Junius, Karsten

Working Paper — Digitized Version

The determinants of urban concentration

Kiel Working Paper, No. 835

Provided in Cooperation with:
Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Kieler Arbeitspapiere
Kiel Working Papers

The Determinants of Urban Concentration

Kiel Working Paper No. 835

by

Karsten Junius

Institut für Weltwirtschaft an der Universität Kiel
The Kiel Institute of World Economics

ISSN 0342 - 0787

Kiel Institute of World Economics

Department IV

24100 Kiel

The Determinants of Urban Concentration

Kiel Working Paper No. 835

by

Karsten Junius

September 1997

The Determinants of Urban Concentration

Abstract

Urban concentration differs across countries. One determinant of these differences is economic development, which first increases and subsequently decreases urban concentration. I condition the degree of urban concentration on the potential of countries to develop a balanced urban system. These conditions are approximated by the land area, population density and density of the transportation system, which all decrease urban concentration. It is also found that countries with a long independent urban history have lower degrees of urban concentration than countries with a recent colonial past. Furthermore, I assess the impact of some historic variables and historic patterns of urban concentration on current patterns of concentration.

JEL classification: R11, R12

Keywords: Industrialization, Urban Concentration, Economic Development

Karsten Junius

Kiel Institute of World Economics

P.O. Box 4309

D-24100 Kiel

Phone: +49 431 8814497

Fax: +49 431 8814500 or +49 431 85853

e-mail: Junius@ifw.uni-kiel.de

1. Introduction*

For long urban economists have tried to explain why the distribution of the urban population differs so much between countries. The share of the urban population that lives in the largest city of a country – the so called primacy ratio – is, for instance, 38 percent in Argentina, 33 percent in Korea, but only 6 percent in India and 8 percent in the Netherlands. This paper tries to explain these different levels of urban concentration across countries.[1]

Development economists have tried to link primacy and economic development. One hypothesis is that economic development first increases and then decreases regional concentration – thus exhibiting an inverted U-shape

---

* I would like to thank Erich Gundlach, Rolf Langhammer, Volker Nitsch, and conference participants at the ERSA in Rome and the EEA in Toulouse for helpful comments on a previous version.

[1] Note that I am not explaining absolute population concentration in the urban giants like Manila, Sao Paulo or Tokyo. One recent paper that focuses on this issue is Ades and Glaeser (1995). They find that high tariffs, high costs of internal trade and low levels of international trade favor absolute population concentration in giant cities. Concentration is fortified by dictatorships and political instability. While external economies of scale can well explain the absolute concentration of the population and industrial production in large scale agglomerations (Junius, 1997), it is less clear why the relative concentration within a country is so different between countries. Note again that some other authors focus on the concentration of the total population, while I analyze the concentration of the urban population. This is because concentration of the total population in the first place is determined by the level of urbanization instead of that of economic development.

relationship between the two variables (Alonso, 1980; Williamson, 1965). Considerable attempts have been made to show such a pattern in time series and cross country analyses. While time series analyses often showed the existence of an inverted U-curve, cross country evidence is mixed. This paper makes a new attempt to show the existence of an inverted U-curve in a cross country setting, using different country samples and a number of conditional variables that influence the likelihood that a country develops a dispersed or concentrated system of urban settlement. The conditional variables are the population density, the size of the country, the transportation infrastructure and political variables like openness, political and economic freedom and a colonial past.

Alternatively, the distribution of the urban population may simply reflect historic developments. The fact that Vienna by so much dominates the urban system of Austria is more likely to be traced back to its century long role as a capital of a huge empire than to current GDP levels. Similarly, France is more concentrated than Germany because it has no federal political history but has been centrally governed from Paris for the longest part of the recent history. Additionally, endowment differences or natural advantages like those of port cities may have led to unbalanced urban concentration. A persistent influence of ancient population distributions would point to the importance of path dependencies, but also to ongoing natural advantages of certain locations in a country. Thus, I test the impact of ancient urban population distributions in addition and against the above mentioned economic and political variables.

I use primacy ratios as a measure for urban concentration, because its computation requires only the size of the one to four largest cities and the total urban population which are available for a large number of countries. An inverted U-curve relationship between primacy and GDP exists if primacy depends positively on GDP and negatively on the squared GDP value, as implied by the following form: $PR = aGDP + bGDP^2 + cX$, where PR is the primacy ratio, GDP is the output measure, X is a vector of further explanatory variables, and a, b and c are the parameters to be estimated.

Section 2 surveys the existing empirical literature. Section 3 discusses possible explanations for the existence of an inverted U-curve and other determinants of urban concentration. Section 4 discusses the empirical procedure and the results. Section 5 summarizes.

2. Empirical Evidence for the Inverted U-Curve

Time series studies find ample evidence for the inverted U-curve in single countries. El-Shakhs (1972) shows the existence of an inverted U-curve for the UK, and Alperovich (1992) shows the existence for the curve for Israel. Parr (1985) estimates Pareto coefficients for 12 countries and several years between

1850 and 1981.[2] He finds strong evidence for the inverted U-curve in the high income countries Austria, France, Sweden and the US. Concentration peaks around 1910 in Austria, 1930 in the US and Sweden, and 1954 in France. The evidence for the curve is weaker for Brazil, Japan, Spain and the USSR, while in the low income countries Egypt, India, Nigeria and Turkey no clear reversal of the trend towards increasing concentration can be observed yet. Eaton and Eckstein (1997) calculate Lorenz curves for city size distributions for France

---

2 The estimation of Pareto coefficients its widely used in this literature. It goes back to Singer (1936) who postulated that similar to Pareto's law of income distribution the size distribution of cities within a country can be described by a "Pareto-function": $y = Ax^{-\alpha}$, where x is a particular population size, y indicates the rank of a city, which equals the number of cities with a population of more than x, and A and α are parameters to be estimated. The higher α, the smaller is the concentration of the urban population in the largest cities. Thus, the existence of an inverted U-curve is shown by a first decreasing and then increasing value of α in the course of economic development.

For a "Pareto coefficient" $\alpha = 1$ one gets the so called rank size rule, which states that the rank times the population of a city equals a constant, which is the size of the largest city. Most authors test a log-linear form of the relationship, which implies a constant Pareto coefficient. Hsing (1990) argues that the degrees of concentration may vary in different stages of development and growth such that a log linear form is inappropriate. He uses a general functional form, which includes the log-linear form as a special case. The flexible functional form turns out to be a much better method for the estimation of Pareto coefficients. The log-linear form is rejected at the 5 percent level. Furthermore, he finds that Pareto coefficients decline with urban size, such that small cities have more population than indicated by the rank-size rule. Alperovich and Deutsch (1995) develop this approach further and also refute the Pareto distribution.

since 1876 and Japan since 1925. In contrast to Parrs' evidence, these authors argue that relative city sizes in France and Japan did not vary that much to exclude the possibility of parallel growth in the course of economic development.[3] [4]

---

3 Related evidence for an inverted U-curve of industrial specialization and GDP is provided by Kim (1995). He analyzed the regional distribution of economic activities in the US between 1860 and 1987. He finds diverging regional specialization until the turn of the century and converging patterns of industrial production since the 1930s.

