Urban Studies (1986) **23**, 285–305

Factors of Urbanisation in the Nineteenth Century Developed Countries: A Descriptive and Econometric Analysis

Paul Bairoch and Gary Goertz

[*First received January 1985; in final form October 1985*]

Summary. This paper describes the situation from the beginning of the industrial revolution when levels of urbanisation were dependent on geography and the historical situation and when the general level of urbanisation was low, through the nineteenth century in which agricultural productivity and industrialisation determined the levels of urbanisation. This period represents a dramatic increase in the levels of urbanisation during which the present urban structures were put into place. A comparative econometric study, finds that economic growth pushed urbanisation, with industrialisation being the most important factor for Europe and agricultural productivity being quite important for the European settled countries. Other important factors are found to be, trade, total population, topography, and form of industrialisation. Railroad networks more or less extensive than normal were not found to be influential.

1. Introduction and overview

The 19th century, and particularly the period between 1820–30 and 1914, constitutes a turning point between a society still essentially rural and a developed, urbanised society prefiguring that of the end of the the 20th century, a society in which over 90 per cent of the population would no longer be involved in agriculture, even if some of these might still live in the countryside. It is no exaggeration to say that, with the exception of the USSR and also Japan,[1] the present urban system was almost in place at the eve of World War I.

Certainly around 1830, England, the cradle of the industrial revolution, had already behind it a large part of what one might call the 'urban transition', since from around 1700 to 1830 the English population living in cities of more than 5,000 inhabitants rose from 15 per cent to 34 per cent. Importantly, this level of urbanisation was achieved without the massive imports of foodstuffs from other countries. The England of the 1830s was certainly no longer a region with a surplus of cereals despite being called the 'breadbasket' of Europe by contemporaries in the middle of the 18th century. However, with its imports consituting 4–6 per cent of its food needs, it was also not the England of the end of the 19th century which, like the Netherlands of the 17th century or certain Italian city-states of the 13–14th century, fed its urban population largely with imported cereals. Around 1700, London with its 550,000 inhabitants was along with Paris, one of the largest cities of Europe, but that continent and even more so Asia had known cities twice as large. However, towards 1830 London or Greater London (the metropolitan area of London) already had 1.5 million

Paul Bairoch is Professor of Economic History at the University of Geneva, Gary Goertz holds appointments in the University of Michigan and the University of Geneva.

The authors would like to thank Professors H. Glejser, of the University of Brussels and P. Balestra and F. Calevaro of the University of Geneva for having read and criticised earlier versions of this paper. They are, of course, not responsible for the views expressed here.

[1]Japan and South Africa are not included in this study because their social and economic structures are too different from the other developed countries. All references, except when otherwise noted, to developed countries are to be understood as excluding Japan and South Africa.

inhabitants, which made it the largest city of the period and maybe the largest in history up to that time. This was certainly the case by 1851 when a census reported that London had 2.4 million inhabitants.

But because the England of 1830, with a total population of 14 million inhabitants, represented only 5 per cent of the population of the future developed world, the urban structure of this future developed world was but marginally changed. In 1830, the level of urbanisation of the future developed world was close to that of previous centuries (see Table 1). At this date, the level of urbanisation[2] can be estimated at 12.3 per cent, compared with about 10–11 per cent in 1700, the level of urbanisation increasing by only about 0.1 per cent per year. This moderate increase during the 18th century continued a trend already established during the preceding two centuries. However, by the eve of World War I, this level was 36 per cent (43 per cent excluding Russia). Hence, between 1830 and 1914, the level of urbanisation increased by 1.3 per cent per year, the urban population by 2.3 per cent per year, while from 1914 to 1980 these increased respectively by 0.4 per cent and 1.7 per cent. Thus, the developed world of around 1830 had only one city of over a million inhabitants became that of 1914 with twelve such cities, the largest of which being London with more than 7 million inhabitants.

The period studied here is not only a key period in the growth of already existing cities, but also a period 'par excellence' of the emergence of new cities. The long term evolution of the urban structure measured by city size is summarised in Table 2. The number of large cities (more than 100,000 inhabitants) which doubled between 1700 and 1800 increased by a factor of 12 between 1800 and 1914. As for very large cities (more than 500,000) these increased by a factor of 200 during the 19th century. If these changes are essentially the result of the growth of the size of the city, the 19th century is also the period in urban history which saw the birth of the largest number of new cities. Effectively, of the approximately 268 cities of more than 100 000 inhabitants in the developed world around 1910, some 98 did not exist or were villages at the beginning of the 19th century (or in England in the middle of the 18th century). The proportion must be higher for smaller cities so that one can estimate that 40 per cent of the cities of more than 50,000 existing in 1910 were new cities or became cities during the 19th century. After World War I very few new cities were created, except in the USSR, although in 1930 there began the creation of new planned cities, these, however, being created largely in order to alleviate the congestion of large existing urban agglomerations.

It is also important to examine the various factors both economic and non-economic which led to the urbanisation of 19th century Europe as well as those other countries now qualified as developed. It is in the period 1830–1914 that Europe undertook urbanisation at its fastest rate. It is on Europe that this study will concentrate in the econometric analysis in the second section, although the European settled

Table 1

Long term evolution of urbanisation of the developed countries

	Total population (millions)	Urban population (millions)	Level of urbanisation (%)
1300	80–100	7–9	7.0–9.0
1500	85–105	8–9	7.0–9.0
1700	135–160	14–17	10.0–11.0
1800	211	23	10.7
1830	255	31	12.3
1880	405	95	23.6
1914	600	215	35.7
1950	749	353	47.1
1980	992	659	66.4

Source: 1800–1914: see Table 3; other data adapted from Bairoch, P., *De Jéricho à Mexico: Villes et économie dans l'Histoire*, 1985.

Note: The degree of rounding of the figures does not imply a correspondingly low margin of error.

Table 2

Evolution of the number of great cities in the developed world

City sizes (thousands)	1580	1700	1800	1914	1980
100–200	3	9	16	138	457
200–500	1	1	6	84	334
500–1000	–	2	1	47	115
1000–5000	–	–	1	10	93
5000 and more	–	–	–	2	7
Total	4	12	24	281	1006

Sources: Adapted from Bairoch (1985).

[2]From this point on, except when indicated to the contrary, we will consider urban population as defined by cities of 5,000 inhabitants or more. See Appendix A1.

countries (i.e. Australia, Canada, New Zealand and the USA) will also be included. One purpose of this study is to try to move beyond the explanation of the level of urbanisation in terms of the level of Gross National Product per capita and to consider economic factors in somewhat more detail and, in addition, to consider social, demographic and structural factors which may contribute to or constrain the urbanisation of a society. It should be recalled that in 1800, when the disparities in GNP per capita were reasonably small, less than a 2 to 1 range, already the level of urbanisation of the Netherlands was four times that of Germany while its GNP per capita was only roughly one-third higher, and that the level of urbanisation varied in Europe by a factor of 12 (using a 5,000 inhabitant criterion, lower if one uses a 2,000 inhabitant criterion, see below). By the beginning of the 20th century the range of GNP per capita was 4 to 1 and the levels of urbanisation varied only by 5 to 1. Thus 19th century Europe and the other future developed countries present a wonderful 'natural experiment' in the causes of urbanisation, since there exists a variety of societies urbanising and growing at different rates relatively free of major political events such as wars or natural catastrophies such as famines or plagues.

Therefore, it is no exaggeration to consider the period studied here as fundamental in the emergence of the contemporary urban system of the developed countries. In addition, new data collected or calculated by one of the authors in the areas of urbanisation and economic development make such a study timely.

This article consists of two quite distinct parts: the first is devoted to a mostly, but not entirely, descriptive analysis, the second is devoted to a comparative econometric analysis. The description of the data and technical details are to be found in two appendices at the end of the article.

2. A Descriptive Analysis

Let us begin by examining the differences in the level of urbanisation before the upheavals brought about by the industrial revolution. The most striking fact is the large spread in the level of urbanisation of the various countries. Around 1800, between the least urbanised country (Finland) and the most (the Netherlands) the spread is from 1 to 12. The urban situation in these various countries around 1800 is still very marked by their several hundred and even several thousand years of previous history. In the previous several hundred years, if one excludes the United Kingdom, the six most urbanised countries (in decreasing order: the Netherlands, Belgium, Italy, Spain, Denmark and Portugal) have an urban network that is the result of previous periods of commercial power already long past. The level of urbanisation for these countries together is 19 per cent, 8 per cent when compared with the rest of Europe. In the previous thousands of years of history, an important fraction of countries with lower levels of urbanisation, notably Scandinavia and Russia, owe these lower levels to the lateness of their neolithic revolution. It is also obvious that the climate of these countries has an effect on their potential for urbanisation independent of their historical situation. Climate forms a larger constraint for traditional societies in colder regions, most importantly with regard to the supply of combustible materials and to perhaps lower agricultural yields.[3]

Regarding the large spread of levels of urbanisation at the beginning of the 19th century, it is important to note that urban population is defined here as the total population of all cities over 5,000 inhabitants. This is at best the least problematic and most operational definition for the second half of the 19th century, but for the early 19th century, a cutoff of 2,000 is better. Since in relative terms the importance of the population living in cities between 2,000 and 5,000 inhabitants is much larger for the less urbanised countries than for the more urbanised,[4] the spread in the levels of urbanisation is much smaller if the 2,000 inhabitants criterion is used. The extreme values using the 2,000 inhabitants criterion pass from a range of 1 to 12 to 1 to 4–5 and the standard deviation from 0.088 to 0.085 (see

[3]The lack of comparable data on cereal yields during the beginning of the 19th century and different technological levels make any estimations of the impact of climate very risky.

