/urban population are .16 (.12), 20.3 (15.4), and 5.8 (9.2). In addition to exogenous variables instruments are rural/urban population for *t*−2 and *t*−3, as well as income per capita and rural/urban population interacted with the education variable for *t*−2. Dropping the income variable has minimal effect on any coefficients.

and technological change are treated as exogenous. We will report instruments and specification tests below; but given urbanization is a dynamic accumulation process, we rely on lagged values of $(n_a/n_u)_{t-1}$, as strong and valid instruments.

3.1. Results

Results are in Table 7. We start with the simple version of Eq. (2), without political variables and openness measures. Columns 1 and 2 give IV and OLS results, and the presentation focuses on the IV results.¹⁷ Results are second step GMM estimates allowing for within period heteroskedasticity (Arellano and Bond, 1991) with standard errors corrected for small sample bias (Windmeijer, 2000). Serial correlation of errors is rejected decisively (*p*-value of .81) and the Sargan test statistic is excellent (*p*-value of .763). We comment on the comparison of IV and OLS results below. For the results in column (1), we first examine the functional specification. The coefficient on national population growth is 1.02, essentially identical to the hypothesized coefficient of 1.0. Inclusion of a lagged dependent variable ($\ln m_{t-1}$) results in an insignificant coefficient of .014 if added to the column (1) specification. These two results are consistent with a steady-state growth model specification in Henderson and Wang (2005), where growth rates in city numbers have a stationary process. In the specification, because missing value problems eliminate most small African and island countries, most of the sample involves larger countries where city number growth rate measures for a country are based on many cities. Results where we exclude nine outlying observations of small countries with high growth rates in a period are very similar to those in Table 7 (and we note that for a country to appear in the estimating sample in a given decade, it must have cities in the base period for that decade). Finally for the full sample, plots

¹⁷ Other than exogenous variables, instruments are secondary education * n_a/n_u (*t*−2), n_a/n_u (*t*−2), n_a/n_u (*t*−3), and $\ln GDP_{pc}$ (*t*−2).

of standardized residuals look very good, with the distribution conforming tightly to the expected normal in a QQ plot, except at the extreme upper tail.

Turning to the technological change variables, the negative coefficient on education represents the force of increasing city sizes reducing the need for more cities, while the positive coefficient on education interacted with n_a/n_u represents the force of urban technology improvements on increased rural–urban migration pushing for more cities. Both forces are at work. Note the coefficient of the covariate containing n_a/n_u has a sixfold change in moving from OLS to IV estimation, with other coefficients unaffected confirming the anticipated problems with this variable. To assess the effect of technological change on growth in city numbers we look at the net result, which depends on the magnitude of n_a/n_u . As long as the relative rural population is not too big, so this ratio in column (1) is below 5.0 (.00326/.000647) which is roughly the mean in 1990, technological advances increasing existing city sizes reduce the need for more cities. At a low n_a/n_u of, say, 0.5 where most of the population is urbanized, a one standard deviation increase (15) in education reduces the growth rate of new cities by .044, where the mean growth rate is .13. But at a higher n_a/n_u of, say, 12 where the rural–urban ratio is high and therefore the base for potential rural–urban migration is very large, a one standard deviation increase in education increases the decade growth rate of numbers of cities by .068 as more rural workers urbanize. More new cities are needed to accommodate rural–urban migration.

3.1.1. Institutional influences

What is the effect of institutions on the growth rate of city numbers? We experimented with institutional measures such as (1) whether a country is a (former) planned economy (Iron or Bamboo curtain), where migration restrictions limit city sizes, (2) British common law system (La Porta et al., 1998), (3) Kaufman et al.'s (1999) extent of rule of law and the IRIS measure of contract repudiation by governments as proxies for well-defined property rights (Knack and Keefer, 1995), and (4) a measure of federalism for 48 larger countries in Davis and Henderson (2003), as well as the Gastil (1978) measure of federalism. In estimating our models, it became apparent that none of these variables in level form have a consistent effect on growth in city numbers (or sizes). Nor is it clear what the effect should be. While institutions may affect city sizes in levels, if we compare two different regimes, one where all cities are big versus one where they are small, these two different regimes can operate with the same growth rate in city numbers. That is, if n is city size and m is city numbers where $n_u = mn$, so that $\dot{n}_u/n_u = \dot{m}/m + \dot{n}/n$, then if \dot{n}/n and \dot{n}_u/n_u are the same in the two regimes, so will be \dot{m}/m . However based on the discussion in the introduction, a change in institutions should have at least a short-term effect on growth in city numbers. Removing constraints on the degree of local autonomy helps the development of non-primate and non-capital cities, by increasing their ability to form, to invest to grow from towns into metro areas and, more generally, to compete with primate and capital cities. Thus a change in such institutions, as measured by the extent of democracy, should lead to a jump in the number of metro areas.