4 See Carroll (1982) or Kasarda and Crenshaw (1991) for a survey of further early time series and cross section studies on the distribution of city sizes. For an overview of determinants of urban concentration see Mutlu (1989, p. 612) and Sheppard (1982, p. 139). Sheppard also provides a thorough discussion of rank size estimates. Using deviations from the rank size rule as a primacy measure and Coles' economic development index, he does not find evidence for an inverted U-curve, but also concludes that this might be so because his measure for primacy is of little empirical value (p. 140). Mutlu (1989) presents a number of variables used in previous studies, but finds only weak evidence for an impact of economic variables on concentration measures. GNP per capita reduces concentration in his sample of up to 90 countries. The size of the absolute urban population and absolute total population, as well as the size of the arable land reduce concentration while greater income inequality increases concentration measures. Mera (1973) uses a sample of 46 developing countries and finds a positive relationship between concentration and development. He argues that urban concentration is a precondition of economic development in early stages. He shows that growth of primacy explains overall GDP growth rates over a 7 years period. This relationship is stronger if the analysis is restricted to larger countries, indicating that in small countries other than economic variables may determine the distribution of urban population. However, as primacy measures, he uses the share of the largest and the three largest cities in the total population instead of in the urban population as primacy measures.

Cross-country studies find mixed evidence for the inverted U-curve. Williamson (1965), Kamerschen (1969), El-Shaks (1972), Whiton and Shishido (1981) and DeCola (1984) find evidence for the inverted U-curve between urban concentration and economic development. Kamerschen (1969) uses data for 80 countries and the share of the largest city in percent of the four largest cities as a measure of concentration. He finds a negative relationship of urban concentration on the one hand and GDP and industrialization on the other hand for developing countries and a positive relationship for developed countries. Wheaton and Shishido (1981) use the inverse H index for measuring urban concentration and test the inverted U-curve relationship against a linear relationship.[5] Using a sample of 38 countries with at least three metropolises, they find that the inverted U-curve much better explains the data.

Rosen and Resnick (1980), Mutlu (1989), Lemelin and Polèse (1995), and Moomaw and Shatter (1996) find a negative relationship between spatial concentration and economic development. Rosen and Resnick (1980) use a

---

These increase with urbanization and do not solely indicate the concentration of the urban population.

5 The inverse H index: $\frac{1}{H} = \sum_{i=1}^{n}\left(\frac{P_i}{P}\right)^2$, where P_i is population in city i, P is total population and n is the number of categories. This index has been criticized for that it is correlated with the number of cities in the sample.

sample of 44 developing and developed countries. They test for the significance of several variables on calculated Pareto coefficients. Higher GNP per capita and total population lead to a more even distributed population. The density of the rail-network, the export to GNP ratio, the percentage of non-agricultural labor force as well as a dummy for former colonies are not found to be significant in the overall sample. In a sample of 61 countries, Lemelin and Polèse find that primacy is indirectly linked to level of economic development. Primacy falls monotonically with the degree of development. They mention, albeit do not report, that their results are robust to alternative measures of primacy and development and alternative country subsets. Moomaw and Shatter (1996) use panel data for 90 countries and find that primacy ratios increase if the largest city is additionally the capital of a country. They also find that GDP per capita, literacy, population and export orientation reduce primacy.

Richardson and Schwartz (1988) find no support for the economic development and primacy link at all. For the 116 countries under study they show that demographic factors are more important and render economic variables insignificant. They show that 40 percent of the variation of primacy can be explained by national population, urban population and a Latin America dummy. Their results are criticized by Lemelin and Polèse (1995) partly for not checking for multicollinearity. Urbanization is strongly correlated with economic development and GDP levels. Hence, insignificant coefficients for GDP levels in the presence of urbanization or industrialization data should not come as a surprise. Another problem might be the large sample size, which propably includes a large number of very small countries, where no economic rationale predicts systematic domestic economic forces to be able to unfold and to play a dominant role for the city size distribution.

Arbitrary sample selection is often the reason why the previous empirical results are very sensitive to the group of countries included in the analysis. Therefore, the economic rationale for the selection of countries and explanatory variables is thoroughly discussed in Section 3 and Section 4.2.

3. Explaining the Inverted U-curve and Further Determinants of Primacy

The review of the empirical literature of cross-country studies shows that the evidence is mixed as to whether an inverted U-curve exists or not. Thus, the current status and the appropriate attitude on the inverted U-curve is somewhat similar to what Williamson (1997, p. xxii) writes about the Kuznets Curve: "now you see it, now you don't. The important inference of that fact, however, is *not* to reject the Kuznets Curve, but to ask why we sometimes see it and sometimes not." Most important for the understanding of changes in the concentration and dispersion of economic activities over time is to understand the interplay between centrifugal and centripetal forces that drives this process. In general, positive economies of scale foster agglomerations and industrial clusters, and negative spillovers and higher factor costs foster population dispersion. From that, some authors have tried to theoretically derive economic explanations of an inverted U-curve.

Wheaton and Shishido (1981) argue that economic development increases capital intensity in industrial production. As capital intensity increases fixed costs compared to variable costs, scale intensity increases. This favors larger cities. Therefore, efficient city size increases with economic development until some sort of capital saturation sets in as scale economies are not exploitable without bound. This again levels the population concentration in later stages of economic development.

Parr and Jones (1983) suggest a five stages approach of economic development and primacy. The pre-urban stage is characterized by a low quality transportation system, which limits the extent of regional markets, intraregional trade and the exploitation of scale economies. Improvements in the transportation system allow for more intraregional trade, which pushes some cites above the critical mass of production. This leads to rapid growth of certain specialized cities. In a third stage stronger interindustry linkages further allow exploitation of scale economies in larger cities. In the fourth stage improvements of interurban transportation systems and a high income elasticity of land demand lead to suburbanization and levels economic conditions between cities. This development is extended in the last stage, where several regional markets achieve a sufficient size for the production of a large number of goods. Together with negative externalities from concentration in the core this leads to a leveling of the population distribution within a country. Thus, falling transport costs have ambiguous effects. They strengthen centripetal forces at high income levels and centrifugal forces at low levels.

Krugman (1991) formalizes the impact of transport costs on the pattern of production in economic geography models. In his basic model, high transportation costs between a center and a periphery may lead to the development of two distinct industrial or urban centers. For intermediate transport costs, the possible realization of stronger backward and forward linkages make the center a more attractive place to locate, while for low transport costs linkage effects can be realized from all possible locations, but negative externalities from urban concentration increase the attractiveness of the periphery.

The economic geography framework has been used by Junius (1996) to derive the U-curve between concentration and economic development. In this model, pecuniary externalities provide forward and backward linkages that strengthen the core of a country in the course of economic development. Core regions benefit from the sectoral change to industrial production that usually accompanies economic development (Syrquin 1989). As a consequence, a higher share of the work force starts working in the footloose industrial sector under increasing returns to scale. The demand from these workers reinforces centripetal forces that make it even more profitable for firms to cluster. Increasing concentration, however, also leads to negative externalities due to crowding and pollution. These congestion effects decrease the advantages of being in the core. This levels economic conditions between core and peripheral regions. Thus, peripheral regions gain after some levels of concentration is reached in the core. Then, the production of goods that can realize only small EOS or those that are very costly to transport have few reasons to cluster and are the first to set up business in peripheral cites.[6] The resulting U-curve between urban concentration and economic development is intensified by a fall of transport costs.

Economic factors can only play an important role if they are able to unfold themselves without being dominated by other factors. Neighboring countries

---

6 Hanson (1996) further describes the process of agglomeration and subsequent dispersion of non-external economies sectors to the periphery, when agglomeration drives up wages, land rents and congestion costs in the center.

and geographic or demographic factors may especially determine the population distribution of small countries like Luxembourg or Monaco, where two independent metropolises additionally would not be distinguishable statistically. Also internal transport costs do not matter much in small countries such that any location would have an advantage for just in time production. Small countries are unlikely to form a functional system of cities and pattern of industry distribution independent from their neighbors and geographic factors.

