A Database of World Infrastructure Stocks, 1950–95

David Canning

This paper describes an annual database of physical infrastructure stocks for a cross section of 152 countries for the period 1950–95. The database contains six measures: *kilometers* of road, of paved road, and of railway line; *number* of telephones and telephone main lines, and *kw* of electricity generating capacity. Some measures of *infrastructure quality* are also included. The paper then relates infrastructure stocks to country population, GDP per capita, land area, and urbanization rate, and looks at the relationship between infrastructure and economic growth.

The paper is a product of the World Bank-funded research on *Infrastructure and Growth: A Multi-Country Panel Study*, a project sponsored jointly by the Public Economics Division of the Development Research Group, and the Transport, Water, and Urban Development Department of the World Bank.

The author thanks Esra Bennathan for his contribution to the data collection, and for his comments.

Summary

This paper contains a description of an annual database of physical infrastructure stocks constructed for a cross-section of 152 countries for the period 1950–95. The database contains 6 measures: kilometers of road, kilometers of paved road, kilometers of railway line, number of telephones, number of telephone main lines, and electricity generating capacity. The data are available on the accompanying diskette "Infra 1.0." Some measures of infrastructure quality, such as the percentage of roads in poor condition, percentage of local telephone calls unsuccessful, percentage availability of diesel locomotives, and percentage of electricity lost from the distribution system are also included. However, these quality measures are usually available only for recent years.

As well as describing the construction of the data, I look at correlation patterns and report regressions relating infrastructure stocks to countries' population, GDP per capita, land area, and percentage urbanization.

Nontransportation infrastructure stocks tend to increase one for one with population but increase more than proportionately with GDP per capita. Geographical factors appear to affect nontransportation infrastructure provision in poor countries but not in rich countries. Transportation infrastructure appears to increase less than proportionately with population and increases with income only after a middle income threshold has been reached. Geographical factors seem to influence total roads and rail line length, but not paved roads length.

I carry out panel unit root tests and find that the log infrastructure stock per capita series are nonstationary and have a unit root. Cross section growth regressions indicate that our common worldwide estimated regression results for infrastructure stocks are stable long run relationships. Preliminary regression results suggest that the telephone main lines per capita have a positive effect on subsequent economic growth.

David Canning
Harvard Institute for International Development
One Eliot Street
Cambridge, MA 02138

dcanning@hiid.harvard.edu

1. Introduction

The development of a large empirical literature generated much debate about what factors influence economic growth. One factor that has been suggested is the provision of physical infrastructure. The World Bank's *World Development Report 1994*, Gramlich (1994) and Jimenez (1995) provide surveys of why infrastructure is important in economic development and evaluate recent empirical results estimating the contribution of public capital and infrastructure to growth. For example, Aschauer (1989) finds very large returns to public capital using U.S. data, while Canning, Fay, and Perotti (1992, 1994) use Barro (1991)-style regressions and 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 of consistent electricity supply, to be a major constraint on Nigerian firms. Antle (1983) finds a significant role for infrastructure in agricultural productivity in developing countries.

A recurrent problem in this literature, as noted by Jimenez (1995), is the lack of data. The main aim of this paper is to provide, and describe, a data set on physical infrastructure stocks: roads, paved roads, railway lines, electricity generating capacity, telephones, and telephone main lines. The emphasis on physical measures of infrastructure stocks is due to the fact that using investment data to estimate infrastructure capital produces some serious problems. Summers and Heston (1991) and Pritchett (1996) argue that the same investment flows in different countries may have very different effectiveness in actually producing infrastructure, due to differences in the efficiency of the public sector and differences in the price of infrastructure capital. In addition, investment figures are annual flows into infrastructure; if we wish to derive estimates of the infrastructure stock at a point in time, we need to use perpetual inventory methods, which may introduce systematic errors in stock estimates. In addition to describing the data I examine patterns between infrastructure provision, economic activity, and geography.

I have tried to construct as complete as possible an annual time series for infrastructure stocks, going back to 1950, the earliest year for which the Summers and Heston (1991) Penn World Tables data on purchasing power parity GDP figures are available. The 152 countries reported are the same as those 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. While for many purposes quinquennial data are sufficient, a full annual data set 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 and reconciling differences by reference to more detailed national sources, and by linking together series when there are breaks due to redefinition or changes in coverage. This data set should therefore be seen as superseding the original data set. The data diskette is described in appendix 1.

The data reported here are of two types. The first reports raw data, with a minimum of manipulation. All data reported are as they appear in the sources, except for road lengths where miles have been converted to kilometers. Where different sources report different figures, we select the series that seems closest to our definition of the relevant variable. Often, it took study of disaggregated national sources to decide which figures best represent the infrastructure stock we wish to measure. When there is little information to choose between sources, our principle has been to use official national sources 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, and some instances of unbelievable jumps in reported stocks.

A problem with these raw data is that within one time series, the reporting source may change, and with it possibly the definition or coverage of the stock variable. However, it seems desirable to make the raw data available to researchers so that any interpolation or transformation can be carried out as desired.

In addition to the raw data sets, for some variables I construct data sets by systematically linking together the data series where there are breaks due to changes in definition or coverage. Where overlapping data are available, I use them to construct proportional indices. Where it has not proved possible to link series, I have deleted the less appropriate series. This produces data series that are more consistent over time than the raw data. In addition, some countries only measure infrastructure stocks at infrequent intervals, reporting the same figure for a number of years until a new measurement is made. In these cases the constructed data set deletes repeated instances of the same value so that the values given relate to the actual year of measurement.

The infrastructure stocks are usually very slow moving, and there is scope for interpolation in some data series with very few gaps in the data. Most gaps in the data are short, one to two years, but in the constructed data I have interpolated over gaps of up to five years. All interpolation is carried out so as to be linear in the logs of the infrastructure stock; that is, I assume exponential growth over the intervening period. The data for telephones, telephone main lines, railways line length, and paved road length are manipulated in this way to produce constructed data sets. For these variables, the constructed data are recommended for use rather than the raw data and the results reported in this paper use only the manipulated data. In the case of electricity generating capacity, the raw data appear so good that no manipulation was carried out. For total roads, however, the problem of producing consistent series seem almost insurmountable; the definition and coverage of the data vary too much over time and across countries.