Table 8 reports results. First we treat such changes in democracy as exogenous, which does pass specification tests; and then we try a specification where we instrument for these changes with 1960 levels of democracy and the Gastil federalism dummy in 1960. Results are in Table 8, columns 1 and 2. In column 1, an increase in the democracy index leads to an increase in numbers of cities. The effect is strong. A one standard deviation (6.8) increase in the democracy index increases the growth rate in numbers of cities by .048, from a mean of .13. When we instrument for changes in democracy the effect doubles, as shown in column 2. One might have strong priors

Table 8
Political and policy variables: growth in city numbers^a (standard errors in parentheses)

	(1)	(2)	(3)	(4)
	IV	IV	OLS	IV
Growth rate of national population (<i>t</i>)	1.10** (.180)	1.08** (.230)	.859** (.197)	1.24** (.208)
% Adults w/ secondary education (<i>t</i> −1)	−.00219* (.00117)	−.00170 (.00147)	−.00241* (.00127)	−.00183 (.00117)
% Adults w/ secondary education * (rural/urban pop.) (<i>t</i> −1)	.000636** (.000239)	.000568** (.000263)	.000130 (.000145)	.000637** (.000230)
Change in democracy (<i>t</i>)	.00701** (.00272)	.0140** (.00711)	.00313 (.00301)	.00717** (.00309)
Openness (<i>t</i> −1)				−.00247** (.000836)
Time dummies	yes	yes	yes	yes
<i>N</i> [countries]	198 [73]	198 [73]	198 [73]	190 [73]
Sargan test stat. [<i>p</i> -value]	10.4 [.844]	21.8 [.295]		13.5 [.762]
<i>R</i> ²			.224	

** Significant at 5% level. * Significant at 10% level.

^a Openness is treated as exogenous. In columns 1 and 4 democracy is treated as exogenous. In column 2, it is instrumented with the 1960 values for democracy and the Gastil dummy for being a federal country.

that unobservables affecting the number of cities, such as investments spurring greater development of hinterland regions, might also affect democratization with increased calls for regional representation, as well as being correlated with the rural to urban allocation of population.

We also look at openness, as a policy related variable, inducing technology transfers and growth. We treat it as exogenous, which could be a problem if trade policy affecting openness is affected by unobservables driving growth in city numbers. However results where we instrument for openness(*t*−1) with its lagged value are virtually the same as those reported. Column (4) shows the results with the openness variable included. All other coefficients are almost the same as in column (1). Openness reduces significantly the growth in numbers of cities, perhaps because it leads to improved technology for cities per se and increased city sizes. A one standard deviation increase in openness (14) decreases the growth rate in number of cities by .04. We do note that the addition of institution and policy variables weakens the education coefficient in the models. Our sample size and ability to instrument are limited, so it is difficult to tease out precise estimates.¹⁸

4. Growth in city sizes

Now we come to our main results, shifting from an analysis of country aggregates to city level data describing growth rates of individual cities. The specification we use is a differenced and linearized version of [Au and Henderson \(2006\)](#) under an assumption of perfect population

¹⁸ Having said that, we also looked at the effect of infrastructure investment in roads, as measured by kilometers of roads normalized by national land area. Estimation cuts the sample size to 144 observations for 63 countries. We estimate by GMM, with lagged road density values as instruments for current log changes. While the road coefficient is positive and significant, other coefficients weaken with the loss in sample size.

mobility.¹⁹ But the estimating equation is intuitive. As noted earlier, in urban models, real income per worker is an inverted-U shape function of city size (population). Then city size is determined by the intersection of that inverted-U with the supply of population to the city, horizontal or not. The usual stability condition requires the equilibrium be to the right of the peak of the inverted-U. In terms of the rising portion of the inverted-U, scale economies in [Au and Henderson \(2006\)](#) include not just the usual external economies of scale and local diversification, but also the new economic geography's home market and market potential effects, under differentiated products and monopolistic competition. Cities have specific locations. Those in denser markets have higher market potential. Increases in external market potential spur own city development through increased demand and prices for local products. But then there is a feedback where one city's growth spurs nearby city development which feedbacks as increased market potential to the own city. In addition to these forces, as noted above, city sizes are influenced by institutions affecting new city formation, the ability of cities to finance public services and infrastructure investment, and the ability of cities to plan and potentially limit in-migration (through zoning and land use regulation). Institutions affect how many towns have been able to grow into metro areas and weaker institutions generally lead to too few cities, all of which are oversized as discussed in the introduction.

In the econometric specification, we use growth rates, rather than absolute city sizes. Increases in market potential will increase own city size, by driving up received prices and inducing profitable expansion and in-migration. Technological improvements driven by the base period level of knowledge as measured by education will shift inverted-U's up and out, leading to increases in city size. And as noted for city numbers, changes in institutions will affect city sizes through their affect in enhancing or retarding the numbers of competitor cities.