In countries with a low absolute (urban) population the distribution of the population is likely to be determined by endowment factors. Neither positive external economies of scale nor negative external effects of population concentration are strong enough to determine the population distribution between different parts of a country. This means that congestion effects are unlikely to push people out of Iceland's Reykjavik with 145,000 inhabitants. Iceland's population distribution is more likely to reflect resource endowments and climatic conditions than scale economies. That is, in order to test for the influence of economic determinants on primacy only these countries can be used that are not too small in terms of the area or the number of inhabitants.

Concluding from the above models one would expect a positive coefficient of GDP per capita and a negative coefficient of the squared GDP per capita value on a measure of urban concentration, provided that countries are equal in all other respects besides GDP per capita. However, countries differ in several respects that influence the likelihood of developing a balanced urban system other than the pure economic forces described by the models. The map in the Appendix indicates primacy ratios. Countries in light colors are less concentrated than countries in dark colors. The map shows especially strong concentration in several African and Latin American countries and low concentration in large countries. Thus, there might be other determinants besides GDP per capita that influence the degree of concentration in a country.[7] To correct for the different potential to develop several urban agglomerations, the degree of concentration should be made conditional on the following variables.

A large land area (*LAND*) increases the probability of the formation of several metropolises, and, thus, leads to lower levels of concentration in a country. First, it increases the number of possible sites for potential cities that emerge because of the availability of certain resources or natural advantages like ports or transportation nodes. Second, it implies large distances between different parts of a country. Such differences may have favored the development

---

7 The map indicates the share of the largest city in the total urban population. Data are for 1995 or the most recent estimate for 180 countries. Of the visible countries only for Eritrea, Greenland and Former Spanish Sahara no data was available.

of different urban systems. This is, because historically, the extent of the market for perishable goods has been much more limited than it is today. Additionally, certain services and administrative functions had to be located close to the generally dispersed rural population, such that large countries developed a larger number of medium size cities.

Densely populated countries (*DENSE*) are likely to be less concentrated for two reasons. First, along with population concentration come negative externalities like congestion effects and pollution. They constitute an upper bound to city size after which no or only few scale economies can be realized, and, thus, provide strong incentives to disperse. Second, high densities provide the possibilities to disperse since a large number of workers will be available in several parts of the country such that scale economies can be exploited not in the primate city only. They constitute the critical mass of workers and local demand for the production of certain goods that is needed for the formation of a city. Thus, through the existence of a lower and an upper bound to city size, densely populated countries are likely to exhibit several optimal sized cities. This reduces urban concentration.

The extent of the transportation system as, for instance, indicated by road kilometers per land area (*ROADLAND*), has an ambiguous effect on the degree of concentration. As theoretical economic geography work has shown, on the one hand lower transport costs increase the degree of competition that the few firms in the peripheral region face from the large number of firms in the core region. This reduces the attractiveness of peripheral sites and increases the advantages of the primate city as a location of production and consumption. On the other hand, lower transport costs mean that distance is getting less important, such that the need to cluster in order to realize scale economies is less severe. Then industries are able to escape higher land prices and congestion effects by shifting production away from the core and still benefit from agglomeration and scale economies. The coefficient of the variable measuring the extent of the transportation system may, therefore, be interpreted as indicating which effect prevails.

In addition to these variables indicating the possibilities of a country to disperse, the degree of concentration can be influenced by politics. An open trade regime (*OPEN*) favors industry and population dispersion, because being close to the center of the home market becomes less important. Instead being close to the market of the neighbouring countries, transportation nodes and harbors becomes decisive, which disperses industries and population. Openness also means better opportunities for non-industrial activities like agriculture. Agriculture is often discriminated in less open regimes, which prevents the development of dispersed food processing and other rural industries.

Undemocratic institutions, the deprivation of civil or political rights, property rights and domestic unrest (*POLITICS*) are likely to favor urban concentration. In countries with these properties, there often is a strong central and primate bias of government spending. This reflects that in such political systems, spatial politics are often used to assure maximum political control over a country, population and administration. Together with a tendency to nepotism in government spending, this increases the size of the primate city.

Finally, a colonial history is likely to favor strong primate cities. For administrative reasons, in this case the capital city had strong links with the colonial power and, consequently, less strong links with the rest of the national urban system. Production and trade was often more oriented towards the demand of the colonial power than towards the demand of the domestic population. Thus, it often took place in export eclaves and delinked the capital from the rest of the country and prevented the development of a dispersed domestic urban system. That also prevented the establishment of a transport system with neighbouring countries. In addition, innovations and other positive growth shocks diffused more slowly to secondary centers as they did in countries without a colonial history. The result is a persistent dominance of the capital city with few and much smaller rival cities. This means different historic or initial conditions, for which the colonial dummies are used.

4. Empirical Results

4.1 Estimation Equation

In order to identify an inverted U-curve, I estimate variants of the general equation

(4.1) $$PR = aGDP + bGDP^2 + cX + \mu,$$

where *PR* is the primacy ratio, *GDP* is the output measure, *X* is a vector of further explanatory variables, *a*, *b* and *c* are the parameters to be estimated and μ is the error term. Since primacy measures are expressed in percentage points and, hence, can only vary between 0 and 1, they are limited dependent variables. Thus, to avoid estimation and interpretation problems, I assume that *PR* is distributed according to the logistic function . That is:

(4.2) $$PR = \frac{e^z}{1+e^z},$$

where $z = aGDP + bGDP^2 + cX$. By the use of (4.2), primacy can be transformed into a variable that is not limited to the 0 to 1 range anymore:

(4.3) $$\frac{PR}{1-PR} = e^z$$

Taking logarithms on both sides, one arrives at an estimable equation:

(4.4) $$L = \ln(PR/(1-PR)) = z + \mu \quad \text{or}$$

(4.5) $$L = aGDP + bGDP^2 + cX + \mu\ .$$

4.2. The Data

The sample of countries is selected as follows. I start with all 209 countries listed in the World Bank CD-ROM (1996) and use their 1990 values unless stated otherwise. I first exclude all small countries. That is, I exclude countries with a land area below 30,000 square kilometers, which is slightly below the size of Belgium and slightly above the size of Haiti. I also exclude countries with an urban population below 1,000,000 and a total population below 3,000,000. The reason is that small countries are more likely to represent special or genuine developments rather than systematic economic regularities.

From the remaining countries, 16 countries are excluded because no PPP adjusted GDP-levels are available for either 1989, 1990 or 1991. The Dominican Republic, Myanmar, Saudi-Arabia, Nicaragua and Sudan are excluded because data is missing for other variables. Observations for Togo, Somalia, Sierra Leone, Senegal, Mozambique, Madagascar, Honduras, Guinea, Chad and Angola are excluded because the last census from which urban population data for 1990 is estimated or interpolated is older than 15 years. Further excluded are the observations for Costa Rica and Zimbabwe as they show large differences according to different sources, and Malawi for which recent estimates are only available for the capital but not the primate city of the country. The remaining sample of 70 countries is listed in the Appendix.[8]

Data for the size of urban agglomerations is available from the World Urbanization Prospects of the UN (UN 1995). I complete the urban data by data from the UN Demographic Yearbooks 1991 and 1994 (UN 1992, 1995a), because the World Urbanization Prospects report urban data only for capital cities and cities with a population of more than 750,000 inhabitants. The advantage of the World Urbanization Prospect Data is that it provides estimates for all countries in a single common year. However, this comes with the drawback that only estimated or interpolated data can be used, which is likely to inhibit measurement errors. Another problem arising from these data is that estimates for urban agglomerations may differ between the two sources. The statistical concept of measuring city size is different in different countries. Some of them report city proper data, others metropolitan area or urban agglomeration data. I use urban agglomeration data where possible, because they better measure the true concentration of a country. On the basis of equation (4.4), I construct, four dependent variables (*PRIMA*1, *PRIMA*2, *PRIMA*3, *PRIMA*4) indicating the share of the one, two, three or four largest cities in the total urban

---

8 The former CSFR takes the place of the Czech Republic. Data for Czechoslovakia is taken from the World Bank (1992).

concentration. These measures are frequently used in addition to *PRIMA*1, in order to consider the size distribution of cities below the largest city.