[4]According to our estimations based on data from various countris and for various periods (data which have a certain margin of error), one can determine by interpolation or extrapolation the difference between the two definitions of urban population. The equation used for calculating the 2,000 inhabitant level of urbanisation is the following: 2,000 level = 0.0453 + 0.965* (level by 5,000 criterion), thus, for example, a level of urbanisation of 5% using the 5,000 inhabitant criterion becomes 9.4% by the 2,000 criterion.

Table 3

Level of urbanisation for individual countries (percentage of total population living in cities of 5,000 or more inhabitants)

	1800[a]	1830	1850	1880	1900	1910
EUROPE	10.9	12.6	16.4	23.5	30.4	32.8
Austria-Hungary	6.5	7.1	9.7	16.0	25.6	28.5
Belgium	20.5	(25.0)	33.5	43.1	52.3	56.6
Bulgaria	5.5	(5.5)	(6.0)	(11.0)	15.0	22.1
Denmark	15.6	14.1	14.6	23.0	33.5	35.9
Finland	3.5	3.5	3.7	6.1	10.4	12.6
France	12.2	15.7	19.5	27.6	35.4	38.5
Germany	8.9	(9.1)	(15.0)	29.1	42.0	48.8
Greece	11.5	(12.0)	(14.0)	(16.0)	(21.0)	22.0
Italy	18.0	(19.0)	(23.0)	(28.0)	(35.5)	(40.0)
Netherlands	37.4	35.8	35.6	44.5	47.8	50.5
Norway	7.0	7.2	9.0	16.0	24.3	25.1
Portugal	15.5	(15.0)	(15.0)	(15.0)	15.7	15.6
Romania	7.5	(7.5)	(11.0)	(14.0)	17.3	16.0
Russia	5.9	(6.0)	(7.2)	10.6	13.2	14.3
Serbia	10.0	(10.0)	(10.0)	(10.0)	9.8	10.0
Spain	17.5	(17.5)	(18.0)	(26.0)	(34.0)	(38.0)
Sweden	6.6	6.6	6.8	12.5	19.3	22.6
Switzerland	7.0	7.5	11.9	20.4	30.6	37.1
United Kingdom	19.2	27.5	39.6	56.2	67.4	69.2
OTHER DEVELOPED COUNTRIES[b]	5.5	7.9	13.9	24.4	35.6	41.6
Canada	6.5	(7.0)	9.5	15.0	35.9	41.6
United States	5.3	7.8	5.3	25.0	35.9	41.6
TOTAL	10.7	12.3	16.2	23.6	31.3	34.4

[a]More approximate data.
[b]Australia, Canada, New Zealand and United States.
Sources: Level of urbanisation according to the 5,000 criterion elaborated by Bairoch (1985), see Appendix A (Point A1).
Notes: The degree of rounding of the figures does not imply a correspondingly low margin of error.
Figures in parentheses have a higher return margin of error than other figures for the same period.

Appendix A1 for a discussion of the theoretical and practical problems of defining urbanisation).

As strange as it may seem, in spite of the important differences in the starting date of development, the spread of the level of urbanisation increases only slightly during the first half of the 19th century and only moderately thereafter (see Section 3 where this variable was found to be important in explaining the level of urbanisation). Thus, the standard deviation of the level of urbanisation of Europe increases only from 0.078 to 1800 to 0.100 in 1850 and 0.159 in 1910. This is due to the fact that the process of economic development and modernisation began first in the less urbanised countries (Bairoch, 1985, Chap. 16).

As one might expect, the 'take-off' in the process of urbanisation — clearly tied to the beginning of economic development and modernisation — began at quite different dates in different countries. These differences are not only due to varying dates of economic take-off but also to the characteristics of this take-off. If the margin of error of the data is taken into account (especially those for 1800 which are probably under-estimated for many countries) only four countries (Belgium, France, the United Kingdom and the United States) out of 23 had a higher level of urbanisation in 1830 than in 1800. The absence of Switzerland from this group, despite the fact that it was probably one of the first countries to have imitated the United Kingdom, is probably due to a process of industrialisation which was not spatially concentrated. Nevertheless, already in 1840–50 Switzerland had started to urbanise, as was also the case in Austria-Hungary and Canada.[5]

[5]The absence of good data for Australia and New Zealand does not permit very certain estimations; but it is likely that until 1850 their level of urbanisation was not very different from that of previous decades, already quite elevated. In addition, around 1850, these countries represent only 0.1% of the total population of the developed countries.

Table 4

Level of urbanisation according to the criterion used to define urban population

	1800[a]		1850		1910	
	5000 criter.	2000 criter.[b]	5000 criter.	2000 criter.[b]	5000 criter.	2000 criter.[b]
EUROPE	10.9	15.0	16.4	20.3	32.9	36.2
Austria-Hungary	6.5	10.8	9.7	13.9	28.5	32.0
Belgium	20.5	24.3	33.5	36.9	56.6	59.2
Bulgaria	5.5	9.8	(6.0)	(10.3)	22.1	25.9
Denmark	15.6	19.6	14.6	18.6	35.9	39.2
Finland	3.5	7.9	3.7	8.1	12.6	16.7
France	12.2	16.3	19.5	23.4	38.5	41.7
Germany	8.9	13.1	(15.0)	(19.0)	48.8	51.6
Greece	11.5	15.5	(14.0)	(18.0)	22.0	25.8
Italy	18.0	21.9	(23.0)	(26.7)	(40.0)	(43.1)
Netherlands	37.4	40.6	35.6	38.9	50.5	53.3
Norway	7.0	11.3	9.0	13.2	25.1	28.8
Portugal	15.5	19.5	(15.0)	(19.0)	15.6	19.6
Romania	7.5	11.8	(11.0)	(15.2)	16.0	20.0
Russia	5.9	10.2	(7.2)	(11.5)	14.3	18.3
Serbia	10.0	14.2	(10.0)	(14.2)	10.0	14.2
Spain	17.5	21.4	(18.0)	(21.9)	(38.0)	(41.2)
Sweden	6.6	10.9	6.8	11.1	22.6	26.3
Switzerland	7.0	11.3	11.9	16.0	37.1	40.3
United Kingdom	19.2	23.1	39.6	42.8	69.2	71.3
OTHER DEVELOPED COUNTRIES[c]	5.5	9.8	13.9	18.0	41.6	44.7
Canada	6.5	10.8	9.5	13.7	41.6	44.7
United States	5.2	9.6	13.9	18.0	41.6	44.7
TOTAL	10.7	14.9	16.2	20.1	34.4	37.8

[a]More approximate data.
[b]Figures calculated according to a sliding ratio (see text); real figures are not available for most of the countries, and could in many cases be different.
[c]Australia, Canada, New Zealand and United States.
Sources: Level of urbanisation according to the 5,000 criterion elaborated by Bairoch (1985), see Appendix A (Point A1).
Notes: The degree of rounding of the figures does not imply a correspondingly low margin error.
Figures in parentheses have a higher margin of error than other figures for the same periods.

This means that around 1850 already seven countries, one-third of all developed countries, had been affected by the process of urbanisation, these countries representing 54 per cent of the total population of the developed countries.

In the first half of the 19th century, three former commercial powers, the Netherlands, Portugal and Denmark, saw their level of urbanisation decrease. It is not impossible that given the margin of error that this slight de-urbanisation also affected Spain, Italy and several colonial regions of the Ottoman Empire.

Between 1850 and 1880 the process of urbanisation, associated with the upheavals of the 19th century, began in the rest of the developed world with the exception of Serbia and Portugal where it began only after World War I. In most cases, the beginning of this process corresponded approximately to the period of economic take-off. It is worth mentioning the case of Denmark whose delay in urbanisation was due to the fact that its economic development was based until 1860–1870 on the export of raw agricultural products. It was only after this period that processed agricultural predominated; and furthermore in 1890 the development of some industries began.

As will be seen in the next section, in spite of the changes brought about by the process of economic development, the level of urbanisation in the various countries before these changes subsequently influenced the urban structure of the developed world and especially of Europe for a long time. These

highly urbanised countries at the beginning of the 19th century which started their economic development late had a level of urbanisation that was not reached by countries less urbanised in 1800, but which had grown economically much more and had started much earlier. Thus, for example, the Scandinavian countries, more developed and industrialised than the Mediterranean countries, by 1910 had lower levels of urbanisation, although the spread had narrowed (see Table 5).