The main aim of this paper is not to look directly at the link between infrastructure and economic growth but to present the data set and describe some of its properties. In particular, we can think of the provision of infrastructure as being the result of demand and supply forces, together with the effect of public policy. Public policy may in fact play a very large role, since in many cases the price mechanism in infrastructure provision works imperfectly, or is absent. In this paper I present only quantity variables, and in the absence of price variables for infrastructure, a complete model of infrastructure provision is not possible. However, there do seem to be stable relationships between infrastructure provision and economic development. Hardy and Hudsen (1981) investigate the relationship between telephones per capita and GDP per capita while Querioz and Gautman (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 road provision and indicators of economic development both at the national and urban levels.

A problem with our method of measuring physical infrastructure capital is that we do not correct fully for the quality of the capital services provided. We do give figures for paved roads as well as total roads, and telephones main lines, as well as the number of telephones, but these figures still hide what may be very large quality differences across countries. Hulten (1996) argues that the management and efficient use of infrastructure may be more important than the quantity. I include some measures of infrastructure quality and efficiency of use in the data set; however, these are only available for a limited number of years.

When we turn to look at patterns in the infrastructure data we find a very strong relationship between infrastructure and measures of economic development and geography. It is important to realize however that these relationships are probably equilibrium outcomes and do not reflect simple demand or supply functions directly. For example, for telephones, and electricity generating capacity, I find that the infrastructure stock rises one for one with population but increases more than proportionately with income per capita. Our geographical variables have a significant impact on provision in poorer countries but have little effect on provision in richer countries. The increase in provision with income cannot be interpreted as an income elasticity of demand, unless the price of infrastructure is constant across countries. Preliminary price data, for road construction, shows that prices do vary systematically across countries, with middle income countries having real prices about two thirds those in rich countries and poor countries, which have equally high price levels. This suggests that the relationship between GDP per capita and infrastructure stocks, while stable, may be the result of a complex interaction of demand and supply effects.

Transportation infrastructure tends to have a different pattern from that of electricity and telephones. The provision of total roads and rail lines rises less than proportionately with population and the level of income per capita, though perhaps they are more sensitive to income in richer countries. This may mean that additional demand does not require additional supply. Below full capacity, the fact that someone else is using a road or rail line does not prevent my use. Both rail lines and total roads increase significantly with area. These results suggests that these forms of infrastructure may serve to link places together, with traffic flows that are generally below their capacity levels, and are consistent with the idea that they provide transportation services which are, to some extent, of a nonrival, public good nature.

Paved roads, on the other hand, increase almost one for one with population and more than proportionally with income on average, though the association between income and paved road provision seems small in poor countries and very large in rich countries. The area of the country has no significant relationship with paved road provision. This suggest that, at least in rich countries, paved roads are generally near full capacity and that increases in demand require additional supply. Despite, in general, being nonexcludable public goods, congestion may render paved road use a rival good. These interpretations assume, of course, that the level of infrastructure provision is close to the efficient level, which is by no means obvious. Given the large scale involvement of government, patterns in infrastructure stocks may be better explained by arguments from political economy than by economic efficiency.

As well as looking at cross section patterns in levels, I also look at what determines infrastructure growth rates over the period 1965–85. These growth regressions test the robustness of the cross section infrastructure relationships; if the cross-country relationships represent equilibrium conditions, the growth rates of the infrastructure stocks should respond to disequilibrium in the relationship. We find significant disequilibrium adjustment for every type of infrastructure.

Simple cross sections regressions aimed at explaining economic growth show a significant positive effect for the initial level of telephones per capita, though not for the other types of infrastructure. This result appears to be robust. However, it must be regarded as preliminary, since cross section regression of this type are sensitive to omitted variable bias, and do not allow for causality between the explanatory variables.

2. The Data

Telephones and Telephone Main Lines

The first approach to data collection took the form of collating numbers from various issues of publications from different international authorities. The main international data sources for each variable are given in appendix 2. For telephones, the various data sources are consistent with each other and produce time series that appear very consistent. The basic sources of telephone data are the International Telecommunications Union's *Yearbook of Common Carrier Statistics* and Associated Telephone and Telegram's publication *World Telephones*. These sources provide almost identical numbers and are in agreement with data from the United Nations' *Statistical Yearbook*. The only suspicious jump in the data is for Dominica, which loses 75 percent of its telephones in one year. Investigation revealed that this jump in the data was real, due to the effects of a hurricane.

The only difference between the data sources appears to be the point within the year at which the stock of telephones is measured. *World Telephones* measures stocks at 1st January of the relevant year, while the *Yearbook of Common Carrier Statistics* appears to measure at different dates for different countries. For many countries the number given by the *Yearbook of Common Carrier Statistics* appears as the 1st January stock of the succeeding year in *World Telephones*. I have used the *Yearbook of Common Carrier Statistics* as my basic source, since it has the most comprehensive coverage. Where other sources have been used to extend the series, or fill gaps, I have adjusted the year reported so that the overlapping part of the series agrees with the *Yearbook of Common Carrier Statistics* data.

Total telephones measures the number of telephone sets, and includes cases where subscribers share a line, while main lines are the number of lines connected to local telephone exchanges. The number of telephone main lines seems to be a better measure of the capacity of a telephone system, though in practice the two measures are highly correlated. In theory, a better measure of infrastructure stock might be the capacity of telephone exchanges. The development of the cellular phone from 1982 onwards, which does not require a main line, means that the infrastructure stock should reflect the area over which cellular calls are possible, as well as the number of such phones. No attempt has been made to distinguish cellular phones from others in the data reported here though data on this topic are available from the International Telecommunications Union.

Data on total telephones are fairly comprehensive for the period 1950–95, while data for 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. As a quality indicator I use the percentage of local calls which are unsuccessful in 1990. Where data for 1990 are not available I have used data for the nearest available later year (up to 1995).

Electricity

For electricity generating capacity the basic source is the United Nations' Energy Statistics, and the Statistical Yearbook. The time series seem good and are reported without adjustment, yielding a fairly complete data set for the period 1950–95. No manipulation of the raw data is carried out, since there is little to be gained by interpolation. Data for Namibia, Swaziland, Botswana, and Lesotho are included in the data for South Africa. These capacity data do not take into account the extent the electricity distribution system. As a quality indicator I give data on percentage of generated electricity lost in the system for 1971, 1980, and 1990.

Total Roads

A more troublesome case is that of the data for roads and paved roads. There are two international sources. The first is the International Roads Federation's *World Road Statistics* which 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 in the 1970s. A second source are the Statistical Yearbooks of the regional commissions of the United Nations. Where the two sources disagree, I have tended to use the data from the United Nations, since this is reported from official government sources.