The following variables are taken or calculated from the World Bank CD-ROM (1996): Population density per square kilometers (*DENSE*), land in square kilometers (*LAND*), density of the transportation system measured as the ratio of road length to land area (*ROADLAND*), total population (*POPTOT*), and urban population (*POPURB*). For Russia, the surface instead of the land area has to be taken. This includes also the surface of interior seas or lakes. Data on road length for Bulgaria, Poland, Rumania is from 1989, for Hungary from 1988 and for Iran from 1985.

Data for PPP adjusted real per capita GDP (*CGDP*) is taken from the Penn World Tables (Mark 5.6). See Summers and Heston (1991) for a description. For Niger, Romania and Russia *CGDP* data is from 1989. Data for openness (*OPEN*) and the political regime (*POLITICS*) is taken from Sachs and Warner (1995). According to their definitions (p. 10-11), *OPEN* is a dummy variable that takes the value one if "a very high proportion of imports [is] covered by quota restrictions," "for Sub-Saharan Africa, [if] a high proportion of exports [is] covered by state export monopolies and state-set prices", the country has "a socialist economic structure", or if "a black market premium over the official exchange rate of 20 percent or more, on average, [prevailed] either for the decade of the 1970s or the decade of the 1980s". Politics is a dummy variable that takes the value one if the country has "a socialist economic structure", "extreme domestic unrest, caused by revolutions, coups, chronic civil unrest, or a prolonged war with a foreign country that is fought on domestic territory", or "extreme deprivation of civil or political rights" (Sachs and Warner, 1995, p. 9).

Data on the colonial history and the date of independence is taken from Fischer Weltalmanach (1997). I use two dummy variables to distinguish different lengths of colonial rule of a country (*COLONY*1815 and *COLONY*1950). *COLONY*1815 takes the value 1 if the country was a colony in the year 1815. Colonial powers, independent countries and countries with a long urban history like China, Germany, Iran or Thailand take the value „0" even if they did not existed as a political entity at that time. *COLONY*1950 indicates countries with a very short history of independence. It takes the value "1", if a country was still a colony in 1950, and "0" otherwise. In case the date of the proclaimed and final or recognized independence differ, I use the year of the proclaimed independence.[9]

---

9 I use the years 1815 and 1950, because they are two landmarks in world history. 1815 was the end of the Napoleonic Wars and the Congress of Vienna. 1950 roughly marks the end of World War II and its postwar turmoils.

Data on historic variables is available from Banks (1971) for 45 countries. I use 1919 data for railroad mileage per square mile (*RAIL*1919), telegraph mileage per square mile (*TE*L1919) and population density (*DENSE*1919). Density data for China is from 1911, for Ireland from 1922, and from Korea for 1904.

Data on historic urban concentration is taken from two sources. For 41 countries data for the size of the largest in percent of the three largest cities is available from Jefferson (1939). Concentration is measured around the year 1935 according to availability. For European countries data for the size of cities for the years 1800 and 1850 is available from Bairoch et al. (1988). From this data, I calculate primacy ratios for 1800 and 1850 (*HISTO*1850 and *HISTO*1800). It should be noted that the data used for the construction of today's primacy ratios is based on the population of urban agglomerations, which are likely to encompass cities and districts that are not included in the population of Bairoch et al.'s city proper data for the 19th century.[10]

GDP data for 1913 (*GDP*1913) is available from Maddison (1995) for 44 countries. For 26 countries, the share of industry in total industrial and agricultural employment in 1870 is taken from Mitchell (1993). The agricultural

10 Therefore, I also use the percentage of the largest city in 1850 and 1800 instead of the percentage of today's largest city in total 1800's or 1850's urban population.

sector encompasses agriculture, forestry and fishing. The industrial sector encompasses extractive industry, manufacturing industry, construction and services.

I also standardize *LAND* by the size of the smallest country, which was Belgium with 30,260 square kilometers. The resulting variable *NLAND* varies between 1 and 564. I standardize CGDP on the value of the poorest country, namely Zaire. The resulting new variables *NCGDP* and *NCGDP*2 now range from 1 to 46.3 and from 1 to 2147. I standardize *TEL*1919 by the smallest value of the sample, such that the new variable *NTEL*1919 varies between 1 and 339.

4.3 Results

As indicated in the last section, data for the size of the largest cities is not collected in a standardized way, but with different statistical concepts and different sources. Additionally, not all explanatory variables are available for the full sample of 70 countries. Hence, I use several different country samples to estimate the likely determinants of urban concentration. The results are reported in Table 2 (Appendix). The reported standard errors are corrected by White's heteroscedasticity consistent variances and covariances.[11]

---

11 I also report the White test without cross terms that shows the probability of no heteroscedasticity in the data, which is, for instance, 12.6 percent in the first regression. In the presence of heteroscedasticity the coefficients are still unbiased, but not efficient, such that t-test and F-test are not interpretable.

I start with the most reliable sample of countries. This sample includes 23 countries for which the World Urbanization Prospects reports at least 4 cities with a population of more than 750,000 inhabitants each.[12] For these countries all four primacy ratios can be calculated from the same data source without relying on additional data from the UN Demographic Yearbook. This sample is not only most reliable in terms of the data. In these countries economic forces have the highest potential to unfold and not to be suppressed by idiosyncratic or genuine influences. This is because this sample only consists of large countries with a large urban population and a system of several cities.

In regression 1, I start with the full set of current independent variables except for *COLONY*1950 because the sample does not contain countries that were a colony in 1950. Except for *OPEN*, all variables have the expected sign , but the equation only explains 20 percent of the observed cross-country variation of urban concentration as measured by *PRIMA*1. The coefficients of *NCGDP* and *NCGDP*2 are statistically significant at the 5 percent level. The coefficient of *NLAND* is significant at the 1 percent level. The coefficients of *OPEN* and *POLITICS* have large standard deviations. t-tests and a redundant variable F-test do not support the hypothesis that trade policy and "bad" politics influence the degree of urban concentration. The standard deviations of

12 These countries are marked with a # in Table 1 in the Appendix.

ROADLAND and *DENSE* are large. Together with a high correlation coefficient between the variables, this points to the plausible multicollinearity between population density and density of the road system.

Multicollinearity means that it is impossible to isolate the individual impact of the variables. OLS estimates remain unbiased and consistent, but have larger standard errors. Thus, confidence intervals get larger, so that the sample data is consistent with a large set of hypotheses. Therefore, in the following regressions, I exclude one of the variables if collinearity between them is high in that sample and report preferred estimates only. Omitting a variable that should be included on theoretical grounds leads to a specification bias, because the remaining variable then measures the combined effect of the correlated variables. Its coefficient gets biased and inconsistent. Since I am mainly interested in the coefficients of *NCGDP* and *NCGDP2*, I accept this bias and focus on the identification of the inverted U-curve.

