Economics and Environment Network

THE AUSTRALIAN NATIONAL UNIVERSITY

http://een.anu.edu.au/

# Productivity, factor accumulation and social networks: theory and evidence

R. Quentin Grafton, Tom Kompas and P. Dorian Owen

Australian National University
Economics and Environment Network Working Paper
EEN0401

March 2004

# Productivity, Factor Accumulation and Social Networks: Theory and Evidence

*R. Quentin Grafton*§
Centre for Resource and Environmental Studies
The Australian National University

*Tom Kompas*
Asia Pacific School of Economics and Government
The Australian National University

*P. Dorian Owen**
Department of Economics
University of Otago

## Abstract

The paper analyzes how social barriers to communication affect economy-wide productivity and factor accumulation. Using a dynamic model of an economy that includes a reproducible capital stock (physical or human) and effective labor, a negative relationship is shown to exist between social barriers to communication and total factor productivity (TFP), per capita consumption and reproducible capital. Robust estimates obtained from cross-country data are consistent with the model's predictions. The theory and empirical results help explain cross-country differences in TFP, the high productivity performance of leading industrialized countries and how productivity 'catch up' may be initiated. (*JEL* O41, C61, C21)

## March 2004

§**Contact Author**:
R. Quentin Grafton
Centre for Resource and Environmental Studies (CRES)
Room 502, 5th Floor Hancock Building (#43)
Institute of Advanced Studies
The Australian National University
Canberra, ACT 0200, AUSTRALIA
email: qgrafton@cres.anu.edu.au,
tel: +61 2 6125 0663,
fax: +61 2 6125 0757

The existence of enormous differences in the levels of productivity and factor accumulation across countries constitutes one of the most perplexing issues in economics. Many explanations have been offered for the large disparities, including the initial level of capital stocks (physical, natural and human), human capital externalities, macroeconomic stability, quality of institutions, geography, trade openness and rules over foreign investment. Increasingly, economists are exploring the ways that public and civic institutions, social mores and norms of behavior, and social networks influence economic activity. Such analysis recognizes that economic growth goes beyond factor accumulation and is linked to another economic dimension that some have labeled social capital.

We model the effects of social networks, specifically social barriers to communication, on total factor productivity (TFP) and (human or reproducible) capital accumulation. In a dynamic model we show that social barriers to communication inhibit linkages that both create and diffuse productivity-enhancing knowledge. We subsequently test the model's propositions using cross-country data to determine whether social differences that hinder communication, such as linguistic diversity, have a negative impact on the level of TFP and factor accumulation.

Our approach is consistent with a number of important stylized facts at an economy-wide level and emphasizes the importance of initial conditions in terms of social barriers to communication. In particular, our paper helps explain the importance and persistence of differences in cross-country TFP, the existence of country growth laggards and leaders, and also provides policy implications in terms of cross-country productivity 'catch up'.

The paper is organized as follows. The following section solves a dynamic optimization model of social networks with barriers to communication and analyzes the implications for TFP and factor accumulation. Section II describes the data and the empirical model used to test the theoretical propositions. Section III reports the empirical results. The economic implications of impediments to communication across social networks are examined in section IV. Concluding remarks are offered in section V.

## I. Economic Modeling of Social Networks

Our modeling of social networks is novel, but we owe a debt to those who have tested for the interaction between economic performance and various characterizations of social capital, social infrastructure or social capability (William Easterly and Ross Levine, 1997; Robert E. Hall and Charles I. Jones, 1999; John F. Helliwell and Robert D. Putnam, 1995; Stephen Knack and Phillip Keefer, 1997; Jonathan Temple and Paul A. Johnson, 1998; Paul J. Zak and Knack, 2001). Our work has similarities to contributions by Edward P. Lazear (1999), Daniel Nettle (2000) and James E. Rauch (2001) that emphasize the importance of linguistic and cultural diversity for, respectively, exchange and trade between individuals, aggregate per capita GDP and international trade. Roland Bénabou (1996) also stresses the importance of heterogeneity, especially with respect to inequality and school funding, while Mark Gradstein and Moshe Justman (2002) present a related concept called social polarization. We explore the wider implications of what others have identified as social or cultural differences and the effects of social relations on economic outcomes (Brian Uzzi, 1996). In particular, we develop a dynamic model of social networks and use it to provide empirical tests of the effects of social barriers to communication on cross-country differences in productivity and factor accumulation.

The model we develop incorporates three key ideas. One, cooperation and group interactions enable economies to use large amounts of specialized knowledge (Gary S. Becker and Kevin M. Murphy, 1992; Robert E. Lucas Jr., 1988; Luis A. Rivera-Batiz and Paul M. Romer, 1991). Two, although knowledge is inherently nonrival, the creation and transfer of tacit knowledge or 'know-how' is highly dependent on communication links both within social networks (John Seely Brown and Paul Duguid, 2000; James S. Coleman, Elihu Katz and Herbert Menzel, 1966; Alfred Marshall, 1916; Walter W. Powell, 1990; Bryce Ryan and Neal C. Gross, 1943; Annalee Saxenian, 1994) and by 'weak ties' (Mark S. Granovetter, 1973) across social networks (Everett M. Rogers, 1995; Thomas W. Valente, 1995). Three, individuals communicate more easily the greater the similarity between them (Gabriel Tarde, 1895; Paul F. Lazarsfeld and Robert K. Merton, 1954), and communication and cooperation across social networks is often much more limited than within networks (Bénabou, 1996; George J. Borjas, 1992 and 1995; Ronald S. Burt, 2002; James A. Davis, 1967; Thomas W. Schelling, 1978; Muzafer Sherif, O.J. Harvey, B. Jack White, William R. Hood and Carolyn W. Sherif, 1961; Robert A. Solo, 1967).

We posit that economy-wide output is increasing in the level of a reproducible capital stock (physical or human), the level of effort devoted to production and the number of communication links between social networks. Communication links help in the creation of productivity-enhancing ideas and also in the transmission of tacit knowledge that raises economy-wide productivity. Differences across social groups make communication and interaction worthwhile via 'cross-fertilization' of knowledge and ideas — complementary knowledge — but social barriers that inhibit communication or interchange (such as linguistic differences) raise the cost of mutually beneficial and productivity-enhancing communications.

To capture the effects of social barriers to communication we assume that a representative agent's utility function depends positively on per capita consumption at time *t*, *c*(*t*), and negatively on the effort required to establish social connections across social networks, *s*(*t*), where the latter variable has a lower bound of zero, i.e.,

$$U(c,s)=\int_0^\infty \left[\frac{c(t)^{1-\theta}}{1-\theta}-\frac{s(t)^z}{z\beta}\right]e^{-\rho t}\,dt \qquad (1)$$

where $\theta$ is the intertemporal elasticity of substitution (assumed to lie between zero and one), $z$ is an effort or communication disutility coefficient that is greater than one, $\rho$ is the rate of time preference and $\beta$ is an economy-wide parameter that affects the ease of establishing social networks.

Factors that reduce social barriers to communication, such as a common national school curriculum or a common language, may be interpreted as a shift in the value of $\beta$. The $\beta$ parameter is taken to be sufficiently positive to ensure that $U(c,s)$ is jointly concave and is bounded from above by the assumption that, even in the absence of social barriers, communication costs between individuals are always positive.

Equation (1) is consistent with an intertemporal consumption/leisure model of individual preferences, and the negative effect of $s(t)$ on utility incorporates the notion that the time involved in making connections across social networks is privately costly. It follows that an increase in the policy parameter $\beta$ (a policy change that makes it easier for agents to form social connections) lowers the 'utility-cost' of forming connections.

To complete the model, aggregate output is determined by

(2) $$Y(t) = \alpha_0 (s(t)N(t))^{\alpha_1} K(t)^{\alpha_2}$$

where $\alpha_0$ is economy-wide productivity, $N(t)$ is the size of the labor force, $S(t)N(t)$ is effective units of labor where the productivity of labor is increasing in the number of economy-wide connections across social networks, $K(t)$ is the reproducible capital stock (physical or human) and, for convenience, we assume constant returns to scale although this is not essential to derive our results.

In per capita form, and suppressing *t*, the economy's aggregate production function is given by

(3) $$y = \alpha_0 s^{\alpha_1} k^{\alpha_2}$$

where $y = Y / N$ and $k = K / N$. The change in the reproducible capital stock with respect to time is governed by

(4) $$\dot{k} = y - c$$

To solve the optimization problem we maximize (1) subject to (4), the initial condition $k(0) = k_0$ and the necessary feasibility constraints. We note that *c* and *s* are both control variables, and define $\lambda$ as the co-state variable.

