The Manchester School Vol 76 No. 5 Special Issue 2008
1463–6786 504–527

INFRASTRUCTURE, LONG-RUN ECONOMIC GROWTH AND CAUSALITY TESTS FOR COINTEGRATED PANELS

by
DAVID CANNING
Harvard University
and
*PETER PEDRONI**
Williams College

We investigate the consequences of various types of infrastructure provision in a panel of countries from 1950 to 1992. We develop new tests which enable us to isolate the sign and direction of long-run effects in a manner that is robust to the presence of unknown heterogeneous short-run causal relationships. We show that while infrastructure does tend to cause long-run economic growth, there is substantial variation across countries. We also provide evidence that each infrastructure type is provided at close to the growth-maximizing level *on average* globally, but is under-supplied in some countries and over-supplied in others.

1 Introduction

We address the issue of whether stocks of infrastructure are at, above or below their growth-maximizing levels. Our approach is based on the growth model of Barro (1990). Infrastructure capital is an input into aggregate production, but it comes at the cost of reduced investment in other types of capital. In this approach there is an optimal level of infrastructure which maximizes the growth rate; if infrastructure levels are set too high they divert investment away from other capital to the point where income growth is reduced.

This model implies a simple 'reduced-form' relationship between income per capita and infrastructure stocks per capita. Below the growth-maximizing

* An earlier working paper version of this paper circulated with the title 'The Effect of Infrastructure on Long Run Economic Growth'. This paper has been funded in part by the World Bank and in part by the USAID CAER II project whose assistance is gratefully acknowledged. The views and interpretations expressed in this paper are those of the authors and should not be attributed to either the World Bank or USAID. Part of the research for this paper was completed while Peter Pedroni was appointed as a non-resident Clayton Fellow for the James A. Baker III Institute for Public Policy, and the author gratefully acknowledges the Institute's support. We would like to thank Esra Bennathan, George von Furstenberg, Eric Leeper, Gary Madden, Salvador Ortigueira and Lucie Schmidt for their comments. We also thank seminar participants at the University of Manchester Conference on Growth and Business Cycles in Theory and Practice, The Bank of England, Indiana University, Rice University, SUNY Albany, University of New Hampshire, Williams College, the North American Summer Meetings of the Econometric Society, the Cornell-Penn State Annual Macro Conference and the Midwest Macro Meetings at the University of Pittsburgh.

Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

infrastructure level positive shocks to infrastructure will tend to increase the level of output, while above the optimal level positive infrastructure shocks will tend to reduce the level of output; this can identify where each country's infrastructure stock stands relative to the growth-maximizing level.

Other empirical works such as Kocherlakota and Yi (1996, 1997) study the relationship between shocks to public capital and subsequent changes in GDP in the USA and the UK over the last 100 years. Our approach builds upon similar methodologies in the sense that we use a similar appeal to growth models for the causal connection between infrastructure and growth. But our empirical approach is distinct in that we exploit a cointegrated panel framework to isolate the direction of long-run causality as well as the sign of the long-run causal effect. Our approach allows us to address the question of whether infrastructure levels have been too low, too high or about right over the sample period, which runs from 1950 to 1992 for a large panel of emerging and developed economies.

We also use a number of innovations in terms of the actual data. First, we use physical measures of infrastructure, kilometers of paved roads, kilowatts of electricity-generating capacity and number of telephones rather than constructing stock estimates from investment flows. While simple physical measures do not correct for quality, monetary investment in infrastructure may be a very poor guide to the amount of infrastructure capital produced (see Pritchett, 1996).

We find that in general both short-run and long-run causality is bi-directional, with infrastructure responding to GDP per capita but GDP per capita also responding to infrastructure shocks. Most importantly, we find evidence of a long-run impact of infrastructure on GDP per capita. For telephones and paved roads, the sign of the effect of an increase in provision on GDP per capita varies across countries, being positive in some but negative in others. *On average* telephones and paved roads are supplied at around the growth-maximizing level, but some countries have too few while others have too many. It follows that the appropriate policy at the country level will depend on country-specific studies. Our finding that some countries actually have too much infrastructure is consistent with Devarajan *et al.* (1996) and Ghali (1998), who find evidence of over-provision of public capital in a number of developing countries. We find that long-run effects of investment in electricity-generating capacity are positive in a large number of countries, with negative effects being found in only a few. This suggests that, on average, electricity may be under-provided.

Our approach allows us to make inferences based on a very simple reduced-form relationship. This avoids the problems of estimating the effect of infrastructure on output in a more fully specified structural growth model (e.g. Easterly and Rebelo, 1993; Gramlich, 1994; Holtz-Eakin, 1994; Holtz-Eakin and Schwartz, 1995; Turnovsky and Fisher, 1995; Garcia-Mila *et al.*, 1996;

Holtz-Eakin and Lovely, 1996; Morrison and Schwartz, 1996; Ghali, 1998).[1] While our approach has the advantage of simplicity, it tells us only the direction of the net effect of infrastructure on growth, not its magnitude.

In the next section we present a stylized growth model to motivate the empirical approach that is undertaken in this study. In particular, we derive our simple estimated relationship as a reduced form of a growth model. In Section 3, we carry out panel-based unit root and cointegration tests to characterize the time-series properties of our data that are relevant for our subsequent analysis. In Section 4 we discuss how we test for the sign and the direction of long-term effects between our variables and present the results of these tests. Section 5 concludes.

2 A Stylized Growth Model with Infrastructure Capital

Our model is adapted from Barro (1990). We add stochastic disturbances to his structural equations and investigate how this affects the reduced form. The simple model form presented here is for illustrative purposes. As we shall see, our estimation procedure actually allows for a somewhat more general structure. Aggregate output Y, at time t, is produced using infrastructure capital G, other capital K and labor L such that

$$Y_t = A_t K_t^{\alpha} G_t^{\beta} L_t^{1-\alpha-\beta} \tag{1}$$

where A_t is total factor productivity at time t. For simplicity we assume a constant savings rate, s, and that both types of capital fully depreciate each period. Next period's infrastructure is a proportion of savings (perhaps through a tax system or by a private sector mechanism that channels investment based on private returns), so that

$$G_{t+1} = \tau_t Y_t \tag{2}$$

Investment in non-infrastructure capital is determined by

$$K_{t+1} = (1-\tau_t) s Y_t \tag{3}$$

Substituting the capital accumulation equations (2) and (3) into the production function (1) produces a difference equation for the evolution of per capita output

$$(Y/L)_{t+1} = A_{t+1} s^{\alpha} (1-\tau_t)^{\alpha} \tau_t^{\beta} (Y/L)_t^{\alpha+\beta} (L_t/L_{t+1})^{\alpha+\beta} \tag{4}$$

To complete the model, we need to describe the evolution of technical progress, A_t, the share of investment going to infrastructure, τ_t, and the size of the workforce, L_t. We assume that each of these is determined by an

[1]For more detailed discussions, see the longer working paper version of this study, Canning and Pedroni (1999).

exogenous stochastic process. We model the log of total factor productivity, a_t, as

$$a_t = a_0 + \sigma t + \varepsilon_t \tag{5}$$

where $\varepsilon_t = \delta\varepsilon_{t-1} + w_t$ for some $0 \leq \delta \leq 1$, and w_t is a stationary random variable with $E(w_t) = 0$. Thus, log total factor productivity depends on a constant, a_0, a trend rate of growth, $\sigma \geq 0$, which we take to be zero when $\delta = 0$, plus a random term that is stationary if $\delta < 1$ and non-stationary if $\delta = 1$.

