77279

THE WORLD BANK ECONOMIC REVIEW, VOL. 12, NO. 3: 529-47

A Database of World Stocks of Infrastructure, 1950-95

David Canning

This article describes an annual database of physical infrastructure stocks for a cross-section of 152 countries for 1950-95. The database includes estimates of six measures of infrastructure: the number of telephones, the number of telephone main lines, kilowatts of electricity-generating capacity, kilometers of total roads, kilometers of paved roads, and kilometers of railway lines. Both raw and manipulated data sets, in which series have been linked to overcome changes in definition and coverage, are reported. Some measures of infrastructure quality, such as the percentage of roads in poor condition, the percentage of local telephone calls that do not go through, the percentage of diesel locomotives available for use, and the percentage of electricity lost from the distribution system, are included. The data on all series except total roads are of reasonably good quality and should prove useful to researchers.

The article also presents regression results relating stocks of infrastructure to population, per capita gross domestic product, land area, and level of urbanization. It shows that stocks of telephones, electricity-generating capacity, and paved roads tend to increase proportionately with population and more than proportionately with per capita gross domestic product. Both the length of total roads and the length of total rail lines rise with country size and are relatively insensitive to population and income.

Physical infrastructure has long been considered an important determinant of economic growth. Aschauer (1989), for example, finds very large returns to public capital in the United States. Canning, Fay, and Perotti (1992, 1994) estimate large growth effects of physical infrastructure. Easterly and Rebelo (1993) find that public investment in transportation and communication is consistently correlated with economic growth. Lee and Anas (1992) find lack of infrastructure, particularly lack of a consistent supply of electricity, to be a major constraint on firms in Nigeria. Antle (1983) finds a significant role for infrastructure in agricultural productivity in developing countries. (For a review of the literature on the importance of infrastructure to economic development and an evaluation of empirical results estimating the contribution of public capital and infrastructure to growth, see World Bank 1994, Gramlich 1994, and Jimenez 1995.)

David Canning is with the Department of Economics at Queen's University of Belfast and is currently visiting Harvard University. The author gratefully acknowledges research funding from the World Bank and the comments and assistance with data collection of Esra Bennathan.

A recurrent problem in this literature, as noted by Jimenez (1995), is the lack of data. The main aim of this article is to provide and describe a data set on the stock of physical infrastructure—telephones, telephone main lines, electricity-generating capacity, roads, paved roads, and railway lines. Physical measures of infrastructure stocks are used because of the problems associated with using investment data to estimate infrastructure capital. Because of differences in the efficiency of the public sector and the price of infrastructure capital, the same level of investment in infrastructure may yield very different results across countries (Pritchett 1996). Moreover, investment figures represent annual flows into infrastructure. Deriving estimates of infrastructure stock at a point in time requires the use of perpetual inventory methods, which may introduce systematic errors in stock estimates.

The annual time series for infrastructure stocks presented here goes back as far as 1950, the earliest year for which the Summers and Heston (1991) Penn World Tables data on purchasing power parity gross domestic product (GDP) are available.[1] The 152 countries reported are the same as those covered in the Penn World Tables. An earlier data set constructed by Canning and Fay (1992) and published by the World Bank (1994) gives infrastructure data on a quinquennial basis back to 1960. Although for many purposes quinquennial data are sufficient, a full set of annual data allows more detailed investigation of the time-series properties of the data. In addition, the data reported here have been improved by cross-checking different sources, reconciling differences by referring to more detailed national sources, and linking series in which breaks occur as a result of changes in definition or coverage. This data set should therefore be seen as superseding the original data set.

Two types of data files are included. The first reports raw data. All of the data in this set are reported exactly as they appear in the original sources (except for road lengths in miles, which have been converted to kilometers, and adjustments to deal with border changes). Where different sources report different figures, the series that seems closest to our definition of the relevant variable is used. In many cases, disaggregated national sources have been studied to determine which figures best represent a particular stock of infrastructure. When different sources appear equally plausible, official national sources are used first, international sources collected from government agencies second, and international sources collected from nongovernmental agencies last. The only manipulation of these raw data is the suppression of what appear to be misprints in data sources or implausibly large year-to-year changes in reported stocks.

A problem with these raw data is that within a single time series, the reporting source, the definition of the stock variable, or the coverage of the variable may change. It nevertheless seems desirable to make the raw data available to researchers, so that they can interpolate or transform the data as desired.

1. The database is available on the Internet at http://www.worldbank.org/html/dec/Publications/Workpapers/WPS1900series/WPS1929/canning1.xls.

For some variables, additional data sets have been created by systematically linking the data series where breaks occur as a result of changes in definition or coverage. Overlapping data are used where available to construct proportional indexes; that is, all of the data before the break have been changed proportionately to bring them into line with data after the break. Where it is not possible to link series, the less appropriate series has been deleted. Some countries measure infrastructure stocks only at infrequent intervals, reporting the same figure for a number of years. In cases where this practice is documented, the constructed data set deletes repeated instances of the same value, so that the values given relate only to the actual year of measurement.