Along the optimal path, both (4) and the following necessary conditions must be satisfied for all *t*:

(5) $$c^{-\theta} = \lambda$$

(6) $$s^{z-\alpha_1} = \lambda \alpha_0 \alpha_1 \beta k^{\alpha_2}$$

(7) $$\frac{\dot{\lambda}}{\lambda} = \rho - \alpha_0 \alpha_2 s^{\alpha_1} k^{\alpha_2 - 1}$$

The first-order conditions imply that

(8) $$s = \left( c^{-\theta} \alpha_0 \alpha_1 \beta k^{\alpha_2} \right)^{1/(z-\alpha_1)}$$

Thus a policy increase in $\beta$, which *reduces* the barriers to communication, results in an increase in social connections across networks. Equation (8), along with the necessary conditions, can be used to derive the following transition paths:

(9) $$\dot{c} = \frac{c}{\theta}\left[ \alpha_0^{\frac{z}{z-\alpha_1}} \alpha_1^{\frac{\alpha_1}{z-\alpha_1}} \beta^{\frac{\alpha_1}{z-\alpha_1}} k^{\frac{z\alpha_2}{z-\alpha_1}-1} c^{\frac{-\theta\alpha_1}{z-\alpha_1}} \alpha_2 - \rho \right]$$

and

(10) $$\dot{k} = \alpha_0^{\frac{z}{z-\alpha_1}} \alpha_1^{\frac{\alpha_1}{z-\alpha_1}} \beta^{\frac{\alpha_1}{z-\alpha_1}} k^{\frac{z\alpha_2}{z-\alpha_1}} c^{\frac{-\theta\alpha_1}{z-\alpha_1}} - c .$$

At the steady state, per-capita consumption ($c^*$) is a function of the steady-state reproducible capital ($k^*$) and is expressed as follows,

(11) $$c^* = k^* \left[ \frac{\rho}{\alpha_2} \right].$$

Thus the steady-state values for consumption and reproducible capital are,

(12) $$c^* = \left( \frac{\rho}{\alpha_2} \right) \left[ \alpha_0^z \alpha_1^{\alpha_1} \alpha_2^{z-\alpha_1+\theta\alpha_1} \rho^{\alpha_1-z-\theta\alpha_1} \beta^{\alpha_1} \right]^{\frac{1}{\alpha_1(z+\theta-1)}}$$

and

(13) $$k^* = \left[ \alpha_0^z \alpha_1^{\alpha_1} \alpha_2^{z-\alpha_1+\theta\alpha_1} \rho^{\alpha_1-z-\theta\alpha_1} \beta^{\alpha_1} \right]^{\frac{1}{\alpha_1(z+\theta-1)}} .$$

These results yield the following proposition.

PROPOSITION 1: *The transition paths and the steady-state values of* $c^*$ *and* $k^*$ *are increasing in the policy parameter* $\beta$ *that reduces the social barriers to communication.*

It follows immediately that, if $z > 1$, which is required for convexity in the effort-disutility relationship, and 0 < $\theta$ < 1, Proposition 1 holds true.

The intuition for the results is that increases in $\beta$, or lowering the costs of forming connections across social networks, increases the *network-augmented* rate of return given in the first term of the brackets of equation (9). In turn, this induces the accumulation of reproducible capital and raises the steady-state values of consumption and capital.

Finally, expressing equation (8) as a function of $k^*$ and substituting into (2) yields a closed-form solution for aggregate output:

(14) $$Y = AN^{1-\frac{\alpha_1(\alpha_2-\theta)}{z-\alpha_1}+\alpha_2} K^{\frac{\alpha_1(\alpha_2-\theta)}{z-\alpha_1}+\alpha_2}$$

where

(15) $$A = \alpha_0(\alpha_0\alpha_1)^{\frac{\alpha_1}{z-\alpha_1}} \left(\frac{\rho}{\alpha_2}\right)^{\frac{-\theta\alpha_1}{z-\alpha_1}} \beta^{\frac{\alpha_1}{z-\alpha_1}}$$

Equation (15) yields the following proposition.

PROPOSITION 2**:** *Increases in the policy parameter* $\beta$ *reduce the social barriers to communication and raise economy-wide total factor productivity, A.*

Our results can be tested using cross-country data and measures of TFP, social barriers to communication and other variables that may reduce communication costs. In testing the propositions, it should be emphasized that the policy parameter $\beta$ affects TFP and transitory and steady state levels of consumption and reproducible capital. Given that the number of social connections across social networks cannot be quantified at a national level, $\beta$ proxies social factors or variables that are both measurable and influence the number and quality of connections across social networks.

## II. Tests of the Propositions

To test Proposition 2 we estimate a model where the regressand is a measure of TFP. The regressors include proxies for $\beta$ and are consistent with a rich literature that shows people socialize with persons with similar characteristics (Lazarsfeld and Merton, 1954). We also include as regressors factors that might mitigate the effects of social barriers to communication on productivity, such as the level of social infrastructure, trade openness, measures of mass communications and population density.

The chosen measure for TFP comes from Hall and Jones (1999) and is solved as a labor-augmenting measure of productivity from the following output per worker production function:

$$(16) \qquad y_i = (K_i / Y_i)^{\alpha/(1-\alpha)} h_i A_i$$

where $y_i$ is output per worker, $Y_i$ is total output, $K_i$ is the physical capital stock, and $h_i$ is human capital per worker in country *i*. Human capital per worker is defined as $e^{\phi(E_i)}$ where $E_i$ is the years of schooling and the function $\phi(E_i)$ is the relative efficiency of labor in country *i* with $E_i$ years of schooling. Given data on $y_i, K_i, L_i$ and $E_i$ for 1988, assuming $\phi(\cdot)$ is piecewise linear and a value of $\alpha$ equal to one third, Hall and Jones solve for values of $\ln A_i$ for 127 countries, which hereafter we denote ln*TFP*.[1]

Equation (16) is not identical to (3) or (14), but it does provide a measure of the unexplained differences in cross-national economic performance after accounting for changes in labor and both physical and human capital. Thus if social barriers to communication do affect productivity, as predicted by proposition 2, we would expect the proxies for $\beta$ to be statistically significant in a regression where ln*TFP* is the regressand.

Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat and Romain Wacziarg (2003) is the source of data on three potentially important measures of social barriers to communication based on linguistic, ethnic and religious fractionalization. Their measures are calculated as follows:

(17) $$FRAC_i = 1 - \sum_{k}^{n} f_{ki}^2$$

where $f_{ki}$ is the share of (linguistic, ethnic or religious) group *k* in country *i*. Using data from the Encyclopedia Britannica (2000), with other sources used for cross-checking, they calculate their measures of ethnic (*Ethnic*), linguistic (*Language*) and religious (*Religion*) fractionalization for up to 215 countries. We also use James D. Fearon's (2003) cultural fractionalization measure, which is based on the structural distance between languages.

These fractionalization indexes do not directly measure the quality or quantity of communication between social groups, but do reflect the number and relative sizes of distinct social groups within a country. Cross-country summary statistics of the fractionalization measures and other key variables are provided in Table 1.

Proposition 1 implies that the higher is the initial level of social barriers to communication, the lower will be the transitory and steady-state levels of reproducible capital (physical or human). We test this proposition by estimating the following equations:

(18) $$\Delta AYS_i = \delta_0 + \delta_1 FRAC_i + \delta_2 AYS60_i + \delta_3 \ln RGDPW60_i + \mu_i$$

(19) $$\Delta \ln KAPW_i = \gamma_0 + \gamma_1 FRAC_i + \gamma_2 \ln KAPW65_i + \gamma_3 \ln RGDPW60_i + \nu_i$$

where $\Delta AYS$ is the change in Robert J. Barro and Jong-Wha Lee's (2001) measure of the average years of schooling in the total population aged 15 years and over between 1960 and 1999, $\Delta \ln KAPW$ is the change in the natural log of real physical capital stock per worker between 1965 and 1990 (from the Penn World Tables) and subscript $i$ denotes observations for country $i$. For *FRAC*, as well as the measures constructed by Alesina et al. and Fearon, we also use an ethnolinguistic fractionalization index for 1960, *ELF*, obtained from Rafael La Porta, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny (1999). The sources for Alesina et al.'s fractionalization measures are mainly from the early to mid 1990s although they argue that they will display considerable time persistence. *ELF* provides a base-period index to assess the sensitivity of the results to the timing of the *FRAC* measure. *AYS60* and ln*KAPW65* are base-period values for the respective capital stock proxies. ln*RGDPW60* is (the natural log of) real gross domestic product per worker in international prices in 1960. The appended error terms, $\mu_i$ and $\nu_i$, are country specific and assumed to be independently and normally distributed. Consistency with Proposition 1 requires that the estimated coefficient on the fractionalization measure be negative and statistically significant.

To test whether social barriers to communication have a negative effect on TFP, as predicted, we estimate variants of the following reduced-form model.

$$(20) \qquad \ln TFP_i = \pi_0 + \pi_1 Ethnic + \pi_2 Language + \pi_3 Religion_i + \psi \ Control_i + \varepsilon_i$$

where *TFP* is the Hall-Jones proxy for TFP and $\varepsilon_i$ is the country-specific error term. The term $\psi$ is a vector of parameters to be estimated and *Control* is a vector of regressors to control for variables such as institutional quality, trade openness and measures of mass communication that may influence TFP. Proposition 2 implies that the estimated coefficients for at least some of the fractionalization regressors be negative and statistically significant.