We assume that the proportion of investment going to infrastructure is $\tau_t = \bar{\tau} + \mu_t$ where μ_t is a zero mean stationary series. Finally we assume that the growth rate of population is given by $\log(L_{t+1}/L_t) = \bar{n} + n_{t+1}$, where n_t is a zero mean stationary series. We further assume that we can identify the workforce with the total population. Alternatively, we can easily weaken this to an assumption that the labor force participation rate is a stationary series. Under these assumptions, our difference equation can then be written in terms of log income per capita, y, as

$$y_{t+1} = c + (\alpha + \beta)\, y_t + v_{t+1} \tag{6}$$

where $c = \alpha_0 + \sigma t + \alpha \log \quad s - (\alpha + \beta)\bar{n}$ and $v_{t+1} = \varepsilon_{t+1} + \alpha \log(1 - \bar{\tau} - \mu_t) + \beta \log(\bar{\tau} + \mu_t) - (\alpha + \beta)\, n_{t+1}$.

Note that all the random terms in equation (6) are stationary, except possibly total factor productivity, ε_{t+1}. According to equation (6) the process for y_t contains a unit root whenever $\delta = 1$ and $\alpha + \beta < 1$, or $\delta < 1$ and $\alpha + \beta = 1$. We require that one of these two mechanisms operates to explain the very persistent unit-root-type behavior in per capita income that we observe in the data, but we remain agnostic as to which one is appropriate for any particular country of our sample.

Similarly, the process for infrastructure formation can be written in log per capita form as

$$g_{t+1} = \bar{\tau} + y_t + \mu_t - n_{t+1} \tag{7}$$

We can rewrite this as

$$g_{t+1} - \bar{\tau} - y_{t+1} = \Delta y_{t+1} + \mu_t - n_{t+1} \tag{8}$$

If y_t has a unit root, Δy_t is stationary, as are the remaining error terms in the relationship. In this case, g and y are cointegrated, since a linear combination of g and y produces a stationary variable. This will be true regardless of which assumption we use to generate the unit root in y. However, in the exogenous growth version, the driving force behind growth is technical progress, and long-run infrastructure levels simply follow income levels. In the endogenous growth model, on the other hand, there is the possibility that shocks to infrastructure investment have permanent effects on the level of income.

Furthermore, the sign of this permanent effect may be positive or negative, depending on whether $\bar{\tau}$ has been set above or below the tax rate that maximizes expected growth. Note that expected growth is maximized when the average share of investment in infrastructure is set at the level τ^* that maximizes the expected value of $\alpha\log(1-\bar{\tau}+\mu_t)+\beta\log(\bar{\tau}+\mu_t)$. In general this depends on the distribution of the shocks. However, without shocks, setting $\tau^* = \beta/(\alpha+\beta)$ maximizes the growth rate, as in Barro (1990).[2] We assume a fixed savings rate so that investment in infrastructure represents a diversion from other types of capital. In practice, setting a suboptimal level of τ_t reduces the rate of return to capital as a whole and may reduce the savings rate and further lower the growth rate. Furthermore, it is important to notice that by treating the savings rate as fixed, we are in effect taking the key margin over long periods to be the allocation between different types of investment, rather than between the total level of investment and consumption.

We now summarize each of these results in the following proposition.

Proposition 1: For the model specified by equations (1)–(8), we have the following.

(i) If $\delta = 1$ and $\alpha + \beta < 1$, or $\delta < 1$ and $\alpha + \beta = 1$, then log per capita output, y_t, and log per capita infrastructure series, g_t, will each be non-stationary and integrated of order one, but there will exist a cointegrating vector (possibly different for each country) such that some linear combination of g_t and y_t will be stationary. Shocks to productivity have a long-run positive effect on log per capita output y_t.
(ii) If $\delta = 1$ and $\alpha + \beta < 1$, then shocks to per capita infrastructure, μ_t, have no long-run effect on per capita output, y.
(iii) If $\delta < 1$ and $\alpha + \beta = 1$, then shocks to per capita infrastructure, μ_t, will have a non-zero long-run effect on per capita output, y_t. For small shocks, the sign of this effect will be positive if $\bar{\tau} < \tau^*$, and negative if $\bar{\tau} > \tau^*$.

The proof of Proposition 1 is straightforward and is located in the Appendix. In the neoclassical version of the model, shocks to infrastructure have no long-run effect. In the endogenous growth version of the model, a positive shock to infrastructure increases income per capita when $\bar{\tau} < \tau^*$, and decreases income per capita when $\bar{\tau} > \tau^*$. It should be noted that all of our results are for small changes to infrastructure investment, since large changes could conceivably move the system across the optimal infrastructure level into a different regime.

[2]In Barro's model this is also the welfare-maximizing infrastructure level. However, in the presence of shocks, increasing expected growth may also increase the volatility of the growth rate. If agents are risk averse, maximizing expected growth need not maximize expected welfare.

Given these results for the reduced-form structure of the model, we can estimate a bivariate relationship between income per capita and infrastructure stocks per capita, and test which version of the model best describes the long-run properties of the data. The model described in this section represents a typical country of our data set. To apply the model to a panel of countries we assume that all variables and innovations terms in the model carry a double index i, t to represent the value of the variable in country i at time t. Furthermore, any parameters of the model are assumed to be indexed by an i subscript, so that we allow all of these to vary across countries. These include, for example, the income share parameters of the production function, α and β, the savings rate, s, the average share of infrastructure investment, $\bar{\tau}$, and the persistence of the technology shock δ.

Finally, we should emphasize that for our empirical implementation we simply require that the data be characterized by the properties described in the results of Proposition 1. This characterization can be expected to apply to a broad class of similar models.

3 The Data

Our data are annual and cover the period 1950–92. We use GDP per worker from the Penn World Tables 5, 6 (see Summers and Heston, 1991). The infrastructure data are from Canning (1998), who gives physical infrastructure measured on an annual basis, in kilometers of paved road, kilowatts of electricity-generating capacity and the number of telephones.

We deflate each variable by population so as to obtain per capita values, and then take logs of these per capita values. This means we have variables representing log GDP per capita, log paved roads per capita, log electricity-generating capacity per capita and log telephones per capita. If the services provided by the infrastructure stocks are a rival good, then these simple measures can be thought of as the average consumption of infrastructure services per capita. We begin by investigating the time-series properties of the data.

3.1 Testing for Unit Roots

We wish to test for non-stationarity against the alternative that the variable is trend stationary, where we allow different intercepts and time trends for each country. We use the unit root test proposed by Im *et al.* (2003), which allows each panel member to have a different autoregressive parameter and short-run dynamics under the alternative hypothesis of trend stationarity. To carry out the unit root and cointegration tests, we select countries and time periods for each variable to construct a balanced panel, which entails

TABLE 1
PANEL UNIT ROOT TESTS

Variable	*Period*	*Number of countries*	*Average ADF*	*Test statistic*
log GDP per capita	1950–92	51	−2.164	−1.116
log EGC per capita	1950–92	43	−1.908	0.160
log TEL per capita	1960–90	67	−1.333	4.192
log PAV per capita	1961–90	42	−1.815	0.291
Δ log GDP per capita	1951–92	51	−2.465	−3.465***
Δ log EGC per capita	1951–92	43	−2.688	−4.863***
Δ log TEL per capita	1961–90	67	−2.172	−2.310**
Δ log PAV per capita	1962–90	42	−2.889	−5.992***

Notes: The test statistics are distributed as N(0, 1) under the null hypothesis of non-stationarity. The statistics are constructed using small sample adjustment factors from Im *et al.* (2003). ** and *** denote significance at 5 per cent and 1 per cent levels.
EGC represents kilowatts of electricity-generating capacity. TEL represents the number of telephones. PAV represents kilometers of paved roads.

a trade-off between the time span and number of countries in the sample.[3] For income per capita and electricity-generating capacity we take as our period 1950–92. However, for telephones and paved roads we limit the period to 1960–90 and 1961–90, respectively, in order to get a reasonable number of countries into the sample. When we come to look at the bivariate relationship the coverage of the data set is always the same as for the infrastructure variable. As suggested by Im *et al.*, before carrying out the tests the data are purged of any common effects across countries by regressing each variable on a set of time dummies and taking residuals.