Infrastructure stocks usually move very slowly, and there is scope for interpolation in some series with very few gaps in the data. Most gaps are short (one to two years), but we have interpolated over gaps of up to five years in the constructed data. All interpolation is carried out so as to be linear in the logs of the infrastructure stock (that is, growth is assumed to be exponential over the intervening period). The data for telephones, telephone main lines, length of paved road, and length of railway line have been manipulated in this way to produce constructed data sets. For these variables, the constructed data rather than the raw data are recommended for use, and the results reported here are based only on the manipulated data. In the case of electricity-generating capacity, the raw data appear so good that no manipulation has been carried out. For total roads, the definition and coverage of the data vary too much over time and across countries to produce a consistent series. If, for a particular purpose, interpolation and linking of series are not desirable, it is possible to derive a data set that contains only high-quality reported data by imposing the condition that the same figure be reported in both the raw and the manipulated data sets.

The provision of infrastructure reflects the forces of demand and supply and the effect of public policy. Public policy may, in fact, play a very large role, because in many cases the price mechanism works imperfectly or is absent in the provision of infrastructure. In the absence of price variables for infrastructure, a complete model of infrastructure provision is not possible. It is nonetheless possible to note the positive relationships between infrastructure provision and economic development. For example, Queiroz and Gautam (1992) find a significant correlation between kilometers of paved roads and GDP in a cross-country study. Ingram and Liu (1997) find stable relationships between provision of roads and indicators of economic development, both at the municipal and national levels.

A problem with using measures of physical infrastructure capital is that they do not reflect the quality of the services provided. Although figures for paved roads as well as total roads and for telephone main lines as well as telephones are given, they may still conceal possibly large differences in quality across countries. Hulten (1996) argues that the management and efficient use of infrastructure may be more important than the quantity. To address this issue, some measures of infrastructure quality and efficiency of use are included. These measures are available for only a limited number of years, however.

The data reveal a very strong relationship between infrastructure and measures of economic development and geography. For telephones and electricity-generating capacity, the stock of infrastructure rises proportionately with population and more than proportionately with per capita income. Urbanization is positively correlated with the number of telephones in poorer countries, while country size is negatively correlated with the number of telephones in richer countries.

For transportation infrastructure, a different pattern emerges. The provision of total roads and rail lines rises less than proportionately with population and level of income per capita but more than proportionately with area.

In contrast, on average paved roads increase almost proportionately with population and more than proportionately with income, with the association between income and paved roads being stronger in rich countries. Area appears to have no significant relationship with the provision of paved roads. These results suggest that rail lines and unpaved roads operate below capacity and represent nonrival public goods, while, at least in rich countries, paved roads generally operate near full capacity. Although paved roads are generally considered to be nonexcludable public goods, congestion effects may transform them into rival goods. These interpretations assume, of course, that the level of infrastructure provision is close to the efficient level, an assumption that may not be valid. Given the large-scale involvement of government, patterns in infrastructure stocks may be explained better by political economy than by economic efficiency.

However, all these relationships are probably equilibrium outcomes and do not directly reflect either the demand or supply functions. The positive relationship between infrastructure provision and income cannot be interpreted as reflecting income elasticity of demand unless the price of infrastructure is held constant across countries. Preliminary price data for road construction show that prices vary systematically across countries, with real prices in middle-income countries averaging about two-thirds those in rich countries and poor countries. This suggests that the stable relationship between per capita GDP and infrastructure stocks may be the result of a complex interaction of demand and supply effects.

The determinants of infrastructure growth rates during 1965–85 are also examined. These growth regressions test the robustness of the cross-sectional infrastructure relationships and their stability over time: if the cross-country relationships represent equilibrium conditions, the growth rates of the infrastructure stocks should respond to disequilibrium in the relationship. In fact, significant disequilibrium adjustment is found for every type of infrastructure.

In this article, the emphasis is on describing the construction of the data set and some correlations between the infrastructure data and economic development. The central questions of the direction of causation and the size of any causal effects are left to future research.

I. The Data

Data on six measures of physical infrastructure stocks—the number of telephones, the number of telephone lines, kilowatts of electricity-generating capacity, kilometers of road, kilometers of paved road, and kilometers of railway line—from various international publications were collected for 152 countries for the period 1950–95. (The main international data sources for each variable are given in the appendix.) These data were then supplemented with data from national sources as necessary.

Telephones and Telephone Main Lines

For telephones the various data sources are consistent with one another and produce series that appear continuous over time. The basic sources of data on telephones are the International Telecommunications Union's *Yearbook of Common Carrier Statistics* and the American Telephone and Telegram Company's publication *The World's Telephones*. These sources provide almost identical figures and are in agreement with data from the United Nations' *Statistical Yearbook*.

The only difference between the data sources appears to be the point within the year at which the stock of telephones is measured. *The World's Telephones* measures stocks as of 1 January, while the *Yearbook of Common Carrier Statistics* measures stocks at different dates in different countries. For many countries, the figure that appears in the *Yearbook of Common Carrier Statistics* is the same figure that appears in *The World's Telephones* for the following year. I have used the *Yearbook of Common Carrier Statistics* as my main source, because its coverage is more comprehensive. Where other sources are used to extend the series or fill the gaps, the year reported is adjusted so that the overlapping part of the series agrees with data from the *Yearbook of Common Carrier Statistics*.