## III. Empirical Results

The tests for proposition 1 and 2 are presented separately because they require different data. Our primary focus is on the effects of social barriers to communication, proxied by fractionalization measures, on productivity because we hypothesize that it is links across networks that make labor more effective, which, in turn, induces factor accumulation.

### A. *Factor Accumulation*

Table 2 provides the ordinary least squares (OLS) estimates of (18) and (19) that test Proposition 1 using alternative fractionalization indexes to proxy the effects of social barriers to communication. The reported diagnostics include Jurgen A. Doornik and Henrik Hansen's (1994) $\chi^2$ test for normality of the errors (denoted *Normality*) and an *F*-form of an asymptotic test for heteroskedasticity (denoted *White-Hetero*) based on regressing the squared residuals on the original regressors and their (non-redundant) squares (Halbert White, 1980). For two of the models estimated for $\Delta\ln KAPW$ the heteroskedasticity test (in columns (5) and (6)) is statistically significant; however, the use of heteroskedasticity-consistent standard errors has little effect on the statistical significance of the coefficients.

In all models, the relevant base-period capital stock measure has a significant negative coefficient at the 5-percent level of significance, or better. Base-period real per capita GDP has a positive coefficient that is statistically significant at the 10-percent level for the model in column (3) and at the 5-percent level or better in the other models. When the set of three Alesina et al. measures is included, ethnic, but not linguistic or religious, fractionalization has a statistically significant negative coefficient at the 5-percent level in both the $\Delta AYS$ and $\Delta\ln KAPW$ equations. Fearon's fractionalization index, *Culture*, has a negative coefficient that is statistically significant in the human capital equation at the 5-percent level, and *ELF* has a significant negative coefficient at the 1-percent level of significance in both the human capital and physical capital equations. Although measurement of human and physical capital stocks is problematical, these results do support the hypothesis that the initial level of social barriers to communication reduces capital accumulation.

### B. *Total Factor Productivity: OLS Results*

Table 3 provides OLS estimates of variants of equation (20) that tests Proposition 2. In column (1), which includes only the fractionalization measures and no control regressors, the coefficients on *Ethnic* and *Language* have the predicted negative signs and are both statistically significant at the 5-percent level.

In addition to the fractionalization measures, other factors are also likely to influence TFP. Consequently, column (2) gives the results of a model that includes, separately, the two components of Hall and Jones' (1999) social infrastructure index. The two components are *GADP*, an index of government antidiversion policies, which incorporates equally-weighted measures of law and order, bureaucratic quality, corruption, risk of expropriation and government repudiation of contracts, and *YrsOpen*, an index of the extent to which countries are open to international trade.[2] In the model in column (2), the coefficient on *Ethnic* is no longer statistically significant, but the results for *Language* are robust to the addition of these controls. Given our hypotheses about the nature of the transmission of productivity-enhancing ideas, we would expect linguistic differences to be the most important barriers to communication across networks.

Diagnostic tests suggest the presence of heteroskedasticity (with the *White-Hetero* test statistically significant at the 5-percent level for the models in columns (2), (3) and (4)). Heteroskedastic-consistent standard errors are also reported, although these give qualitatively similar results to the conventional standard errors.

Given the hypothesized importance of linguistic barriers, we re-estimated the initial model, but included only a measure of linguistic differences (Fearon's fractionalization index, *Culture*) along with the controls *GADP* and *YrsOpen*. Column (3) reports these results; the coefficient on Fearon's index is negative, as predicted, and statistically significant at the 5-percent level.

The results in columns (4) to (6) provide further evidence on the robustness of the initial results. Studentized residuals and leverage statistics were calculated for the model in column (2) in order to identify potential outliers and/or influential observations.[3] Column (4) presents the results from re-estimating the model with the observations identified by the above statistics removed from the sample, in order to check the sensitivity of the results to the omission of outliers and/or influential observations. While the overall goodness of fit improves and the coefficient on *Language* increases in absolute size, the results are qualitatively unchanged.

To test whether the effects of fractionalization vary between rich and poor countries, we also re-estimated the model in column (2) excluding OECD countries; these results are given in column (5) and are very similar to those in columns (2) and (4). As a further check on the sensitivity of the results, column (6) provides estimates using an alternative estimate for the logarithm of TFP derived by Nazrul Islam (1995). As predicted by Proposition 2, the coefficient for *Language* is negative and statistically significant at the 5-percent level throughout.

### C. *Total Factor Productivity: Robustness Results*

As a check on the robustness of the results in Table 1, we applied a general-to-specific (Gets) algorithm implemented in PcGets (David Hendry and Hans-Martin Krolzig, 2001) to select a preferred model for TFP.[4] The essence of Gets modelling is to start from a general unrestricted model that is 'congruent' with the data, i.e., displays no evidence of misspecification. Variables with coefficients that are not statistically significant are eliminated so to obtain a simpler congruent model that encompasses rival models in the sense that no important information is lost (e.g., Hendry, 1995, p. 365). The basic Gets approach has been significantly enhanced by recent developments by Hendry and Krolzig (1999) and Krolzig and Hendry (2001). These include: examining multiple search paths, considering only model reductions that do not fail diagnostic tests in order to retain congruence, employing 'pre-search simplification', using overlapping sub-sample testing to aid in the overall assessment of the 'reliability' of the significance of the coefficients, implementing encompassing tests to distinguish between competing candidate congruent models that emerge from different search paths, and using an information criterion to make a final selection if encompassing tests fail to pick a unique dominant final model.[5]

Monte Carlo evidence to date (e.g., Krolzig and Hendry, 2001; Hendry and Krolzig, 2001) suggests that the different elements of the overall algorithm combine to give impressive properties: the size of the model selection process is close to the nominal size of the tests used in the search such that the power approaches that obtained if the process started from the data generating process.[6] In particular, Kevin D. Hoover and Stephen J. Perez (2001), in a Monte Carlo study designed to reflect the 'realistic' setting of cross-country growth regressions, show that a cross-section version of a Gets algorithm outperforms Ross Levine and David Renelt's (1992) and Xavier X. Sala-i-Martin's (1997) versions of Edward E. Leamer's (1983) extreme-bounds approach to model selection.[7]

Table 4, column (1) reports results for the model specified in equation (20) where, in addition to *GADP* and *YrsOpen*, the control variables include measures of mass communication, population density and interaction effects. Given that social barriers to communication impede the exchange of productivity-enhancing ideas, we hypothesize that physical infrastructure that aids in communications may mitigate the negative impact on TFP. We also test whether increased proximity between people, as measured by population density (*Popn Density*) and road density (*Road Density*), reduces the effect of social communication barriers. Interaction effects are included to test the hypothesis that increases in mass communications or population density reduce the negative partial effect of linguistic fractionalization on TFP. Due to the heavily parameterized nature of the model given in column (1) of Table 4, it is not surprising that few of the individual coefficients are statistically significant at conventional levels.[8] Nevertheless, we use this initial model as a starting point for the application of a general-to-specific simplification process.

The results in column (2) of Table 4 are the final specific model selected using the general-to-specific model selection algorithm applied to the model in column (1), Table 4. Two measures of social barriers to communication, *Language* and *Religion*, are selected and have coefficients that are statistically significant at the 1-percent level and have the hypothesized negative sign. One of the measures of mass communications, the number of telephones per capita (*Telephones*) has a coefficient with the hypothesized positive sign that is also statistically significant at the 1-percent level. Another mass communication measure is included in the selected interaction term *Language*Radios*. Its coefficient is positive and statistically significant at the 5-percent level, implying that the negative effects of linguistic fractionalization are reduced with improvements in mass communication, proxied by the number of radios per capita.

Further robustness tests are provided in columns (3) and (4) in Table 4. Column (3) contains median regression (least absolute errors) estimates for the final selected model to assess the robustness of the results to potential outliers. Point estimates and standard errors based on the design-matrix-bootstrapping estimator (Moshe Buchinsky, 1998) produce qualitatively similar conclusions to column (2) with the estimated coefficients for linguistic and religious diversity both negative and statistically significant at the 1-percent level. Column (4) presents the results of the final model selected from a general-to-specific search applied to a model of the form in column (1) of Table 4, except that Fearon's *Culture* index replaces the three Alesina et al. (2003) measures and the *Language* variable in the interaction terms. Again, the linguistic diversity measure (*Culture*) is selected in the final model and has a negative coefficient that is statistically significant at the 5-percent level. In addition, both the trade openness measure and telephones per capita are also selected in the final model.

Overall, the robustness tests indicate that the estimated coefficients for the linguistic fractionalization measures, which proxy social barriers to communication, have a negative and statistically significant on TFP. Thus the results are consistent with the proposition that social barriers to communication have a negative impact on productivity.