The results of these unit root tests for each of our variables are shown in Table 1. The test is based on the average of the adjusted Dickey–Fuller (ADF) test statistics calculated independently for each member of the panel, with five lags to adjust for autocorrelation. The adjusted test statistics (adjusted using the tables in Im *et al.*, 2003) are distributed as N(0, 1) under the null of a unit root, and large negative values lead to rejection of a unit root in favor of stationarity.

In no case can we reject the null hypothesis that every country has a unit root for the series in log levels. We then test for a unit root in first differences, although in this case the alternative hypothesis is stationarity without a trend, since any time trend in levels is removed by differencing. When we use first differences, the test statistic is negative and significant in each case. This indicates that we have stationarity in first differences and each of the four variables can be regarded as I(1). In what follows we will proceed on the

[3]As shown in the Monte Carlo studies reported in Pedroni (2004), nuisance parameters that are associated with the serial correlation properties of individual member country time series are eliminated asymptotically as T grows large relative to N, which suggests that we should give more weight to the time dimension when balancing the panel in order to avoid size distortion. The power of the tests, on the other hand, rises most dramatically with the N dimension, and rapidly approaches 100 per cent against stationary but near unit root alternative hypotheses for the estimated residuals, even in relatively short panels.

assumption that all log level variables are I(1) and all log differenced variables are I(0).

3.2 Testing for Cointegration

Now we turn to the question of possible cointegration between each infrastructure variable and GDP per capita. Given the possibility of reverse causality between the variables we use the panel cointegration technique from Pedroni (1999, 2004) which is robust to causality running in both directions and allows for both heterogeneous cointegrating vectors and short-run dynamics across countries. In particular, the cointegrating regression that we estimate is

$$g_{it} = a_i + b_t + \beta_i y_{it} + e_{it} \tag{9}$$

so that each country has its own relationship between g_{it}, the log per capita infrastructure variable, and y_{it}, log per capita income. The variable e_{it} represents a stationary error term. Note that we allow the slope of the cointegrating relationship to differ from unity and to vary across countries. This reflects the fact that in practice the relationship between infrastructure investment, infrastructure stocks and income per capita may be more complex than set out in equation (2). Furthermore, this allows for the possibility that, in practice, growth need not be balanced, so that the ratio of capital stocks to output need not be one. The common time dummies, b_t, capture any common worldwide effects that would tend to cause the individual country variables to move together over time. These may be either relatively short-term business cycle effects, or longer-run effects such as worldwide changes in technology that may affect the relative costs or benefits of infrastructure and thus the equilibrium relationship.

The residuals of this regression are then used to construct the group mean ADF test for the null of no cointegration from Pedroni (2004). The lag length for the ADF-based tests is allowed to vary by country, and is chosen by the step-down procedure starting from a maximum of five lags. Under the null hypothesis of no cointegration, the test has a normal distribution, and diverges to the left under the alternative of cointegration. As the results in Table 2 indicate, we are able to reject the null of no cointegration in each of

Table 2
Panel Cointegration Tests

	Period	*Countries*	*Average ADF*	*Test statistic*
GDP and TEL	1960–90	67	–2.33	–3.04***
GDP and EGC	1950–92	43	–2.28	–2.02**
GDP and PAV	1961–90	42	–2.27	–1.90**

Notes: The test statistics are distributed as N(0, 1) under the null hypothesis of no cointegration. The statistics are constructed using adjustment values from Pedroni (2004). All the data (GDP, TEL, EGC, PAV) are in log per capita form. ** and *** denote significance at 5 per cent and 1 per cent levels.

the three cases. Consequently, in what follows we will proceed on the assumption that each of our series is non-stationary, but that there is cointegration between each infrastructure variable and GDP per capita.

4 Long-run Effects; Empirical Implementation and Econometric Issues

Having established a long-run relationship between infrastructure and income we now turn to the issue of causality. In particular, we are interested in whether innovations to infrastructure stocks have a long-run effect on GDP per capita and what the sign of such an effect is. We begin this section by setting out tests for the presence and sign of such long-run effects and then proceed to carry out these tests on our data.

4.1 The Direction of Long-run Causality, and the Sign of the Long-run Effect

Since in each country the series g and y are individually non-stationary but together are cointegrated, we know from the Granger representation theorem (Engle and Granger, 1987) that these series can be represented in the form of a dynamic error correction model. To estimate the error correction form we use a two-step procedure. In the first step, we estimate the cointegrating relationship between log per capita income and log per capita output given in equation (9) for each country, using either the Johansen (1988, 1991) maximum-likelihood procedure or fully modified ordinary least squares.[4] In the second step, we use this estimated cointegrating relationship to construct the disequilibrium term, $\boxed{\hat{e}_{it} = g_{it} - \hat{\alpha}_i - \hat{b}_t - \hat{\beta}_i y_{it}.}$ We then estimate the error correction model

$$\boxed{\begin{aligned} \Delta g_{it} &= c_{1i} + \lambda_{1i}\hat{e}_{it-1} + \sum\nolimits_{j=1}^{K} \varphi_{11ij}\Delta g_{i,t-j} + \sum\nolimits_{j=1}^{K} \varphi_{12ij}\Delta y_{i,t-j} + \varepsilon_{1it} \\ \Delta y_{it} &= c_{2i} + \lambda_{2i}\hat{e}_{it-1} + \sum\nolimits_{j=1}^{K} \varphi_{21ij}\Delta g_{i,t-j} + \sum\nolimits_{j=1}^{K} \varphi_{22ij}\Delta y_{i,t-j} + \varepsilon_{2it} \end{aligned}} \quad (10)$$

The variable e_{it} represents how far our variables are from the equilibrium relationship and the error correction mechanism estimates how this disequilibrium causes the variables to adjust towards equilibrium in order to keep the long-run relationship intact. The Granger representation theorem implies that at least one of the adjustment coefficients λ_{1i}, λ_{2i} must be non-zero if a long-run relationship between the variables is to hold.

By Proposition 1, shocks to income have a persistent, positive component. Furthermore, the Granger representation theorem places restrictions

[4]Specifically, we use parametric maximum-likelihood methods in those cases where we construct time-series-based tests and we use non-parametric corrections in the form of fully modified ordinary least squares in those cases where we construct panel-based tests. See Pedroni (2000, 2001) for discussions of fully modified ordinary least squares based methods for cointegrated panels.

on the singular long-run response matrix of the moving average representation for the data in differences. This restricts the relationship between the long-run response matrix and the speed of adjustment coefficients λ_{1i}, λ_{2i} in the error correction representation. We can exploit these two pieces of information to test for the existence, and the sign, of any long-run causal effects running from innovations in log per capita infrastructure to log per capita output. We summarize our results in the following proposition.

Proposition 2: Under the specification of our growth model, we have the following.

(i) The coefficient, λ_2, on the lagged equilibrium cointegrating relationship in the dynamic error correction equation for Δy_t is zero if, and only if, innovations to log per capita infrastructure have no long-run effect on log per capita output.
(ii) The ratio of the coefficients, $-\lambda_2/\lambda_1$, on the lagged equilibrium cointegrating relationship in the dynamic error correction equations for Δy_t and Δg_t has the same sign as the long-run effect of innovations to log per capita infrastructure on log per capita output.

The proof of Proposition 2 is also located in the Appendix. It follows from Proposition 2 that we can test hypotheses about the long-run effect of infrastructure on output by testing restrictions on the estimated coefficients in the dynamic error correction equations. According to Proposition 2, a test for the significance of λ_{2i} for any one country can be interpreted, conditional on our growth model, as a test of whether innovations to per capita infrastructure have a long-run effect on per capita output, and a test for the sign of the ratio $-\lambda_{2i}/\lambda_{1i}$ can be interpreted as a test of the sign of the long-run effect of innovations to per capita infrastructure on per capita output. Note that Proposition 2(ii) need not necessarily hold for cointegrated systems in general; the proof relies on both the Granger representation theorem and specific features of the growth model set out in Section 2.