Because of the many problems with these data, a detailed study of national sources was carried out. This increased the coverage of the data substantially over that found in the international sources and also produced 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 being filled from international sources where possible. However, in many cases study of the national sources brings to light the problems underlying discontinuities in the data.

The international data on total roads are patchy, with frequent gaps in the series and many large jumps in particular years, jumps that are quickly reversed. It seems that different countries have different definitions of "road," and this definition can change within countries over time. As well as variations in the definition of minimum quality standards for roads, there are differences in reporting that reflect the functional split of road management between central and local government. Large jumps in the series are often due to the source switching between jurisdictions, for example, from roads controlled by central government only to those controlled by central and provincial governments. In practice, intra-urban roads are often centrally controlled (when above a certain quality threshold), while urban roads are controlled by municipal authorities, leading to an underreporting of urban roads and low quality rural roads by the central authority.

As far as possible I have tried to ensure that the reported time series for total roads includes urban roads, and reflect total public road length in the country, independently of the controlling authority. Some results using total roads are reported here for comparison purposes, but, generally, the raw data reported for total roads seem to be too unreliable for practical use.

Paved Roads

I define "paved roads" to be concrete or bitumen-surfaced roads. It excludes stone, gravel, water-bound gravel, oil-bound gravel, and earth roads. This accords with the definition used by most countries. Where it does not, and there is sufficient information, the paved road data have been adjusted to fit this definition. For example, in some years, the United States includes gravel roads as paved roads and reports these in national and international sources. This increases the reported paved road stock by a factor of 2 over the actual paved road stock by my stricter definition. Similarly, China, in its official publications, uses a much wider definition of paved road than used here. The data I report for China for paved roads follow my narrower definition and come from a World Bank country report. In many cases detailed data on the type of road is not available and the national definition is used. When national sources give figures for "paved" or "hard-surfaced" and no other information is available, these categories have been accepted as equivalent to my definition of paved roads; not so, however, “all-weather roads.”

For many countries the international sources report only nonurban roads; for countries where both data sets are available, urban roads make up about 15–30 percent of the total paved road stock. Using national sources we have been able to construct, for some countries, total paved road stocks by adding urban and nonurban paved kilometers.

However, after all these adjustments there are still a number of large jumps and splits in the series. The raw data for paved roads again seem too inconsistent for practical use and further adjustment is required if we are to have series that are consistent over time. This involves linking together series when there are changes in definition to produce a series for total paved roads. Where this is not possible the series is for nonurban roads, if these data are available. In some countries, such as Ireland, road stocks are measured infrequently and reported road length is constant between measurements. In these cases repeat values have been deleted. The processed data are then interpolated over gaps of up to five years. The resulting processed data, documented as to source and coverage, are the best available at the moment.

Within the “paved road” category there may still be large variations in quality. In particular, no allowance is made for the width of the road, which varies from single lane to multiple lane highways. As a quality measure, I give the percentage of the main paved and unpaved road network considered to be in good, fair and poor condition by the World Bank in 1984 and 1988. These quality 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.

Rail

For railway track length the basic sources are Mitchell's *International Historical Statistics* for the continents, until 1980, and the World Bank thereafter, supplemented by national sources. The data refer to line length; note that 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 that are owned by companies for industrial use and are not open to the public (for example, the sugar industry railways in Latin America). To produce a consistent series, I have concentrated on railroads open to the public. The raw data are provided and the processed data again link series over data breaks and interpolates. In the processed data, countries which report no railways in national or international sources at any time over the period 1950–95 are assumed to have zero line length. As a quality index, the percentage availability of the stock of diesel locomotives measured between 1990 and 1995 is given.

Assessment and Recommendations

The data sets for telephones, telephone main lines, and electricity generating capacity seem to be excellent, and both the raw data sets and the manipulated (interpolated) data can be used without worry. The data for railways seem good, but for a number of countries series have had to be linked together to achieve consistency over time. Despite this, the data seem good enough to use; it is recommended that the processed data be used, particularly for time series work. On the other hand, the data for total roads are unreliable. Jumps in the series seem too numerous, and unexplained, to allow the construction of consistent series. Since most of the large jumps in the series are due to changes in definition, rather than actual changes in the stock of roads, these data are unsuitable for time series work. They may be used for cross section work if allowance is made for the fact that different definitions across country mean that the series contains large measurement error from an ideal measure of total roads. Similar comments can be made about the raw data series for paved roads. However, the processed paved road data overcome most of these problems. The series are consistent over time. In cross section, countries which report only an administrative sub-category of paved roads should not be used (labelled 4 in the data coverage table). This leaves countries which report either total paved roads or nonurban paved roads, or do not make clear precisely which they are reporting. Using all these data in cross section introduces an under reporting error of around 15 percent to 30 percent of the total road stock for those reporting only nonurban roads. Some adjustment to the data to try to allow for this underreporting could be carried out, or the data can be used as it is, provided measurement error in this range is acceptable.

3. Cross Country Relationships: Infrastructure, Income, and Geography

I begin by looking at cross sectional patterns in the data. Taking 1985 as a base year for comparison since it has a good coverage of our infrastructure measures, table 1 reports the correlation between infrastructure stocks. As expected, the infrastructure stocks are all positively correlated. The correlation between total telephones and telephone main lines is over 99 percent, suggesting that in practice using telephones rather than telephone main lines may not be too bad a proxy. Paved roads and total roads are correlated but not very closely indicating that one series may not be a good proxy for the other.

It seems likely that the stock of infrastructure in a country varies with population and GDP per capita. We can think of these variables affecting the demand for infrastructure, as well perhaps as the cost of providing it. In addition the geography of the country may matter. For example, Hong Kong and Singapore both have very low stocks of paved roads relative to their population size and their level of GDP per capita. It may be that in such city states the need for kilometers of road is low, given the high density of population. We can proxy the geography of a country using the percentage of the population living in urban centres, and the total area of the country. Of course these may not be good proxies for geography; 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 matters. However, for the purposes of aggregate analysis at the country level, we need to reduce these multidimensional factors into simple aggregate statistics. Clearly, more can be done along these lines, if we formulate hypotheses that can be tested using aggregate summary statistics of geography.

Table 2 reports results for OLS regressions on a cross section of countries in 1985, explaining infrastructure levels with these factors. All variables, other than the ratio of people in urban centres, are in logarithms, so the coefficients can be interpreted as elasticities. The t-ratios given are heteroskedastic consistent. For nontransportation infrastructure, the coefficient on population is significant and close to one, indicating that holding other factors constant, infrastructure rises in line with population. Except for total roads and railways, the coefficient on GDP per capita is above one, indicating that infrastructure stocks 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, other things being equal (population, GDP per capita, percentage population urbanized), a large country with more land area, has more roads, railway track, and electricity generating capacity, and fewer telephone main lines than a smaller country. We can explain this by noting that our infrastructure stock data measures subtly different things in each case.