### D. *Total Factor Productivity: IV Results*

A possible concern with the estimates reported in Tables 3 and 4 is that, while it may be reasonable to treat the fractionalization measures as exogenous, as do Alesina et al. (2003), several of the controlling variables may be endogenous. If this is the case, then OLS estimates will be inconsistent. To address this issue, we use instrumental variables that should be uncorrelated with $\varepsilon_i$, but strongly correlated with the potentially endogenous variables.

Table 5 presents results obtained using instrumental variables (IV) estimation in which all variables other than the fractionalization measures are treated as potentially endogenous. We follow Hall and Jones (1999) in including Jeffrey A. Frankel and David Romer's (1999) (natural log) predicted trade share (based on a trade model including exogenous gravity variables), ln*FraRom*, and the fraction of the population speaking a European language, *EurFrac*, in the instrument set. Hall and Jones also use distance from the equator as an instrument, but, following Jeffrey D. Sachs's (2003) argument that this is a poor proxy for geographical factors such as climate, we instead use mean annual temperature, *MeanTemp*, which provides better fits for the first-stage regressions, as well as the proportion of land area within 100km of the coast, *LT100km*, and total land area, *LandArea*. In addition, we include a measure of 'state antiquity', *StateHist*, constructed by Valerie Bockstette, Areendam Chanda and Louis Putterman (2002), which their empirical results suggest is a significant predictor of Hall and Jones' composite social infrastructure measure.[9] We also include the interactions between linguistic fractionalization and a subset of the geographical instruments in some of the instrument sets to allow for the endogeneity of interaction terms involving fractionalization and the other right-hand-side variables, such as *Language*×*Radios*.

Table 5 provides evidence on the suitability of the sets of instruments used. To check on the explanatory power of the instrument sets, the values of $R^2$ for the first-stage regressions of each right-hand-side endogenous variable on the instruments, including a constant, were calculated. We also calculated $p$-values of the $F$-statistics for the joint null hypothesis that the coefficients on all the instruments (including the exogenous regressors) are zero; these are not reported in Table 5 because they are all 0.000. These $p$-values, reflecting the high $R^2$ values, indicate that the instrument sets are strongly associated with the endogenous right-hand-side variables. To check on the correlation between the residuals and the instruments we calculated J. Denis Sargan's (1964) general misspecification test for instrumental variables estimation of over-identified models. The test statistic, denoted Sargan $\chi^2$ in Table 5, is obtained as $NR^2$ from the regression of the IV residuals on the set of all instruments and is asymptotically distributed as a central chi-square with degrees of freedom equal to the number of over-identifying restrictions. The hypothesis that the over-identifying instruments are independent of the error terms is not rejected for any of the models. We also report a Hausman test of the consistency of the OLS estimates by comparison with IV based on the selected instrument set(s); under the null that the OLS estimates are consistent, the test is asymptotically distributed as a central chi-square with degrees of freedom equal to the number of potentially endogenous right-hand-side variables. The results imply that OLS estimates are not significantly affected by endogeneity for the models in columns (1) and (3) of Table 5, but are inconsistent when compared to the IV estimates in column (5), using a 5-percent significance level, and more marginally, at the 10-percent significance level, for columns (2) and (4).

Columns (1) and (2) in Table 5 report the IV estimation results for the models corresponding to the OLS estimates in column (2) and (3) in Table 3. Again, both sets of results are consistent with the prediction that linguistic diversity has a negative impact on TFP. The results presented in column (3) in Table 5 correspond to the model estimated in column (2) of Table 4, i.e., including those variables retained in the final model from the OLS-based general-to-specific selection process. Apart from a reduction in the statistical significance of the coefficient on *Telephones*, the IV results are very similar to those obtained using OLS, an interpretation supported by the non-rejection of the Hausman Test. Column (4) in Table 5 is the final model obtained by commencing with the general model in Table 4, column (1) and applying the general-to-specific simplification, but based throughout on IV estimation, using the specified instrument set, rather than OLS. One component of Hall and Jones' social infrastructure proxy, *YrsOpen*, and *Road Density* are selected, in place of *Telephones*, but linguistic and religious fractionalization continue to have a significant negative effect. In addition, the role of communications, proxied by *Radios*, in reducing the effect of linguistic fractionalization remains significant through the interaction term. Finally, to illustrate the robustness of the results for the fractionalization and communications variables to the inclusion of social infrastructure proxies, column (5) of Table 5 reports the results obtained by again applying the general-to-specific simplification based on IV estimation commencing from a general model excluding *GADP* and *YrsOpen*. The variables selected are, apart from the excluded *YrsOpen* variable, identical to those in column (4), reinforcing the robustness of these results.

### E. *Economic Significance of Total Factor Productivity Results*

To assess the economic significance of the effect of social barriers to communication, we carried out a simple simulation. Taking the results from Table 4, column (2) as representative, the coefficients, which being statistically significant at the 5-percent level or better are all relatively precisely estimated, were used to predict the values of ln*TFP* for each country and these were transformed into levels. The 110 countries in the sample were then sorted in ascending order on the basis of their values for *Language*. The means of the predicted values of TFP in levels for the lower and upper quartile countries (defined as the bottom 27 and top 27 countries in terms of the ranking with respect to *Language*) were then calculated. The ratio of the mean predicted TFP values for the quartile with the lowest measure of linguistic fractionalization, relative to the mean predicted TFP values for the quartile with the highest measure of linguistic fractionalization, is greater than two (2.293). This implies that the effects of social barriers to communication are economically as well as statistically significant in explaining cross-country variation in TFP levels.

Overall, the results provide support for propositions 1 and 2. Together with other explanatory factors, our results provide a plausible explanation for the large disparity in productivity across countries, and why these differences may not necessarily decline over time.

## IV. Economic Implications of Social Networks

Our model emphasizes the *social* dimension of cross-country economic differences rather than simply differences in levels of capital (human and physical). It also explains or supports a number of important stylized facts at an economy-wide level, and thus goes further than the existing literature on social cohesion and polarization (Bénabou, 1996; Gradstein and Justman, 2002). Indeed, our approach provides a partial explanation for the large differences in total factor productivity, and thus income, across countries (Easterly and Levine, 2001).

Our results address three key features of economic performance. One, the on-going high performance of leading industrialized countries; two, the ability of a few countries to initiate 'catch up' with economic leaders; and three, the reason why some countries remain growth laggards (Lant Pritchett, 1997). To the extent that increased connections contribute to higher levels of trust and cooperation between individuals, our results also provide a possible explanation for the positive empirical relationship between social capital and human capital accumulation (Edward L. Glaeser, David Laibson and Bruce Sacerdote, 2002).

### A. *High Productivity Performance*

We emphasize that diversity, *per se*, is not detrimental to productivity because it is diversity that provides the basis for mutually beneficial exchanges and the 'cross-fertilization' of knowledge and ideas — a point made by John Stuart Mill (1848, p. 594) over 150 years ago. Rather, it is the associated higher costs of and barriers to group-to-group communication that act as an impediment to increases in productivity and factor accumulation that diversity would otherwise bring. Indeed, radial, spanning or bridging connections across networks, on an individual level, are strongly associated with early adoption of technologies (Valente, 1995, p. 42). Our results support this finding on a national level with evidence that factors that inhibit radial links across social networks, such as linguistic barriers, lower economy-wide productivity.

We speculate that social barriers to communication may, in part, explain the high productivity of the United States (US) that has a common language and is also a multicultural, pluralistic and geographically and socially mobile society (Borjas, 1992). In other words, countries like the US that have a common language and a unifying culture can reap the benefits from complementary knowledge sets inherent in different social and professional networks. Our thesis is also supported by recent empirical evidence that uses data from US cities and finds that, after controlling for endogeneity, the greater is the cultural diversity, the more productive are US-born citizens (Gianmarco I.P. Ottaviano and Giovanni Peri, 2004). By contrast, countries that are less socially diverse and mobile than the US, or that are diverse but have major impediments (social, physical and institutional) to group-to-group communications, may suffer from a lower level of productivity because they have less effective radial links across social networks.

### B. *Productivity 'catch up'*

The social networks perspective offers insights as to how countries might engineer a 'catch up' in terms of productivity by fostering approaches that mitigate barriers to communication across social groups. For example, the offering of common national curricula that engender a shared identity and heritage and reduce social distance (Gradstein and Justman, 2002), citizenship and native language classes for immigrants, a common and actively promoted official language, investments in mass communications such as internet access and communication links, are all approaches that may generate a substantial pay-off in productivity by reducing the costs of communication across social networks.

To some extent, such measures have been adopted to varying degrees by countries, but usually without a proper recognition of their economic benefits for both productivity and factor accumulation. In sum, national policies can positively influence economic growth if they lower social communication costs that impede the creation and diffusion of productivity enhancing ideas.