The advantage of our two-step estimation procedure, first estimating the cointegrating relationship and then the error correction mechanism, is that all the variables in equation system (10) are stationary. Asymptotically, the fact that we use the estimated disequilibrium rather than the true disequilibrium in (10) does not affect the standard properties of our estimates, due to the well-known superconsistency properties of the estimator of the cointegrating relationship.[5] It follows that we can carry out standard hypothesis tests on the coefficients estimated in the system.

[5]Toda and Philips (1993) study these properties in the context of more conventional dynamic Granger causality tests in cointegrated systems. See also Urbain (1992) for a related discussion on testing causality in error correction models.

By exploiting the cointegrating relationship we are able to summarize the long-run effects of the innovations in the variables in terms of two parameters, λ_{1i} and λ_{2i}. This contrasts with using the differenced variables in a stationary vector autoregressive representation to estimate the impulse responses over long horizons. The trade-off is that we only test for the existence and sign of long-run effects rather than obtaining a quantitative measure of the size of these effects. On the other hand, as is well known, the standard errors for vector autoregression based estimates of impulse responses over long horizons are notoriously large and unreliable, making inference difficult.[6] In essence, by exploiting the cointegrating relationships present in the data, and summarizing the long-run effects of our growth model in a small number of parameters, we avoid the problems of inference that are typically associated with summing sequences of impulse response coefficients over a long horizon. This is particularly important when we apply the test in a panel context, given the large number of parameters that would otherwise need to be estimated. This allows us to construct group mean and group median based tests for the direction of long-run causality in panels with heterogeneous dynamics, and similarly allows us to construct group median based tests for the sign of the long-run causal effect in such panels.

However, before implementing these tests for long-run causal effects, we begin by asking a simpler question. We first test whether the coefficients on lagged infrastructure changes and the error correction adjustment parameter in the regression explaining income changes are all zero. This is a test of no causal effect from infrastructure shocks to income at any time horizon, which does not distinguish between short-run and long-run effects. We also test for such causality running in the other direction from income to infrastructure. These tests correspond to the usual Granger causality tests in that they are tests for the null hypothesis of whether one variable evolves exogenously with respect to the other at all non-contemporaneous time horizons.

Column 3 of Table 3 reports the percentage of countries that reject an F test of the hypothesis of no causality at the 10 per cent significance level. One interpretation of these results is that causality seems to occur in some countries, but not in others. However, if there really were no causality we would expect to reject this hypothesis, and accept causality in 10 per cent of the countries if we use the 10 per cent significance level for our test. Rejection in a larger number of countries can be taken as evidence against the hypothesis that there is no causality in any country. Using this criterion, we have strong evidence in favor of causality running in both directions between each of our

[6]See, for example, Faust and Leeper (1998) for a discussion of these issues. Furthermore, as Phillips (1998) demonstrates, inferences for such long-horizon impulse responses are very sensitive to mis-specification of the underlying unit root and cointegration properties of the data.

TABLE 3
GRANGER CAUSALITY TESTS

Null hypothesis: no causality	*Number of countries*	*Countries rejecting null at the 10 per cent significance level (percentage)*	*Full sample likelihood ratio test*
GDP does not cause TEL	67	37.7***	850*** (335)
GDP does not cause EGC	43	51.2***	504*** (215)
GDP does not cause PAV	42	45.2***	695*** (210)
TEL does not cause GDP	67	46.3***	801*** (335)
EGC does not cause GDP	43	30.2***	368*** (215)
PAV does not cause GDP	42	42.9***	424*** (210)

Notes: Under the null hypothesis of a parameter value of zero in every country, the percentage rejecting at the 10 per cent significance level has an expected value 10 with a standard error of $30N^{-1/2}$. The likelihood ratio test is distributed as χ^2 with the degrees of freedom given in parentheses. All the data (GDP, TEL, EGC, PAV) are in log per capita form. *** denotes significance at the 1 per cent level.

infrastructure variables and GDP, since we find rejections of no causality in a great deal more than 10 per cent of countries.[7]

A test of the joint hypothesis of no causality in any country is given in column 4 of Table 3. This is a likelihood ratio test of the hypothesis that all the relevant parameters are zero in every country. Under the null of no causality the test statistic is distributed as χ^2 with degrees of freedom equal to the number of restrictions imposed, which is given in parentheses beneath the statistic. Large values of this statistic lead to rejection of the null hypothesis of no causality. Again, evidence supports two-way causality between these variables and GDP per capita and each of our infrastructure variables. The fact that non-causality is rejected in a significant number of countries supports the idea that the results for the likelihood test of non-causality in any country are not being driven by a small number of extreme estimates in a few countries.

4.2 Panel Tests for the Direction of Long-run Causality and the Sign of the Long-run Effect

The conventional Granger causality results indicate two-way feedback. However, the causality associated with this feedback may be only of a short-run nature, so that innovations to infrastructure have an impact on GDP per

[7]Under the null of no causality, the percentage of countries rejecting at the 10 per cent significance level has an expected value of 10 with a standard deviation of $30N^{-1/2}$ (for N large). Using this distribution, the number of countries in which we reject no causality is significantly greater than expected even at the 1 per cent level.

capita from business cycle or multiplier effects that eventually die out and do not have a lasting effect on long-run growth.

Therefore, we now turn to the issue of whether infrastructure investment affects long-run economic growth. Since our variables are cointegrated, Proposition 2 gives us some simple tests for the direction of long-run causality. In principle, these can be implemented on a country-by-country basis, and an example of such results is presented in Table 4 for the case of paved roads.[8] Column 2 reports the point estimate for λ_{2i}, column 3 reports the associated t test for the null hypothesis that $\lambda_{2i} = 0$ and column 4 reports the corresponding p value for the test result from column 3. Columns 5 through 7 report analogous results for λ_{1i}, and column 8 reports the estimate for the 'sign' ratio $-\lambda_{2i}/\lambda_{1i}$. However, in practice the reliability of these various point estimates and associated tests for any one country is likely to be poor given the relatively short time sample over which the data are observed. Consequently, the majority of our tests will be panel based.

Before proceeding to the panel tests, we first examine a few key issues based on the individual parameter estimates. First, we ask whether the parameters are homogeneous across countries. In Table 5 we report the results for tests of homogeneity of the long-run adjustment parameters across countries. The test that we use for homogeneity is a Wald test. For a parameter θ, the test statistic is calculated as $\Sigma_{i=1}^{N}\sigma_{\theta}^{-2}\left(\hat{\theta}_i - \overline{\theta}\right)^2$, where $\overline{\theta}$ is the sample mean of the country-specific parameter estimates and σ_{θ}^{-2} is the inverse of the sample variance of the country-specific parameter estimates.[9] Using this test, we decisively reject homogeneity of λ_{2i} across countries. Furthermore, when we test the ratio $-\lambda_{2i}/\lambda_{1i}$, which we call the sign parameter, we reject homogeneity for telephones and paved roads, though in this case only at the 10 per cent significance level. However, it is interesting to note that we do not reject homogeneity of the sign parameter across countries in the case of electricity. These results are potentially informative when we interpret our tests for the sign of the long-run effects.[10]

The next test that we consider is simply a joint test of the hypothesis that the adjustment parameter λ_{2i} is zero in every country, so that the null hypothesis is that there is no long-run effect of infrastructure on income in any country of the panel. We report the results of this test in column 2 of Table 6. This likelihood ratio test provides strong evidence against the long-run effect being uniformly zero among all countries, and easily rejects the null of no long-run effect at the 1 per cent significance level in each case.

[8]Individual country results for the infrastructure types are also available upon request.

[9]This is only a test that $\theta_i = \overline{\theta}$ for all i but it is easy to check that the test statistic is larger for any other test of the form $\theta_i = \theta$. It follows that if we reject $\theta_i = \overline{\theta}$ for all i, we reject $\theta_i = \theta$ for all i, for any choice of θ.

[10]However, we should interpret these sign ratio tests with caution given that the mean and variance of the ratio are likely to be poorly defined. See the discussion on panel-based tests of the sign ratio for further details.