Provision of telephone services is measured by both the number of telephone sets and the number of main lines connected to local telephone exchanges. The number of telephone main lines seems to be the better measure of the capacity of a telephone system (although in practice the two measures are almost perfectly correlated).

In theory, a better measure of infrastructure stock might be the capacity of telephone exchanges. To reflect the use of cellular phones, which have been in use since 1982, the infrastructure stock should reflect the area over which cellular calls are possible, as well as the number of cellular phones. The data reported here include cellular phones; detailed data on this topic are available from the International Telecommunications Union.

Data on the number of telephone sets are fairly comprehensive for the period 1950–95; data on telephone main lines are sparse in the earlier years. In addition to the raw data, a data set with interpolation of periods up to five years is also provided. Interpolation is carried out linearly in the log of the variable. The percentage of local calls that were unsuccessful in 1990 is used as an indicator of quality. Where data for 1990 are not available, data for the nearest available later year (up to 1995) are used. The data on quality indicate high variations in the percentage of calls that fail, from none up to 98 percent, indicating that a pure quantity measure may be deficient as a proxy for telecommunications services.

Electricity

The main sources of data on electricity-generating capacity are the United Nations' *Energy Statistics* and *Statistical Yearbook*. The time series seem good and are reported without adjustment, yielding a fairly complete data set for the period 1950–95. Data for Botswana, Lesotho, Namibia, and Swaziland are included in the data for South Africa. The raw data are not manipulated because there appear to be no breaks in the series and there is little to be gained by interpolation. These data on capacity do not take into account the extent of the electricity distribution system. The percentage of generated electricity lost in the system is used as an indicator of quality for 1971, 1980, and 1990.

Total Roads

The data on roads come from two international sources, the International Road Federation's *World Road Statistics* and the statistical yearbooks of the regional commissions of the United Nations, as well as various national sources. *World Road Statistics* is based on data supplied by the contracting industry in each country. The earliest data available are for 1958, and the coverage of the data set expands with time, particularly during the 1970s. Where the two international sources disagree, I tend to use the United Nations data, which are reported from official government sources.

The international data on total roads are patchy, with frequent gaps and many large changes that are often quickly reversed. Different countries define roads differently, and the definition of a road often changes within countries over time. The definition of minimum quality standards for roads varies, and differences in reporting reflect the functional split of road management between central and local government. Large changes in the series occur, for example, when the reporting source changes the nature of its coverage (from coverage only of roads controlled by the central government, for example, to roads controlled by central and provincial governments). Intraurban roads above a certain quality threshold are often centrally controlled, while urban roads are controlled by municipal authorities, leading to an underreporting of urban and low-quality rural roads controlled by the central authority.

National sources are used to supplement the data from international sources. Use of these sources increases the coverage of the data substantially and produces more consistent numbers. When the national sources agree broadly with the international sources, the national sources are used as the primary data source, with gaps filled from international sources where possible. Study of national sources also reveals, to some extent, the causes of discontinuities in the internationally reported data.

As far as possible, the reported time series for total roads includes urban roads and reflects total length of public roads in the country, regardless of the controlling authority. Some results using total roads are reported here for comparison, but for the most part the raw data seem too unreliable to be useful. Even with the use of national sources, it appears impossible to construct data that are consistent either across countries or over time. No attempt is made to manipulate the raw data since there is a lack of documentation on basic questions of definition.

Paved Roads

Paved roads are defined as concrete or bitumen-surfaced roads. This definition excludes stone, gravel, water-bound gravel, oil-bound gravel, and earthen roads and is consistent with the definition used by most countries. Where a country defines paved roads differently, the data are adjusted to fit our narrow definition. China, for example, defines paved roads much more broadly. The data reported here for China are therefore taken from a World Bank country report rather than from official Chinese sources. In many cases, detailed data on the type of road are not available, and the national definition is used. When national sources give figures for "paved" or "hard-surfaced" roads and no other information is available, these categories are accepted as equivalent to the narrow definition of paved roads used here.

For many countries, international sources report only rural roads. Using national sources, we are able to construct, for some countries, the total stock of paved roads by adding urban and rural paved kilometers. However, for many countries, it is unclear even in national sources exactly what the coverage is. In countries for which both urban and rural data are available, urban roads make up about 15–30 percent of the total stock of paved roads.

Even using national data, however, the series still exhibits large changes and splits, making the raw data for paved roads too inconsistent for practical use. To make the series consistent over time, it is necessary to link series following changes in definition and coverage where these are documented. The processed data are then interpolated over gaps of up to five years. The resulting processed data are the best currently available.