### C. *Stylized facts*

We have explored the concept and consequences of social barriers to communication across networks at an economy level, but the idea is also consistent with a number of important stylized facts at a regional and global level. For instance, Lazear (1999), Rauch (2001), Rauch and Vitor Trindade (2002), and others, have identified the importance of networks and a common language in trade, but their explanation is that networks alleviate the difficulties of enforcing contracts, provide information on trade opportunities and help match buyers and sellers. Our work emphasizes the importance of trade and migration flows in the transmission of ideas whereby connections across networks provide a basis for productivity gains.

Our model has parallels to network theory where key agents (individuals or groups of individuals) that link across networks provide the means of knowledge transmission within a network via positive interactions (Durlauf, 1997) and by acting as role models (Rogers, 1995). Our thesis that economy-wide productivity is affected by connections across social networks also has empirical support — Jungsoo Park (2004) finds, using OECD data on cross-country student flows, that the return of foreign-educated workers is an empirically important channel for research and development spillovers. An example of such a spillover is reported by Easterly (2002, pp.145-148); he describes how a single individual played a lead role in developing the garment industry in Bangladesh following the transfer of tacit knowledge via South Korea.

At a local or regional level a number of distinguished thinkers, including Schelling (1978) and Tarde (1895), have observed the tendency for ‘like-with-like’ interactions, known as homophily (Lazarsfeld and Merton, 1954, p. 23). This is consistent with our model where establishing links across social networks is costly. Locations where people ‘connect’ also exemplify low social barriers to communication (or a high $\beta$) that promote knowledge transfer and innovation. These localized effects have been measured in the creation of localized spillovers in the household adoption of information technology via social networks (Austan Goolsbee and Peter J. Klenow, 2002). The positive effects of localized social networks have also been observed in spatial patterns of patents (Adam B. Jaffe, Manuel Trajtenberg and Rebecca Henderson, 1993; Laura Bottazzi and Giovanni Peri, 2003) and help to explain the higher rents and wages (Rauch, 1993) found in knowledge ‘hot spots’.

The existence of positive economic payoffs from connecting across social networks is supported by evidence that cities and social interactions promote the accumulation of human capital (Borjas, 1995; Glaeser and David C. Maré, 2001; Lucas, 1988; Marshall, 1916, p. 271). Our result of a network-augmented rate of return for capital (human or physical) in locations where people ‘connect’ is also consistent with the stylised fact that factors of production agglomerate and provides an explanation as to why capital might flow from poor to rich countries (Lucas, 1990).

## V. Concluding Remarks

This paper addresses the question: what explains the huge variation in productivity across countries? Using an optimal growth model, social barriers to communication are shown to have a negative effect on both productivity and the accumulation of reproducible capital. The model generates important and testable propositions: a policy parameter that reduces the barriers to communication across social networks raises economy-wide productivity, and also increases transitional and steady-state levels of per-capita consumption and the reproducible capital stock.

Propositions from the model are tested using cross-country data from over a 100 countries. The results obtained from OLS and instrumental variable estimation, and with an exhaustive set of robustness tests, are both statistically and economically significant and support the hypothesis that social barriers to communication, as measured by linguistic diversity, reduce total factor productivity. Some evidence is also found to support the idea that the effects of social barriers to communication may be mitigated by improvements in mass communications. In addition, changes in the stocks of human capital and physical capital are shown to be decreasing in social barriers to communication, after controlling for the initial level of income and capital stock.

Our results are broadly consistent with a number of important stylized facts including the importance of social networks in research and development spillovers, the creation of localized spillovers in technology adoption and the flow of capital to locations where people 'connect'. The theory and empirical evidence together provide an important and novel explanation for the large cross-country differences in total factor productivity, the high productivity performance of leading industrialized countries such as the United States and the ability of some countries to initiate productivity 'catch up'.

## APPENDIX A: COUNTRIES INCLUDED IN SAMPLE

The following countries are included in the sample for the regressions in Table 4, columns (1) to (3): Algeria, Argentina, Australia, Austria, Bangladesh, Barbados, Belgium, Benin, Bolivia, Botswana, Brazil, Burkina Faso, Burundi, Cameroon, Canada, Central African Republic, Chad, Chile, China, Colombia, Comoros, Congo, Costa Rica, Côte d'Ivoire, Cyprus, Denmark, Dominican Republic, Ecuador, Egypt, Fiji, Finland, France, Gabon, The Gambia, Germany, Greece, Guatemala, Guinea, Guinea-Bissau, Guyana, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia, Iran, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, South Korea, Lesotho, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Myanmar, Namibia, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Saudi Arabia, Senegal, Seychelles, Sierra Leone, Singapore, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Kingdom, United States of America, Uruguay, Venezuela, Zambia and Zimbabwe.

## APPENDIX B: DATA SOURCES AND DEFINITIONS

ln*TFP*: Hall and Jones measure of total factor productivity (in natural logs). Source: Hall and Jones (1999)

*Ethnic*, *Language*, *Religion*: Fractionalization indexes for ethnic, linguistic and religious groups. Source: Alesina et al. (2003)

*Culture*: Cultural fractionalization index accounting for cultural distances between groups based on language. Source: Fearon (2003)

*GADP*: index of 'government antidiversion policies' calculated as the average of five International Country Risk Guide measures (1985-1995) law and order, bureaucratic quality, corruption, risk of expropriation, government repudiation of contracts, [0-1] range. Source: Hall and Jones (1999)

*YrsOpen*: Sachs and Andrew M. Warner (1995) index of fraction of years open during 1950 to 1994 period. [0, 1] range. Source: Hall and Jones (1999)

*Telephones*: Telephone mainlines (per 1,000 people) in 1988. Source: World Bank (2000).

*Popn Density*: Population density (people per sq km) in 1988. Source: World Bank (2000).

*Radios*: Radios (per 1,000 people) 1989. Source: World Bank (2000).

*Road Density*: Roads/Land Area in 1988 or nearest year. Source: Total roads (kms) in 1988, or nearest year, from David Canning (1998); Land Area (in km) from World Bank (2000).

*MeanTemp*: Mean annual temperature (degrees Celsius) in 1987. Source: John W. McArthur and Sachs (2001, Appendix)

*LT100km*: Proportion of land area within 100km of the sea coast. Source: McArthur and Sachs (2001, Appendix)

*LandArea*: Land area (sq km). Source: World Bank (2000).

*EurFrac*: Fraction of population speaking a major Western European language: English, French, German, Portuguese, or Spanish. Source: Hall and Jones (1999)

ln*FraRom*: Natural log of the Frankel-Romer predicted trade share (computed from a gravity model based on population and geography). Source: Hall and Jones (1999)

*StateHist*: Measures the length and and coverage of formal states in current geographical borders from 1 to 1950. Source: *Statehist5* from Bockstette et al. (2003)

*ELF*: Ethnolinguistic Fractionalization – Average value of five different indices (range 0 to 1). Source: La Porta, Lopez-de-Silanes, Shleifer and Vishny (1999, Appendix B).

ln*RGDPW60*: Real GDP (chain) per worker (1996 international prices) (in natural logs). Source: Penn World Tables 6.1

ln*KAPW*: Real non-residential capital stock per worker (1985 international prices) (in natural logs). Source: Penn World Tables 5.6

*AYS*: Average schooling years in the total population (aged 15 years and over). Source: Barro and Lee (2001)

## REFERENCES

**Alesina, Alberto; Devleeschauwer, Arnaud; Easterly, William; Kurlat, Sergio and Wacziarg, Romain.** "Fractionalization." *Journal of Economic Growth*, June 2003, *8*, pp. 155-194.

**Barro, Robert J. and Lee, Jong-Wha.** "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers*, July 2001, *53*(3), pp. 541-563.

**Becker, Gary S. and Murphy, Kevin M.** "The Division of Labor, Coordination Costs, and Knowledge." *Quarterly Journal of Economics*, 1992, *107*(4), pp. 1137-1160.

**Belsley, David A.; Kuh, Edwin and Welsch, Roy E.** *Regression diagnostics: identifying influential data and sources of collinearity*. New York: John Wiley, 1980.

**Bénabou, Roland.** "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review*, 1996, *86*(3), pp. 584-609.

**Bockstette, Valerie; Chanda, Areendam and Putterman, Louis** "States and Markets: The Advantage of an Early Start." *Journal of Economic Growth*, December 2002, *7*(4), pp. 347-369.

**Borjas, George J.** "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics*, 1992, *107*(1), pp. 123-150.

**___________.** "Ethnicity, Neighborhoods, and Human Capital Externalities." *American Economic Review*, 1995, *85* (3), pp. 365-390.

**Bottazi, Laura and Peri, Giovanni.** "Innovation and Spillovers in Regions: Evidence from European Patent Data." *European Economic Review*, 2003, *47*, pp. 687-710.

**Brock, William A. and Durlauf, Steven N.** "Growth Empirics and Reality." *The World Bank Economic Review*, 2001, *15*(2), pp. 229-272.

**Brown, John Seely and Duguid, Paul.** *The social life of information*. Boston, MA: Harvard Business School Press, 2000.