TABLE 4
KILOMETERS OF PAVED ROADS (EXAMPLE OF INDIVIDUAL COUNTRY RESULTS)

	λ_2: $g_{it} \rightarrow y_{it}$			λ_1: $y_{it} \rightarrow g_{it}$			$-\lambda_2/\lambda_1$
Country	*Estimate*	*Test*	*p value*	*Estimate*	*Test*	*p value*	*Estimate*
Congo	0.15	1.41	(0.16)	–0.74	–4.30	(0.00)	0.20
Ghana	0.26	1.00	(0.32)	–0.16	–1.85	(0.06)	1.61
Ivory Coast	–0.04	–0.32	(0.75)	–0.30	–1.52	(0.13)	–0.14
Kenya	0.28	0.68	(0.50)	–0.80	–2.87	(0.00)	0.35
Malawi	–0.12	–2.36	(0.02)	–0.29	–1.79	(0.07)	–0.42
Mauritius	–0.12	–1.28	(0.20)	–0.01	–0.31	(0.76)	–9.73
Morocco	–0.05	–1.26	(0.21)	–0.06	–2.05	(0.04)	–0.89
Mozambique	–0.53	–3.77	(0.00)	–0.28	–1.91	(0.06)	–1.92
Nigeria	–0.31	–1.11	(0.27)	–0.26	–1.67	(0.09)	–1.17
South Africa	0.33	2.10	(0.04)	0.16	2.21	(0.03)	–2.10
Tunisia	–0.23	–1.56	(0.12)	–0.26	–1.93	(0.05)	–0.89
Costa Rica	0.09	1.41	(0.16)	–0.42	–3.55	(0.00)	0.21
El Salvador	0.07	1.13	(0.26)	–0.13	–1.05	(0.29)	0.53
Guatemala	0.16	2.95	(0.00)	–0.02	–0.38	(0.71)	8.53
Honduras	0.08	2.10	(0.04)	–0.34	–2.38	(0.02)	0.23
Panama	0.21	1.76	(0.08)	0.15	1.30	(0.19)	–1.44
USA	0.13	1.92	(0.05)	–0.07	–1.13	(0.26)	1.85
Argentina	–0.01	–0.02	(0.98)	–0.34	–2.01	(0.04)	–0.02
Brazil	0.02	0.28	(0.78)	–0.23	–2.48	(0.01)	0.08
Chile	0.05	0.24	(0.81)	–0.24	–1.81	(0.07)	0.20
Colombia	–0.06	–2.04	(0.04)	–0.42	–2.44	(0.01)	–0.14
Ecuador	0.00	0.04	(0.97)	–0.31	–2.37	(0.02)	0.01
Venezuela	0.16	0.99	(0.32)	–0.14	–1.48	(0.14)	1.16
Hong Kong	–0.06	–0.91	(0.36)	–0.05	–1.36	(0.17)	–1.18
India	0.19	1.05	(0.30)	–0.10	–0.78	(0.44)	1.87
Indonesia	0.13	1.35	(0.18)	–0.47	–2.27	(0.02)	0.28
Japan	0.08	1.83	(0.07)	–0.03	–1.04	(0.30)	2.40
Korea	0.01	0.13	(0.90)	–0.18	–1.52	(0.13)	0.04
Malaysia	–0.14	–1.07	(0.29)	–0.15	–1.90	(0.06)	–0.93
Pakistan	0.35	3.41	(0.00)	–0.02	–0.29	(0.77)	18.59
Philippines	0.71	5.21	(0.00)	0.12	0.65	(0.52)	–5.83
Taiwan	0.03	0.32	(0.75)	–0.14	–1.51	(0.13)	0.20
Thailand	–0.05	–0.56	(0.58)	–0.19	–2.21	(0.03)	–0.25
Denmark	0.09	1.58	(0.11)	–0.09	–2.59	(0.01)	1.03
Greece	–0.08	–0.61	(0.54)	–0.51	–1.93	(0.05)	–0.15
Italy	0.01	0.17	(0.87)	–0.08	–1.20	(0.23)	0.15
Luxembourg	–0.06	–0.94	(0.35)	–0.14	–4.62	(0.00)	–0.40
Norway	0.03	1.50	(0.13)	–0.03	–1.18	(0.24)	1.15
Sweden	–0.02	–0.88	(0.38)	–0.11	–1.84	(0.07)	–0.22
UK	0.05	0.82	(0.41)	–0.05	–1.64	(0.10)	1.00
Australia	0.05	0.40	(0.69)	–0.06	–0.44	(0.66)	0.78
New Zealand	0.40	3.56	(0.00)	–0.17	–1.51	(0.13)	2.31

However, one might question whether such a test is too strict, in the sense that we would like to know more about the pervasiveness of a long-run causal effect in the panel rather than simply finding that there is at least some long-run causality present in at least one member of the panel. For this, we construct both a group mean based test and a lambda-Pearson based test.

TABLE 5
TESTS OF PARAMETER HOMOGENEITY FOR LONG-RUN EFFECTS ACROSS COUNTRIES
Null Hypothesis: Homogeneity of Parameters across Countries

	Test of λ_2	*Test of $-\lambda_2/\lambda_1$*
	Wald test	*Wald test*
TEL to GDP	232***	101***
	(67)	(67)
EGC to GDP	124***	46
	(43)	(43)
PAV to GDP	153***	57*
	(42)	(42)

Notes: Test statistics are distributed as χ^2 under the null hypothesis, with the degrees of freedom given in parenthesis. All the data (GDP, TEL, EGC, PAV) are in log per capita form. * and *** denote significance at 10 per cent and 1 per cent levels.

TABLE 6
TESTS FOR PRESENCE OF LONG-RUN EFFECTS
Null Hypothesis: No Long-run Effects from Infrastructure to Income

	Test of λ_2
	Likelihood ratio test
TEL to GDP	325***
	(67)
EGC to GDP	164***
	(43)
PAV to GDP	211***
	(42)

Notes: All test statistics are distributed as χ^2 under the null hypothesis, with the degrees of freedom given in parenthesis. All the data (GDP, TEL, EGC, PAV) are in log per capita form. *** denotes significance at the 1 per cent level.

The group mean test is based on the sample average of the individual country λ_{2i} tests and will allow us to ask whether the long-run causal effect is zero on average for the panel. Specifically, the group mean panel estimate is computed as $\bar{\lambda}_2 = N^{-1}\Sigma_{i=1}^{N}\hat{\lambda}_{2i}$ and the group mean panel test for the null of no long-run causal effect from infrastructure to per capital income is computed as $\bar{t}_{\lambda_2} = N^{-1}\Sigma_{i=1}^{N}t_{\lambda_{2i}}$ where $t_{\lambda_{2i}}$ is the individual country test for the null hypothesis that $\lambda_{2i} = 0$. The t_{λ_2} test has a standard normal distribution under the null hypothesis of no long-run causal effect for the panel, and rejects in favor of the alternative on either tail of the distribution.

In contrast, the lambda-Pearson panel test uses the p values associated with each of the individual country t tests to compute the accumulated marginal significance associated with these. Specifically, the lambda-Pearson test takes the form $P_{\lambda_2} = -2\Sigma_{i=1}^{N}\ln p_{\lambda_{2i}}$ where $\ln p_{\lambda_{2i}}$ is the log of the p value associated with individual country i's t test for the null hypothesis that $\lambda_{2i} = 0$. The P_{λ_2} is distributed as a χ^2 with $2N$ degrees of freedom under the null hypothesis of no long-run causal effect for the panel and rejects in favor of the alternative on the right tail of the distribution.