Before adding in institutional considerations, our base equation is

$$\Delta \ln n_{it} \equiv \ln n_{it} - \ln n_{it-1} = a_0 \Delta \ln(\text{MP}_{it}) + a_1 h_{it-1} + \beta_t + v_{it}, \quad (3)$$

where n_{it} is the population size of city i at time t . Individual city growth is affected by city specific conditions such as a city's own internal market potential, as well as country conditions such as the national level of technology (absent city specific education attainment measures). For market potential, in the trade literature, specifications have iceberg transport costs for all urban exports, in a differentiated products and monopolistic competition framework. Any variety of export produced in city i is sold and shipped to all other cities in the country and potentially abroad. From trade theory (see [Head and Mayer, 2004](#); [Overman et al., 2003](#), for a review), we can then obtain a measure of market potential, MP, for each city. Specifically the price facing any firm in city i is $p_i = \text{MP}_i^{1-\rho} (x_i)^{-(1-\rho)}$, where x_i is the output of a firm in city i , MP_i is city i 's market potential, and $\text{MP}_i = \sum_k E_k I_k / (1 + \tau_{ik})^{\rho/(1-\rho)}$. The sum is over all locations, E_k is disposable income at k , τ_{ik} is the shipping costs from i to k , and I_k is a price index.²⁰ Note we are looking for the moment at just domestic, or within country j market potential. We don't have the

¹⁹ Specifically we take Eq. (8) in Au and Henderson for real income per worker in a city, equate it to the national real income in national labor markets (under perfect mobility), takes logs, difference, and then linearize.

²⁰ For example $I_k = \sum_l (p_l f_l \tau_{kl})^{-\rho/(1-\rho)}$. f_l is the number of firms in city l and there is only one type of differentiated product.

income or price index data to implement this measure of market potential. Instead we reply on a gravity index

$$MP_i(t) = \sum_{\substack{k \in j \\ i \neq k}}^{m_j} \frac{n_k(t)}{d_{ik}} \tag{4}$$

which is the measure for city *i* in country *j*. The measure is the distance discounted sum of populations (n_k) of all other cities in the country (m_j) at time *t*, excluding *i*. d_{ik} is the distance from city *i* to *k* in 100's of miles. In estimation we use $\ln(MP_i(t))$. While this population measure is a crude representation of the extent of market demand for a city's products, compared to the proper definition of market potential, it is viewed as generally performing well in trade econometrics (Head and Mayer, 2004). We also tried extending or adding as a separate component market potential extending across borders, but results are insignificant. However we do use border information.

4.1. Error structure

The main problem in estimation of Eq. (3) is error structure. There are unobservables related to individual cities in a country that affect city growth, such as climate, not to mention local culture and business climate. These may operate at a regional level and be correlated with market potential growth. There are also country conditions that relate to unobserved aspects of institutions that affect both urban growth and human capital levels. And certainly covariates are measured with error. From a practical perspective, in any city growth estimation, there is the issue of enormous time persistence in the error structure. Any IV estimation of (3) for any set of plausible instruments fails basic Sargan specification tests, which are a joint test of orthogonality of instruments and of model specification. We believe specification tests fail due to failure to model this persistence. In simple OLS or IV estimation of an equation like Eq. (3), a common procedure is to control for all these effects by introducing as a control, the lagged value of city size. So the equation is then asking, given a base value, what covariates influence growth from that base. We examined estimates of the basic equations of the model using that standard procedure.²¹ But the specification faces two problems. First it doesn't directly deal with modeling persistence in the error structure. Second, again IV estimation usually fails specification tests.

Thus we focused on finding a plausible error structure to (3). We started with a simple AR (1) error structure, where in Eq. (3) $v_{it} = \rho v_{it-1} + e_{it}$. By substituting for v_{it-1} from the equation for $\Delta \ln n_{it-1}$ in

²¹ IV and OLS results are quite similar, itself raising a suspicion that lagged city size isn't a proper control for all unobservables. IV results are $\Delta \ln n_{it} = -.0519 \ln n_{it-1} + .811 \Delta \ln MP_{it} - .0106 h_{jt-1} + .000815 h_{jt-1} \ln n_{it-1}$, where all coefficients are significant at the 5% level. The base period value of city population indicates modest mean reversion, as usual. The education variables show a very basic result of the paper, that human capital effects are heterogeneous across cities by size. We expect increased education levels improve local technology, especially in commuting, and hence accelerate city growth. However under all specifications education entered on its own is usually insignificant and sometimes negative. The result here suggests effects are heterogeneous — improved education increases the growth rate of bigger cities more than for smaller cities. This specific estimation passes the Sargan specification test (*p*-value = .38), but only for one set of instruments (*t*-2 values for $\ln(\text{city pop.})$, percent high school, $\ln(\text{market potential})$, percent labor force in agriculture and all pair-wise combinations except percent high school*percent labor force in agriculture, as well as $\ln(\text{distance from the city to the nearest sea coast})$). We tried about 10 other instrument lists, adding on other instruments from *t*-2 (like national land area or democracy measures) or using just 1960 variables (which should pass if *t*-2 ones pass). No specification which added on democracy or national scale covariates ever passed. We believe these all fail because of an incorrect model specification. As part of this, error terms still show high persistence; the results cannot reject 1st or 2nd order serial correlation.