The relative weight of Russia, little urbanised at the beginning of the century and with a late economic development, significantly influences the level of urbanisation in Europe. Furthermore, the difference between Europe with and Europe without Russia increased during the period studied. Thus, by 1910, this difference reached nine percentage points, with Europe with and Europe without Russia having a urbanisation of 32.9 per cent and 41.9 per cent respectively (a relative difference of 28 per cent). Conversely the precociousness of the industrial revolution in the United Kingdom meant that, in spite of its relatively modest size, it had a strong influence on the level of urbanisation of Europe. In this case, the most important effect occurs around 1830 when the United Kingdom alone increased the European level of urbanisation from 10.4 per cent to 12.6 per cent.

The largest migration in history, that of some 46 million Europeans leaving the continent between 1846 and 1914 to essentially the non-European developed countries (Bairoch, 1976, p. 111–122) was also not with out effect on the process of urbanisation. This movement is concentrated particularly in the period 1865–1914 when 36–39 million Europeans emigrated overseas (of which about one-half returned). These migrations were a factor accelerating urbanisation in the receiving countries and a restraining factor in the home countries. If, as we will see in the next section, the econometric analysis does not allow us to assign an important role to migration for Europan countries, it is very likely that in the absence of this migration the process of urbanisation would have been even more rapid. This is because in most of the European countries the proportion of peasants among the immigrants was higher than the national average.[6] These peasants

Table 5

Level of urbanisation by region (percentage of total population living in cities of 5,000 or more inhabitants)

	1800	1830	1850	1880	1900	1910
EUROPE TOTAL	10.9	12.6	16.4	23.5	30.4	32.8
Nordic countries[a]	7.6	7.5	7.9	13.9	21.1	23.4
Mediterranea countries[b]	17.2	17.1	20.1	25.3	31.8	35.3
Early industrialised[c]	14.8	20.5	28.0	40.7	51.1	54.2
Later industrialised[d]	8.9	9.3	13.4	24.0	35.2	40.4
Europe less United Kingdom	10.0	10.4	13.7	19.8	26.2	28.9
Europe less Russia	12.6	14.9	19.6	29.0	38.1	41.9
Europe less United Kingdom & Russia	11.9	12.9	16.5	24.5	33.3	37.4
OTHER DEVELOPED[e]	5.5	7.9	13.9	24.4	35.6	41.6
TOTAL	10.7	12.3	16.2	23.6	31.3	34.5

[a]Denmark, Finland, Norway, Sweden.
[b]Greece, Italy, Portugal, Serbia, Spain.
[c]Belgium, France, Switzerland, United Kingdom.
[d]Austria-Hungary, Germany, Netherlands.
[e]Australia, Canada, New Zealand, United States.
Sources: See Table 3.

[6]The data in the area are far from being perfect, mostly because of the use of different criteria in the collection of statistics for emigrants and for censuses. In addition, these criteria vary from one country to another. As a general rule the rural emigrants are over-represented for Western Europe and for Northern Europe while for the countries of Eastern Europe and for Mediterranean countries this tendency is less clear. For statistics in this area, see especially Ferenczi, I. and Wilcox, W. (eds.) *International Migrations, Vol. I: Statistics*, New York, 1929, especially pp. 334–337 and Thomas, B. *Migration and Economic Growth*, Cambridge, 1954, especially pp. 60–62. For the data on the structure of the population by occupation, see Bairoch, P. (under the supervision of), Deldycke, T., Gelders, H. and Limbor, J.-M. *The Working Population and its Structure* Brussels and New York, 1968.

were as strongly attracted by the economic possibilities of the overseas countries as they were repulsed from the rural areas because of problems caused by increases in population density. In the absence of the possibility of migration, a noticeable proportion of these potential emigrants would probably have installed themselves in European cities. Without considering it as definitive proof, it is clear that the level of urbanisation in Europe increased between 1830 and 1860 by 1.3 per cent per year and from 1860 to 1910 by 1.1 per cent per year. Of course, this latter period is also one of reduced economic growth: the years 1872–1892 are described as the 'great European depression'; but globally from 1860 to 1910 economic growth was close to and even slightly more rapid than in 1830–1860 (1.0 per cent relative to 0.9 per cent per year per inhabitant).

Within the large population growth in the overseas developed countries (a growth of 1,900 per cent between 1800 and 1910), the relative impact of immigration was greatest between 1846 and 1860. For example, the net increase in the total population of the USA over the decades 1840–1860 was 10 per cent of the total population at the beginning of these decades as opposed to 8 per cent for later decades. The period from 1840–1860 is also that of the fastest growth in urban population. The econometric study in the following part confirms the importance of immigration in the overseas countries. This remains valid even if the statistical analyses of Gallaway and Vedder (1971) have completely discredited the idea that the new immigrants who arrived after 1890, largely from southern and eastern Europe, were more inclined to install themselves in cities than the immigrants who preceded them. It remains true that the 'foreign born' were generally more likely to live in the cities than were the 'native born'. This was natural behaviour since even in new countries a newcomer is doubly new in a rural milieu.

Because of this rapid urbanisation in the overseas developed countries, after 1880 Europe loses its place as the most urbanised region of the developed

Table 6

Urban population and level of urbanisation

	Urban Population and Urbanisation level			Yearly Growth Rate		
	Europe	Other develop. count.[a]	Total	Europe	Other develop. count.[a]	Total
URBAN POPULATION (in millions)						
1800	22.4	0.3	22.7	–	–	–
1830	30.4	1.1	31.5	1.0	4.3	1.1
1840	36.8	1.9	38.6	1.9	5.5	2.1
1850	45.0	3.7	48.8	2.0	7.2	2.4
1860	55.2	6.3	61.5	2.0	5.4	2.3
1870	66.8	9.1	76.9	1.9	4.8	2.3
1880	81.6	14.1	95.7	2.0	3.4	2.2
1890	100.8	22.3	123.1	2.1	4.7	2.5
1900	125.9	30.6	156.5	2.2	3.2	2.4
1910	152.5	43.7	196.2	1.9	3.6	2.3
LEVEL OF URBANISATION (%)						
1800	10.9	5.5	10.7	–	–	–
1830	12.6	7.9	12.3	0.5	1.2	0.5
1840	14.3	9.8	14.0	1.3	2.3	1.3
1850	16.4	13.9	16.2	1.4	3.6	1.5
1860	18.8	17.5	18.6	1.4	2.3	1.4
1870	20.9	22.2	21.0	1.1	2.4	1.2
1880	23.5	24.4	23.6	1.2	1.0	1.2
1890	26.6	31.1	27.4	1.3	2.4	1.5
1900	30.4	35.6	31.3	1.3	1.3	1.3
1910	32.8	41.6	34.4	0.8	1.6	1.0

[a]Australia, Canada, New Zealand and United States.

Source: See Table 3.

Notes: The growth rates have been calculated on less rounded off figures.
The degree of rounding off of the figures does not imply a correspondingly low margin of error.

world. These countries, especially North America, contain an increasing proportion of the developed countries' city-dwellers. Around 1800, only 0.1 per cent of the city-dwellers lived in the overseas developed countries but this increased to 10.2 per cent in 1860 and 22.4 per cent in 1910. Within this group of countries, the USA alone is responsible for 90 per cent of the urban population.

Different starting dates for development combine with varying rates of demographic and urban growth to create a growth curve of urban population for the developed countries that is quite regular from 1830 to 1910 (Table 6). The total urban population increased between 1830 and 1910 by an average rate of 2.3 per cent, with the lowest rate being 2.1 per cent (1830–40) and the highest being 2.5 per cent (1880–90). The level of urbanisation increased on average by 1.3 per cent per year — the lowest growth rate being 1.0 per cent (1900–10), the highest being 1.5 per cent (1840–50, 1880–90).

In the Introduction we have already emphasised the unique character of urban development in the nineteenth century, not only with regard to previous centuries but also with regard to the twentieth century. It is true that the existence of an absolute limit (100 per cent) to the level of urbanisation implies automatically a reduction in the possibility for urban expansion for recent decades. With regard to the nineteenth century, as we have seen at the beginning of this section, it is appropriate to note that the criterion used here (5,000 inhabitants) in order to determine the urban population is better adapted than the 2000 inhabitants criterion to the middle and the end of the century than to the beginning. This implies that the 'real' growth of urban population was slower than that calculated using the 5,000 inhabitants criterion. Table 7 presents a tentative estimation of the probable real level of urbanisation. Based on factors which are, in part, arbitrary but based on knowledge of the data in the area, we have estimated the real urban population in 1800 as that calculated using the 5,000 inhabitants criterion to which we have added 100 per cent of the population living in cities or administrative units of 2,000–5,000 inhabitants. For 1830, 90 per cent of this second group has been added, while for 1910 20

Table 7

Alternative definitions of the level of urbanisation for the developed countries (excluding Japan and South Africa)

	Urban Population and Urbanisation Level			Yearly Growth Rate		
	5000 criter.	2000 criter.	'Real level'	5000 criter.	2000 criter.	'Real level'
URBAN POPULATION (in millions)						
1500	8.5	13.0	13.4	–	–	–
1700	10.5	14.6	14.8	0.11	0.06	0.05
1800	22.7	31.5	31.6	0.77	0.78	0.76
1830	31.5	42.0	40.9	1.09	0.96	0.86
1850	48.6	60.8	56.2	2.21	1.87	1.60
1880	95.7	110.8	101.0	2.27	2.02	1.97
1900	156.5	173.7	160.7	2.49	2.28	2.35
1910	196.2	215.2	200.0	2.29	2.16	2.21
1980	659.0	682.0	658.7	1.75	1.66	1.72
LEVEL OF URBANISATION (%)						
1500	8.0	12.2	12.6	–	–	–
1700	10.5	14.6	14.8	0.13	0.09	0.08
1800	10.7	14.9	14.9	0.02	0.02	0.01
1830	12.3	16.4	16.0	0.46	0.32	0.24
1850	16.2	20.1	18.6	1.36	1.02	0.75
1880	23.6	27.4	24.9	1.28	1.03	0.98
1900	31.3	34.7	32.1	1.41	1.20	1.27
1910	34.4	37.8	35.1	0.97	0.85	0.90
1980	66.4	68.7	66.4	0.94	0.85	0.91

Source: 1800–1910: see Table 3; other data: adapted from Bairoch, P., *De Jéricho à Mexico: Villes et économie dans l'histoire*, 1985.