Under the maintained assumption that the numerical value for the λ_{2i} parameter is the same for all countries of the panel, the null and alternative hypotheses for the two tests are the same, namely that $\lambda_{2i} = 0$ for all i members under the null, and $\lambda_{2i} \neq 0$ for some non-negligible portion of the members of the panel under the alternative. However, an important point to note here is that, when the values for λ_{2i} are potentially heterogeneous among the members of the panel, the interpretation of the two tests differs. In particular, since the $\bar{t}_{\lambda_2}$ is a two-tailed test that can take on positive and negative values under the alternative depending on whether $\hat{\lambda}_{2i}$ is positive or negative, the associated group mean test can be interpreted as testing whether λ_{2i} is zero on average for the panel against the alternative that it is non-zero on average. In contrast, the P_{λ_2} test is a one-sided test that only takes on positive values under both the null and the alternative regardless of whether the individual values for $\hat{\lambda}_{2i}$ and $\bar{t}_{\lambda_2}$ are positive or negative.

Consequently, the combination of the group mean $\bar{t}_{\lambda_2}$ and the lambda-Pearson P_{λ_2} can be particularly informative when the underlying parameters of interest are heterogeneous. For example, when $\bar{t}_{\lambda_2}$ fails to reject the null while P_{λ_2} succeeds in rejecting the null, this can be interpreted as a situation in which we do not reject that the average value for λ_{2i} is zero, even though we reject that it is pervasively zero in the panel. This can occur when the value for λ_{2i} is significantly positive for some fraction of the panel and significantly negative for another fraction of the panel. In this case we can say that a long-run causal effect is present, even if for some members of the panel it is positive while for others it is negative.

Finally, we also construct panel-based tests for the sign ratio $-\hat{\lambda}_{2,i}\left(\hat{\lambda}_{1,t}\right)^{-1}$, as described in Proposition 2, part (ii). However, we need to be cautious in using group mean or lambda-Pearson based tests for the sign ratio. The reason for this is that the numerator and denominator of the sign ratio follow normal distributions, so that the ratio follows a Cauchy distribution. The moments corresponding to the mean and variance do not exist in general for the Cauchy distribution. A practical implication of this for finite samples is that group mean estimators and tests are likely to be overly influenced by the presence of statistic outliers in the form of occasional extreme test values for individual countries, and p values required for the lambda-Pearson based tests are likely to be inappropriate. Indeed, a quick glance at column 8 of Table 4 reveals the frequent occurrence of statistical outliers for the sign ratio estimates.

Fortunately, however, the median is well defined for the Cauchy distribution, and the estimator for the median in turn has a well-defined variance. Consequently, we can construct group median based estimators and associated standard errors for the panel. It should be noted, however, that the individual distributions are likely to be fairly complex since they are formed from the ratio of dependent non-identical normal distributions. Therefore, it is necessary to simulate the variance. We do this by resampling from the estimated vector error correction mechanism corresponding to equation (10) from each country for each infrastructure type in order to simulate the variance of the corresponding group median estimate.

4.3 *Discussion of the Panel Long-run Causality Results*

The results for each of these panel tests for the direction of long-run causality and the sign of the long-run causal effect as described in the previous section are presented in Tables 7–9. We divide the results of these tests into separate tables for each infrastructure type. Specifically, Tables 7–9 report the results for electricity-generating capacity, number of telephones and paved roads, respectively. For each infrastructure type, results are reported for the panel as a whole, as well as for various regional subgroups, which were chosen subject to the criterion that the subgroup must contain at least 10 countries. For each of these subgroups there are two rows, one for the group mean based tests, and one for the lambda-Pearson based tests. Columns 2–4 report these for tests based on the parameter λ_{2i}, which reflects the presence or absence of long-run causality running from the infrastructure type to the per capita incomes. The second column reports the panel point estimate, which exists only for the group mean, not for the lambda-Pearson. The third column reports the corresponding panel test statistics and the fourth column reports the p value for outcome of the panel test statistic. The next three columns repeat this same pattern for analogous tests based on the parameter λ_{1i}, which reflects the presence or absence of long-run causality running from per capita

Table 7
Kilowatts Electricity-generating Capacity

	λ_2: $g_{it} \rightarrow y_{it}$			λ_1: $y_{it} \rightarrow g_{it}$			$-\lambda_2/\lambda_1$
	Estimate	*Test*	*p value*	*Estimate*	*Test*	*p value*	*Median*
All 43							
Group mean	–0.01	–0.09	(0.47)	–0.22	–1.68	(0.05)	–0.05
Lambda-Pearson		118.27	(0.01)		249.13	(0.00)	(0.08)
OECD 21							
Group mean	–0.01	0.02	(0.51)	–0.19	–1.54	(0.06)	–0.02
Lambda-Pearson		69.51	(0.00)		118.41	(0.00)	(0.13)
Non-OECD 22							
Group mean	–0.02	–0.19	(0.42)	–0.25	–1.80	(0.04)	–0.07
Lambda-Pearson		48.76	(0.29)		130.72	(0.00)	(0.09)
Latin America 12							
Group mean	–0.01	–0.04	(0.49)	–0.30	–1.72	(0.04)	0.11
Lambda-Pearson		28.47	(0.24)		66.71	(0.00)	(0.10)

TABLE 8
NUMBER OF TELEPHONES

	λ_2: $g_{it} \rightarrow y_{it}$			λ_1: $y_{it} \rightarrow g_{it}$			$-\lambda_2/\lambda_1$
	Estimate	*Test*	*p value*	*Estimate*	*Test*	*p value*	*Median*
All 67							
Group mean	0.04	0.29	(0.61)	–0.26	–1.48	(0.07)	0.14
Lambda-Pearson		282.65	(0.00)		324.62	(0.00)	(0.12)
OECD 10							
Group mean	0.02	0.71	(0.76)	–0.19	–1.02	(0.15)	0.06
Lambda-Pearson		37.91	(0.01)		42.56	(0.00)	(0.42)
Non-OECD 57							
Group mean	0.04	0.21	(0.58)	–0.28	–1.56	(0.06)	0.14
Lambda-Pearson		244.74	(0.00)		282.06	(0.00)	(0.12)
Africa 25							
Group mean	0.06	0.43	(0.67)	–0.30	–1.41	(0.08)	0.19
Lambda-Pearson		116.70	(0.00)		106.92	(0.00)	(0.22)
Latin America 17							
Group mean	0.08	0.44	(0.67)	–0.27	–1.44	(0.08)	0.28
Lambda-Pearson		75.16	(0.00)		77.22	(0.00)	(0.20)
Asia 11							
Group mean	–0.05	–0.55	(0.29)	–0.21	–1.95	(0.03)	–0.08
Lambda-Pearson		37.59	(0.02)		70.87	(0.00)	(0.26)

TABLE 9
KILOMETERS OF PAVED ROADS

	λ_2: $g_{it} \rightarrow y_{it}$			λ_1: $y_{it} \rightarrow g_{it}$			$-\lambda_2/\lambda_1$
	Estimate	*Test*	*p value*	*Estimate*	*Test*	*p value*	*Median*
All 42							
Group mean	0.05	0.49	(0.69)	–0.19	–1.59	(0.06)	0.12
Lambda-Pearson		194.15	(0.00)		250.68	(0.00)	(0.14)
OECD 12							
Group mean	0.06	0.79	(0.79)	–0.13	–1.72	(0.04)	0.89
Lambda-Pearson		43.69	(0.01)		70.04	(0.00)	(0.47)
Non-OECD 30							
Group mean	0.05	0.37	(0.65)	–0.21	–1.54	(0.06)	–0.00
Lambda-Pearson		150.46	(0.00)		180.63	(0.00)	(0.16)
Africa 11							
Group mean	–0.04	–0.59	(0.28)	–0.27	–1.63	(0.05)	–0.89
Lambda-Pearson		53.32	(0.00)		78.26	(0.00)	(0.42)
Latin America 11							
Group mean	0.07	0.80	(0.79)	–0.22	–1.70	(0.04)	0.20
Lambda-Pearson		39.33	(0.01)		70.91	(0.00)	(0.20)

incomes to the infrastructure type. Finally, the last column reports the group median point estimate of the sign ratio in the first row, with the simulated standard error reported in parentheses in the second row.