the equation for $\Delta \ln n_{it}$ (with v_{it} written as $\rho v_{it-1} + e_{it}$), we get the dynamic representation $\Delta \ln n_{it} = \rho \Delta \ln n_{it-1} + a_1 \hat{h}_{jt-1} + a_2 \Delta \ln MP_{it} - \rho a_1 \hat{h}_{jt-2} - \rho a_2 \Delta \ln MP_{it-1} + \beta_i^* + e_{it}$. This model performs quite well in the sense of passing Wald tests for the basic structure (implied constraints on coefficients across covariates) in unconstrained linear IV estimation, passing specification tests for most plausible sets of instruments, and yielding good final estimates that incorporate the cross term constraints on coefficients. However the estimates in this model for ρ are close to 1. In the most preferred specification for this model, point estimates of ρ are .993 and .988 in respectively constrained and unconstrained IV estimation, and in other specifications an estimate of 1 is never rejected. Given this, a less complicated error specification seemed in order where in Eq. (3) we write $v_{it} = \tilde{v}_i + e_{it}$, where \tilde{v}_i is a simple city fixed effect for a growth equation (we of course allow for time fixed effects in e_{it}).

Thus we simply difference Eq. (3), estimating a double differenced version of a population levels equation. Differencing (3) leaves a remaining error term, $e_{it} - e_{it-1}$. Given we are concerned with contemporaneous city or country level shocks that affect growth and covariates as well, we need to instrument with variables from either $t-2$ or $t-3$. Use of the former assumes covariates like h_{it-2} are predetermined relative to e_{it-1} , the shocks affecting $\ln n_{it-1} - \ln n_{it-2}$. Specification tests reject this. As such we utilized two sets of instruments. First note in estimation, with differencing there are two equation-years for data going from 1970 to 2000 with the last being $(\ln n_{i1990} - \ln n_{i1980}) - (\ln n_{i1980} - \ln n_{i1970})$. For both equation years, as one set of instruments, we used instruments from 1960 for both equations while for the other we used ones from 1970 for the first equation and ones from 1960 for the second. This latter is probably more efficient and performed better — better specification test results and better first stage regression results — and we rely on it.

The remaining question is what instruments to use. The general approach here follows [Arellano and Bond \(1991\)](#) and [Blundell and Bond \(1998a,b\)](#), where lagged level values of variables are used as instruments for covariates in change form. But why should lagged levels should be strong instruments for change variables? There are two rationales. First, the covariates involved represent accumulation processes (as in education or market potential) and one can specify underlying dynamic processes where changes in covariates have as arguments past levels. An example is dynamic human capital accumulation, where investment flows follow a growth path where current investment is a function of past levels. A second argument is mean reversion: there are positive temporary shocks to the local or national economy which give high level covariates, which are then associated with subsequent negative changes. These arguments leave the practical question of whether lagged levels are strong instruments. Our instruments are lagged values of city population, market potential, education, and national percent of the labor force in agriculture which is potentially correlated with emerging measures of (urban) market potential and education. We use these variables and the last three interacted with the first (instruments are listed at the bottom of relevant tables). When we introduce political considerations and considerations of national scope, the list will expand as detailed later. But for the basic model in [Table 9](#) in column (1) for example, for the instruments listed, first stage regressions for the three covariates have average F 's and R^2 's of respectively 72.8 and .19 with minimums of 58.9 and .15, which is strong for differenced (and double differenced) covariates.

4.2. Basic results

Our results are in [Table 9](#). The base results without institutions and policy variables are in column (1), which estimates Eq. (3) in a differenced form (but the table specifies variables as in the model in Eq. (3)). This readily passes specification tests for the listed instruments, as well as other instrument lists such as all 1960 variables (the $t-4/t-3$ combination), or adding in other

Table 9
City size growth: main models^a

	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Change in $\ln(\text{market potential})_{ij}(t)$.896** (.199)	.936** (.200)	.786** (.120)	.670** (.188)
Percent high school _{<i>j</i>} (<i>t</i> - 1)	-.0646** (.0123)	-.0566** (.0139)	-.0560** (.0139)	-.0458** (.0205)
Percent high school _{<i>j</i>} (<i>t</i> - 1) * $\ln n_{ij}(t - 1)$.00526** (.000988)	.00474** (.00115)	.00460** (.00112)	.00376** (.00164)
Change in democracy _{<i>j</i>}		.183** (.0627)	.152** (.0481)	.136** (.0505)
Change in democracy _{<i>j</i>} * $\ln n_{ij}(t - 1)$		-.0148** (.00521)	-.0126** (.00391)	-.0111** (.00414)
Port * change in openness _{<i>j</i>}			.00970** (.00408)	
$\ln(\text{dist. to coast})$ * change in openness _{<i>j</i>}			-.00132** (.000666)	
$\ln(\text{nat. urban pop.}_j(t - 1))$				-.924 (.633)
$\ln(\text{nat. urban pop.}_j(t - 1)) * \ln n_{ij}(t - 1)$.0694 (.0481)
Time dummies	Yes	Yes	Yes	Yes
Sargan [<i>p</i> -value]	2.46 [.651]	5.67 [.340]	18.9 [.062]	8.84 [.183]
<i>N</i> (cities)	2452 (1339)	2275 (1279)	1730 [1044]	1991 (1104)