Regression 2 reports my preferred estimates for the sample of 23 countries, excluding *DENSE*, *POLITICS* and *OPEN*. An F-test shows a probability of 84 percent that the omitted variables can indeed be excluded from the regression. I also use the regression specification error test (RESET) to check whether the structural specification is subject to the problem of omitted variables. It tests the null hypothesis that the expected value of the disturbance term conditional on the regressors equals zero.[13] The probability that this is the case in this regression is 67.7 percent. Additionally, I test the normality assumption of the classical normal linear regression model. I report the Jarque-Bera statistic and the corresponding probability that the residuals are normally distributed, which is 51 percent in this regression. As before, I find that the coefficients of *NCGDP* and *NCGDP2* have the expected sign. They are statistically significant at the 1 percent level. The adjusted R^2 rises to 0.30.

In regressions 3-7, I test whether the results are sensitive to the sample of countries included in the regression. I construct the samples according to the statistical concept that has been used to measure the size of the primate cities. In regression 3, I restrict the previous sample of 23 large countries with 4 cities above 750,000 inhabitants to those 19 countries that use urban agglomeration as the statistical concept. In regression 4, I also use countries that use metropolitan area as the statistical concept. In regression 5, I use all 44 countries that use urban agglomeration as the statistical concept, and in regression 6, I add those

---

13 The test augments the original regression by a matrix of test variables and tests the H_0 that the elements of the coefficient vector of these variables is jointly zero. The additional variables can be the original independent variables raised to the 2nd, 3rd, 4th ... power, depending on the degrees of freedom remaining in the regression. This makes the test also powerful in detecting non-linearities and a wrong functional form. The drawback of the test is that it cannot discriminate between omitted variables and wrong functional form. See Zietz (1988) for a discussion of several adequacy and misspecification tests.

10 countries that report data for metropolitan areas.[14] In regression 7, I use the full set of 70 countries, including those that report city proper data only.

Due to the different statistical concepts, the true degree of urban concentration is measured with an error that is increasing with the sample size. OLS estimates remain unbiased and consistent for measurement errors in the dependent variable, but they are not efficient. Therefore, lower levels of statistical significance are to be expected for the larger samples that use primacy measures on the basis of different statistical concepts.

Nevertheless, I find empirical evidence for an inverted U-curve relationship between urban concentration and GDP in all five regressions. The coefficients of *NGDP* vary between 0.1139 and 0.0488. The coefficients of *NGDP2* vary between -0.0021 and -0.0009. Both decrease in absolute values if the sample gets larger. The coefficients indicate that a maximum level of urban concentration is reached at per capita GDP levels between 11,371 and 13,062 US$.

As expected, the dummy variable for colonies in 1815 still colonies in 1950 increases urban concentration. It is remarkable that the coefficient of *NLAND* is significant at the 1 percent level in all regressions 2-7. *ROADLAND* enters all equations where it is included with a negative coefficient. This indicates that a

14 The 44 and 54 country samples are marked in Table 1 in the Appendix.

better transportation infrastructure, which means lower transport costs, mainly benefits non-core regions. Thus, the negative effect of a higher degree of competition is more than offset by the positive effect of better access of the periphery to the core-market.

In regression 8-10, I use the original sample of 23 countries to test whether the results are sensitive to the measurement of urban concentration. I.e., I use *PRIMA*2, *PRIMA*3 and *PRIMA*4 as dependent variables. The coefficients and the statistical fit of regressions with different dependent variables cannot be compared. However, it can be seen that the existence of the inverted U-curve is robust for different primacy measures, because for all three regressions, I find that the coefficients of *NCGDP* and *NCGDP*2 have the expected sign and are statistically significant.[15]

In regression 11 and 12, I test the hypothesis that countries that industrialized and, thus, urbanized relatively early are less concentrated today. This would reflect that industrial location decisions in the last century were dictated by natural endowments to a larger degree than nowadays. High cost of internal

---

[15] This confirms e.g., Rosen and Resnick (1980) who find a high correlation between different primacy measures. In my sample of 70 countries, the correlation between *PRIMA1* and *PRIMA2*, *PRIMA3* and *PRIMA4* is 97 percent, 93 percent and 89 percent, respectively. In my sample of 23 countries, that is used in regressions No. 1, 2, 8, 9, 10 the correlation is 95 percent, 91 percent and 88 percent, respectively.

transportation led to the development of, for instance, steel production close to natural resources and transportation nodes in "the Ruhr area" in Germany and in the "Great Lake region" in the US. Consequently, early developers set up resource intensive industrial production not necessarily close to existing centers.[16] If so, industrialization might have led to population dispersion. Countries that industrialized relatively late could rely on more advanced transportation technologies. Therefore, they could set up industrial production close to the existing centers and benefit from the availability of a larger number of workers, intermediate goods suppliers and final demand. In this case, industrialization would increase urban concentration. I test this hypothesis adding *INDUSTRY*1900 in regression 11 and *GDP*1913 in regression 12 as measures of early development. As expected, both the share of the industrial sector as well as the per capita GDP level in 1913 have negative coefficients. However, they have large standard errors. The presence of the inverted U-curve is not destroyed by the inclusion of these variables.

Regression 13 and 14 further investigate the role of ancient conditions on current urban concentration. I exclude *DENSE* and *ROADLAND* from these

---

16 For instance, Mokyr (1995: 25) points out that it is striking that "neither Brussels nor Paris nor Berlin nor Amsterdam, nor any other major capital city in Europe, became a center of modern industry".

regressions and include *DENSE*1919, *NTEL*1919 and *RAIL*1919. The density of the telegraph system and the railroad system indicate the cost of overcoming distance. Population density is used as before. Due to a correlation coefficient of 85 percent between *DENSE*1919 and *RAIL*1919 individual effects of these variables cannot be distinguished and, hence, they are not included together in the equations. The results show that all variables enter the equation with a negative coefficient. Thus, early conditions seem to matter, again leaving intact the previously established U-curve.[17]

Regression 15-17 explore whether historic patterns of urban concentration can add to the explanation of different degrees of urban concentration nowadays. This might correct for some idiosyncratic or genuine differences between countries resulting from a historic accident[18] or geographic, climatic or endowment differences. In turn, this might improve the significance of the other economic and conditional variables. However, it is also possible that adding historic degrees of concentration renders all other variables insignificant. In this

---

17 Note that the high possibility of having omitted relevant variables in regressions 13 and 14, as indicated by the F-statistic is most likely due to a 85 percent correlation between *OPEN* and *NGDP*. Therefore, *OPEN* would indicate some of the effects that actually should be assigned to *NGDP*.

18 An example for such a historic accident is the dissolution of the Austrian-Hungarian empire which left a huge capital with a relatively small hinterland. This results in a very high primacy ratio for Austria.

case current patterns would be entirely determined by historic patterns. Population growth and further urbanization would then only have resulted in a proportional growth of all cities. Degrees of concentration would then be determined by early historic accidents, climatic geographic or endowment differences. It seems reasonable that much of today's concentration can be explained by degrees of concentration in the 1980s and 1970s and may be even in the 1960s and 1950s, but it is unclear how long the effects take to peter out.

Therefore, I include, as a further independent variables, measures of urban concentration at three points in time. In regression 15, I use the primacy ratio in 1800 (*HISTO*1800). The same ratio is calculated for 1850 (*HISTO*1850) and used in regression 16. In regression 17, I use the percentage of the largest in the three largest cities (*JEFFRATIO*), which is available for 41 countries in the 1930s from Jefferson (1939). As geographic and demographic differences of countries are reflected in the historic measures of urban concentration, only the political and economic explanatory variables are included in the regressions.