Note: The degree of rounding off of the figures dos not imply a correspondingly low margin of error.

per cent is added (the other values are calculated by linear interpolation). For 1500 and 1700 we have considered this limit of 2,000 inhabitants too high and therefore have added for 1500 110 per cent of the urban population of agglomerations of 2,000–5,000 inhabitants and 105 per cent for 1700.

Based on these corrections (see Table 7), one can estimate that between 1830 and 1910 the real growth of urban population was on the order of 2.0 per cent per year rather than the 2.3 per cent per year derived from the 5,000 inhabitants criterion. Likewise, the level of urbanisation probably increased by 1.0 per cent per year rather than 1.3 per cent. However, these rates constitute a fact without precedent in the history of occidental societies. Here as well as in almost every social and economic aspect, the industrial revolution constitutes a brutal rupture; a rupture that resulted in a rapid acceleration of some pre-existing trends. From 1500 to 1830, the level of urbanisation of western countries increased by 0.06–0.08 per cent while from 1830 to 1910 it increased by 0.95–1.05 per cent, a rhythm 13–17 times more rapid. It is in order better to understand this veritable 'urban revolution' that we have tried to determine, through an econometric approach, some of the principal factors which might explain this explosion in urban population.

3. An Econometric Analysis

In the previous section, we described the growth of the urban population of Europe and the rest of the developed world. Of course, at the same time that these countries were urbanising they were growing rapidly economically and demographically. These phenomena are clearly closely related. In ths section, we will try to determine which factors were most closely related to this increase in urban population.

We will include a variety of factors (some 14 all together) that we feel are important in explaining the level of urbanisation of a country in the nineteenth century, some of which are general factors such as GNP per capita or the level of urbanisation in 1800, some that are specific economic factors such as the level of industrialisation, trade and agricultural productivity, some that are demographic such as migration and total population, some that are geographic such as the density of population and the topography, and some that are social/economic such as the centralisation of industry. Of course, not all these factors are of equal importance and in fact one of the goals of this study is to determine their general relative importance although exact estimates are impossible due to the multicollinearity of the factors. For a brief description of these data and their sources, see Appendix A.

Another aspect of this problem is the changing relative importance of the factors as well as the changing values of parameters estimated over the 80 year period (1830–1910). It can be hypothesised that certain factors are more important when countries are less urbanised in contributing to urbanisation than later when a country is more developed. To examine these possible changes separate regression analyses are developed for the early and late nineteenth century.

At the same time that we hypothesise about the changing relative importance of various factors over time, we are assuming that over space they do not change; that idiosyncratic characteristics of a state have minor importance. Since we include data for approximately 20 countries, but have no country specific variables in our models, we have made the rather strong assumption that in addition to the same factors being important in each country, the effect of each factor is also the same across countries as different as Belgium and Bulgaria, that the increase in, for example, GNP per capita of one unit results in the same increase in the level of urbanisation in all countries; this being a standard assumption made in cross-sectional studies. It is an assumption that we will implicitly test with regard to European countries, and more explicitly test with regard to the difference between Europe and European-settled countries.

The method that we have used in order to examine the success with which we are able to explain levels of urbanisation is ordinary least squares, with an adjustment for serial correlation. No cross-sectional adjustments were made since they were not found to contribute very much to improving the model (see Appendix B for further details). In the best of all possible worlds, we would have used orthogonal independent variables and hence be able to state unambigously whether a factor is important or not, and what is its importance relative to other factors. However, since the correlations between economic factors are high, we are unable to attribute unambiguously influence although estimates remain unbiased. Table 8 gives the intercorrelations between all the real value variables, which can be

Table 8

Correlation matrix (for European countries)

	GNP	IND	EXPO	AGRI	CERE	MIGR	TPOP	RAIL	GOVE	URBL
GNP/CAP	1.00									
INDUST	0.58	1.00								
EXPORT	0.61	0.02	1.00							
AGR PROD	0.70	0.45	0.46	1.00						
CEREAL	−0.65	−0.17	−0.54	−0.18	1.00					
MIGRATI	−0.08	0.15	−0.26	−0.05	0.24	1.00				
TOT POP	0.09	0.64	−0.23	0.09	0.13	0.19	1.00			
RAILR	0.23	0.01	0.19	0.15	−0.25	0.27	−0.01	1.00		
GOVERN	0.02	−0.07	0.06	−0.02	−0.21	−0.04	−0.50	−0.09	1.00	
URB LEV	0.82	0.59	0.42	0.59	−0.55	0.07	0.15	−0.05	0.45	1.00

Number of observations: 127
GNP/CAP —gross national product per capita
INDUST —level of industrialisation
EXPORT —exports as a percentage of GNP
AGR PROD—agricultural productivity
CEREAL —cereal production as percentage of domestic consumption
MIGRAT —migrations as percentage of total population
RAILR —railroads
GOVERN —importance of central government
URB LEV —level of urban population.

divided basically into two groups: 1. the economic variables, e.g. GNP per capita, agricultural productivity, 2. demographic variables, e.g. urban population, total population, migration. As one can see, GNP per capita, the level of industrialisation, exports and agricultural productivity are all highly correlated with one another, with agricultural productivity being very highly correlated with GNP per capita (0.70). The correlation between the demographic variables themselves and these variables and the economic ones are relatively small, except for urban population. There are quite a few interesting aspects to this table which do not have much to do with levels of urbanisation, but generally the correlations seem plausible and accord with common sense. Apart from the high correlation between the economic variables, another quite high correlation is the positive one between the level of industrialisation and total population. The rest of the correlations are generally relatively modest.

One of our concerns in this study is to try to establish the relative importance of factors which might explain the urbanisation of a society. The problem with economic factors is that they all tend to grow together. Of the factors we consider, GNP per capita is a general measure of the wealth of a country and is clearly the most important factor in determining the level of urbanisation. Any attempt to determine the importance of, say, agricultural productivity is rendered difficult because of its high correlation with GNP per capita.

We can suppose, however, that GNP per capita itself is in large measure determined by other more specific economic factors, such as agricultural productivity or the level of industrialisation. Thus, we try to deal with this problem by regressing our important economic factors, level of industrialisation, agricultural productivity and exports, on GNP per capita ($r^2 = 0.74$). We then calculate the regression residuals which represent the wealth of a nation not explained by these three factors and is not correlated with them. The results given in Table 9 for GNP per capita are for this residual GNP per capita and in general when we refer to GNP per capita we mean this GNP per capita residual.

Table 9 presents the results of a regression with all variables both for Europe and all developed countries (i.e. adding the USA, New Zealand, Australia and Canada). The percentage of the variation explained for Europe is 94 per cent and if we include the European settled countries, it is 87 per cent. Thus, especially when considering that between 19 and 23 different countries are included, we feel that we have included most of the important factors. This is even more astonishing since, as will be recalled, we made rather strong assumptions about the cross-national stability of the various factors. A regression run with the 2,000 inhabitant level crite-

rion must give the same result since the 2,000 level data are a linear function of the 5,000 level data (see note 4).

Looking at the various individual factors we find the economic ones (GNP per capita not explained by other economic factors, industrialisation, exports and agricultural productivity) explain the major share of the variation in the levels of urbanisation. If these four factors are used alone, they account for 70–80 per cent of the variance in the level of urbanisation. This can also be seen by comparing the estimated coefficients in Table 9 which are given in standardised form (beta coefficients).