Within the category of paved roads, there may be large variations in quality. In particular, the data do not reflect the width of the road, which varies from single-lane roads to multiple-lane highways. The percentage of the main paved and unpaved road network that the World Bank considers to be in good, fair, and poor condition in 1984 and 1988 is used as a measure of quality. These data cover most developing countries. The quality measures refer to the main road network and may not be representative of the total road network. In addition, they make no allowance for the age of the road stock and thus may not be good indicators of maintenance levels. The data file on infrastructure quality contains information on the data source used for paved roads in each country and on the coverage of the data, where this information is available.

Railways

The main sources of data on length of railway lines are Mitchell's *International Historical Statistics* (1992, 1993, 1995) until 1980 and World Bank data thereafter. National sources are also used to supplement these data. The data refer to length of line, where a line may consist of two or more tracks. The only problem with these data seems to be changes in coverage due to the treatment of rail lines owned by companies for industrial use and not open to the public (such as railways owned by the sugar industry in Latin America). To produce a consistent series, the data are limited to railroads open to the public.

The processed data link the series over breaks in the data. Interpolation is also carried out, although the gaps in the railway line series tend to be very minor. Countries reporting no railways in national or international sources at any time during 1950–95 are assumed to have zero line length in the processed series. The percentage of the stock of diesel locomotives available for use measured between 1990 and 1995 is given as a measure of quality.

Assessment and Recommendations

The data sets for telephone sets, telephone main lines, and electricity-generating capacity seem excellent, and both the raw data sets and the manipulated (interpolated) data can be used without worry. The data for railways seem good, although for a number of countries the series have to be linked to achieve consistency over time. For this series, the processed data should be used, particularly for time-series work.

The data for total roads are unreliable. Unexplained fluctuations in the series seem too numerous to allow the construction of consistent series. Because most of the large changes reflect changes in definition rather than changes in the stock of roads, these data are not suitable for time-series work. They may be used for cross-sectional work if allowance is made for the fact that different definitions across countries mean that the series contain large measurement errors.

This problem also affects the raw data series for paved roads. There, however, most of these problems are overcome by processing, and the series are consistent over time. For cross-sectional work, countries that report only an administrative subcategory of paved roads should not be used. Using the remaining countries—countries that report either total paved roads or rural paved roads or do not define clearly what they are reporting—in cross-sectional studies introduces an under-reporting error of about 15–30 percent of the total stock of roads for countries reporting only rural roads. Measurement error of this magnitude should be acceptable if the data are to be used for cross-sectional studies.

II. Cross-Country Relationships: Infrastructure, Income, and Geography

For each data set, the mean, standard deviation, minimum, maximum, and number of observations per 1,000 population are reported for each country. Table 1 provides summary descriptive statistics. As we shall see, this may not be an appropriate normalization for all the variables in the data set. The year 1985 is used as a base year for comparison, because it has good coverage of all the relevant measures of infrastructure.

The correlation between stocks of infrastructure is also examined (table 2). As expected, all of the infrastructure stocks are positively correlated. The correlation between the number of telephones and the number of telephone main lines is greater than 99 percent, indicating that the number of telephones can be used as a proxy for the number of telephone main lines. Paved roads and total roads are correlated only loosely, suggesting that one series cannot be used as a proxy for the other.

Table 1. *Infrastructure per 1,000 Population, 1985*

Type of infrastructure	*Mean*	*Standard deviation*	*Minimum*	*Maximum*	*Number of observations*
Number of telephone main lines	99.1	144.3	0.45	627.8	146
Number of telephones	137.9	196.5	0.59	840.1	129
Kilometers of paved roads	2.82	4.38	0.06	24.51	116
Kilometers of total roads	6.71	7.99	0.11	54.13	137
Kilometers of rail lines	0.29	0.48	0.00	3.74	150
Kilowatts of electricity-generating capacity	0.62	0.95	0.002	5.59	146

Source: Author's calculations.

Table 2. *Correlation of Infrastructure Levels, 1985*

Type of infrastructure	*Number of telephones*	*Number of telephone main lines*	*Kilowatts of electricity-generating capacity*	*Kilometers of paved roads*	*Kilometers of roads*	*Kilometers of rail lines*
Number of telephones	1.00					
Number of telephone main lines	0.99	1.00				
Kilowatts of electricity-generating capacity	0.96	0.96	1.00			
Kilometers of paved roads	0.69	0.70	0.83	1.00		
Kilometers of roads	0.54	0.53	0.71	0.83	1.00	
Kilometers of rail lines	0.36	0.37	0.59	0.79	0.84	1.00

Source: Author's calculations.

Table 3. *Cross-Country Patterns of Infrastructure, 1985*

Item	*Log telephones*	*Log telephone main lines*	*Log electricity-generating capacity*	*Log paved roads*	*Log total roads*	*Log rail lines*
Constant	-7.93	-8.22	-13.0	-8.10	-0.771	-5.58
	(11.0)	(13.2)	(21.0)	(11.1)	(11.1)	(7.63)
Log population	1.001	0.997	0.931	0.798	0.544	0.552
	(17.5)	(21.5)	(21.5)	(9.98)	(9.14)	(9.36)
Log per capita GDP	1.479	1.442	1.398	1.243	0.583	0.820
	(13.9)	(14.8)	(17.6)	(10.8)	(6.57)	(8.42)
Percentage of population	0.891	1.221	1.167	-0.661	-0.217	-0.634
living in urban area	(2.07)	(3.10)	(3.96)	(1.25)	(0.56)	(1.08)
Log area	-0.107	-0.095	0.081	0.108	0.335	0.352
	(2.50)	(2.60)	(2.41)	(1.50)	(6.72)	(5.24)
Number of observations	126	144	142	112	133	106
Adjusted R^2	0.925	0.938	0.932	0.842	0.866	0.721

Note: Heteroscedastic-consistent t-ratios are in parentheses.
Source: Author's calculations.