The coefficients of the primacy ratios of 1800 and 1850 in regressions 15 and 16 are significant at the 1 percent level. However, the *NGDP* and *NGDP*2 coefficients are loosing its significance if the historic variables are included in this sample. Whether this is due to the small sample size of 22 countries, the different statistical concepts of measuring city size in this sample and the evolving measurement errors in the dependent variable or the non-existence of an inverted U-curve in this sample remains unclear. Also the F-test and the RESET-test point to a misspecification if *HISTO*1850 or *HISTO*1800 are included and the previous explanatory variables are exclude in the sample. Using a sample of 41 countries in regression 17, the *JEFFRATIO* turns out to be significant at the 1 percent level. The inclusion of this ratio does not touch the existence of the inverted U-curve and the coefficient of COLONY1815. Apparently it reduces standard errors as indicated by the exceptionally high significance levels in this regression. Therefore, future research should continue in this line and better identify the role of path dependencies, and how long an impact historic patterns of urban concentration have on current patterns of urban concentration.

5. Conclusion

In this paper, I test whether the relationship of urban concentration and economic development takes the form of an inverted U-curve, where concentration first increases and then decreases in the course of economic development. I find evidence for this hypothesis, using different samples of countries. The relationship is conditional on the size of the land area, population density and the density of the transportation system. I also find that countries with a long independent urban history today have lower degrees of urban concentration and countries with a relatively recent colonial past have higher degrees of urban concentration. Openness and political and economic freedom add little to the explanation of urban concentration. In none of the regressions, their coefficients are significant at the 10 percent level. This might be due to the measurement of this variables. Future empirical research should better identify the relationship of these variables and urban concentration. Historic values of population density and the extent of the transportation system also add to the explanation of urban concentration. This indicates that some determinants can have long lasting effects. 19th century degrees of urban concentration were found to have an impact on current patterns of concentration. The degree of concentration of the 1930s also improves the explanation of current concentration significantly.

Understanding the determinants of urban concentration and transition is important for being confident in future projections of urban growth and regional inequalities. These in turn, are essential to formulating proper regional, social and economic policies. The strong appeal of an inverted U-curve pattern between population concentration and economic development and the strong efforts of urban and development economists to find such a relationship, lies not only in the fact that economists like to describe empirical phenomena by economic instead of demographic or geographic determinants. The appeal of the inverted U-curve is also rooted in the hope that some problems especially developing countries have with regional inequalities and the excessive growth of their primate cities will vanish in the course of further development. This paper yields some evidence from cross-country regressions that this hope is well-founded.

References

Ades, A. and E. Glaeser (1995). Trade and Circuses: Explaining Urban Giants. *Quarterly Journal of Economics* 110 (1): 195-227.

Alonso, W. (1980). Five Bell shapes in Development. *Papers of the Regional Science Association* 45: 5-16.

Alperovich, G. (1992). Economic Development and Population Concentration. *Economic Development and Cultural Change* 41: 63-74.

Alperovich, G., and J. Deutsch (1995). The Size Distribution of Urban Areas: Testing for the Appropriateness of the Pareto Distribution Using a Generalized Box-Cox Transformation Function. *Journal of Regional Science* 35 (2): 267–276.

Bairoch, P., J. Batou, and P. Chèvre (1988). The Population of European Cities: Data Bank and Short Summary of Results. Geneva.

Banks, A. (1971). Cross-polity Time-Series Data. Cambridge, MA.

Carroll, G. (1982). National City-Size Distributions: What do we know after 67 Years of Research?. *Progress in Human Geography* 6 (1): 1-43.

DeCola, L. (1984). Statistical Determinants of the Population of a Nation's Largest CityLognormal. *Economic Development and Cultural Change* 33(1): 71-98.

Eaton, J., and Z. Eckstein (1997). Cities and Growth: Theory and Evidence from France and Japan. *Regional Science and Urban Economics* 27: 443-474.

El-Shakhs, S. (1972). Development, Primacy, and the System of Cities. *Journal of Developing Areas* 7: 11-36.

Fischer-Weltalmanach (1997). Zahlen, Daten, Fakten. Fischer, Frankfurt.

Hanson, G. (1996). Localization Economies, Vertical Organization, and Trade. *American Economic Review* 86 (5): 1266-1278.

Hsing, Y. (1990). A note on Functional Forms and the Urban Size Distribution. *Journal of Urban Economics* 27: 73-79.

Jefferson, M. (1939). The Law of the primate City, *The Geographical Review* 29: 226-232.

Junius, K. (1996). Economic Development and Industrial Agglomeration: An Inverted U-Curve. Kiel Working Papers, 770. Kiel Institute of World Economics.

Junius, K. (1997). Economies of Scale: A Survey of the Empirical Literature. Kiel Working Papers, 813, Kiel Institute of World Economics.

Kamerschen, D. (1969). Further analysis of overurbanization. *Economic Development and Cultural Change* 17: 235-253.

Kasarda, J., and E. Crenshaw (1991). Third World Urbanization: Dimensions, Theories and Determinants. *Annual Review of Sociology* 17: 467–501.

Kim, S. (1995). Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987. *Quarterly Journal of Economics* 110: 881-908.

Krugman, P. (1991). Increasing Returns and Economic Geography. *Journal of Political Economy* 99: 483–499.

Lemelin, A., and M. Polèse (1995). What about the Bell-shaped Relationship between Primacy and Development? *International Regional Science Review* 18 (3): 313-330.

Maddison, A. (1995). Monitoring the World Economy, 1820-1992. OECD 1995, Paris.

Mera, K. (1973). On the Urban Agglomeration and Economic Efficiency. *Economic Development and Cultural Change* 21: 309-24.

Mitchell, B. (1993). International Historical Statistics, 1750-1988. The Americas 1750-1988. Second edition. Basingstoke, Hants.

Mokyr, J. (1995). Urbanization, Technological Progress and Economic History. In: H. Giersch (ed.). Urban Agglomeration and Economic Growth, Springer.

Moomaw, R., and A. Shattler (1996). Urbanization and Economic Development: A Bias Toward Large Limits? *Journal of Urban Economics* 40: 13-37.

Mutlu, S. (1989). Urban Concentration and Primary Revisited: An Analysis and Some Policy Conclusions. *Economic Development and Cultural Change*: 611–639.

Parr, J. (1985). A Note on the Size Distribution of Cities Over Time. *Journal of Urban Economics* 18: 199–212.

Parr, J., and C. Jones (1983). City Size Distributions and Urban Density Functions: Some Interrelationships. *Journal of Regional Science* 23 (3): 283–307.

Richardson, H.W., and G. Schwartz (1988). Economic Development, Population, and Primacy. *Regional Studies* 22: 467-75.

Rosen, K., and M. Resnick (1980). The Size Distribution of Cities: An Examination of the Pareto Law and Primacy. *Journal of Urban Economics* 8: 165-186.

Sachs, J., and A. Warner (1995). Economic Convergence and Economic Policies. NBER Working Papers, 5039.

Sheppard, E. (1982). City Size Distributions and Spatial Change. *International Regional Science Review* 7 (2): 127–151.

Singer, H.W. (1936). The "Courbe des Populations": A Parallel to Paretos Law. *Economic Journal* 46: 254–263.

Summers, R., and A. Heston (1991). The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988. *Quarterly Journal of Economics* 106: 328-368.

Syrquin, M. (1989). Patterns of Structural Change. In: H. Chenery and T. Srinivasan (eds.), Handbook of Development Economics, 1, Amsterdam.