Table 9

Regression analyses: Coefficient with level of urbanisation

Variable	Estimated coefficient	Standard error	T-statistic
EUROPE			
GNP/capita#	0.21	0.037	5.66***
Agr. prod.	0.23	0.102	2.30*
Indust.	0.99	0.114	8.63***
Exports	0.43	0.072	5.86***
Cereals	0.044	0.047	0.92
Migration	0.026	0.021	1.24
Tot pop.	−0.20	0.05	−4.02**
Ini. urb. level	0.61	0.084	7.33***
Pop. density	−0.049	0.07	−0.69
Start of modern	0.22	0.10	2.27*
Form of indust. con.	0.32	0.09	3.50**
Form of indust. aver.	0.32	0.067	4.67***
Topogr. mod. flat	0.04	0.056	0.65
Topogr. hilly	−0.21	0.071	2.89**
Railroads	−0.002	0.0012	−1.76
Form of gov't	−0.019	0.065	−0.29
r^2 0.94			
Number of observations 127			
Durbin-Watson 1.76			
*=p<0.05. **=p<0.01. ***=p<0.001.			
ALL DEVELOPED COUNTRIES			
GNP/capita#	0.15	0.024	6.60***
Agr. prod.	0.35	0.063	5.55***
Indust.	0.30	0.049	6.22***
Exports	0.23	0.052	4.39**
Cereals	−0.12	0.047	−2.62**
Migration	−0.046	0.019	−2.45*
Tot. pop.	−0.011	0.059	−0.18
Ini. urb. level	0.59	0.08	7.93***
Start of modern	−0.10	0.074	−1.39
Pop. density	−0.038	0.078	−0.50
Form of indust. con.	0.075	0.061	1.19
Form of indust. aver.	0.097	0.043	1.68
Topogr. mod. flat	0.21	0.062	3.33**
Topogr. hilly	0.044	0.054	0.81
r^2 0.88			
Number of observations 147			
Durbin-Watson 1.71			
*p<0.05. **=p<0.01. ***=p<0.001.			

#Note: 1. All coefficients are 'Beta coefficients', i.e. standardised coefficients.

2. 'p<0.05' indicates that the probability of getting this value if there is no relationship between the variable and the level of urbanisation is less than 5 in 100.

3. There is no constant in the regression because standardisation variables were used.

4. GNP/capita is residual GNP that is not explained by the level of industrialisation, agricultural productivity, and exports, see p. 13.

Of the four factors, the level of industrialisation is clearly the most important, with a coefficient of 0.99, followed by the level of exports, with a coefficient of 0.43. Agricultural productivity and GNP per capita residual follow with coefficients of 0.23 and 0.21 respectively. Clearly it is the industrialisation of these countries which drew a large part of the population to European cities. In another study (Bairoch, 1985) it was calculated that during the nineteenth century industrial employment in European cities had been multiplied by between nine and ten times, bringing the share of industrial employment in total employment from 35–45 per cent to 50–55 per cent. International trade also constituted an important force attracting people to the cities, a factor, as we have already seen, that was already important in earlier centuries. These forces were relatively independent, since generally the large port cities were not centres of manufacturing, e.g. London, Rotterdam, Antwerp, Copenhagen, etc. This is also revealed by the very low correlation between the level of industrialisation and the level of exports (Table 8), although they are both highly related to GNP per capita.

When one includes the European settled countries agricultural productivity remains significant and its importance increases. That this factor becomes more important when including the European-settled countries is not surprising since these countries had a much higher level of agricultural productivity than did the European countries. One might interpret these results by saying that for the European settled countries urbanisation was 'pushed' by increases in agricultural productivity and that in Europe it was more 'pulled' by industrialisation.

The same regression as in Table 9 was performed for Europe with GNP per capita instead of the residual GNP per capita. The results were that the other three economic factors decreased in importance, and GNP per capita becomes the most important factor ($B=1.07$, $p<0.001$). The level of

industrialisation ($B = 0.57$, $p < 0.001$) is now the next most important economic factor, and exports ($B = 0.16$, $p < 0.01$) and agricultural productivity ($B = -0.25$, $p < 0.05$) and follow in importance. The negative coefficient for agricultural productivity is certainly related to the fact that this factor is very highly correlated with GNP per capita. One possible interpretation is that successful agriculture also retains people in rural areas. Denmark is an example of a country that, given its GNP per capita, had a lower level of urbanisation than other countries as we have seen. The fact that this inverse relationship is obtained when comparing 19 countries may indicate that this is a more general phenomenon than might be imagined. This is largely the case in the Third World today where one of the factors accelerating urbanisation has been by the failure of agriculture. Otherwise, the global R-squared and the other coefficients must remain the same due to the procedure used.

Turning to the demographic variables, total population, population density before modernisation and migration, we find that total population is inversely related to the level of urbanisation. The total population of a country is one measure of the size of the country. Thus, it is not surprising that the level of urbanisation is inversely related to the size of the country, since smaller countries tend to be more urbanised. This should not be confused with the density of population before modernisation which is marginally ($B = -0.27$, $p < 0.01$) inversely correlated with the level of urbanisation and is not related to the size of the country.

Migration is not statistically related to levels of urbanisation in Europe, but it is related to them in the European-settled countries. This is quite reasonable since Europeans generally migrated from the countryside, but entered their future homes through cities where many of them remained.

Turning to geographic variables, the form of industrialisation (concentrated to dispersed) and the topographical character of the country, we find that the more concentrated a country's industry the higher the level of urbanisation. Countries that concentrate their industry in a few places have higher levels of urbanisation than those which have their industry dispersed throughout the country. This factor is relatively important (coefficient 0.54) when compared to the economic factors (industrialisation coefficient 0.57).

We found that the topography of the country plays a role in the level of urbanisation. In Europe, mountainous countries had a higher level of urbanisation, when all other factors are taken into account. This is contrary to what our intuition would lead us to expect, since we imagine mountains as barriers to the movement of people. The countries that fell into this category (topographically hilly) were generally small countries with a large percentage of mountainous land. This may be a case of spurious correlation since these countries are also small and smaller countries are more urbanised, in spite of the fact that we have included the total population of a country as a factor.

The other variables we included were the level of urbanisation in 1800 and the starting date of modernisation. With the level of urbanisation in 1800, we tried to measure the effect of the particular history of a country and in a sense to standardise the model (see the Appendix for a brief description of our use of a lagged dependent variable as an independent variable). This factor is quite important in determining the level of urbanisation of a society. We find that countries with high levels of urbanisation remain 'over-urbanised' for their economic and demographic position and likewise for those countries that started the century with small urban populations. This is also evidence for the irreversible nature of urbanisation and the difficulties of 'de-urbanisation'. Cities exercise a definite attraction in terms of life-style, as evidenced by the very popular World War I American song 'How can you keep them home on the farm once they've seen gay Paris'.

We also find that those countries that began their modernisation later have higher levels of urbanisation that did the early industrialising countries. We also found that neither the level of cereal imports, nor the importance of the central government were related to the level of urbanisation.

In spite of a widely held belief that the extension the railroad system significantly contributed to urbanisation we found this variable had no significant effect on the level of urbanisation. There are several possible explanations for this, of which the most prominent is the definition of this factor. The problem consists in trying to construct some comparative measure of the size of the railroad network that takes into account factors of size, population topography etc. We did this with a regression technique (see Appendix, A11). This seemed to us a plausible approach and the results were not aberrant

to the naked eye, but they remain disputable. Our measure for railroads is not, however, a measure of the length of the railroad network *per se*, but a measure of the extent to which a state had a more or less developed railroad network for a certain period given its population, topography, population density and form of industrialisation. That this was not important does not mean that railroads themselves were not important but that whether the railroad network was over- or under-developed seemed to have little effect. One might suggest that the often noted relation of railroads to urbanisation may well be spurious, both of them being a function of industrialisation and economic growth. Thus, once other factors such as industrialisation, economic growth and population are included in the analysis, the effect of railroads would be quite small.

Comparing the results of the regressions for Europe with those for all developed countries (adding Australia, Canada, New Zealand and the USA), we find that the addition of the European-settled countries changes significantly the picture from that of Europe alone. The main differences lie in the greater effect of agricultural productivity, the importance of cereal production, the lesser importance of industrialisation, and the effects of immigration. A formal test (Chow, $p < 0.001$) of whether these four countries belong to the European regression indicates that they do not. It is clear that the factors and mechanisms leading to the urbanisation of the European-settled countries are quite different from those operating in European context. Data for those parameters that vary radically are those for which the European settled countries represent outliers with regard to Europe, e.g. agricultural productivity, cereal production, migration. Thus, in many ways it makes no sense to estimate the equation with the European settled countries, together with the European countries since the results are a mix which fits neither subpopulation well. Unfortunately, we did not have enough data to make estimates for the European-settled countries separately. Therefore, our assumption that there are no cross-national differences now has to be revised, since the European-settled countries present some significant differences from the European situation.

A major constraint on the size of a city is the availability of sufficient food resources and capacity to move them cheaply to urban areas. As noted above, an over- or under-developed railroad system was not related to levels of urbanisation. One could argue that it is not for the movement of people that railroads are important, but for the transport of foodstuffs. By this argument, a better measure of the importance of railroads would perhaps be the price per ton for freight. As for cereal production itself, a simple correlation finds it quite strongly negatively correlated with levels of urbanisation. In the complete model it is not statistically significant for Europe but is significant for the European-settled centuries. One argument could be that cereal production as a percentage of domestic consumption is a function of agricultural productivity, but Table 8 indicates they are not highly correlated for Europe. The place of this nexus of factors, agricultural productivity, cereal production and transportation remains to be explained, but it seems to be one of the major areas of difference between Europe and the European-settled centuries.