Consider a country with two population centres that have to be linked by communication infrastructure. In a large country these are likely to be further apart, so the length of each link will be longer, and the cost of a link will be higher. It is quite likely that due to the higher cost per link, the total number of links in a large country will be smaller, while since the length of each linkage goes up, total length of links rises. Notice that we measure the number of telephone main lines, while we measure the length of roads in kilometers. It is quite possible that while large countries have fewer telephone main lines, the "length" of main lines (which we do not measure) is greater than in small countries.

Again, in the case of electricity, we measure generating capacity and not the number of connections, or the total length of the distribution system, so the geographical effects on this measure of infrastructure are likely to be quite different from the other cases. One explanation of the rise in generating capacity with area is that electricity distribution systems suffer leakage, which depends on the length of connection. In large countries, with low density of population, leakage can be avoided by having small local plants. This, however, may reduce the scope for economies of scale and increase the need for reserve capacity for peak periods, if transfers within the system are difficult.

Electricity and telephone provision tend to increase with urbanization, which can be explained by the lower cost of providing these services in an urban environment due to lower connections costs per consumer. Alternatively, the large urbanization effect may be due to urbanization acting as a proxy for industrial structure, with higher rates of urbanization being associated with more production in manufacturing and less in agriculture. If manufacturing output has a greater need for electric power than agriculture, industrial structure may be very important. The degree of urbanization does not seem to influence transportation infrastructure significantly, but the signs are as expected.

The large differences between the results for paved roads and for total roads leads to some interesting conclusions. One way of interpreting the results is to think of roads as simply connecting places, while paved roads handle large volumes of traffic. Holding other factors constant, an increase in area significantly increases road length while it has a statistically insignificant impact on paved road provision. Note that the average distance between two points increases by the square root of the increase in area, so we expect a coefficient of 0.5 on log area to correct for distances.

On the other hand, paved roads increase one for one (or more) with income while total roads increase much more slowly. This can be explained by the idea that unpaved roads link places, and these roads increase less than proportionately with population and income because they typically have spare capacity. Paved roads are built to handle large traffic flows and must increase in line with these flows. In many ways the results for railways are similar to those for total roads, suggesting they may serve similar functions.

However, the regression results for railways only cover those countries with a positive stock of railways; a more sophisticated approach would also include those countries with zero rail length.

The relationships may be more complex than set out in table 2. This is investigated in table 3, which reports regressions with GDP per capita squared and with interactive effects, which allow the geographical factors to vary with the level of GDP per capita. Only statistically significant variables are reported.

Using interactive effects the influenced of a variable on the stock of infrastructure depends on the level of GDP per capita. The range of log GDP per capita varies from around 5 to 10 in the sample. Substituting in a value for log GDP of 10 it appears that in the richest countries the effect of area and urbanization on electricity generating capacity and telephone provision disappears. It seems that the relationship between these types of infrastructure and population density and geography may be an issue only in poorer countries.

For roads we do find significant nonlinear effects of income on infrastructure levels. The elasticity of the roads paved stock with respect to income is estimated to be

$$-1.051 + 0.270 \text{ log GDP per capita}$$

In the poorest countries, with log GDP per capita of around 5, this gives an elasticity which is positive, but close to zero, while in the richest countries, with log GDP per capita over 10, it suggests an elasticity of nearly 2. The results for total roads are similar. This suggests the hypothesis that congestion effects lead to an increased need for roads, only after countries have reached middle income levels. This may mean that roads in developed countries are rival and should be treated more like a private good, than those in developing countries which are nonrival public good. For paved roads no significant geographical effects were found, though for total roads we still have large geographical effects. As in the case of other infrastructure, urbanization rates are important in poor countries, but not in rich countries. However, area seems important for total roads at all income levels. Railways again seem to have a threshold, rising with income only after middle income levels have been reached; again, land area seems an important factor.

While the R^2 for each of the five regressions is quite high, I make no claim that these regressions explain infrastructure levels. The regressions are merely designed to look at patterns in the data. In particular, they have nothing to say about directions of causation. It may well be the case that GDP per capita and urbanization rates depend on infrastructure provision. The important point, however, is that infrastructure provision is significantly correlated with geography, particularly for poorer countries, and it seems likely that this is 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 when we come to study these effects, geographical considerations should be taken into account.

To show that we have not explained infrastructure stocks fully, we need only look at table 4. This shows the correlation between the residuals of the six infrastructure regressions in table 3. Each correlation is positive, indicating that a country that has above average infrastructure of one type, given its characteristics, tends to have above average provision of the other types of infrastructure.[1] This is most likely due to some country specific variables that affect infrastructure but are excluded from our analysis; in particular, government policy is a possible source of such correlation.

When we come to applications, it is usual to normalise quantity variables so as to make them independent of the size of the country. This leaves open the appropriate size variable. For telephones, telephone main lines, electricity generating capacity, and perhaps paved roads, it seems reasonable to normalise by population, since each of these variables appears to increase one for one with population, on average. That is to say, we could rerun our regression in tables 2 and 3 using infrastructure stock per capita on the left, and not including population on the right, without changing the other coefficients. However, this does not hold for total roads and railways. For rival goods normalisation by population seems appropriate since the quantity of the good divided by the population indicates average consumption. However, for nonrival goods, normalising by population does not give average per capita consumption; increases in population, with a fixed stock of nonrival infrastructure need not reduce average consumption. If transportation infrastructure is really nonrival, then normalising by population is unlikely to be appropriate. Ingram and Liu (1997) normalize by area, but again this appears to have a coefficient less than one in our regressions; explaining total roads per km^2 of area requires area as an explanatory variable. A case could be made for using the square root of area on theoretical grounds, but the data appears to support a figure nearer to the cube root of

[1] While the residuals are correlated, running our four regressions as a system of a seemingly unrelated regressions (SUR) gives no efficiency gain over OLS since the regressors are the same in each case.

area. It is probably better to say for the moment there is no obvious normalization for our transportation infrastructure stocks.

4. Time Series Properties of the Infrastructure Data

In addition to the cross sectional pattern of the data, we can also look at the data in time series. An important issue in applications is whether time series variables are stationary or nonstationary. We could test our variables for a unit root directly. However, it is more interesting to test the level of infrastructure per capita, since this is the variable that will be used more frequently in applications. I carry out these tests for nontransportation infrastructure and paved roads per capita, excluding total roads and railway lines since it is not clear that per capita measures of these variables are meaningful.