**Significant at 5% level; *significant at 10% level.

^a OLS errors are clustered by country. In column 1, instruments are *t* - 3 values for $\ln(\text{city pop.})$, percent high school, $\ln(\text{market potential})$, percent labor force in agriculture and each of the last three variables interacted with $\ln(\text{city pop.})$. In column 2 and beyond we add $\ln(\text{national land area})$, democracy, and $\ln(\text{national land area}) * \ln(\text{city pop.})$. Additionally in column 4 we add $\ln(\text{distance from the city to the nearest sea coast})$, $\ln(\text{market potential}) * \text{percent high school}$, and $\ln(\text{market potential}) * \text{percent labor force in agriculture}$. Additional to that, in column 3, we add port, whether a country is landlocked, openness, openness squared and openness * democracy. Having port interacted with openness as an instrument, while not substantially affecting coefficients, does lead the Sargan test to fail.

instruments used in columns (2) and (3) of the Table. In column (1), a 1% increase in a city's market potential increases city growth by .9% over a decade, supporting the new economic geography emphasis on local markets. While, for point estimates the education effect starts off negative, by a metro area size of 215,000, it is positive. At a size of about 1 million, a one standard deviation (15) increase in education increases city size by 12% over a decade, and at 2.5 million the effect is 19%, a large magnitude. The education variable entered on its own has an insignificant coefficient, given its differing effects for cities of different sizes. Increases in human capital benefit bigger cities substantially more than small ones. In smaller, uncongested cities, people can walk or bike easily and intra-city expressways may not be needed; improvements in commuting technology have less of an effect. In addition, bigger cities with their higher scale economies and wider product mix may benefit more from knowledge externalities which enhance scale effects. In column (4), putting aside the democracy term for the moment, we explore the positive correlation between spatial Gini's and national urban scale found in Table 5. The results in column (4) are suggestive, as coefficients are only significant at about a 15% level. They suggest that increases in national urban population spread the size distribution, increasing the growth rates of bigger cities. This could be some type of national economies of scope, where as

economy size grows there is greater diversity in the types of cities and the potential to support the largest types. But we note the result is weak in all specifications.

4.2.1. *Institutions and globalization*

Column (2) of Table 9 contains a main result of the paper, incorporating institutions. As noted earlier, the anticipated effect on city growth rates of institution level measures is unclear. But a move to greater democratization and implied regional representation helps even the playing field across cities, reducing the advantage in securing local public goods of more favored and, typically, bigger cities which are the home to political elites. It should also spur the development of more cities (which hence retards growth of existing cities), given the greater ability of hinterlands to establish major cities which are competitors to existing cities, without consent from the center. In column (2) as hypothesized, the effect of improved democracy becomes negative at existing city size of 235,000 and becomes increasingly negative as we move up the hierarchy. At a city size of 1 million, a one standard deviation increase (6.8) in the democracy index²² reduces city growth by 15% in decade; at 2.5 million the effect is 24%, which obviously is huge.

Recent political–economic development has had two opposing indirect effects on the city size distribution, which seem to cancel each other out as reflected in the relatively time invariant spatial Gini's in Table 4. As human capital accumulates, that spreads the size distribution out, as bigger cities benefit more from the technology improvements in commuting and traffic management. On the other hand to the extent democratization is associated with increased development, that reconcentrates the size distribution, increasing the ability of new hinterland cities to form and compete for migrants and industry with existing larger cities.

We also examined whether capital cities grow faster than other cities in our time interval. We have two basic results. First, treating being a national capital as exogenous causes specification tests to fail. Second, whether being a capital is treated as exogenous or endogenous, the coefficient is never significant. Capital cities while being larger don't grow faster than other cities in the time interval. What dominates the results for 1970–2000 is the changes in institutions, or the force of democratization.

In column (3) openness favors ports. A one standard deviation (14) increase in openness leads to a 14% increase in the growth of a port city. The effects of increased openness decline with distance to the coast, but remain positive for virtually all distances to the coast in the data (given the high effect at zero distance on the coast). The results are very sensitive to specification and choice of instruments and even this “best” specification barely passes the Sargan text. While the pattern of results generally holds and is suggestive, under different instrument choices, some coefficients are statistically weak and most specifications fail Sargan tests.

5. Conclusions

In this paper we show that there is no growing imbalance to urban hierarchies and much of urbanization is accommodated in smaller and medium size cities. We show that city formation and the growth in numbers of cities can be explained in large part by a simple model and a few covariates — national population growth, inferred technological change, and changes in institutions. For the last, increasing democratization facilitates the formation of new cities.

²² The standard deviation for changes in the democracy index is also high, at 5.1.