As we mentioned earlier, one of our interests was in the changing importance of various factors over time. We supposed that the demographic and geographic factors would be more important in earlier periods and that economic ones would be more important later. To examine this we divided the period 1830–1910 into two parts, 1830–1860 and 1870–1910, analyzing only the data for Europe.

In the period 1830–1860 none of the economic variables are significant (GNP per capita residual, $p < 0.1$), with the exception of the level of agricultural productivity ($B = 0.3$, $p < 0.05$) and the level of exports ($B = 2.25$, $p < 0.05$), which are positively related to the level of urbanisation. These two factors represent two important aspects influencing the level of urbanisation in the early to middle nineteenth century as have already seen with regard to trade in the first section. This is a period when increases in agricultural productivity began to be felt. Also those countries with a long tradition of commerce, e.g. the Netherlands and Italy, had generally higher levels of urbanisation than did other countries in the early nineteenth century. In the second period agricultural productivity remained significant, but GNP per capita and industrialisation became much more important. Thus, it seems that in the early stages of urbanisation, agricultural productivity and commerce played a much stronger role than later when industry and other forms of wealth became increasingly important.

Another important variable for the first period is

an average topography, those countries not flat, not mountainous being slightly less urbanised. None of the demographic variables were significant, including the size of the country. With the large number of insignificant estimates, the R-squared remained at a modest 0.72.

Looking at the parameter estimates for the period 1870–1910, they resemble very strongly the results for the whole period. The variables that are important for the whole period are also important in this period and their relative importance remains unchanged. An interesting non-change was in the importance of the level of urbanisation in 1800. It was as important in the second period as in the first, indicating that the past history of the country continued to play an important role in determining the level of urbanisation throughout the century.

A formal test (Chow $p < 0.001$) confirms what is already clear, that the earlier period has a different set of parameters from the one which followed. It is the economic factors which become more important and we see a drift away from trade and agricultural productivity to the benefit of industrialisation in determining the level of urbanisation of a society. This however remains conditioned by history, geography and demography.

In order to compare how well this regression line fits various countries, we calculated the average absolute distance from the actual data to the regression line. There appeared to be no pattern in these results, either by region, wealth, or size (measured by near zero correlations), indicating that the regression is not overtly biased towards one category of country. One common factor of the worst fitting cases is that they are poorer peripheral countries of Europe, e.g. Spain, Greece and Bulgaria, however, for some of these peripheral countries the fit is quite good, e.g. Russia, Portugal, Finland. In conclusion, we see no obvious factors that characterise the poor, average and good fitting countries, hence no factor which might at first glance improve the fit for the worst cases. The model seems to fit well for Europe as a whole and is not obviously biased in one direction or another.

4. Conclusion

We now summarise briefly the principal conclusions of our analyses of the urbanisation of developed countries in the nineteenth century.

The Patterns of Urbanisation in the 19th Century

The level of urbanisation of the future developed countries was not really influenced by economic modernisation until 1830–40, except for the United Kingdom. The point of departure (1800) is characterised by a level of urbanisation around 12 per cent, with a rather large spread between individual countries going from a minimum of 3–4 per cent to a maximum of 35–40 per cent for the 5,000 inhabitants criterion and from 8–9 per cent to 38–42 per cent for the 2,000 inhabitants criterion.

The levels of urbanisation in various countries around 1830 are as much a function of the geographic and economic structure of the moment as they are of economic functions (industrial and commercial) which have not existed for a long time or are strongly attenuated. Particularly, trade and agricultural productivity continued to play an important role in the early and middle nineteenth century. In spite of the upheavals of the nineteenth century, the past continues to have a strong influence throughout the nineteenth century.

The years 1830–40 to 1914 are not only the period of the most rapid expansion ever recorded, but also the period when the present urban structures were almost entirely put into place.

The rate of growth of urban population was about 2.1–2.3 per cent per year (depending on the criterion used) and each year saw an increase in the level of urbanisation of about 1.1–1.3 per cent.

Since, in general, the less urbanised countries at the beginning of the nineteenth century began their economic development earlier, the spread in the levels of urbanisation grew only moderately during the century.

The combination of different economic take-off dates with a variety of rates of urban and demographic growth permitted from 1830–1914 a quite regular growth rate of the level of urbanisation of the developed countries as a whole.

What are the Factors that Might Explain this Growth?

It is clear that it was economic growth that pushed the process of urbanisation in the developed countries. These factors accounted for 70–80 per cent of

the growth in the level of urbanisation. Of the economic factors industrialisation was the most important force drawing the rural population to the cities. Also exports, residential GNP per capita and agricultural productivity (in decreasing order) were important in promoting urbanisation. The case of increasing agricultural productivity is more ambiguous. In the non-European countries, it was much more important than in Europe in promoting urbanisation. In Europe, successful agriculture may have also contributed to retaining population in rural areas. In Europe, urbanisation was pulled more by industrialisation and in the non-European developed countries it was more pushed by increasing agricultural productivity.

The level of urbanisation is inversely related to the size of the country as measured by its total population. Countries with concentrated industries were also more likely to have a higher level of urbanisation than countries with dispersed industrialisation.

Those countries that started their modernisation later were found to have relatively higher levels of urbanisation than those which began earlier.

The importance of the central government was found not to be significant nor was the population density strongly associated with the level of urbanisation. An over- or under-developed railroad network was also found not to be important, but this is a more tentative finding due to the difficulty of constructing a good comparative indicator.

In the early to middle nineteenth century, it was the previous history of the country together with its level of exports and its agricultural productivity that basically determined the level of urbanisation.

The non-European developed countries, Australia, Canada, New Zealand and the USA present a quite different situation from Europe, especially with regard to the increased importance of agricultural productivity and migrations. They deserve detailed future study.

Appendix A

Description of the data

(A1) Urban Population

In this study, we have mainly used as the definition of urban population the share of the total population living in cities (or other administrative agglomerations) of 5,000 inhabitants or more. Before describing the methods used to gather these data, it is appropriate that we justify this criterion, which is partially arbitrary, but which is finally the only possible operational one.

If, in theoretical terms one can establish a valid definition of urban population, practically the statistical definition of a large geographical group presents numerous problems and only very approximate solutions. Theoretically, one can consider that there is a consensus on the three criteria that together define an urban agglomeration (at least for periods before the twentieth century):

(1) size of the agglomeration
(2) the density of population
(3) the dominance of non-agricultural activities

In practice, the problem is to define the parameter for each of these criteria; this in relation to the availability of relevant statistics for geographical areas. This is why most censuses use only the notion of size. This is all the more justified because there is in general an excellent relation between size and density (Guest, 1973; Best et al., 1974; Bairoch, 1977). There is, as well, a relation between the proportion of the population engaged in agriculture living in the agglomeration and the size of cities. Of course, with the development of public transportation and the automobile, we have seen the explosion of cities, which has certainly modified the problematic, inspite of which, these definitions of a city remain generally unchanged.

The notion of size remains the most often used, but the limits vary radically.[7] The criteria used in the

[7]Here are the definitions used by European countries for the census of 1960:

Austria	5,000	Greece	10,000	Romania	A
Belgium	5,000	Hungary	A	Spain	10,000
Bulgaria	A	Italy	E	Sweden	200 + B
Denmark	200 + B	Netherlands	5,000 + E	Switzerland	10,000
Finland	200 + B	Norway	2,000 + B	United Kingdom	A
France	2,000 + B	Portugal	2,000	Yugoslavia	E mobile
Germany	2,000				

A: administrative criterion; E: employment criterion (percentage of the population not involved in agriculture); B: urban criterion (usually distance between houses). Taken from United Nations *La croissance de la population mondiale urbaine et rurale, 1930–2000*, New York, 1970, pp. 89–90.

nineteenth century are not significantly different from those used in the middle of the twentieth, with the exception of the less frequent use of other criteria than size. In many cases, the agglomerations are divided by size without explicitly considering the larger categories as urban. Generally, these limits are between 2,000 and 10,000 inhabitants, with the average being 6,700.[8] In addition, in many cases the definition varies over time, however, these changes are often not related to socio-economic realities.

Comparative studies almost always choose a single criterion to define urban population — the least bad solution — which is what we have done. The limit of 5,000 inhabitants that we have chosen is the result of two considerations: (1) it is close to the average of those used in the nineteenth century, (2) it is close to the final results when multiple criteria are used.[9] If this limit is valid starting in the years 1840–50 with a decreasing under-estimating bias as one moves towards 1910, it is clear that for the beginning of the nineteenth century 2,000 would be better. For certain analyses we have also used this criterion as well as a moving criterion, but this only for large regions.

The methods used to estimate these data have varied according to the period. For 1800, and for a few cases in 1830 and 1850, the data were calculated from a data bank on the population of the cities of developed countries between 1800 and 1850. This data bank includes all cities (to date more than 1,600) that at one time or another had more than 5,000 inhabitants. We have estimated that this data bank is practically complete for all cities above 20,000 inhabitants in the nineteenth century. For smaller cities we have estimated the population using 'Davis' Law' (Davis, 1969 and 1972), but with the parameters adjusted for the nineteenth century (Bairoch, 1985). For certain countries, notably Germany and Switzerland, for which our data bank was more complete for smaller cities, we have used these data to calculate the level of urbanisation.