It seems likely that the stock of infrastructure in a country varies with population and per capita GDP. These variables affect the demand for infrastructure, as well as the cost of providing it. Geography may also matter. Hong Kong (China) and Singapore, for example, both have very low stocks of paved roads relative to the size of their population and the level of per capita GDP. It may be that in such city states, where population density is high, the need for roads is low. This relationship can be tested by developing a measure that reflects the percentage of the population living in urban centers and the total area of the country. Such a measure may not be a good proxy for geography, however, because infrastructure often has network effects, and the precise shape of a country, as well as the location of mountain ranges and rivers and the distribution of population, may affect the provision of infrastructure. For lack of a better measure of geography, however, this measure is used here.

Ordinary least squares regressions have been run for infrastructure levels on these factors using data from a cross-section of countries in 1985 (table 3). All variables other than the ratio of people in urban centers are in logarithms, so the coefficients can be interpreted as elasticities. The t-ratios given are heteroscedastic consistent. For nontransportation infrastructure, the coefficient on population is significant and close to 1, indicating that holding other factors constant, infrastructure rises proportionately with population. The coefficient on GDP per capita is greater than 1, indicating that stocks of nontransportation infrastructure rise more than proportionately with income.

The most interesting variables in the regression are the geographical factors—urbanization and area—which have different effects for different types of infrastructure. For example, holding constant population, per capita GDP, and percentage of the population living in urban areas, a country with more land area has proportionately more electricity-generating capacity and fewer telephone main lines than a smaller country.

To understand why these relationships may hold, consider a country with two population centers that have to be linked by communication infrastructure. In a large country, the two centers are likely to be farther apart, so the length of each link will be longer and the cost of each link will be higher. Because of the higher cost per link, the total number of links in a large country is likely to be smaller, while the total length of the links is likely to be greater. It is possible that while large countries have fewer telephone main lines, the length of their main lines (which is not measured) is greater than in small countries.

In the case of electricity, generating capacity, not the number of connections or the total length of the distribution system, is measured, so the effects of geography are likely to be quite different from the effects on the other infrastructure measures. One explanation for the rise in generating capacity with area is that electricity distribution systems suffer leakage, which depends on the length of the connection. In large countries with low population density, leakage can be avoided by operating small local plants. Doing so, however, may reduce the

Table 4. *Cross-Country Patterns of Infrastructure in Less- and More-Developed Countries, 1985*

Item	*Log telephones*	*Log telephone main lines*	*Log electricity-generating capacity*	*Log paved roads*	*Log total roads*	*Log rail lines*
Less-developed countries						
Constant	–6.23	–7.49	–14.3	–5.16	1.32	–2.70
	(4.57)	(5.88)	(9.54)	(3.39)	(1.17)	(1.67)
Log population	0.906	0.946	0.940	0.807	0.503	0.490
	(12.3)	(15.4)	(13.9)	(7.11)	(6.07)	(5.45)
Log per capita GDP	1.291	1.358	1.473	0.739	0.318	0.395
	(6.72)	(7.3)	(6.42)	(3.36)	(1.96)	(1.89)
Percentage of population	1.953	2.051	2.563	0.762	1.002	0.129
living in urban area	(2.60)	(3.10)	(3.34)	(1.03)	(1.69)	(0.101)
Log area	–0.086	–0.093	0.141	0.120	0.292	0.429
	(1.30)	(1.68)	(1.95)	(1.19)	(4.29)	(3.79)
Number of observations	63	72	69	57	66	52
Adjusted R^2	0.863	0.893	0.867	0.765	0.826	0.640
More-developed countries						
Constant	–6.81	–6.94	–13.2	–11.0	–2.16	–7.73
	(5.27)	(5.32)	(16.6)	(6.60)	(1.60)	(5.04)
Log population	1.179	1.124	0.941	0.832	0.642	0.619
	(16.4)	(17.9)	(18.7)	(7.59)	(7.97)	(6.98)
Log per capita GDP	1.277	1.263	1.472	1.606	0.707	1.076
	(7.97)	(7.25)	(16.5)	(7.12)	(4.23)	(6.01)
Percentage of population	0.093	0.409	0.396	–1.337	–0.893	–1.095
living in urban area	(0.22)	(0.71)	(2.21)	(1.68)	(1.94)	(1.46)
Log area	–0.177	–0.140	0.039	0.086	0.313	0.281
	(3.27)	(3.00)	(1.15)	(0.89)	(4.72)	(3.27)
Number of observations	63	72	73	55	67	54
Adjusted R^2	0.925	0.931	0.964	0.840	0.886	0.724
F-test for parameter equality	$F(5,116) = 4.31$	$F(5,134) = 3.64$	$F(5,132) = 3.39$	$F(5,102) = 1.70$	$F(5,123) = 2.15$	$F(5,96) = 1.21$
between subsamples	$p = 0.004$	$p = 0.001$	$p = 0.006$	$p = 0.141$	$p = 0.063$	$p = 0.311$

Note: Less-developed countries are those with annual per capita income in 1985 less than $2,500; more-developed countries are those with annual per capita income in 1985 of $2,500 or more. Heteroscedastic-consistent t-ratios are in parentheses.