In examining the details of Tables 7–9, the first pattern to note is that tests based on the λ_{1i} parameters as reported in columns 5 through 7 of each of the tables indicate that long-run causality that runs from income to infrastructure is present for all infrastructure types. The results hold pervasively among individual countries and on average for the entire panel as well as for the various regional subgroups. This result conforms to our expectations based on both the neoclassical and endogenous growth specifications which predict that infrastructure levels should rise as per capita incomes rise.

Next, examining the results in Table 7, we see that we cannot reject the hypothesis that electricity-generating capacity has a zero average long-run effect both globally and on average for the various geographical subgroups. At the same time, however, we see that, at least for the Organization for Economic Cooperation and Development (OECD) subgroup, we do rule out that the long-run effect of electricity is pervasively zero, although the sign of the effect is mixed among OECD countries so that the average is still zero. The implication of these results is that changes in electricity-generating capacity do appear to induce permanent changes in long-run per capita income among OECD countries, though possibly not in other countries. Within the OECD the sign tests appear to indicate that some countries are over-invested while others are under-invested in electricity-generating capacity, so that on average the marginal long-run impact is zero among OECD countries, even though it is positive in some and negative in others.[11]

The results for the other infrastructure types are even stronger and fairly unambiguous in their interpretation. For example, in Table 8 we see that, for number of telephones, although the average long-run effects are again zero globally and for all regional subgroups as indicated by the group mean tests, the lambda-Pearson tests clearly indicate that the long-run effects are pervasively non-zero individually for countries globally and in all subgroups. Furthermore, the group median sign ratio tests in column 8 indicate that the number of telephones is associated with a positive causal effect pervasively among Latin American countries, with mixed positive and negative effects in other regional subgroups. The implication of these results is that, although the average long-run effect is zero globally, telephone prevalence is nevertheless associated with permanent long-run causal effects on per capita income throughout the world. At the margin some countries are over-invested in telephone infrastructure while others are under-invested, so that the net effect is approximately zero at the margin. However, specifically for the Latin American subgroup, telephone infrastructure is pervasively under-invested, so that at the margin the long-run impact of telephone infrastructure investment will be positive on average for the Latin American subgroup.

Next we discuss the results for kilometers of paved roads. The group mean results in columns 2–4 of Table 9 again show that on average the effect

[11]On the other hand, if we take the homogeneity test for the sign ratio on electricity-generating capacity in Table 5 at face value, then this could also be consistent with approximately optimal investment in electricity-generating capacity at the individual country level throughout the OECD.

of paved roads is zero at the margin, both for the panel as a whole and for the various regional subgroups. But again this disguises the fact that paved roads nonetheless do have a non-zero long-run effect pervasively among individual countries, globally as well as within each of the regional subgroups, as demonstrated by the results of the lambda-Pearson based tests. Among the regional subgroups, the sign of the long-run effect appears to be pervasively positive among OECD countries and pervasively negative among African countries, with mixed sign effects elsewhere. The implication again is that investment in paved roads is efficient only on average, but that many countries are over-invested in paved roads relative to other investments and many other countries are under-invested in paved roads relative to other investments. An implication of the sign ratio tests is that countries of the OECD are as a group under-invested in roads, meaning that when additional road construction is undertaken it tends at the margin to have a positive effect. Conversely, the implication for African countries is that as a group they tend to be over-invested in roads in the sense that when additional road construction is undertaken the drain on resources for other potential investments overwhelms the positive effect of the presence of the roads, so that the net long-run effect is negative.

5 Conclusions

Infrastructure must be paid for. According to our model, there is a growth-maximizing level of infrastructure above which the diversion of resources from other productive uses outweighs the gain from having more infrastructure. Below this level, increases in infrastructure provision increase long-run income, while above this level an increase in infrastructure reduces long-run income. It follows that we can use the effect of shocks to infrastructure provision on long-run income levels as a test of where a country's infrastructure stock stands relative to its optimum level from a growth-maximizing perspective. This is conceptually a very simple test since it does not rely on knowing the full structure of the system being examined.

Our results are interesting from the point of view of economic policy. Rather than simply asking whether there is evidence for a strong relationship between public infrastructure and long-run incomes, we are able to isolate the presence of an effect of infrastructure on income while controlling for the reverse effect that income levels are likely to have on infrastructure provision. Furthermore, by identifying the sign of this long-run effect our approach allows for the fact that infrastructure provision may divert resources from other forms of non-infrastructure investment and asks whether the level of provision is likely to be above or below the optimum from a growth-maximizing perspective. In this context, it will be interesting in future research to explore where other forms of public investment such as education stand relative to their growth-maximizing levels.[12] Finally, we are able to show that allowing for heterogeneity across countries is also very important for policy purposes; average results for groups of countries tend to disguise large differences between countries.

For all infrastructure types we find that long-run causality from infrastructure investment to per capita income is pervasive despite the fact that it is often zero on average for large groups of countries due to the fact that some countries are over-invested while others are under-invested in the particular infrastructure type. Geographically, for telephones, we find that Latin America is on average under-invested. For paved roads, we find that Africa as a group tends to be over-invested in the sense that when additional road construction is undertaken the drain on resources for other potential investments overwhelms the positive effect of the presence of the roads, so that the net long-run effect is negative. In contrast, for the OECD as a group, we find evidence for under-investment in roads, meaning that when additional road construction is undertaken it tends at the margin to have a net positive effect on long-run income. Globally, however, we find no evidence of a worldwide infrastructure shortage. Rather, we find that *on average* countries are near the growth-maximizing levels of infrastructure provision globally, although a significant number of countries are over-providing while in others there is under-provision.

In some ways our results are not surprising. If infrastructure were provided in competitive markets, and there were no externalities present, this optimality result would be exactly what we would expect. However, in practice, infrastructure has often been supplied by the public sector, and we have the possibility of large externalities, perhaps leading to misallocation of resources. In this context it could be said that the finding of optimality, even if just on average, is more surprising. For policy purposes our results point to the need for detailed country studies of the type used by Fernald (1999) in order to find appropriate rates of return to infrastructure.

Appendix

Proof for Proposition 1

(i) Using equation (6) it is easy to show that y has a unit root under either specification, and cointegration of y and g follows directly from equation (8). In equation (6), when $\delta = 1$, exogenous technology, ε_t, follows a random walk, and innovations to productivity have a permanent effect on y even when $\alpha + \beta < 0$. When $\alpha + \beta = 1$, the endogenous process for output accumulation is no longer

[12] For example, Bils and Klenow (2000) provide strong evidence for the quantitative importance of controlling for the effect of growth on education when examining the effect of education on growth and Pedroni (2007) provides evidence for the importance of accounting for production function heterogeneity for empirical growth and convergence studies that involve education and other poorly observed factors of production.

mean reverting, so that when exogenous technology is mean reverting, with $\delta < 1$, innovations to productivity have a permanent effect on y. Finally, since $\alpha + \beta > 0$, positive innovations to productivity lead to positive long-run effects.

(ii) Shocks to infrastructure, μ_t, only affect the steady state through their effect on y. But when $\alpha + \beta < 0$, variations in y eventually dissipate since the parameter in the difference equation (6) is less than one.