**Buchinsky, Moshe.** "Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research." *Journal of Human Resources*, 1998, *33*(1), pp. 88-126.

**Burt, Ronald S.** "Bridge Decay." *Social Networks*, 2002, *24*, pp. 333-363.

**Canning, David.** "A Database of World Infrastructure Stocks, 1950-95." World Bank (Washington DC) Policy Research Working Paper 1929, 1998.

**Coleman, James S.; Katz, Elihu and Menzel, Herbert.** *Medical Innovation: A Diffusion Study*. Indianapolis: The Bobbs-Merrill Company, Inc., 1966.

**Davis, James A.** "Clustering and Structural Balance in Graphs." *Human Relation*, 1967, *20*, pp. 181-187.

**Doornik, Jurgen A. and Hansen, Henrik.** "An Omnibus Test for Univariate and Multivariate Normality." Nuffield College, Oxford, Working Paper, 1994.

**Durlauf, Steven N.** "The Memberships Theory of Inequality: Ideas and Implications." Santa Fe Institute (Santa Fe, New Mexico) Working Paper 95-05-047, Abstract, 1997.

**__________.** "On the Empirics of Social Capital." *Economic Journal*, 2002, *112*, pp. F459-479.

**Easterly, William.** *The elusive quest for growth: economists' adventures and misadventures in the tropics*. Cambridge, Mass: The MIT Press, 2002.

**Easterly, William and Levine, Ross.** "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics*, 1997, *111*(4), pp. 1203-1250.

**_________.** "It's Not Factor Accumulation: Stylized Facts and Growth Models." *World Bank Economic Review*, 2001, *15*(2), pp. 177-219.

**Encyclopedia Britannica**. *Encyclopedia Britannica 2000*. Chicago: Encyclopedia Britannica, 2000.

**Fearon, James D.** "Ethnic and Cultural Diversity by Country." *Journal of Economic Growth*, June 2003, *8*(2), pp. 195-222.

**Frankel, Jeffrey A. and Romer, David.** "Does Trade Cause Growth?" *American Economic Review*, June 1999, *89*(3), pp. 379-399.

**Glaeser, Edward, L.; Laibson, David and Sacerdote, Bruce.** "An Economic Approach to Social Capital." *Economic Journal*, 2002, *112*, pp. F437-F458.

**Glaeser, Edward L. and Maré, David C.** "Cities and Skills." *Journal of Labor Economics*, April 2001, *19*(2), pp. 316-342.

**Goolsbee, Austan and Klenow, Peter J.** "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *Journal of Law and Economics*, October 2002, *45*, pp. 317-341.

**Gradstein, Mark, and Justman, Moshe.** "Education, Social Cohesion, and Economic Growth." *American Economic Review*, September 2002, *92*(4), pp. 1192-1204.

**Granovetter, Mark S.** "The Strength of Weak Ties." *American Journal of Sociology*, May 1973, *78*(6), pp. 1360-1380.

**Hall, Robert E. and Jones, Charles I.** "The Productivity of Nations." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5812, 1996.

**_________.** "Why Do Some Countries Produce so Much More Output Per Worker than Others?" *Quarterly Journal of Economics*, February 1999, *114*(1), pp. 83-116.

**Hendry, David F.** *Dynamic econometrics*. Oxford: Oxford University Press, 1995.

**Hendry, David and Krolzig, Hans-Martin.** "Improving on 'Data mining reconsidered' by K.D. Hoover and S.J. Perez." *Econometrics Journal*, 1999, *2*, pp. 202-219.

**_________.** *Automatic econometric model selection using PcGets*. London: Timberlake Consultants Ltd, 2001

**Helliwell, John F. and Putnam, Robert D.** "Economic Growth and Social Capital in Italy." *Eastern Economic Journal*, Summer 1995, *21*(3), pp. 295-307.

**Hoover, Kevin D. and Perez, Stephen J.** "Evaluating Specification Search in Cross-Country Growth Regressions: Extreme Bounds vs. General-to-Specific Methodologies". Department of Economics, University of California, Davis, mimeograph, 2001.

**Islam, Nazrul.** "Growth Empirics: A Panel Data Approach." *Quarterly Journal of Economics*, 1995, *110*(4), pp. 1127-1170.

**Jaffe, Adam B.; Trajtenberg, Manuel and Henderson, Rebecca.** "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, 1993, *108*(3), pp. 577-598.

**Knack, Stephen and Keefer, Phillip.** "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *Quarterly Journal of Economics*, November 1997, *112*(4), pp. 1251-88.

**Krolzig, Hans-Martin and Hendry, David F.** "Computer Automation of General-to-Specific Model Selection Procedures." *Journal of Economic Dynamics and Control*, 2001, *25*, pp. 831-866.

**La Porta, Rafael; Lopez-de-Silanes, Florencio; Shleifer, Andrei and Vishny, Robert.** "The Quality of Government." *Journal of Law, Economics and Organization*, 1999, *15*(1), pp. 222-282.

**Lazarsfeld, Paul F. and Merton, Robert K.** "Friendship as Social Process: A Substantive and Methodological Analysis," in Monroe Berger, Theodore Abel and Charles H. Page, eds., *Freedom and control in modern society*. New York: D. Van Nostrand Company, Inc., 1954, pp. 18-66.

**Lazear, Edward P.** "Culture and Language." *Journal of Political Economy*, December 1999, *107*(6), Part 2, pp. S95-S126.

**Leamer, Edward E.** "Let's Take the Con Out of Econometrics." *American Economic Review*, 1983, *73*(1), pp. 31-43.

**Levine, Ross and Renelt, David.** "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review*, 1992, *82*(4), pp. 942-963.

**Lucas, Robert E. Jr.** "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 1988, *22*, pp. 3-42.

**Lucas, Robert E. Jr.** "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review,* May 1990, *80*(2), pp. 92-96.

**Marshall, Alfred.** *Principles of economics: an introductory volume*. London: Macmillan and Co., Limited, 1916.

**McArthur, John W. and Sachs, Jeffrey D.** "Institutions and Geography: Comment on Acemoglu, Johnson and Robinson (2000)." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 8114, 2001.

**Mill, John Stuart.** *Principles of political economy with some of their applications to social philosophy*, with an introduction by V.W. Bladen, Toronto: University of Toronto Press, Routledge and Kegan Paul, (first published in 1848), 1965.

**Nettle, Daniel.** "Linguistic Fragmentation and the Wealth of Nations: The Fishman-Pool Hypothesis Reexamined." *Economic Development and Cultural Change*, January 2000, *49* (2), pp. 335-48.

**Ottaviano, Gianmarco I.P. and Peri, Giovanni.** "The Economic Value of Cultural Diversity: Evidence from US Cities." CESifo Working Paper No. 1117, 2004

**Park, Jungsoo.** "International Student Flows and R&D Spillovers." *Economics Letters*, 2004, *82*, pp. 315-320.

**Powell, Walter W.** "Neither Market nor Hierarchy: Network Forms of Organization." *Research in Organizational Behavior*, 1990, *12*, pp. 295-336.

**Pritchett, Lant.** "Divergence, Big Time." *Journal of Economic Perspectives*, Summer 1997, *11*(3), pp. 3-17.

**Rauch, James E.** "Productivity Gains from Geographic Concentration of Human Capital: Evidence from Cities." *Journal of Urban Economics* November 1993, *34*(3), pp. 380-400.

**__________.** "Business and Social Networks in International Trade." *Journal of Economic Literature*, December 2001, *39*, pp. 1177-1203.

**Rauch, James E. and Trindade, Vitor.** "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics*, February 2002, *84*, pp. 116-130.

**Rivera-Batiz, Luis A. and Romer, Paul M.** "Economic Integration and Endogenous Growth." *Quarterly Journal of Economics*, May 1991, *106*(2), pp. 531-555.

**Rogers, Everett M.** *Diffusion of innovations*. Fourth Edition, New York: The Free Press, 1995.

**Ryan, Bryce and Gross, Neal C.** "The Diffusion of Hybrid Seed Corn in Two Iowa Communities." *Rural Sociology*, 1943, *8*(1), pp. 15-24.

**Sachs, Jeffrey D.** "Institutions Don't Rule: Direct Effects of Geography on Per Capita Income." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 9490, 2003.

**Sachs, Jeffrey D. and Warner, Andrew M.** "Economic Reform and the Process of Global Integration." *Brookings Papers on Economic Activity*, Spring 1995, *1*, pp. 1-95.

**Sala-i-Martin, Xavier X.** "I Have Just Run Two Million Regressions." *American Economic Review*, May 1997, *87*(2), pp. 178-183.

**Sargan, J. Denis**. "Wages and Prices in the United Kingdom: A Study in Econometric Methodology," in P. E. Hart, G. Mills and J. K. Whitaker, eds, *Econometric analysis for national economic planning*, London: Butterworths, 1964, pp. 25-63.