Similarly growth of individual city sizes is explained by changes in local market conditions, technological change, and changes in national institutions and policies. Technology, policy and institutional effects on individual city growth are heterogeneous. Technology improvements help bigger cities more, while increasing democratization levels the playing field across cities, allowing smaller cities to compete more freely for firms and residents.

Appendix A

Table A.1

Size distributions by parts of the world (1960 versus 2000 world cities)

World	1960			2000		
	Count	Mean	Share ^a	Count	Mean	Share ^a
12,000,000 ≤ x	1	14,164,000	2.34	11	15,848,182	9.65
4,000,000 ≤ x < 12,000,000	17	6,448,941	18.15	42	7,041,262	16.36
2,000,000 ≤ x < 4,000,000	32	2,600,125	13.77	116	2,780,428	17.85
1,000,000 ≤ x < 2,000,000	65	1,374,756	14.79	249	1,370,773	18.89
500,000 ≤ x < 1,000,000	161	705,902	18.82	355	703,043	13.81
250,000 ≤ x < 500,000	263	346,121	15.07	644	349,695	12.46
100,000 ≤ x < 250,000	681	151,224	17.05	1267	156,706	10.99
Total	1220	495,101	100.00	2684	673,340	100.00

^a A ratio of total population in the group to total population of cities with ≥ 100,000.

Latin America and the Caribbean

Latin America and the Caribbean	1960			2000		
	Count	Mean	Share	Count	Mean	Share
12,000,000 ≤ x	0	.	0.00	3	16,163,667	18.45
4,000,000 ≤ x < 12,000,000	4	5,453,500	37.89	5	6,807,800	12.95
2,000,000 ≤ x < 4,000,000	1	2,032,000	3.53	14	3,091,214	16.47
1,000,000 ≤ x < 2,000,000	7	1,292,571	15.72	29	1,416,187	15.63
500,000 ≤ x < 1,000,000	10	669,800	11.64	42	721,139	11.52
250,000 ≤ x < 500,000	21	356,140	12.99	85	332,659	10.76
100,000 ≤ x < 250,000	71	147,811	18.23	244	153,142	14.22
Total	114	504,961	100.00	422	622,765	100.00

Sub-Saharan Africa

Sub-Saharan Africa	1960			2000		
	Count	Mean	Share	Count	Mean	Share
12,000,000 ≤ x	0	.	0.00	1	13,427,000	9.74
4,000,000 ≤ x < 12,000,000	0	.	0.00	1	5,064,000	3.67
2,000,000 ≤ x < 4,000,000	0	.	0.00	11	2,659,104	21.22
1,000,000 ≤ x < 2,000,000	1	1,147,000	8.45	21	1,449,632	22.09
500,000 ≤ x < 1,000,000	5	661,000	24.35	25	725,510	13.16
250,000 ≤ x < 500,000	10	337,390	24.85	44	350,925	11.20
100,000 ≤ x < 250,000	38	151,303	42.35	163	159,861	18.91
Total	54	251,397	100.00	266	518,117	100.00

(continued on next page)

Appendix A (continued)

North Africa and Middle East

	1960			2000		
	Count	Mean	Share	Count	Mean	Share
$12,000,000 \leq x$	0	.	0.00	0	.	0.00
$4,000,000 \leq x < 12,000,000$	0	.	0.00	4	6,671,750	24.91
$2,000,000 \leq x < 4,000,000$	1	3,712,000	20.20	7	2,747,710	17.95
$1,000,000 \leq x < 2,000,000$	3	1,465,333	23.93	13	1,377,569	16.72
$500,000 \leq x < 1,000,000$	4	658,500	14.34	23	675,811	14.51
$250,000 \leq x < 500,000$	5	321,462	8.75	43	368,558	14.79
$100,000 \leq x < 250,000$	39	154,421	32.78	75	158,702	11.11
Total	52	353,302	100.00	165	649,235	100.00

South Asia

	1960			2000		
	Count	Mean	Share	Count	Mean	Share
$12,000,000 \leq x$	0	.	0.00	3	14,433,667	19.06
$4,000,000 \leq x < 12,000,000$	2	4,780,000	21.38	7	7,534,286	23.22
$2,000,000 \leq x < 4,000,000$	1	2,283,000	5.10	11	2,510,000	12.15
$1,000,000 \leq x < 2,000,000$	6	1,402,500	18.82	29	1,308,207	16.70
$500,000 \leq x < 1,000,000$	5	738,600	8.26	39	693,668	11.91
$250,000 \leq x < 500,000$	27	360,845	21.79	78	341,140	11.71
$100,000 \leq x < 250,000$	72	153,166	24.66	75	158,757	5.24
Total	113	395,768	100.00	242	938,668	100.00