Source: Author's calculations.

scope for economies of scale and increase the need for reserve capacity for peak periods if transfers within the system are difficult.

The provision of electricity and telephones tends to increase with urbanization, a result that is consistent with the lower cost of providing these services in cities, where the cost of connecting a consumer is lower. Alternatively, the large urbanization effect may reflect the fact that urbanization is a proxy for industrial structure: higher rates of urbanization are associated with more manufacturing and less agricultural production. If manufacturing output has a greater need for electric power than agriculture, industrial structure may be very important. Turning to transportation infrastructure, we see somewhat different relationships.

The results for paved roads and total roads differ significantly from each other, possibly reflecting differences in the nature of the services provided by the two types of infrastructure. One reason for this difference may be the fact that paved roads handle large volumes of traffic, while unpaved roads link places together that have low levels of traffic flowing between them. Holding other factors constant, an increase in area significantly increases the length of road, while it has a statistically insignificant impact on the provision of paved roads. Because the average distance between two points increases by the square root of the increase in area, a coefficient of 0.5 on log area is expected to correct for distances. The fact that paved roads increase proportionately (or more than proportionately) with income and almost proportionately with population while total roads increase less than proportionately with both factors may reflect the fact that paved roads are rival goods due to congestion effects, while unpaved roads typically have excess capacity.

The results for railways are similar to those for total roads, suggesting that they may serve similar functions. The regression results for railways cover only those countries with a positive stock of railways, however; a more sophisticated approach would also include those countries with zero rail length.

Regression results for countries with 1985 per capita income of more than and less than \$2,500 dollars a year are reported separately to reveal how these relationships differ across the two groups of countries (table 4). Results of an F-test for parameter stability across the two subsamples are also shown.

For nontransportation infrastructure, parameter stability is rejected at the 1 percent level of significance. The main difference in parameters between the two subsamples appears to be that urbanization has a stronger positive effect on infrastructure in less-developed countries than in more-developed countries. The number of telephones and telephone main lines seems to decline with area in richer countries, but not in poorer ones. In the case of transportation infrastructure, the parameter stability cannot be rejected, even at the 5 percent level of significance, although there does seem to be some indication that the elasticity of transportation infrastructure with income is higher in more-developed countries.

While the R^2 coefficients for each of the regressions in tables 3 and 4 are high, the regressions may not explain infrastructure levels. In particular, they have

Table 5. *Infrastructure Growth Regressions, 1965–85*
(growth rates)

Item	*Telephones*	*Telephone main lines*	*Electricity-generating capacity*	*Paved roads*	*Total roads*	*Rail lines*
Constant	-2.929	-1.978	-5.071	-2.668	-2.015	-0.492
	(4.70)	(2.41)	(5.44)	(2.27)	(3.35)	(0.88)
Growth of population	1.034	1.072	1.341	-0.326	0.967	0.254
	(4.03)	(3.22)	(4.10)	(0.96)	(3.34)	(0.94)
Growth of per capita GDP	0.926	0.859	0.982	0.887	0.040	0.389
	(8.29)	(6.23)	(6.74)	(5.72)	(0.36)	(3.74)
Change in urbanization ratio	2.645	0.028	2.030	-0.813	1.228	-0.964
	(4.33)	(3.84)	(2.77)	(0.86)	(1.97)	(1.59)
Log population, 1965	0.320	0.316	0.323	0.305	0.276	0.008
	(5.30)	(4.17)	(4.63)	(3.87)	(6.72)	(0.18)
Log per capita GDP, 1965	0.572	0.341	0.592	0.619	0.323	0.085
	(5.88)	(2.69)	(5.39)	(4.02)	(4.31)	(1.17)
Percentage of population	-0.200	0.000	0.011	-1.372	-0.292	-0.192
living in urban area, 1965	(0.787)	(0.08)	(0.03)	(2.87)	(1.12)	(0.74)
Log area	-0.094	-0.128	0.036	0.201	0.041	0.106
	(3.45)	(3.91)	(1.04)	(4.70)	(1.04)	(2.83)
Log 1965 stock of	-0.271	-0.202	-0.343	-0.513	-0.304	-0.117
relevant infrastructure	(4.75)	(2.58)	(5.71)	(9.00)	(5.27)	(2.09)
Number of observations	105	79	113	80	79	85
Adjusted R^2	0.682	0.642	0.553	0.712	0.574	0.182

Note: Heteroscedastic-consistent t-ratios are in parentheses.
Source: Author's calculations.

nothing to say about the directions of causation; it may be that per capita GDP and urbanization rates depend on the provision of infrastructure. The important point is that the provision of infrastructure is significantly correlated with geography, particularly for poorer countries, probably because the costs and benefits of infrastructure vary with geography. This implies that the impact of infrastructure on economic growth may depend on geography and that geographical considerations should be taken into account when analyzing these effects.