**Saxenian, Annalee.** *Regional advantage: culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press, 1994.

**Schelling, Thomas, C.** *Micromotives and macrobehavior*. New York: W.W. Norton and Company, 1978.

**Sherif, Muzafer, Harvey, O.J.; White, B. Jack; Hood, William R. and Sherif, Carolyn W.** *Intergroup conflict and cooperation: the Robbers Cave experiment*. Norman, Oklahoma: Institute of Group Relations, University of Oklahoma, 1961.

**Solo, Robert A.** *Economic organizations and social systems*. Indianapolis: The Bobs-Merrill Company, Inc., 1967.

**Tarde, Gabriel.** *Les lois de l'imitation*. Second edition, Félix Alan: Paris, 1895.

**Temple, Jonathan and Johnson, Paul A.** "Social Capability and Economic Growth." *Quarterly Journal of Economics*, August 1998, *113*(3), pp. 965-990.

**Uzzi, Brian**. "The Sources and Consequences of Embeddedness for the Economic Performance of Organizations: The Network Effect." *American Sociological Review,* August 1996, *61*(4), pp. 674-698.

**Valente, Thomas W.** *Network models of the diffusion of innovations*. Cresskill, New Jersey: Hampton Press, Inc., 1995.

**White, Halbert.** "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica*, May 1980, *48*(4), pp. 817-838.

**World Bank.** *World development indicators on CD-ROM*. Washington DC: The World Bank, 2000.

**Zak, Paul J. and Knack, Stephen.** "Trust and Growth." *Economic Journal*, 2001, *111*, pp. 295-321.

TABLE 1 – SUMMARY STATISTICS FOR KEY VARIABLES

| | $N$ | Mean | Standard Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Ln*TFP* | 110 | 7.9570 | 0.7195 | 6.2845 | 9.0154 |
| *Ethnic* | 110 | 0.4424 | 0.2763 | 0.0000 | 0.9302 |
| *Language* | 110 | 0.3771 | 0.3028 | 0.0021 | 0.9227 |
| *Religion* | 110 | 0.4217 | 0.2500 | 0.0028 | 0.8603 |
| *Culture* | 106 | 0.2951 | 0.2156 | 0.0000 | 0.7330 |
| *GADP* | 110 | 0.6167 | 0.1958 | 0.3080 | 1.0000 |
| *YrsOpen* | 110 | 0.3581 | 0.3453 | 0.0000 | 1.0000 |
| *Telephones* | 110 | 128.19 | 176.86 | 0.6224 | 663.94 |
| *Popn Density* | 110 | 189.00 | 680.06 | 1.5527 | 5683.4 |
| *Radios* | 110 | 379.25 | 344.95 | 0.2517 | 2119.3 |
| *Road Density* | 110 | 0.5450 | 0.9323 | 0.0043 | 4.7438 |
| *ELF* | 82 | 0.3451 | 0.2975 | 0.0000 | 0.8902 |
| $\Delta AYS$ | 82 | 2.7558 | 1.2827 | −0.8050 | 6.5910 |
| $AYS60$ | 82 | 3.8318 | 2.4689 | 0.1160 | 9.7260 |
| $\Delta \ln KAPW$ | 57 | 0.8840 | 0.6248 | −0.5495 | 3.0909 |
| $\ln KAPW65$ | 57 | 8.3797 | 1.3194 | 4.6347 | 10.536 |
| $\ln RGDPW60$ | 82 | 8.8213 | 0.9268 | 6.5403 | 10.376 |

$N$ is the number of observations. $N = 110$ corresponds to the sample used in Table 4, columns (1)-(3), $N = 106$ to Table 4, column (4), $N = 82$ to Table 2, column (1), and $N = 57$ to Table 2, column (4).

TABLE 2 – CHANGES IN CAPITAL STOCKS AND SOCIAL BARRIERS TO COMMUNICATION

| Dependent variable | (1) $\Delta AYS$ | (2) $\Delta AYS$ | (3) $\Delta AYS$ | (4) $\Delta \ln KAPW$ | (5) $\Delta \ln KAPW$ | (6) $\Delta \ln KAPW$ |
|---|---|---|---|---|---|---|
| Constant | −1.586 | −1.842 | 0.062 | 1.955 | 0.837 | 2.309 |
| | (2.097) | (1.865) | (2.051) | (0.965) | (0.948) | (0.924) |
| *Ethnic* | −1.672 | | | −0.843 | | |
| | (0.741) | | | (0.355) | | |
| *Language* | 0.160 | | | −0.166 | | |
| | (0.654) | | | (0.342) | | |
| *Religion* | 0.283 | | | −0.034 | | |
| | (0.615) | | | (0.312) | | |
| *Culture* | | −1.168 | | | −0.504 | |
| | | (0.697) | | | (0.428) | |
| *ELF* | | | −1.263 | | | −1.187 |
| | | | (0.566) | | | (0.340) |
| *AYS60* | −0.286 | −0.242 | −0.204 | | | |
| | (0.091) | (0.082) | (0.082) | | | |
| ln*KAPW60* | | | | −0.382 | −0.374 | −0.426 |
| | | | | (0.105) | (0.114) | (0.101) |
| Ln*RGDPW60* | 0.678 | 0.664 | 0.443 | 0.278 | 0.366 | 0.272 |
| | (0.247) | (0.227) | (0.243) | (0.158) | (0.171) | (0.149) |
| **Diagnostics** | | | | | | |
| $R^2$ | 0.185 | 0.168 | 0.158 | 0.301 | 0.198 | 0.331 |
| Regression SE | 1.208 | 1.214 | 1.200 | 0.547 | 0.590 | 0.521 |
| *N* | 82 | 79 | 82 | 57 | 54 | 56 |
| *Normality* | 0.527 | 2.279 | 2.048 | 3.516 | 3.014 | 5.565 |
| [*p*-value] | [0.769] | [0.320] | [0.359] | [0.172] | [0.222] | [0.062] |
| *White-Hetero* | 1.043 | 1.064 | 1.659 | 1.520 | 2.843 | 2.470 |
| [*p*-value] | [0.418] | [0.392] | [0.143] | [0.163] | [0.019] | [0.036] |

Notes: Standard errors are in parentheses and *p*-values for diagnostic tests in square brackets. *Normality* is the Doornik-Hansen test of normal errors and *White-Hetero* is White's test for heteroskedasticity.

TABLE 3 – DETERMINANTS OF TFP: OLS RESULTS

| Dependent variable: ln*TFP* | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Constant | 8.533 | 7.206 | 7.237 | 7.297 | 7.342 | 5.905 |
| | (0.138) | (0.260) | (0.213) | (0.273) | (0.369) | (0.300) |
| | [0.142] | [0.300] | [0.183] | [0.243] | [0.423] | [0.278] |
| *Ethnic* | –0.755 | 0.182 | | 0.311 | 0.148 | –0.233 |
| | (0.301) | (0.283) | | (0.311) | (0.335) | (0.311) |
| | [0.274] | [0.290] | | [0.283] | [0.341] | [0.339] |
| *Language* | –0.567 | –0.532 | | –0.763 | –0.560 | –0.652 |
| | (0.278) | (0.229) | | (0.244) | (0.260) | (0.251) |
| | [0.251] | [0.211] | | [0.259] | [0.231] | [0.297] |
| *Religion* | –0.087 | –0.417 | | –0.465 | –0.502 | –0.070 |
| | (0.254) | (0.220) | | (0.211) | (0.272) | (0.270) |
| | [0.279] | [0.223] | | [0.229] | [0.281] | [0.260] |
| *Culture* | | | –0.618 | | | |
| | | | (0.244) | | | |
| | | | [0.213] | | | |
| *GADP* | | 1.310 | 0.952 | 1.273 | 1.190 | 2.293 |
| | | (0.395) | (0.353) | (0.407) | (0.603) | (0.463) |
| | | [0.364] | [0.280] | [0.367] | [0.614] | [0.432] |
| *YrsOpen* | | 0.644 | 0.853 | 0.655 | 0.588 | 0.672 |
| | | (0.206) | (0.199) | (0.201) | (0.235) | (0.231) |
| | | [0.189] | [0.180] | [0.200] | [0.206] | [0.225] |
| **Diagnostics** | | | | | | |
| $R^2$ | 0.243 | 0.494 | 0.470 | 0.575 | 0.336 | 0.722 |
| Regression SE | 0.644 | 0.531 | 0.527 | 0.479 | 0.578 | 0.516 |
| *N* | 118 | 118 | 113 | 108 | 96 | 88 |
| *Normality* | 7.936 | 2.467 | 0.037 | 2.365 | 1.920 | 0.679 |
| [*p*-value] | [0.019] | [0.291] | [0.982] | [0.307] | [0.383] | [0.712] |
| *White-Hetero* | 1.351 | 2.739 | 5.073 | 3.285 | 1.714 | 0.919 |
| [*p*-value] | [0.241] | [0.005] | [0.0001] | [0.001] | [0.091] | [0.521] |

Notes: Conventional standard errors are in parentheses and heteroskedastic-consistent standard errors in square brackets. *Normality* is the Doornik-Hansen test of normal errors and *White-Hetero* is White's test for heteroskedasticity. The sample used in column (4) omits influential observations and/or outliers, and in column (5) omits OECD countries. In column (6) the dependent variable is Islam's (1995) measure of ln*TFP*.