(iii) In this case all shocks to output are permanent. The long-run effect of an infrastructure shock to log output per capita is the same as the short-run effect and is given by

$$E(\mu) = \alpha \log(1 - \bar{\tau} - \mu) + \beta \log(\bar{\tau} + \mu)$$

Hence

$$E'(\mu) = \frac{-\alpha}{1 - \bar{\tau} - \mu} + \frac{\beta}{\bar{\tau} + \mu}$$

Evaluating this at $\mu = 0$ and setting $\tau^* = \beta/(\alpha + \beta)$ we have

$$E'(\mu)|_{\mu=0} > 0 \Leftrightarrow \bar{\tau} < \tau^* \qquad E'(\mu)|_{\mu=0} < 0 \Leftrightarrow \bar{\tau} > \tau^*$$

It follows that for $\bar{\tau} < \tau^*$ small positive shocks to infrastructure raise output in both the short run and the long run while for $\bar{\tau} > \tau^*$ small positive shocks tend to reduce output. ■

Proof for Proposition 2

Let $\Delta Z_t = F(L)\varepsilon_t$ be the stationary moving average representation for the differenced data $\Delta Z_t = (\Delta g_t, \Delta y_t)'$ in terms of the innovations $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$, so that

$$F(1) = \begin{bmatrix} F(1)_{11} & F(1)_{12} \\ F(1)_{21} & F(1)_{22} \end{bmatrix}$$

represents the matrix of long-run responses of the levels Z_t to innovations in ε_t. So $F(1)_{ij}$ represents the long-run effect of j on i, and we are particularly interested in $F(1)_{21}$, the long-run effect of infrastructure on output. According to the Granger representation theorem (Engle and Granger, 1987), if the individual series of z_t are cointegrated, then the long-run response matrix $F(1)$ will contain a singularity such that $F(1)\lambda = 0$, where $\lambda = (\lambda_1, \lambda_2)'$ is the vector of adjustment coefficients to the error correction term in the error correction mechanism representation given in equation (10). This implies $F(1)_{21}\lambda_1 + F(1)_{22}\lambda_2 = 0$.

According to our Proposition 1, part (i), we know that innovations to per capita output productivity must have a positive long-run effect on per capita output under either parameterization of the model, so that $F(1)_{22} > 0$. Under cointegration, since an error correction mechanism exists, we cannot have both elements of λ equal to zero. Combined with the restriction that $F(1)_{22} > 0$, this implies $F(1)_{21} = 0$ if and only if $\lambda_2 = 0$, which establishes part (i) of the proposition.

Furthermore, suppose $\lambda_1 = 0$ Since $F(1)_{22} > 0$ this implies $\lambda_2 = 0$, which would contradict the fact that there is an error correction mechanism and the series are cointegrated. Hence $\lambda_1 \neq 0$ and we can write $F(1)_{21} = -(\lambda_2/\lambda_1)F(1)_{22}$. The restriction $F(1)_{22} > 0$ implies that the ratio $-(\lambda_2/\lambda_1)$ has the same sign as $F(1)_{21}$, which establishes part (ii) of the proposition. ■

References

Barro, R. J. (1990). 'Government Spending in a Simple Model of Endogenous Growth', *Journal of Political Economy*, Vol. 98, pp. 103–125.

Bils, M. and Klenow, P. (2000). 'Does Schooling Cause Growth?', *Working Paper*, University of Rochester.

Canning, D. (1998). 'A Database of World Infrastructure Stocks, 1950–1995', *World Bank Economic Review*, Vol. 12, pp. 529–547.

Canning, D. and Pedroni, P. (1999). 'Infrastructure and Long Run Economic Growth', *Working Paper No. 99-09*, Center for Analytical Economics, Cornell University.

Devarajan, S., Swaroop, V. and Zou, H. F. (1996). 'The Composition of Public Expenditure and Economic Growth', *Journal of Monetary Economics*, Vol. 37, pp. 313–344.

Easterly, W. and Rebelo, S. (1993). 'Fiscal Policy and Economic Growth: an Empirical Investigation', *Journal of Monetary Economics*, Vol. 32, pp. 417–458.

Engle, R. F. and Granger, C. W. J. (1987). 'Co-integration and Error Correction: Representation, Estimation, and Testing', *Econometrica*, Vol. 55, pp. 251–276.

Faust, J. and Leeper, E. (1998). 'When Do Long-run Identifying Restrictions Give Reliable Results?', *Journal of Business and Economic Statistics*, Vol. 15, pp. 345–354.

Fernald, J. G. (1999). 'Roads to Prosperity? Assessing the Link between Public Capital and Productivity', *American Economic Review*, Vol. 89, pp. 619–638.

Garcia-Mila, T., McGuire, T. J. and Porter, R. H. (1996). 'The Effect of Public Capital in State Level Production Functions Reconsidered', *Review of Economics and Statistics*, Vol. 78, pp. 177–180.

Ghali, K. H. (1998). 'Public Investment and Private Capital Formation in a Vector Error-correction Model of Growth', *Applied Economics*, Vol. 30, pp. 837–844.

Gramlich, E. M. (1994). 'Infrastructure Investment: a Review Essay', *Journal of Economic Literature*, Vol. 32, pp. 1176–1196.

Holtz-Eakin, D. (1994). 'Public Sector Capital and the Productivity Puzzle', *Review of Economics and Statistics*, Vol. 76, pp. 12–21.

Holtz-Eakin, D. and Lovely, M. E. (1996). 'Scale Economies, Returns to Variety and the Productivity of Public Infrastructure', *Regional Science and Urban Economics*, Vol. 26, pp. 105–123.

Holtz-Eakin, D. and Schwartz, A. E. (1995). 'Infrastructure in a Structural Model of Economic Growth', *Regional Science and Urban Economics*, Vol. 25, pp. 131–151.

Im, K. S., Pesaran, M. H. and Shin, Y. (2003). 'Testing for Unit Roots in Heterogeneous Panels', *Journal of Econometrics*, Vol. 115, pp. 53–74.

Johansen, S. (1988). 'Statistical Analysis of Cointegration Vectors', *Journal of Economic Dynamics and Control*, Vol. 12, pp. 231–254.

Johansen, S. (1991). 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Autoregressive Models', *Econometrica*, Vol. 59, pp. 155–158.

Kocherlakota, N. R. and Yi, K. M. (1996). 'A Simple Time Series Test of Endogenous versus Exogenous Growth Models: an Application to the United States', *Review of Economics and Statistics*, Vol. 78, pp. 126–134.

Kocherlakota, N. R. and Yi, K. M. (1997). 'Is There Endogenous Long-run Growth? Evidence from the United States and the United Kingdom', *Journal of Money, Credit and Banking*, Vol. 29, pp. 235–262.
Morrison, C. J. and Schwartz, A. E. (1996). 'State Infrastructure and Productive Performance', *American Economic Review*, Vol. 86, pp. 1095–1111.
Pedroni, P. (1999). 'Critical Values for Cointegration Tests in Heterogeneous Panels with Multiple Regressors', *Oxford Bulletin of Economics and Statistics*, Vol. 61, pp. 653–670.
Pedroni, P. (2000). 'Fully-modified OLS for Heterogeneous Cointegrated Panels', *Advances in Econometrics*, Vol. 15, pp. 93–130.
Pedroni, P. (2001). 'Purchasing Power Parity Tests in Cointegrated Panels', *Review of Economics and Statistics*, Vol. 83, pp. 727–731.
Pedroni, P. (2004). 'Panel Cointegration; Asymptotic and Finite Sample Properties of Pooled Time Series Tests, With an Application to the PPP Hypothesis', *Econometric Theory*, Vol. 20, pp. 597–625.
Pedroni, P. (2007). 'Social Capital, Barriers to Production and Capital Shares: Implications for the Importance of Parameter Heterogeneity from a Nonstationary Panel Approach', *Journal of Applied Econometrics*, Vol. 22, pp. 429–451.
Phillips, P. C. B. (1998). 'Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs', *Journal of Econometrics*, Vol. 83, pp. 21–56.
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*, Policy Research Department, World Bank.
Summers, R. and Heston, A. (1991). 'The Penn World Table (Mark 5): an Expanded Set of International Comparisons, 1950–1988', *Quarterly Journal of Economics*, Vol. 106, pp. 327–336.
Toda, H. Y. and Philips, P. C. B. (1993). 'Vector Autoregression and Causality', *Econometrica*, Vol. 61, pp. 1367–1393.
Turnovsky, S. J. and Fisher, W. H. (1995). 'The Composition of Government Expenditure and its Consequences for Macroeconomic Performance', *Journal of Economic Dynamics and Control*, Vol. 19, pp. 747–786.
Urbain, J. P. (1992). 'On Weak Exogeneity in Error Correction Models', *Oxford Bulletin of Economics and Statistics*, Vol. 54, pp. 187–208.