One approach would be to examine the time series of each country separately and test if it is a unit root, but such tests have notoriously low power; moreover, it is difficult to know what to make of the results. For example, if we test for a unit root in 100 countries, we would expect 5 rejections of the null of a unit root at the 5 percent significance level, even if every series does in fact have a unit root. It is clear that in the event of a small number of rejections, we cannot conclude that the series for which we reject a unit root are, in fact, stationary.

An alternative approach is to follow Im, Pesaran, and Shin (1995 who develop a panel unit root test for the joint null hypothesis that every time series in the panel is nonstationary. This approach is to run a standard augmented Dickey -Fuller unit root test for each country and average the t-values of the test statistic found. If the data from each country are statistically independent then, under the null, we can regard the average t value as the average of independent random draws from a distribution with known expected value and variance (that is, those for a nonstationary series). This provides a much more powerful test of the unit root hypothesis than the usual single time series test; as N, the number of countries, gets large the average t-value converges quite quickly to the expected value, if the data are actually nonstationary. For large N even very small deviations of the average t value from the expected t value will lead us to reject nonstationarity. Note that, by running each augmented Dickey-Fuller regression separately before averaging, we allow each country to have its own short-run dynamics.

Im, Pesaran, and Shin recommend the removal of any common time effects by first regressing the variable on a set of time dummies and taking the residuals. This reduces the risk of correlation across countries, and is carried out for each of our variables. A problem with the approach is that it requires us to use a common time period for every country, each country having a complete run of data over the time period. To keep the number of countries relatively large, I have used shorter time periods in the case of telephones, telephone main lines, and paved roads. In each case the augmented Dickey Fuller regression, with a constant and time trend, and 5 or 7 augmenting lags, depending on the data period[2], is run using the residuals after common time effects have been removed. Results are reported in table 5.

For example, in the case of log GDP per capita we have 51 countries that have a complete data set over the period 1950–92. Running 51 augmented Dickey -Fuller regressions,

[2] These lag lengths seem the longest practicable given our data length and would seem to be sufficient to capture business cycle effects. For 90 percent of countries the Aitken Information Criterion (AIC) suggests these lag lengths are sufficient. For a few countries the AIC chooses implausibly long lag lengths (in excess of 15 years).

after removing common time effects, gives an average t value of -2.011. Under the null of nonstationarity, the t value in each country has an expected value of -2.035 with a variance of 0.728 (as tabulated by Im, Pesaran, and Shin). The test statistic calculated as the difference between the average t value and this expected value, and adjusted for the variance, has a N(0, 1) distribution under the null of nonstationarity, with large negative values indicating stationarity. It is clear that we cannot reject a unit root for log GDP per capita. However, when we repeat the process for changes in log GDP per capita, we find a lower average t statistic in our 51 countries, a value of –2.487, which is very unlikely under the null. The test statistic is -2.911, which gives a decisive rejection of nonstationarity. It follows that we can regard log GDP per capita as an I(1) series.

Turning to our infrastructure series per capita, it is clear we cannot reject a unit root for any of our variables in levels. However, in first differences, we reject a unit root in every case, at the 1 percent significance level for electricity generating capacity, telephone main lines and paved roads and at the 5 percent critical value for telephones. The importance of this result is that with nonstationary series, Pedroni (1995) shows that unless the relationship estimated is a cointegrating one, the estimated parameters in a panel regression with fixed effects converge asymptotically to zero. In addition, with nonstationarity, the usual t statistics are inconsistent in panels as well as in time series, so that inference is likely to be wrong. It follows that simple panel regression using our infrastructure variables in levels may give very misleading results. For example, it is tempting to run the type of regressions reported in tables 3 and 4 for the whole panel of data and allow for each country to have fixed effects. Unless the estimated relationships are cointegrating relationships, this will produce very misleading results. Note, however, that our cross section regressions, reported in tables 2 and 3, are consistent, since they do not use the time series dimension of the data.

If the cross section regressions generate stable relationships they may reflect cointegrating mechanisms for the data. That is, if a country is out of line given the cross section relationship, we might expect it to move towards this relationship in the long run. To address this question, while avoiding the pitfalls of panel estimation, table 6 reports regressions where the dependent variable is the growth rate of infrastructure stock over the period 1965–85. These regressions are similar to those of Barro (1991) for economic growth; a negative coefficient on the initial infrastructure stock indicates convergence of infrastructure stocks to an equilibrium level which depends on the other variables in the regression.

Infrastructure stocks in each country appear to be converging to the same equilibrium relationship, conditional on their values of the other initial condition included in the regressions. We have included the same level variables (but now for 1965) as in table 2. The long run steady state coefficients estimated from table 6 are remarkably similar to those found in table 2. For example, the effect of the initial conditions on the 20 years of growth of telephones can be written as

$$0.27\ (1.18 \log pop + 2.11 \log GDP/pop - 0.74\ urban - 0.35 \log area - \log telephones)$$

This implies that about 27 percent of the deviation between actual telephone stocks and the “equilibrium” telephone stock (when this expression is zero) is made up in a 20 year period. It appears that countries are converging to this relationship.

5. Economic Growth

In table 7, I report the results of simple cross section regressions aimed at explaining growth rates over the period 1970–90. This period was chosen because an earlier initial time period tends to reduce the size of the data set significantly. While a host of possible variables have been suggested for inclusion in such regressions, three commonly accepted factors are initial GDP per capita, initial education levels, and average investment rates over the period. In addition to the infrastructure variables, I include log area and the urbanization ratio in the regressions. The infrastructure variables are correlated with these and leaving them out may introduce omitted variable bias. All the regressions use an instrumental variables approach; I instrument the average investment rate over the period 1970–90 with average investment over the period 1960–70 to allow for reverse causation from growth to investment rates within the period.

Column one in table 7 reports a simple base-line regression and is compatible with the usual results that growth increases with education levels and investment rates and, holding these factors constant, is higher in poorer countries. Adding our infrastructure per capita, and geography variables (in column 2) suggests that telephone main lines per capita have a significant positive impact on subsequent growth rates of GDP per capita, while electricity generating capacity and area have negative impacts. In column 3 we report results after stepwise removal of insignificant variables.