TABLE 4 – DETERMINANTS OF TFP: ROBUSTNESS RESULTS

| Dependent variable: ln*TFP* | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Constant | 8.079 | 8.072 | 8.292 | 7.706 |
| | (0.502) | (0.118) | (0.152) | (0.107) |
| *Ethnic* | 0.122 | | | |
| | (0.305) | | | |
| *Language* | –1.331 | –0.755 | –0.981 | |
| | (0.908) | (0.219) | (0.311) | |
| *Religion* | –0.501 | –0.507 | –0.705 | |
| | (0.258) | (0.217) | (0.328) | |
| *Culture* | | | | –0.570 |
| | | | | (0.245) |
| *GADP* | –0.171 | | | |
| | (0.922) | | | |
| *YrsOpen* | 0.206 | | | 0.722 |
| | (0.393) | | | (0.203) |
| *Telephones* | 0.002 | 0.002 | 0.001 | 0.001 |
| | (0.001) | (0.0004) | (0.0006) | (0.0004) |
| *Popn Density* | 0.00001 | | | |
| | (0.00003) | | | |
| *Radios* | –0.0001 | | | |
| | (0.0005) | | | |
| *Road Density* | 0.017 | | | |
| | (0.117) | | | |
| *Language*Telephones* | –0.002 | | | |
| | (0.003) | | | |
| *Language*Radios* | 0.002 | 0.002 | 0.002 | |
| | (0.002) | (0.0006) | (0.0007) | |
| *Language*Popn* | 0.0003 | | | |
| | (0.001) | | | |
| *Language*Road* | -0.025 | | | |
| | (0.407) | | | |
| *Language*GADP* | 1.032 | | | |
| | (1.859) | | | |
| *Language*YrsOpen* | 0.103 | | | |
| | (0.887) | | | |
| **Diagnostics** | | | | |
| $R^2$ | 0.533 | 0.509 | 0.490 | 0.464 |
| Regression SE | 0.530 | 0.514 | 0.528 | 0.524 |
| *N* | 110 | 110 | 110 | 106 |
| *Normality* | 3.273 | 3.867 | | 0.067 |
| [*p*-value] | [0.195] | [0.145] | | [0.967] |
| *White-Hetero* | 1.524 | 1.412 | | 1.639 |
| [*p*-value] | [0.071] | [0.114] | | [0.049] |

Notes: Standard errors are given in parentheses and *p*-values for diagnostic tests in square brackets. Results in columns (1), (2) and (4) are obtained using OLS. Results in column (3) are median regression estimates.

TABLE 5 – DETERMINANTS OF TFP: IV RESULTS

| Dependent variable: ln*TFP* | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Constant | 7.735 | 7.798 | 8.228 | 8.003 | 8.308 |
| | (0.529) | (0.501) | (0.132) | (0.203) | (0.127) |
| *Ethnic* | –0.001 | | | | |
| | (0.373) | | | | |
| *Language* | –0.540 | | –1.056 | –0.838 | –1.142 |
| | (0.268) | | (0.272) | (0.267) | (0.219) |
| *Religion* | –0.299 | | –0.558 | –0.525 | –0.750 |
| | (0.333) | | (0.237) | (0.268) | (0.246) |
| *Culture* | | –0.800 | | | |
| | | (0.326) | | | |
| *GADP* | 0.112 | –0.468 | | | |
| | (1.136) | (1.078) | | | |
| *YrsOpen* | 1.479 | 1.999 | | 0.871 | |
| | (0.657) | (0.687) | | (0.458) | |
| *Telephones* | | | 0.001 | | |
| | | | (0.0006) | | |
| *Road Density* | | | | 0.147 | 0.362 |
| | | | | (0.158) | (0.113) |
| *Language*Radios* | | | 0.002 | 0.002 | 0.003 |
| | | | (0.0009) | (0.0009) | (0.0007) |
| **Diagnostics** | | | | | |
| $R^2$ | 0.476 | 0.376 | 0.523 | 0.536 | 0.511 |
| Regression SE | 0.544 | 0.610 | 0.506 | 0.499 | 0.510 |
| *N* | 91 | 91 | 99 | 88 | 88 |
| Sargan $\chi^2$ | 1.887 | 2.216 | 5.374 | 2.556 | 6.084 |
| [*p*-value] | [0.596] | [0.529] | [0.146] | [0.923] | [0.638] |
| Hausman $\chi^2$ | 3.386 | 5.819 | 1.990 | 6.849 | 6.612 |
| [*p*-value] | [0.184] | [0.055] | [0.370] | [0.077] | [0.037] |
| **$R^2$ for first-stage regressions** | | | | | |
| *GADP* | 0.741 | 0.667 | | | |
| *YrsOpen* | 0.512 | 0.522 | | 0.571 | |
| *Telephones* | | | 0.703 | | |
| *Road Density* | | | | 0.475 | |
| *Language*Radios* | | | 0.663 | 0.730 | |

Notes: Asymptotic standard errors are given in parentheses and *p*-values in square brackets. $R^2$ for IV regressions is calculated as the squared correlation- between the observed and predicted values of the dependent variable. Sargan $\chi^2$ is Sargan's misspecification test for IV estimation and Hausman $\chi^2$ is a test for the consistency of the corresponding OLS estimates.
Instrument sets: Column (1): *Ethnic*, *Language*, *Religion*, *MeanTemp*, *LT100km*, *StatHist*, *EurFrac*, ln*FraRom*; Column (2): *Culture*, *MeanTemp*, *LT100km*, *StatHist*, *EurFrac*, ln*FraRom*; Column (3): *Language*, *Religion*, *Meantemp*, *LT100km* and the interaction of *MeanTemp*, *LT100km* and *LandArea* with *Language*; Columns (4) and (5): *Ethnic*, *Language*, *Religion*, *StatHist*, *EurFrac*, ln*FraRom*, *MeanTemp*, *LT100km*, *LandArea* and the interaction of each of the last three variables with language.

**End Notes:**

*Authorship is alphabetical. The authors are especially grateful for helpful discussions with Steve Dowrick, Stephen Knowles and Warwick McKibbin in the preparation of this paper. We have also benefited from the comments of Kevin Fox, Tue Gorgens, Jack Pezzey, Rhema Vaithianathan, participants at the Centre for Applied Economic Research (CAER) of the University of New South Wales in November 2002, Department of Economics of the University of Queensland in May 2003, Department of Finance, Canada in July 2003, Conference of the New Zealand Association of Economists in July 2003, Research School of Pacific and Asian Studies, The Australian National University in July 2003, and Conference of Economists of Australia in October 2003. Grafton and Owen acknowledge financial support from the Marsden Fund, administered by the Royal Society of New Zealand, and the research assistance of Clayton Weatherston. Kompas acknowledges the financial support of the Australian Research Council and the research assistance of Tuong Nhu Che.

[1] Hall and Jones note that their estimates are very similar to those obtained in Hall and Jones (1996) where "…the production function is not restricted to Cobb-Douglas, and factor shares are allowed to vary across countries" (Hall and Jones 1999, p. 93).

[2] Given the way the components are measured, high values of *GADP* are conducive to supporting production.

[3] The cut off values used were 2 for the studentized residuals and $2k/N$ for the leverage statistics (David A. Belsley, Edwin Kuh and Roy E. Welsch, 1980).

[4] The diagnostic tests implemented in the search algorithm were the *Normality* and *White-Hetero* tests, discussed above, plus *F*-tests for parameter constancy for breakpoints at the sample mid-point and $90^{th}$ percentile. For the diagnostic tests, a 1-percent significance level was used throughout to help control the overall null-rejection probability, as suggested by the Monte Carlo evidence in Krolzig and Hendry (2001).

[5] A complete listing of the PcGets algorithm is available in Hendry and Krolzig (2001, Appendix A1) or Krolzig and Hendry (2001, Tables 1 and 2).

[6] In this context, power and size relate to the probabilities of retaining in the final model variables that are, respectively, included and not included in the DGP.

[7] Use of a general-to-specific modelling approach also helps address the issue of model uncertainty (William A. Brock and Steven N. Durlauf, 2001; Durlauf, 2002).

[8] Excluding the constant, only the coefficient on *Religion* is statistically significant at the 10-percent level (on a two-tailed test), with the coefficients on *Language* and *Telephones* significant at the 15-pecent level.

[9] This index rates the territory of the current geographical boundaries of a country in terms of whether the government is above tribal level, is colonial or locally based, and the territorial coverage of the government for 50 year sub-periods from 0 to 1950. A single observation for each country is obtained by discounting the effect of past values. We use Bockstette et al's preferred measure corresponding to a discount rate of 5 percent.