In terms of applications, it is usual to normalize quantity variables in order to make them independent of the size of the country. For telephones, telephone main lines, electricity-generating capacity, and perhaps paved roads, it seems reasonable to normalize by population, because each of these variables appears to increase proportionately with population. In other words, the regressions in tables 3 and 4 could be rerun using infrastructure stock per capita on the left and excluding population on the right without changing the other coefficients. Total roads and railways do not increase proportionately with population, however. For rival goods, normalization by population seems appropriate, because the quantity of the good divided by the population indicates average consumption. For nonrival goods, however, normalizing by population does not yield average per capita consumption; with a fixed stock of nonrival infrastructure, increases in population need not reduce average consumption. If transportation infrastructure is really nonrival, then normalizing by population is unlikely to be appropriate. Ingram and Liu (1997) normalize by area, but doing so yields a coefficient of less than 1 in our regressions; explaining total roads per square kilometer of area requires that area appear as an explanatory variable. A case could be made for using the square root of area on theoretical grounds, but the data appear to support a figure nearer to the cube root of area. We are thus left with no obvious normalization for stocks of transportation infrastructure.

If the cross-sectional regressions generate stable relationships, they may reflect cointegrating mechanisms for the data. That is, a country that is out of line given the cross-sectional relationship might be expected to move into line in the long run. To address this question, while avoiding the pitfalls of panel estimation, regressions are run in which the dependent variable is the growth rate of infrastructure stock during 1965–85 (table 5). (The same level variables are used as in table 3.) These regressions are similar to those of Barro (1991) for economic growth; a negative coefficient on the initial stock of infrastructure indicates convergence of infrastructure stocks to an equilibrium level that depends on the other variables in the regression.

Infrastructure stocks in each country appear to be converging to the same long-run equilibrium relationship (defined as the steady state, where all growth rates are zero), conditional on the values of the other initial condition included in the regressions. The long-run steady state can be determined by first postulating that all growth rates on the right-hand side of the regression in table 5 are zero. Doing so yields a growth rate of telephones of

(1) $0.27(1.18 \log \text{population} + 2.11 \log \text{per capita GDP} - 0.74$ $\text{percent urbanized} - 0.35 \log \text{area} - \log \text{telephones})$

(where the parameters have been reduced by a common factor of 0.27). About 27 percent of the gap between actual telephone stocks and the "equilibrium" telephone stock (when this expression is zero) is closed in a 20-year period. Setting this expression to zero (that is, assuming the growth rate of telephones is zero), it appears that telephones in each country are converging to the following relationship:

(2) $\log \text{telephones} = 1.18 \log \text{population} + 2.11 \log \text{per capita GDP}$ $- 0.74 \text{ percent urbanized} - 0.35 \log \text{area}.$

The steady-state relationships for all of the infrastructure variables in table 5 can be determined in a similar way. Estimated in this way, the long-run steady-state coefficients are remarkably similar to those found in table 3.

III. CONCLUSIONS

This article describes a new set of panel data on stocks of infrastructure in a cross-section of countries over time. The stock of infrastructure across countries varies significantly with their population size, income level, and geography, and these relationships appear stable over time.

The data sets for telephones, telephone main lines, electricity-generating capacity, and railway lines for 1950–95 are fairly complete. Some minor additions could be made if data become available to fill some gaps, but the data presented here are not likely to change. The data for total roads and paved roads may change substantially for a number of countries if better national data sources are found. In addition, it is hoped that a more definite classification of the variable for coverage of the stock of paved roads will eventually be possible.

APPENDIX. DESCRIPTION AND SOURCES OF NONSURVEY VARIABLES

Variable	*Source*
GDP per capita	Penn World Table 5.6 (Summers and Heston 1991)
Land area of country	World Bank (1995)
Percentage of population living in urban areas	World Bank (1995)
Kilowatts of electricity-generating capacity	United Nations, *Energy Statistics* (various years); United Nations, *Statistical Yearbook* (various years)
Number of telephones	International Telecommunications Union (various years); American Telephone and Telegram Company (various years)
Number of telephone main lines	International Telecommunications Union (various years); American Telephone and Telegram Company (various years)