United Nations (1992). Demographic Yearbook 1991. New York

United Nations (1995a). Demographic Yearbook 1994. New York

United Nations (1995b). World Urbanization Prospects: The 1994 Revision. New York.

Wheaton, W. and H. Shishido (1981). Urban Concentration, Agglomeration Economics and the Level of Economic Development. *Economic Development and Cultural Change* 50:17-30.

Williamson, J.G. (1965). Regional Inequality and the Process of National Development: A Description of the Patterns. *Economic Development and Cultural Change* 13 (4), Part II: 3-84.

Williamson, J.G. (1997). Industrialization, Inequality and Economic Growth, Edward Elgar, Cheltenham, UK.

World Bank (1992). Social Indicators of Development 1991-1992.

World Bank CD-ROM (1996). World Development Indicators.

Zietz, J. (1988). Adequacy Test for Linear Regression Models: A Nontechnical Overview with Applications. *Jahrbücher für Nationalökonomie und Statistik* 204 (3): 255-272.

Appendix

Table 1 — Country List[a]

		GDP	Primacy			GDP	Primacy
Algeria	*	2957	0.2353	Kenya		1080	0.2756
Argentina	*#	5532	0.3793	Korea	#	8271	0.3337
Australia	#+	17517	0.2452	Malaysia		5997	0.1275
Austria	*	15560	0.4814	Mali	*	714	0.3366
Bangladesh	+	1641	0.3462	Mexico	#+	6896	0.2459
Belgium	*	16533	0.1195	Morocco	*	2554	0.2509
Benin		1128	0.3989	Netherlands	*	16096	0.0794
Bolivia		1890	0.2815	New Zealand	*	14591	0.3078
Brazil	*#	4792	0.1340	Niger	*	563	0.3804
Bulgaria		7529	0.2157	Nigeria	*	1117	0.2287
Burkina Faso		608	0.3941	Norway	*	16345	0.2221
Cameroon		1249	0.2155	Pakistan	*#	1661	0.2205
Canada	*#	20752	0.1771	Paraguay	+	2496	0.2984
Chile	+	5279	0.4160	Peru	+	2603	0.4297
China	*#	1536	0.0452	Philippines	*	2112	0.2686
Colombia	*#	3902	0.2080	Poland	*#	4564	0.1448
Cote d'Ivoire	*	1372	0.4482	Portugal	*	9005	0.5015
Czech Republic	*	5066	0.0990	Romania	*	2656	0.1697
Denmark	+	17217	0.3086	Russian Federation	*#	8780	0.0836
Ecuador	*	3163	0.2653	South Africa	*#	3886	0.1258
Egypt, Arab Rep. of	*	2153	0.3696	Spain	+	11765	0.1409
Finland	*	17080	0.2848	Sri Lanka		2468	0.1671
France	*#	16956	0.2264	Sweden	*	18024	0.2095
Germany	*#	18235	0.0938	Switzerland	*	20729	0.2031
Ghana	+	1101	0.2751	Syrian Arab Rep.	*	4714	0.2888
Greece	*	8203	0.5449	Thailand		4270	0.5708
Guatemala		2535	0.2324	Tunisia	*	3392	0.3930
Hungary		6430	0.3134	Turkey	*#	4489	0.1905
India	*#	1505	0.0567	Uganda	*	625	0.4019
Indonesia	*#	2323	0.1697	United Kingdom	*#	15741	0.1434
Iran, Islamic Rep. of	*#	3577	0.1914	United States	*#	21827	0.0854
Ireland	*	11273	0.4596	Uruguay		5536	0.4679
Italy	*#	15309	0.1210	Venezuela	#+	6859	0.1573
Japan	*#	17625	0.2623	Zaire		471	0.3284
Jordan		3774	0.4284	Zambia		799	0.2860

[a]GDP denotes PPP-adjusted GDP levels. Primacy denotes the share of the largest city in total urban population. Both values are for 1990. — #Country has at least 4 cities with mor than 750,000 inhabitants (23 countries). — *Country reports urban agglomeration data (44 countries). — +Country reports urban agglomeration or metropolitan area data (54 countries).

Source: UN (1995b), World Bank CD-ROM (1996).

Table 2 — Regression Results

	Basic results		Different samples		
Regression	1	2	3	4	5
Dependent variable	*PRIMA1*	*PRIMA1*	*PRIMA1*	*PRIMA1*	*PRIMA1*
NCGDP	0.1120**	0.1139***	0.0829**	0.0833**	0.0676**
	(0.0387)	(0.0388)	(0.0346)	(0.0289)	(0.0287)
NCGDP2	-0.0021**	-0.0021***	-0.0015**	-0.0015**	-0.0013**
	(0.0008)	(0.0007)	(0.0007)	(0.0006)	(0.0006)
NLAND	-0.0027***	-0.0023***	-0.0028***	-0.0027***	-0.0035***
	(0.0008)	(0.0006)	(0.0005)	(0.0005)	(0.0005)
DENSE	-0.0006		-0.0032	-0.0030	-0.0037***
	(0.0020)		(0.0019)	(0.0017)	(0.0010)
ROADLAND	-0.1996	-0.2229			
	(0.3997)	(0.2854)			
COLONY1815	0.4624	0.4224*	0.3644	0.4097*	0.3641**
	(0.2746)	(0.2365)	(0.2186)	(0.1944)	(0.1767)
*COLONY*1950					0.4645*
					(0.2533)
POLITICS	0.2877				
	(0.2937)				
OPEN	-0.1809				
	(0.3649)				
Historic variables					
*INDUSTRY*1900					
*GDP*1913					
*DENSE*1919					
*NTEL*1919					
*RAIL*1919					
JEFFRATIO					
*HISTO*1850					
*HISTO*1800					
$\bar{R}^2$	0.20	0.30	0.38	0.42	0.53
SEE	0.56	0.53	0.49	0.46	0.53
No. of observations	23	23	19	22	44
Jarque-Bera test for	1.08	1.35	0.38	0.25	0.40
normality of residuals[a]	[0.58]	[0.51]	[0.82]	[0.88]	[0.50]
White test for	2.09	1.06	2.32	3.21	0.76
heteroscedasticity[a]	[0.13]	[0.44]	[0.11]	[0.03]	[0.65]
Ramsey reset test (3)[a]	1.86	0.52	2.48	1.93	0.40
	[0.20]	[0.68]	[0.12]	[0.17]	[0.76]
F-test for omitted variables[a]	-	0.27	0.88	0.91	0.45
		[0.84]	[0.48]	[0.46]	[0.72]

Standard deviations in parenthesis: *Significant at 10 percent level; **significant at 5 percent level; ***significant at 1 percent level. — SEE = Standard error of estimates. — [a][ ]Marginal probability values in parenthesis. — Constants are not reported.