East Asia and Pacific

	1960			2000		
	Count	Mean	Share	Count	Mean	Share
$12,000,000 \leq x$	0	.	0.00	1	12,887,000	3.14
$4,000,000 \leq x < 12,000,000$	2	7,554,000	13.65	11	7,561,091	20.24
$2,000,000 \leq x < 4,000,000$	9	2,488,556	20.24	28	2,726,300	18.57
$1,000,000 \leq x < 2,000,000$	10	1,291,900	11.68	78	1,390,749	26.39
$500,000 \leq x < 1,000,000$	52	713,229	33.52	101	699,101	17.18
$250,000 \leq x < 500,000$	42	359,528	13.65	99	376,286	9.06
$100,000 \leq x < 250,000$	52	154,645	7.27	142	156,682	5.41
Total	167	662,596	100.00	460	893,444	100.00

Soviet Bloc

	1960			2000		
	Count	Mean	Share	Count	Mean	Share
$12,000,000 \leq x$	0	.	0.00	0	.	0.00
$4,000,000 \leq x < 12,000,000$	1	6,170,000	9.17	2	7,227,000	9.88
$2,000,000 \leq x < 4,000,000$	2	2,905,000	8.63	5	2,525,600	8.63
$1,000,000 \leq x < 2,000,000$	6	1,316,500	11.74	24	1,292,875	21.21
$500,000 \leq x < 1,000,000$	24	715,148	25.50	33	656,763	14.82
$250,000 \leq x < 500,000$	34	355,049	17.94	97	343,055	22.75
$100,000 \leq x < 250,000$	126	144,316	27.02	216	153,740	22.70
Total	193	348,694	100.00	377	387,980	100.00

Appendix A (continued)

	1960			2000		
	Count	Mean	Share	Count	Mean	Share
12,000,000 ≤ x	1	14,164,000	4.85	3	18,741,333	10.92
4,000,000 ≤ x < 12,000,000	8	7,122,500	19.52	12	6,631,417	15.45
2,000,000 ≤ x < 4,000,000	18	2,609,444	16.09	40	2,854,854	22.17
1,000,000 ≤ x < 2,000,000	32	1,422,974	15.60	55	1,353,763	14.46
500,000 ≤ x < 1,000,000	61	706,045	14.76	92	720,385	12.87
250,000 ≤ x < 500,000	124	335,929	14.27	198	345,966	13.30
100,000 ≤ x < 250,000	283	153,583	14.89	352	158,682	10.84
Total	527	553,771	100.00	752	684,954	100.00

Appendix B

Data and sources.

Education information is from Barro and Lee (1996). GDP per capita data are downloaded from the Penn World Tables website. Openness, total national and total urban population, and land area are from the World Bank, World Development Indicators. Details on the world city data set, how different samples are drawn, and the data themselves are at <http://www.econ.brown.edu/faculty/henderson/worldcities.html>. For some city-years, documentation is listed as incomplete, generally indicating that sometimes conflicting multiple data sources were examined, where the choice of data source was based on consistency over time and across cities.

Means and standard deviations ^a		
	Mean	Standard Deviation
Growth in city numbers	.13	.30
Rural/urban population	5.8	9.2
Growth in national population	.16	.12
ln(city size)	13	1.1
Δln(city size)	.23	.26
Change in Δln(city size)	-.052	.26
ln(distance to coast)	3.9	2.2
Dummy port	.29	
ln(national urban population)	11	1.4
Δln(nat. urban population)	.21	.16
Means and standard deviations ^a		
% adults with secondary education	31	15
Δ% adults with secondary education	3.6	7.4
Openness	21	14
Δopenness	3.9	6.8
Change in Δopenness	1.3	9.4
Democracy	5.0	6.8
Δdemocracy	1.4	5.1
Change in Δdemocracy	.44	8.8
ln(market potential)	16	1.3
Δln(market potential)	.23	.17
Change in Δln(market potential)	-.052	.14

^a Except for first three items, the numbers are calculated for the city size growth rate sample.