The results in table 7 are suggestive, but they are not sufficient to conclude that telephones aid growth or that the other infrastructure variables have no effect. Telephones may simply be a proxy for some unobserved variable and not directly causal. Paved roads and electricity generating capacity may influence growth in more subtle ways than we have tried to estimate here. For example, they may only be important with particular combinations of geographical factors at early stages of development, or may influence growth through investment. More work needs to be done to have any confidence in the results found in this table. In particular, the time dimension of the data constructed here allows Canning (1997) to examine the timing of infrastructure growth and economic growth, and so carry out Granger causality tests.

6. Conclusion

We have presented a new panel data set on stocks of infrastructure in a cross section of countries over time. It is clear that the stock of infrastructure across countries varies significantly with their population size, income level, and geography, and this relationship appears stable over time. The infrastructure stock per capita series appear to be I(1) in every country. Simple cross country growth regressions suggest that the number of telephone main lines per capita does have a significant impact on subsequent growth rates of GDP per capita but that the other infrastructure variables do not, though this conclusion must be regarded as very preliminary.

The data sets for electricity generating capacity, telephones, telephone main lines and railway lines for 1950–90 are fairly complete. They may undergo some minor additions, if data becomes available to fill gaps in the data, but is unlikely that existing data will change. The data for total roads and paved roads may change substantially for a number of countries if better data sources are uncovered at the national level. In addition, it is hoped that at some future date a more definite classification of the coverage of the paved road stock variable will be possible.

APPENDIX 1

World Infrastructure Data
Infra 1.0

The data set is available in downloadable form from the World Bank Website. The specific URL address for the data set is shown in the abstract of this paper.

The data files are in 11 Microsoft Excel 97 spreadsheets in a single file of about 550 kilobytes.

The data runs annually from 1950 to 1995 and cover the same 152 countries as the Penn World Tables 5.6. The codes used for the countries are the same as in the Penn World Tables 5.6.

DATAFILES

Filename	Description, units	Data type
ROADFIN	Total roads, kilometers	Raw data
PAVFIN	Paved roads, kilometers	Raw data
PAVWORK	Paved roads, kilometers	Manipulated data
EGCFIN	Electricity generating capacity, thousand kilowatts	Raw data
RAILFIN	Rail line length, kilometers	Raw data
RAILWORK	Rail line length, kilometers	Manipulated data
TELFIN	Telephones, number	Raw data
TELWORK	Telephones, number	Manipulated data
MAINFIN	Telephone main lines, number	Raw data
MAINWORK	Telephone main lines, number	Manipulated data

Raw data is as it appears in sources. Manipulated data has been linked together over breaks to give data that is consistent over time, and gaps of up to five years have been interpolated (linear interpolation in logs).

The file INFQUAL gives information on quality of the infrastructure data.

The first two columns give the source and coverage of the paved road data.

Code	Data source	Data coverage
0	No source	No coverage
1	National source	All paved roads
2	United Nations	All paved roads / Unclear
3	International Road Federation	Nonurban paved roads
4	World Bank Data	Subclass of paved roads

The later columns give data on infrastructure quality:

% local telephone calls unsuccessful 1990
% paved main roads in fair condition 1984
% paved main roads in poor condition 1984
% unpaved main roads in fair condition 1984
% unpaved main roads in poor condition 1984
% paved main roads in fair condition 1988
% paved main roads in poor condition 1988
% unpaved main roads in fair condition 1988
% unpaved main roads in poor condition 1988
% of diesel locomotives available 1990–95
% of system electricity loses 1971, 1980, 1990

The percentage of the road network in good condition can be calculated as 100 minus the percentages in fair and poor condition.

APPENDIX 2

Data sources

GDP, GDP per capita: Penn World Table (5.6). Available from NBER, Boston.

AREA, land area of country, World Bank, World Tables, 1995.

URBAN, percentage of population living in urban areas, World Bank, World Tables, 1995.

EGC, Kilowatts of Electricity Generating Capacity: United Nations (various years), Energy Statistics, New York. United Nations (various years) Statistical Yearbook.

TEL, Number of Telephones: International Telecommunications Union (various years), Yearbook of Common Carrier Statistics, Geneva. Associated Telephone and Telegram (various years) World Telephones. United Nations (various years) Statistical Yearbook. National Sources.

TELMAIN Number of Telephone Main Lines: International Telecommunications Union (various years), Yearbook of Common Carrier Statistics, Geneva. Associated Telephone and Telegram (various years) World Telephones.

PAV, Kilometers of Paved Roads: International Road Federation (various years), World Road Statistics, Washington DC. United Nations Economic Commissions (various years) African

Statistical Yearbook, Statistical Yearbook for Arab Countries, Statistical Yearbook for Asia and the Pacific, Statistical Yearbook for Latin America, Bulletin of Transport Statistics for Europe. National Sources.

ROAD, Kilometers of Roads: International Road Federation (various years), World Road Statistics, Washington DC. United Nations Economic Commissions (various years) African Statistical Yearbook, Statistical Yearbook for Arab Countries, Statistical Yearbook for Asia and the Pacific, Statistical Yearbook for Latin America, Bulletin of Transport Statistics for Europe. National Sources.

To allow better use of the database, I provide a listing of the data source for each country and the coverage of the data in the processed data file for paved roads. The source indicates the best source used, national data, usually from a national statistical abstract, United Nations data, or World Road Federation data, in that order. Note that data from other sources may also be included in the raw data file; the processed data file uses data overlaps to link series together based on the best source. The column on "Coverage of the data" indicates if the data represents all paved roads, nonurban paved roads, or some sub-class of paved roads, usually based on an administrative classification. The coverage reported is based on national sources when the coverage of the data is made clear by reference to different categories of roads. Unfortunately, in many cases, the national sources simply state the stock of paved roads, and no explicit discussion of coverage takes place. In this case I label coverage as "total paved roads/unclear."

RAIL , Kilometers of Rail Track: Mitchell, International Historical Statistics. World Bank Rail Statistics Database.

% local telephone calls unsuccessful. International Telecommunications Union, World Telecommunications Indicators on Diskette.

% paved main roads in poor condition. United Nations, Survey of Economic and Social Conditions in Africa 1988–89.

% total main roads in poor condition. United Nations, Survey of Economic and Social Conditions in Africa 1988–89. P. Gyamfi, Infrastructure Maintenance in LAC: the cost of Neglect and Options for Improvement, Vol. 4 The Road Sector, Latin America and the Caribbean Technical Department, report 17, World Bank, 1992. Road Deterioration in Developing Countries, A World Bank Policy Study, 1988.

% of diesel locomotives available. World Bank Rail Statistics Database.