For the rest of the nineteenth century, data were based on various national censuses.

(A2) GNP (total and per capita)

This series consists of data for gross national product and market prices expressed in 1960 US dollars and prices. These data have been elaborated by one of the authors and partially published elsewhere (Bairoch, 1976; Bairoch, 1980). The following presents very schematically (see the above cited articles for further details) the procedure used. The correction (base year 1960) of the data for purchasing power parity is a result of a synthesis of various methods currently used. To these adjusted figures, the most recent data on historical rates of growth have been applied. In addition, for 1928 as well as 1900 the data have been adjusted based on an analysis of a dozen indirect indicators of GNP, e.g. mortality rates, the proportion of the labour force in agriculture, the consumption of various food products, letters sent, etc. For those countries for which there exists no retrospective calculations of GNP, we have made our own estimations based on the volume of agricultural and industrial production.

(A3) Agricultural Productivity

This is measured by agricultural production as expressed in millions of direct calories per male working in agriculture. These data for a certain number of periods and countries have already been calculated and published (Bairoch, 1965) elsewhere. However, the data used here are the result of new calculations; in addition to including more countries and more periods, the author used more recent data, extended the range of products considered and

[8]Here are the criteria used for the 1860 or 1870 census for countries that used a size criterion (occasionally used with other criteria).

Austria	2,000	Greece	10,000	Serbia	2,000
Belgium	5,000	Ireland	2,000	Spain	5,000
Bulgaria	10,000	Italy	6,000	Switzerland	10,000
France	2,000	Netherlands	20,000	United States	8,000
Germany	2,000	Portugal	10,000		

From Meuriot, P. *Des agglomerations urbaines dans l'Europe contemporaine*, Paris, 1897, pp. 47–57; Weber, A. *The Growth of Cities in the Nineteenth Century*, New York, 1899, pp. 20–122.

[9]This does not mean that in the case of a definition using multiple criteria that there are no cities with fewer than 5,000 inhabitants and rural units with more than 5,000 inhabitants.

improved the methods of estimation for the production of meat as well as those relating to the estimation of the active male agricultural labour force. These modifications were made in order to render the data more comparable. These data will be subject of a future publication.[10]

(A4) Level of Industrialisation

This series consists of data on the industrial production (manufactures) per inhabitant. As for the preceding series, these data have been published elsewhere (Bairoch, 1982) where the interested reader will find a fuller description of the methodologies used. Broadly, it consists not only of industrial production proper, e.g. textiles, steel, chemicals, cement, etc., but also activities more artisanal such as furniture, food, clothing, etc., are included. The construction industries are excluded.

(A5) Total Population

This is the total population as of the 30th June of the year in question. These data are the result of the collation of the most recent estimations.

(A6) Migration

This series consists of the net migration during the 10 year period expressed as a proportion of the total population of the beginning of the decade in question. These data are essentially those of Sundbarg (1908), with data for 1900–1910 from la Statistique Generale de France (1932), and with data for the USA and several isolated cases provided by one of the authors (see also, Mulhall, 1898).

(A7) Starting Date of Modernisation

This is the approximate date at which the society in question started to be noticeably affected by the process of economic development and not the beginning of this process (usually a difference of 10–20 years).

Four criteria have been used to determine this date. The first is the estimation of one of the authors regarding the beginning of the agricultural and industrial revolution. The second is the date given by Rostow (1962) as that of the 'take off'. The third is the date at which the country attained a certain level of industrial development (Bairoch, 1982) (10 on a scale where the United Kingdom in 1900 equals 100. 7–8 being that of a traditional society). The fourth is the period in which the agricultural productivity began to rise significatly. Finally, for certain countries, exporters of cereals, we have also considered the date of their integration into the system of international trade. Clearly, the final choice is partially arbitrary, but at the same time also rests on the experience of one of the authors on the economic histories of these countries.

(A8) Cereal Imports

This series consists of net importations of all cereals, including flour, expressed as a percentage of the total domestic consumption (including seed and animal consumption). These data are calculations and estimations of one of the authors.

(A9) Population Density

This is population per square kilometre with a correction made for countries with large uninhabitable (or very sparsely inhabited) areas in the extreme north.

(A10) Topography

This variable was an attempt to measure variations in the topography of a country. It consisted in categorising all countries on a three point scale, 1 — flat, 2 — average, 3 — mountainous, based on the study of atlases and encyclopedias. This categorisation was performed independently by the two authors with perfect intercoder agreement. An independent check was performed by P. Guichonnet, University of Geneva.

(A11) Railroads

The railroad mileage of the various countries was taken from various standard sources (Mitchell, 1975; Urquart and Buckley, 1965; US Bureau of the

[10]These data are being finalised and will appear in a forthcoming article on agricultural productivity and agricultural yields for all developed countries in the 19th century.

Census, 1962; Woytinsky, 1927). The problem is how to standardise these data for comparative purposes, taking into account the size and the population of the country. The method used was to regress total population, population density, topography, form of industrialisation and time trend on railroad mileage ($r^2 = 0.78$). This provided a norm as to what a country of a certain size, population, etc. would be 'expected' to have in terms of railroads. The measure then used was the difference between the actual mileage and this 'expected' mileage (the residual). Generally, countries remained above or below the expected level, but there were cases where they crossed from below expected to above and vice versa, e.g. France and Germany. This was not done for all developed countries since the addition of the four non-European countries distorted the regression too much.

(A12) Form of Government

This variable was used to indicate the extent to which a given country had a centralised governmental system. After considering various possible indicators such as the constitutional form of government, the share of the central government budget in GNP, it was decided to simply use the relative size of the capital city (corrected if it was also a port city) as a reflection of the importance of the central government.

(A13) Form of Industrialisation

The countries have been divided according to the geographical concentration of their industries. The first group includes countries where dispersed industries are predominate, that is those industries consisting of relatively small enterprises that need not necessarily be in the same region. The second group concerns those countries in which concentrated industries predominate, that is those industries composed of large enterprises and which often implies the concentration of several industries in one area. The third group consists of intermediary countries. The data used are derived from those described in section (A4) above on the level of industrialisation.

(A14) Initial Level of Urbanisation

For all the countries except the United Kingdom the initial level of urbanisation is the level of urbanisation in 1800. For the United Kingdom it is the level of urbanisation in 1750.

Appendix B

Technical Appendix for the Econometric Analysis

As we have chose levels of urbanisation for our dependent variable, we are confronted with the problem that this dependent variable may only take on values between zero and one, which is due to the fact that these levels are expressed as percentages. This is related to the fact that most models of urban growth do not assume linear growth but rather some sort of logistic growth curve (a kind of S-curve). In fact, it is a fairly well established fact that the percentage of the total population that lives in cities assumes a growth pattern of this sort (Davis, 1965). This is very convenient, since an often proposed solution to the problem of a dependent variable limited between zero and one is a logistic or normal transformation. That we did not use such a transformation can be explained in large part by the fact that graphs of the levels of urbanisation were linear for the period studied. This in turn can be explained by the fact that the range of the data was from 4 per cent to 70 per cent (with few cases at either extreme). We have only two nation-years with a level over 55 per cent (Belgium and the United Kingdom), thus the cases falling in the top part of the logistic curve are absent. Cases in the lower part of the curve were relatively rare (this is a somewhat subjective judgement since the 'lower' part depends on the parameter chosen), this, in addition to the fact that data were gathered every ten years and the generally poorer quality of the data for these cases (generally corresponding to earlier periods and less developed countries, hence poorer statistical sources), make the assumption of a linear trend reasonable, because the middle part of the logistic curve can be approximated quite well by a straight line. It should be noted that in a number of analyses a logistic response function was used with virtually the same results as for a linear response function.

In any longitudinal study one has to deal with the problem of serial correlation and the specification of the nature of the error term. In this study we are limited by the fact that even if we have complete data for a country, this amounts to only nine observations, since we take observations every ten

years. If we specify a one term lag, this leaves us with eight observations, certainly not enough to detect or estimate an complicated structure of the error term (especially with 14 independent variables). At the same time we have 19–23 countries in the study of which we have complete data for only seven, which complicates the problem.

One approach to the serial correlation problem would be to estimate a different p* (auto-regression coefficient) for each country. This is clearly not practicable in our case due to the number of countries with just a few observations and even in cases with complete data the estimates would not be reliable due to the small n. The simplest thing to do is to assume that p* ($e_t = \mu + p^* e_{t-1}$) is the same for all countries, which is what was done. An estimated value of p* (0.74, significant at 0.001) was calculated. OLS with the transformed data gave a Durbin-Watson statistic in the acceptable range. A global Durbin-Watson value that is acceptable does not imply that we have removed the serial correlation from any country (to assume so would be a case of the 'ecological fallacy'). However, examining the residual plots and calculating Durbin-Watson statistics for countries for which the data are complete give no obvious indication that serial correlation remains a problem. Hence, no further efforts were made to estimate other forms of serial correlation.