Table 2 continued

	Different samples		Different measures of urban concentration		
Regression	6	7	8	9	10
Dependent variable	*PRIMA*1	*PRIMA*1	*PRIMA*2	*PRIMA*3	*PRIMA*4
NCGDP	0.0658**	0.0488	0.1039***	0.0985***	0.1008***
	(0.0312)	(0.0340)	(0.0351)	(0.0319)	(0.0331)
NCGDP2	-0.0012*	-0.0009	-0.0018**	-0.0017**	-0.0017**
	(0.0006)	(0.0007)	(0.0006)	(0.0006)	(0.0006)
NLAND	-0.0033***	-0.0035***	-0.0019***	-0.0022***	-0.0023***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
DENSE	-0.0011	-0.0008			
	(0.0010)	(0.0008)			
ROADLAND	-0.2003	-0.2245*	-0.2077	-0.2723	-0.3071
	(0.1323)	(0.1168)	(0.2859)	(0.2687)	(0.2651)
COLONY1815	0.4888***	0.3802**	0.5410**	0.5326***	0.5617***
	(0.1662)	(0.1658)	(0.1985)	(0.1805)	(0.1841)
*COLONY*1950	0.4885*	0.2242			
	(0.2614)	(0.2625)			
POLITICS					
OPEN					
Historic variables					
*INDUSTRY*1900					
*GDP*1913					
*DENSE*1919					
*NTEL*1919					
*RAIL*1919					
JEFFRATIO					
*HISTO*1850					
*HISTO*1800					
$\bar{R}^2$	0.41	0.31	0.30	0.37	0.41
SEE	0.57	0.60	0.49	0.45	0.45
No. of observations	54	70	23	23	23
Jarque-Bera test for	0.79	0.06	1.02	0.84	0.96
normality of residuals[a]	[0.67]	[0.97]	[0.60]	[0.66]	[0.62]
White test for	1.13	1.27	2.08	1.91	1.64
heteroscedasticity[a]	[0.36]	[0.26]	[0.11]	[0.14]	[0.20]
Ramsey reset test (3)[a]	0.49	1.45	1.30	1.45	1.60
	[0.69]	[0.24]	[0.31]	[0.27]	[0.23]
F-test for omitted variables[a]	0.06	0.77	0.13	0.16	0.20
	[0.94]	[0.47]	[0.94]	[0.92]	[0.90]

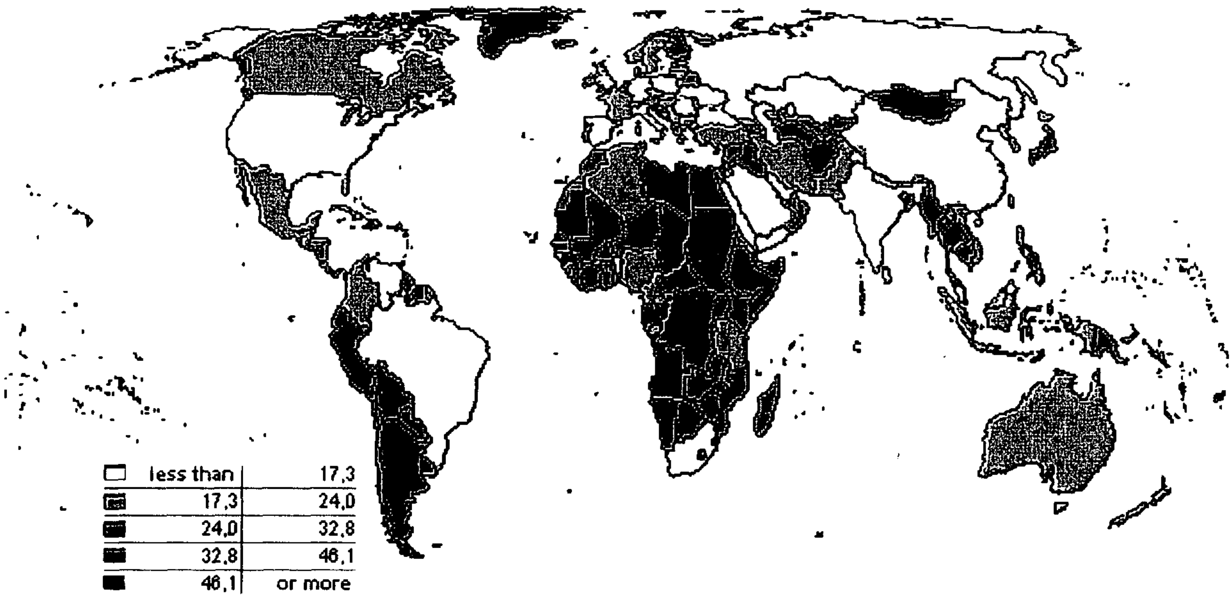
Standard deviations in parenthesis: *Significant at 10 percent level; **significant at 5 percent level; ***significant at 1 percent level. — SEE = Standard error of estimates. — [a][ ]Marginal probability values in parenthesis. — Constants are not reported.

Table 2 continued

	Historic indicators		Historic infrastructure		Historic urban concentration		
Regression	11	12	13	14	15	16	17
Dependent variable	*PRIMA*1	*PRIMA*1	*PRIMA*1	*PRIMA*1	*PRIMA*1	*PRIMA*1	*PRIMA*1
NCGDP	0.0897*	0.0905**	0.0835**	0.0711**	0.0229	0.0225	0.0847**
	(0.0452)	(0.0336)	(0.0320)	(0.0337)	(0.0624)	(0.0650)	(0.0381)
*NCGDP*2	-0.0017*	-0.0016**	-0.0017**	-0.0014**	-0.0008	-0.0007	-0.0017**
	(0.0008)	(0.0006)	(0.0006)	(0.0007)	(0.0013)	(0.0014)	(0.0008)
NLAND	-0.0032***	-0.0033***	-0.0035***	-0.0034***			
	(0.0005)	(0.0005)	(0.0005)	(0.0006)			
DENSE	-0.0041***						
	(0.0012)						
ROADLAND		-0.3021***					
		(0.1025)					
*COLONY*1815	0.2539	0.5266**	0.4162**	0.5116***	0.7835**	0.6693*	0.6213***
	(0.2012)	(0.2182)	(0.1758)	(0.1861)	(0.3617)	(0.3532)	(0.2248)
*COLONY*1950							
POLITICS							
OPEN							
Historic variables							
*INDUSTRY*1900	-0.8028						
	(0.8303)						
*GDP*1913		-0.0001					
		(0.0001)					
*DENSE*1919			-0.0002***				
			(0.0001)				
*NTEL*1919			-0.0018	-0.0022*			
			(0.0013)	(0.0013)			
*RAIL*1919				-0.0003**			
				(0.0001)			
JEFFRATIO							3.0756***
							(0.7086)
*HISTO*1850						2.7574***	
						(0.6449)	
*HISTO*1800					2.3547***		
					(0.6391)		
$\bar{R}^2$	0.41	0.30	0.44	0.38	0.32	0.40	0.38
SEE	0.61	0.67	0.58	0.61	0.66	0.62	0.60
No. of observations	27	44	45	45	22	22	41
Jarque-Bera test for normality of residuals[a]	0.76	0.35	0.73	0.48	1.00	0.87	1.29
	[0.68]	[0.84]	[0.70]	[0.79]	[0.61]	[0.65]	[0.52]
White test for heteroscedasticity[a]	1.36	0.50	0.44	0.50	0.46	1.04	0.39
	[0.28]	[0.88]	[0.91]	[0.88]	[0.82]	[0.44]	[0.88]
Ramsey reset test (3)[a]	0.50	0.21	0.35	0.33	2.17	1.18	0.39
	[0.68]	[0.89]	[0.79]	[0.80]	[0.14]	[0.35]	[0.76]
F-test for omitted variables[a]	0.78	1.50	3.65	2.41	3.40	2.84	3.15
	[0.52]	[0.23]	[0.04]	[0.10]	[0.05]	[0.08]	[0.03]

Standard deviations in parenthesis: *Significant at 10 percent level; **significant at 5 percent level; ***significant at 1 percent level. — SEE = Standard error of estimates. — [a][ ]Marginal probability values in parenthesis. — Constants are not reported.

Population in the largest city (% of urban population)



Source: World Development Indicators (1997).