% of system electricity loses. World Bank, World Development Indicators 1997.

References

Antle, J. 1983. “Infrastructure and Aggregate Agricultural Productivity: International Evidence.” *Economic Development and Cultural Change* 31.

Aschauer, D.A. 1989. "Is Public Expenditure Productive." *Journal of Monetary Economics* 23.

Barro, R.J. 1991. "Economic Growth in a Cross Section of Countries." *Quarterly Journal of Economics* 106:407–44.

Canning, D. 1997. “Does Infrastructure Cause Economic Growth.” Queen's University of Belfast, Department of Economics. Mimeo.

Canning, D., P. Dunne, and M. Moore. 1995. "Testing the Augmented Solow and Endogenous Growth Models." Working Paper 53. Queen's University of Belfast, Department of Economics.

Canning, D., and M. Fay. 1992. “Infrastructure and Economic Growth.” Columbia University, New York. Mimeo.

Canning, D., M. Fay, and R. Perotti. 1992. "Dotazioni Infrastrutturali e Crescita Economica." *Revista di Politica Economica* LXXXII(November):117–54.

_____. 1994. “Infrastructure and Growth.” In M. Baldassarri, L. Paganetto, and E. Phelps, eds., *International Differences in Growth Rates.* Macmillan Press Ltd.

Chow, G.C. 1993. "A Two Step Procedure for Estimating Linear Simultaneous Equations with Unit Roots." *Review of Economics and Statistics* 75:107–11.

Easterly, W., and S. Rebelo. 1993. “Fiscal Policy and Economic Growth: An Empirical Investigation.” *Journal of Monetary Economics* 32:417–58.

Engle, R.F., and C.W.J. Granger. 1987. “Co-integration and Error Correction: Representation, Estimation, and Testing." *Econometrica* 55:251–76.

Gramlich, E.M. 1994. “Infrastructure Investment: A Review Essay.” *Journal of Economic Literature* 32: 1176–96.

Hardy, A., and H. Hudsen. 1981. “The Role of the Telephone in Economic Development: An Empirical Analysis.” International Telecommunication Union, Geneva.

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. Mimeo.

Im, K.S., M.H. Pesaran, and Y. Shin. 1995. "Testing for Unit Roots in Heterogeneous Panels." Cambridge University, Department of Applied Economics. Mimeo.

Ingram, G.K., and Z Liu. 1997. “Motorization, Road Provision, and Economic Growth.” World Bank, Washington, DC. Mimeo.

Jimenez, E. 1995. “Human and Physical Infrastructure: Investment and Pricing Policies in Developing Countries.” In J. Behrman and T.N. Srinivasan, *The Handbook of Development Economics,* vol. 3B.

Johansen, S. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–54.

_____. 1991. "Estimation and Hypothesis Testing of Cointegration Vectors in Gausian Autoregressive Models." *Econometrica* 59:1551–80.

Lee, K.S., and A. Anas. 1992. “The Impact of Infrastructure Deficiencies on Nigerian Manufacturing.” Infrastructure Department Working Paper INU 98. World Bank, Washington DC.

Pedroni, P. 1995. "Panel Cointegration." Working Paper in Economics 95–013. Indiana University.

Pritchett, L. 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. World Bank, Policy Research Department, Washington, DC.

Querioz, C., and S. Gautman. 1992. “Road Infrastructure and Economic Development: Some Diagnostic Indicators.” Policy Research Working Paper 921. World Bank, Washington, DC.

Summers, R., and A. Heston. 1991. "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–88. *Quarterly Journal of Economics* CVI:327–36.

World Bank. 1994. *World Development Report 1994: Infrastructure for Development.* Washington, DC.

TABLE 1
CORRELATION OF INFRASTRUCTURE LEVELS 1985

	TEL	TELMAIN	EGC	PAVED ROAD	ROAD	RAIL
TEL	1					
TELMAIN	0.99	1				
EGC	0.96	0.96	1			
PAVED ROAD	0.69	0.70	0.83	1		
ROAD	0.54	0.53	0.71	0.83	1	
RAIL	0.36	0.37	0.59	0.79	84	1

TABLE 2
CROSS COUNTRY PATTERNS OF INFRASTRUCTURE 1985

Dependent Variable

	LOG TEL	LOG TELMAIN	LOG EGC	LOG PAVED	LOG ROAD	LOG RAIL
constant	-7.93 (11.0)	-8.22 (13.2)	-13.0 (21.0)	-8.10 (11.1)	-0.771 (11.1)	-5.58 (7.63)
log Population	1.001 (17.5)	0.997 (21.5)	0.931 (21.5)	0.798 (9.98)	0.544 (9.14)	0.552 (9.36)
log GDP per capita	1.479 (13.9)	1.442 (14.8)	1.398 (17.6)	1.243 (10.8)	0.583 (6.57)	0.820 (8.42)
Urbanized ratio	0.891 (2.07)	1.221 (3.10)	1.167 (3.96)	-0.661 (1.25)	-0.217 (0.56)	-0.634 (1.08)
log area	-0.107 (2.50)	-0.095 (2.60)	0.081 (2.41)	0.108 (1.50)	0.335 (6.72)	0.352 (5.24)
N	126	144	142	112	133	106
R^2 Adjusted	0.925	0.938	0.932	0.842	0.866	0.721

Heteroskedatic-consistent t ratios in parenthesis

TABLE 3
CROSS COUNTRY PATTERNS OF INFRASTRUCTURE WITH INTERACTIVE EFFECTS 1985
Dependent Variable

	Log TEL	Log TELMAIN	Log EGC	Log PAVED ROAD	Log ROAD	Log RAIL
Constant	-7.48 (4.10)	-7.56 (4.65)	-18.1 (12.8)	1.02 (0.24)	10.9 (2.28)	4.94 (1.03)
log Population	1.025 (18.1)	1.021 (22.3)	0.923 (22.3)	0.873 (18.9)	0.573 (10.1)	0.555 (9.36)
Log GDP per capita	1.408 (5.70)	1.346 (6.04)	2.044 (11.3)	-1.051 (0.97)	-2.344 (1.86)	-1.837 (1.56)
Log GDP per capita Squared				0.135 (1.94)	0.178 (2.14)	0.160 (2.23)
Urbanized ratio	5.560 (2.25)	5.027 (2.28)	6.721 (3.76)		6.673 (2.31)	
Urbanized ratio x log GDP per capita	-0.541 (1.82)	-0.445 (1.65)	-0.699 (3.43)		-0.803 (2.38)	
log area	-0.499 (2.11)	-0.511 (2.58)	0.602 (2.72)		-0.272 (1.04)	0.335 (4.82)
log area x log GDP per capita	0.046 (1.72)	0.049 (2.21)	-0.063 (2.53)		0.070 (2.20)	
N	126	144	142	116	133	106
R^2 Adjusted	0.930	0.941	0.937	0.843	0.876	0.728