Kilometers of paved roads	International Road Federation (various years); United Nations Economic Commission for Africa (various years); United Nations Economic and Social Commission for Western Asia (various years); United Nations Economic and Social Commission for Asia and the Pacific (various years); United Nations Economic and Social Commission for Latin America and the Caribbean (various years); United Nations Economic Commission for Europe (various years); national sources
Kilometers of roads	International Road Federation (various years); United Nations Economic Commission for Africa (various years); United Nations Economic and Social Commission for Western Asia (various years); United Nations Economic and Social Commission for Asia and the Pacific (various years); United Nations Economic and Social Commission for Latin America and the Caribbean (various years); United Nations Economic Commission for Europe (various years); national sources
Kilometers of rail lines	Mitchell (1992, 1993, 1995); World Bank Rail Statistics Database
Percentage of local telephone calls unsuccessful	International Telecommunications Union (1996)
Percentage of paved main roads in poor condition	United Nations Economic Commission for Africa (1993)
Percentage of total main roads in poor condition	United Nations Economic Commission for Africa (1993); World Bank (1988)
Percentage of diesel locomotives available for use	World Bank Rail Statistics Database
Percentage of system electricity losses	World Bank (1997)

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

American Telephone and Telegram Company. Various years. *The World's Telephones.* New York.

Antle, J. M. 1983. "Infrastructure and Aggregate Agricultural Productivity: International Evidence." *Economic Development and Cultural Change* 3(April):609–19.

Aschauer, D. A. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23(2):117–200.

Barro, Robert J. 1991. "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106(2):407–43.

Canning, David, and Marianne Fay. 1992. "Infrastructure and Economic Growth." Department of Economics, Columbia University, New York. Processed.

Canning, David, Marianne Fay, and Roberto Perotti. 1992. "Dotazioni infrastrutturali e crescita economica." *Revista di Politica Economica* 82(November):117–54.

———. 1994. "Infrastructure and Growth." In Mario Baldassarri, Luigi Paganetto, and Edmund Phelps, eds., *International Differences in Growth Rates*, pp. 113–47. New York: Macmillan Press.

Easterly, William, and Sergio Rebelo. 1993. "Fiscal Policy and Economic Growth: An Empirical Investigation." *Journal of Monetary Economics* 32(3):417–58.

Gramlich, E. M. 1994. "Infrastructure Investment: A Review Essay." *Journal of Economic Literature* 32(3):1176–96.

Hulten, C. R. 1996. "Infrastructure Capital and Economic Growth: How Well You Use It May Be More Important Than How Much You Have." University of Maryland, College Park. Processed.

Ingram, G. K., and Zhi Liu. 1997. "Motorization, Road Provision, and Economic Growth." World Bank, Washington, D.C. Processed.

International Road Federation. Various years. *World Road Statistics*. Geneva.

International Telecommunications Union. 1996. "World Telecommunications Indicators [on diskette]." Geneva. Processed.

———. Various years. *Yearbook of Common Carrier Statistics*. Geneva.

Jimenez, Emmanuel. 1995. "Human and Physical Infrastructure: Investment and Pricing Policies in Developing Countries." In Jere Behrman and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 3B, pp. 2773–843. Amsterdam: Elsevier Science, North-Holland.

Lee, K. S., and Alex Anas. 1992. "Costs of Deficient Infrastructure: The Case of Nigerian Manufacturing." *Urban Studies* 29(7):1071–92.

Mitchell, B. R. 1992. *International Historical Statistics, Europe, 1750–1988*. New York: Macmillan Press.

———. 1993. *International Historical Statistics: The Americas, 1750–1988*. New York: Stockton Press.

———. 1995. *International Historical Statistics: Africa, Asia, and Oceania, 1750–1988*. New York: Stockton Press.

Pritchett, Lant. 1996. "Mind Your P's and Q's, The Cost of Public Investment Is Not the Value of Public Capital." Policy Research Working Paper 1660. Policy Research Department, World Bank, Washington, D.C. Processed.

Queiroz, Cesar, and S. P. Gautam. 1992. "Road Infrastructure and Economic Development: Some Diagnostic Indicators." Policy Research Working Paper 921. Policy Research Department, World Bank, Washington, D.C. Processed.

Summers, Robert, and Alan Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–88." *Quarterly Journal of Economics* 106(2):327–68.

United Nations. Various years. *Energy Statistics*. New York.

———. Various years. *Statistical Yearbook*. New York.

United Nations Economic and Social Commission for Asia and the Pacific. Various years. *Statistical Yearbook for Asia and the Pacific*. Bangkok.

United Nations Economic and Social Commission for Latin America and the Caribbean. Various years. *Statistical Yearbook for Latin America and the Caribbean*. Santiago, Chile.

United Nations Economic and Social Commission for Western Asia. Various years. *Statistical Abstract of the Region of the Economic and Social Commission for Western Asia*. Baghdad.

United Nations Economic Commission for Africa. 1993. *Survey of Economic and Social Conditions in Africa, 1988–89*. New York.

———. Various years. *African Statistical Yearbook*. Addis Ababa.

United Nations Economic Commission for Europe. Various years. *Annual Bulletin of Transport Statistics for Europe*. New York.

World Bank. 1988. *Road Deterioration in Developing Countries*. Washington, D.C.

———. 1994. *World Development Report 1994: Infrastructure for Development*. New York: Oxford University Press.

———. 1995. *World Tables*. Washington, D.C.

———. 1997. *World Development Indicators 1997*. Washington, D.C.