References

- Abdel-Rahman, H., Anas, A., 2004. Theories of systems of cities. In: Henderson, J.V., Thisse, J.-F. (Eds.), *Handbook of Regional and Urban Economics*, vol. 4. North Holland.
- Ades, A.F., Glaeser, E.L., 1995. Trade and circuses: explaining urban grants. *Quarterly Journal of Economics* 10, 195–227.
- Arellano, M., Bond, S., 1991. Some tests of specification for panel data. *Review of Economic Studies* 58, 277–297.
- Arzaghi, M., Henderson, J.V., 2005. Why countries are fiscally decentralizing. *Journal of Public Economics* 89, 1157–1189.
- Au, C.-C., Henderson, J.V., 2006. Are Chinese cities too small? *Review of Economic Studies* 73, 549–576.
- Barro, R.J., Lee, J.W., 1996. International measures of schooling years and schooling quality. *American Economic Review* 86, 218–223.
- Benhabib, J., Spiegel, M., 1994. The role of human capital in economic development. *Journal of Monetary Economics* 34, 143–173.
- Black, D., Henderson, J.V., 1999. A theory of urban growth. *Journal of Political Economy* 107, 252–284.
- Black, D., Henderson, J.V., 2003. Urban evolution in the USA. *Journal of Economic Geography* 3, 343–373.
- Blundell, R., Bond, S., 1998a. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.
- Blundell, R., Bond, S., 1998b. GMM estimation with persistent panel data: an application to production functions. IFS Working Paper, vol. W99/4.
- Davis, J., Henderson, J.V., 2003. Evidence on the political economy of the urbanization process. *Journal of Urban Economics* 53, 98–125.
- Dobkins, L., Ioannides, Y., 2001. Dynamic evolution of the US city size distribution. In: Thisse, J.-F., et al. (Eds.), *The Economics of Cities*. Cambridge University Press.
- Duranton, G., 2004. Urban Evolutions: The Still, the Slow, the Fast. <http://158.143.49.27/~duranton/research.html#CitySize>.
- Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of city products. *American Economic Review* 91, 1454–1477.
- Eaton, J., Eckstein, Z., 1997. Cities and growth: theory and evidence from France and Japan. *Regional Science and Urban Economics* 27, 443–474.
- Fujita, M.P., Krugman, A., Venables, A., 1999. *The Spatial Economy*. MIT Press.
- Gabaix, X., 1999. Zipf's law for cities: an explanation. *Quarterly Journal of Economics* 739–767.
- Gabaix, X., Ioannides, Y., 2004. The evolution of city size distributions. In: Henderson, J.V., Thisse, J.-F. (Eds.), *Handbook of Urban and Regional Economics*, vol. 4. North Holland.
- Gastil, R., 1978. *Freedom in the World*.
- Glaeser, E., Scheinkman, J., Shleifer, A., 1995. Economic growth in a cross-section of cities. *Journal of Monetary Economics* 36, 117–143.
- Grossman, G., Helpman, E., 1991. *Innovation and Growth in the Global Economy*. MIT Press.
- Head, K., Mayer, T., 2004. The empirics and agglomeration and trade. In: Henderson, J.V., Thisse, J.-F. (Eds.), *Handbook and Regional Economics*, vol. 4. North Holland.
- Henderson, J.V., 1988. *Urban Development: Theory, Fact, and Illusion*. Oxford University Press.
- Henderson, J.V., 2003. Urbanization and economic growth: the so-what question. *Journal of Economic Growth* 8, 47–71.
- Henderson, J.V., Becker, R., 2000. Political economy of city sizes and formation. *Journal of Urban Economics* 48, 453–484.
- Henderson, J.V., Kuncoro, A., 1996. Industrial centralization in Indonesia. *World Bank Economic Review* 10, 513–540.
- Henderson, J.V., Wang, H.G., 2005. Aspects of the rural–urban transformation of countries. *Journal of Economic Geography* 5, 23–42.
- Im, K., Pesaran, M., Shin, Y., 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics* 115, 53–74.
- Jaggars, K., Garr, T.R., 1995. Tracking democracy's third wave in the Polity III data. *Journal of Peace Research* 32, 469–482.
- Jefferson, G., Singhe, I., 1999. *Enterprise Reform in China: Ownership Transition and Performance*. Oxford University Press, New York.
- Junius, K., 1999. Primacy and economic development: bell shaped or parallel growth of cities? *Journal of Economic Development* 24 (1), 1–22.

- Kaufman, D., Kraay, A., Zoido-Lobaton, P., 1999. Aggregating governance indicators. World Bank Policy Research Working Paper, vol. 2195.
- Knack, S., Keefer, P., 1995. Institutions and economic performance. *Economics and Politics* 7, 207–227.
- Krugman, P., Elizondo, R. Livas, 1996. Trade policy and the Third World metropolis. *Journal of Development Economics* 49, 137–150.
- La Porta, R., Lopez-de-Silanes, F., Shleifer, A., Vishny, R., 1998. The quality of government. NBER Working Paper, vol. 6727.
- Levin, A., Lin, C., Chu, C., 2002. Unit root tests in panel data: asymptotic and finite sample properties. *Journal of Econometrics* 108, 1–24.
- Overman, H., Redding, S., Venables, A.J., 2003. The economic geography of trade, production and income: a survey of empirics. In: Harrigan, J., Choi, K. (Eds.), *Handbook of International Trade*. Blackwell.
- Richardson, H., 1987. The costs of urbanization: a four country comparison. *Economic Development and Cultural Change* 33, 561–580.
- Rossi-Hansberg, E., Wright, M., 2004. Urban Structure and Growth. <http://www.stanford.edu/~erossi/USG.pdf>.
- UN, 1993. *World Urbanization Prospects: The 1992 Revision*. United Nations, NY.
- Wheaton, W., Shishido, H., 1981. Urban concentration, agglomeration economies, and the level of economic development. *Economic Development and Cultural Change* 30, 17–30.
- Williamson, J., 1965. Regional inequality and the process of national development. *Economic Development and Cultural Change* 3–45 (June).
- Windmeijer, F., 2000. A finite sample correlation for the variance of linear two-step GMM estimators. IFS Working Paper, vol. W0019. London.
- World Bank, 2000. *Entering the 21st Century World Development Report 1999/2000*. Oxford University Press.