Heteroskedastic-consistent t ratios in parenthesis

TABLE 4
CORRELATION OF INFRASTRUCTURE RESIDUALS 1985

	TEL	TELMAIN	EGC	PAVED ROAD	ROAD	RAIL
TEL	1					
TELMAIN	0.94	1				
EGC	0.54	0.57	1			
PAVED ROAD	0.50	0.47	0.38	1		
ROAD	0.47	0.41	0.37	0.61	1	
RAIL	0.26	0.24	0.27	0.36	0.44	1

TABLE 5
PANEL UNIT ROOT TESTS

Variable	Period	Number of countries	lags	Average adf	Test Statistic
log GDP per Capita	1950-1992	51	7	-2.011	-0.648
Δ log GDP per Capita	1951-1992	51	7	-2.487	-2.911**
log EGC per Capita	1950-1992	43	7	-1.857	0.077
Δ log EGC per Capita	1951-1992	43	7	-2.671	-3.441**
log TELMAIN per Capita	1960-1990	53	5	-1.656	1.499
Δ log TELMAIN per Capita	1961-1990	53	5	-2.239	-2.514**
log TEL per Capita	1960-1990	67	5	-1.333	4.192
Δ log TEL per Capita	1961-1990	67	5	-2.172	-2.310*
log PAV per Capita	1960-1990	31	5	-1.686	0.987
Δ log PAV per Capita	1961-1990	31	5	-2.768	-4.716**

* (**) significant at 5% (1%) level

TABLE 6
INFRASTRUCTURE GROWTH REGERESSIONS 1965-1985

	Dependent Variable					
	Growth Rate of TEL	Growth Rate of TEL MAIN	Growth Rate of EGC	Growth Rate of PAVED ROAD	Growth rate of ROAD	Growth Rate of RAIL
Constant	-2.929 (4.70)	-1.978 (2.41)	-5.071 (5.44)	-2.668 (2.27)	-2.015 (3.35)	-0.492 (0.88)
Growth of POP	1.034 (4.03)	1.072 (3.22)	1.341 (4.10)	-0.326 (0.96)	0.967 (3.34)	0.254 (0.94)
Growth of GDP per capita	0.926 (8.29)	0.859 (6.23)	0.982 (6.74)	0.887 (5.72)	0.040 (0.36)	0.389 (3.74)
Change in Urbanization ratio	2.645 (4.33)	0.028 (3.84)	2.030 (2.77)	-0.813 (0.86)	1.228 (1.97)	-0.964 (1.59)
Log POP 1965	0.320 (5.30)	0.316 (4.17)	0.323 (4.63)	0.305 (3.87)	0.276 (6.72)	0.008 (0.18)
Log GDP per capita 1965	0.572 (5.88)	0.341 (2.69)	0.592 (5.39)	0.619 (4.02)	0.323 (4.31)	0.085 (1.17)
Urbanization Ratio 1965	-0.200 (0.787)	0.000 (0.08)	0.011 (0.03)	-1.372 (2.87)	-0.292 (1.12)	-0.192 (0.74)
Log area	-0.094 (3.45)	-0.128 (3.91)	0.036 (1.04)	0.201 (4.70)	0.041 (1.04)	0.106 (2.83)
Log 1965 stock of relevant infrastructure	-0.271 (4.75)	-0.202 (2.58)	-0.343 (5.71)	-0.513 (9.00)	-0.304 (5.27)	-0.117 (2.09)
N	105	79	113	80	79	85
Adjusted R^2	0.682	0.642	0.553	0.712	0.574	0.182

t- ratios in parenthesis

TABLE 7
CROSS COUNTRY GROWTH REGRESSIONS
Dependent Variable: Growth in GDP per Capita 1970-1990
Two Stage Least Squares

Constant	1.148 (3.45)	1.061 (0.77)	3.028 (4.31)
log GDP per capita 1970	-0.211 (3.89)	-0.303 (1.92)	-0.449 (4.23)
log education per worker 1970	0.152 (1.67)	0.197 (1.60)	0.069 (0.78)
average investment rate 1970-1990	0.036 (3.11)	0.029 (3.04)	0.030 (3.12)
log telephone main lines per capita 1970		0.226 (2.41)	0.169 (2.51)
log paved roads per capita 1970		-0.012 (0.26)	
log electricity generating capacity per capita 1970		-0.230 (2.05)	
Log area		-0.062 (2.88)	-0.052 (2.64)
Ratio Urbanized 1970		0.185 (0.75)	
N	88	58	72
		72	
R^2 adjusted	0.386	0.472	0.456

heteroskedastic consistent t ratios in parenthesis
investment rate 1970-1990 instrumented with investment rate 1960–70

Summary

This paper contains a description of an annual database of physical infrastructure stocks constructed for a cross section of 152 countries for the period 1950-1995. The data base contains 6 measures: kilometers of road, kilometers of paved road, kilometers of railway line, number of telephones, number of telephone main lines, and electricity generating capacity. The data are available on the accompanying diskette “Infra 1.0.” Some measures of infrastructure quality, such as the percentage of roads in poor condition, percentage of local telephone calls unsuccessful, percentage availability of diesel locomotives, and percentage of electricity lost from the distribution system are also included. However, these quality measures are usually available only for recent years.

As well as describing the construction of the data, I look at correlation patterns and report regressions relating infrastructure stocks to countries’ population, GDP per capita, land area, and percentage urbanization.

Nontransportation infrastructure stocks tend to increase one for one with population but increase more than proportionately with GDP per capita. Geographical factors appear to affect nontransportation infrastructure provision in poor countries but not in rich countries. Transportation infrastructure appears to increase less than proportionately with population and increases with income only after a middle income threshold has been reached. Geographical factors seem to influence total roads and rail line length, but not paved roads length.

I carry out panel unit root tests and find that the log infrastructure stock per capita series are nonstationary and have a unit root. Cross Section growth regressions indicate that our common worldwide estimated regression results for infrastructure stocks are stable long run relationships. Preliminary regression results suggest that the telephone main lines per capita have a positive effect on subsequent economic growth.

David Canning
Harvard Institute for International Development
One Eliot Street
Cambridge MA 02138

dcanning@hiid.harvard.edu