In addition to adjustments across time, there exists the problem of adjustments across space. One possible technique is to include indicator variables representing the different cross-sectional units (Pindyck and Rubenfeld, 1981), this implies that the intercept is the only thing which varies across space. Another solution is to define the error terms $e = u_m + v$ where v represents the random error and u represents the effect of the error attributed to the cross-sectional units m (Balestra and Nerlove, 1966). Under suitable and standard assumptions, one can estimate these u.

An examination of the residuals showed few cases where such an adjustment would be appropriate (in fact, using OLS there were several such cases, but they disappeared after adjustments for serial correlation were made).

When economists have generally been more interested in problems of correlation and independence over time, in any cross-sectional study one must also consider the problem of dependence across space. Just as we have trouble specifying the structure of serial correlation, we have trouble specifying the nature of spatial autocorrelation. This is a problem that has interested anthropologists and sociologists probably more than economists, most commonly because of theories of diffusion or problems of sampling (Galton's Problem). It is interesting to pose the question: what do we feel to be the nature of the interdependence across space of our observations. Using the analogy of serial correlation, one could assume that interdependence across space might be due to geographical proximity (this being an approach more similar to that of the sociologists and anthropologists), like we assume serial correlation to be due to proximity in time. One obvious problem is that of defining geographical proximity. For us the problem of dependence across space will be considered resolved if there are no geographical trends in residuals, by some criterion or other. There are various possible ways to examine the residuals for such interdependence. One might, for example, take some sort of cluster analysis technique to see if residuals fall into geographical groups. We, in addition to the usual subjective visual analysis, use a simple ad-hoc test consisting of testing whether the average residual for contiguous countries is different from that of noncontiguous countries. This can be calculated for individual countries as well as for the whole sample. In general we found these means to be the same, though this was not true for every individual country, although small n's make evaluation difficult. Hence, we conclude that interdependence of the errors across countries was not a severe problem.

With regard to problems of heteroscedasticity, there exists the problem that the variance may not be constant for all cases; one solution is to use a generalised least squares procedure. However, estimating the various equations using a robust estimation technique (White, 1980), gave essentially the same results and therefore we considered this problem to be of minor importance and that such biases were not great.

The percentage of missing data among the independent variables was small, with agricultural productivity having the most, and levels of industrialisation and forms of government being the other variables with missing data. These were estimated by interpolation modified relatively subjectively to take into account ideosyncratic factors. Generally, more data were missing for earlier periods and 'less'

important states. It is well-known that if there is measurement error in the independent variable (due, for example, to interpolation) the error term is not independent of the independent variable. Since we have interpolated data and since we know that certain data are less reliable than others, one might suspect that this would be a problem. One possible solution is least squares using instrumental variables. Considering the fact that we have no suitable instrumental variables led us not to adopt this approach.

In classic least squares, in order to make significance tests the assumption of normality of the error term is made. Monte Carlo studies have shown that OLS is quite robust with regard to violations of this assumption. In addition, asymptotically the residuals are normally distributed (Central Limit Theorem).

One of the variables that we have included in our study is the starting level of urbanisation (level of urbanisation in 1800, 1750 for the United Kingdom). Generally speaking, when the endogenous variables are included as independent variables OLS is not a valid procedure, because of correlation of independent variables with the error.[11] Several considerations lead us to consider this bias as not severe.[12] One would expect the bias to be reduced as one moves away from the starting period. Although there are data for the periods 1830 and 1840, 1850 is the first period for which we have complete data for all countries (except 1800), thus the bias might be expected to be smaller. Secondly, we view the starting level variable as an adjustment factor, in a sense, like a modification of the intercept. It represents the previous history of the country until 1800. The fact that we have chosen 'levels' of urbanisation implies that such considerations are more important than if we had chosen something like the growth of urbanisation. There are two basic ways of viewing a variable such as this one and its effects on the level of urbanisation. One way is to consider this as a factor whose effect decreases over time, i.e. the value of the parameter decreases over time. If one takes this view, then the ordinary least squares procedures would be inappropriate, since it assumes that the parameter remains constant. On the other hand, if there were no changes in any of the important variables, e.g. GNP per capita, one would expect the level of urbanisation to change much from its 1800 level. Now, this is a argument for a constant parameter over time. When performing the regressions over two different time periods, 1830–1870 and 1880–1910, the significance as well as the standardised parameter estimates remained about the same from the first period to the second. In addition, the use of a linear decay function also indicated that the effect was non-decreasing. Thus, we take the option of considering this factor relatively constant over time. The same problems both technical and practical occur with the variable starting date of modernisation.

Appendix C

In order to provide the reader with historical data for the actual definition of developed countries we have elaborated in Table 10 data for the other

Table 10

Urbanisation data for non-European developed countries (including Japan and South Africa) *and all developed countries* (including Japan and South Africa)

	Urban Population (in millions)		Level of Urbanisation (%)	
	Other devel. countries	Total devel. countries	Other devel. countries	Total devel. countries
1800	4.7	27.1	13.2	11.2
1830	5.6	35.9	12.6	12.6
1840	6.4	43.1	12.9	14.1
1850	8.3	53.3	14.4	16.0
1860	11.2	66.3	16.3	18.3
1870	15.6	82.4	19.2	20.5
1880	20.1	101.7	20.9	22.9
1890	30.3	131.1	26.4	26.7
1900	38.7	164.5	28.7	30.0
1910	53.4	205.9	33.3	33.0
1950	133.6	388.4	49.3	46.1
1980	281.5	767.6	68.8	65.8

Sources: 1800–1910: see Table 3; other data: adapted from Bairoch, P., *De Jéricho à Mexico: Villes et économie dans l'histoire*, 1985.

Note: The degree of rounding off of the figures does not imply a correspondingly low margin of error.

[11]Our model is not of the type $Y_t = \beta X + Y_{t-1}$ but rather $Y = \beta X + Y_1$.

[12]If we consider Y_1 fixed as opposed to random and assume no serial correlation then the MLE estimates are OLS, see Wonnacott, R. and Wonnacott, T. *Econometrics* 1970, New York.

developed countries (non-European) and for all developed countries. In both cases Japan and South Africa are included; in view of their specific characteristics those two countries were excluded from the study.

References

BAIROCH, P. (1965). Niveaux de developpement economique au XIXe siecle. *Annales, E.S.C.*, 6: 1091–1117.

BAIROCH, P. (1976). Europe's gross national product, 1800–1975. *The Journal of European Economic History*, Vol. 5, 273–340.

BAIROCH, P. (1976a). *Commerce exterieur et developpement economique au XIXe siecle*. Paris, The Hague: Mouton.

BAIROCH, P. (1977). *Taille des villes, conditions de vie et developpement economique*. Paris: Editions de l'Ecole des hautes etudes en science sociales.

BAIROCH, P. (1980). The main trends in national economic disparities since the industrial revolution. In Bairoch P. and Levi-Leboyer, M. (eds.) *Disparities in economic development since the industrial revolution*. London: MacMillan.

BAIROCH, P. (1982). International industrialisation levels from 1750–1980. *The Journal of European Economic History*, Vol. 11, 269–333.

BAIROCH, P. (1985). *De Jericho a Mexico: Villes et economie dans l'histoire*. Paris: Gallimard.

BEST, R. et al. (1974). The density-size rule. *Urban Studies*, Vol. 11, 201–208.

BALESTRA, P. and NERLOVE, M. (1966). Pooling Cross-section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica*, Vol. 34, 585–612.

DAVIS, K. (1965). The urbanization of the human population. *Scientific American*, Vol. 213.

DAVIS, K. (1969 and 1972). *World urbanisation 1950–1970. two volumes*. Berkley: University of California Press.

GALLAWAY, L. and VEDDER, R. (1971). The Increasing urbanisation thesis — Did 'new immigrants' to the United States have a particular fondness for urban life? *Explorations in Economic History* Spring, Vol. 8: 305–319.

GUEST, A. (1973). Urban growth and population densities. *Demography*, Vol. 10, 53–69.

MULHALL, M. (1898). *The dictionary of statistics*. 4th ed. London: Routledge.

MITCHELL, B. (1975). *European historical statistics, 1750–1970*. London: MacMillan.

PINDYCK, R. and RUBENFELD, D. (1981). *Econometric models and econometric forecasts*. New York: McGraw-Hill.

ROSTOW, W. W. (1962). *The process of economic growth*. (2nd edn.) Cambridge: Cambridge University Press.

LA STATISTIQUE GENERALE DE FRANCE (1932). *Annuaire statistique de la France, vol. 31*. Paris: La Statistique Generale de France.

SUNDBARG, G. (1908). *Apercus Statistique internationaux*. Stockholm: In Primerie Royale (Reprinted by Gordon and Breach, New York, 1968).

URQUART, M. and BUCKLEY, F. (eds.) (1965). *Historical Statistics of Canada, 1965*. Ottawa: Statistics Canada.

US BUREAU OF THE CENSUS (1962). *Historical Statistics of the United States, colonial times to 1957*. 1962 Washington: Bureau of the Census.

WHITE, H. (1980). Using least squares to approximate unknown regression functions. *International Economic Review*, Vol. 21, 149–170.

WOYTINSKY, W. (1927). *Die Welt im Zahlen, Vol. 5*. Berlin: R. Mosse.