Urbanization, Human Capital, and Cross-Country Productivity Differences

Alok Kumar*
Brianne Kober**

Abstract

In this paper, we empirically examine the effects of health, education, and urbanization on the total factor productivity (TFP) of a large number of countries. We find that both urbanization and health indicators (life-expectancy, infant mortality rate, and the risk of malaria) significantly affect TFP. Education has insignificant effect on TFP. Coefficients of indicators of health and urbanization remain highly significant even after controlling for endogeneity.

Keywords: Augmented Solow Growth Model, TFP, Health, Education, Urbanization

JEL Code: F43; E23; N10; N30; O47

Address: $*$ Corresponding Author: Department of Economics, University of Victoria, Victoria, British Columbia, Canada, V8W 2Y2, Email: kumara@uvic.ca. $**$ Department of Economics, University of Victoria, Victoria, British Columbia, Canada, V8W 2Y2, Email: bkober@uvic.ca. This research is supported by the Social Sciences and the Humanities Research Council (Canada). We would like to thank the participants of the CEA 2011 meeting for their comments.

1 Introduction

In this paper, we empirically investigate the effects of urbanization and human capital (both education and health) on total factor productivity (TFP) across countries. There is variety of reasons urbanization, education and health capital can affect TFP.[1] Urbanization directly affects TFP through agglomeration effects, the reduction of transaction costs and economies of scale which allow specialization among firms leading to a lower cost of production. Education determines the ability of a nation to develop new technologies and adopt, implement, and effectively utilize existing technologies. Good health increases the capacity of workers to work. Poor health reduces the availability of workers and works as a disincentive for acquiring and adopting newer technologies.

We derive TFP estimates of 100 countries using an *augmented Solow growth model*. Our analysis shows that both urbanization and health indicators significantly affect TFP. The coefficients of urbanization and health indicators remain significant even after controlling for endogeneity. These results are new to the literature. We find that education has an insignificant effect on TFP.

2 Methodology

Let the production function be

$$Y_{it} = [A_{it}L_{it}]^{1-\alpha-\beta} K_{it}^{\alpha} H_{it}^{\beta} \tag{1}$$

where Y is output, A is technology, L, K, and H are labor, physical and health capital respectively, α and β are the elasticities of output with respect to physical and health capital respectively, and the subscripts denote country (i) and time (t). We do not include education as a factor of production as it is well-known in the literature that education has an insignificant effect on real per-capita income either when a fixed-effect model is used or health capital is included or both. We hypothesize that education affects real per-capita income indirectly through its effect on TFP.

Let g, δ, and n_i be the technology growth rate, the depreciate rate of physical and human capital stock, and the labor force growth rate in country i, respectively. Let $\hat{h}_i^*$, s_i^K and s_i^H denote the steady state level of health capital *per-effective labor* and the investment rates for physical and health

[1] For the extended version of the paper with full set of results see Kumar and Kober (2012).

capitals in country i respectively. Let t_1, t_2, and τ denote the initial period, the current period, and the difference between the current and initial periods respectively. Then, one can derive the following growth regression model (Mankiw et al. 1992, Islam 1995)

$$\ln y_{it_2} = \frac{(1-\exp^{-\lambda\tau})\alpha}{1-\alpha}\ln s^K_{i\tau} - \frac{(1-\exp^{-\lambda\tau})\alpha}{1-\alpha}\ln(n_{i\tau}+g+\delta)+$$

$$\frac{(1-\exp^{-\lambda\tau})\beta}{1-\alpha}\ln h^*_{i\tau} + \exp^{-\lambda\tau}\ln y_{it_1} + g(t_2 - \exp^{\lambda\tau} t_1) + (1-\exp^{-\lambda\tau})\ln A_{i0} \quad (2)$$

where $\lambda = (1-\alpha-\beta)(n+g+\delta)$.

Equation (2) represents a dynamic panel data model with $(1-\exp^{-\lambda\tau})\ln A_{i0}$ as the time-invariant country-effect term and can be written in the following conventional form of the panel data literature:

$$y_{i,t} = \gamma y_{i,t-1} + \sum_{j=1}^{3}\phi_j x^j_{it} + \eta_t + \mu_i + v_{it} \quad (3)$$

with

$$y_{i,t} = \ln y_{it_2};\ y_{i,t-1} = \ln y_{it_1};\ x^1_{it} = \ln s^K_{i\tau};\ x^2_{it} = \ln(n_{i\tau}+g+\delta);$$

$$x^3_{it} = \ln h^*_{i\tau};\ \eta_t = g(t_2 - \exp^{\lambda\tau} t_1)\ \&\ \mu_i = (1-\exp^{-\lambda\tau})\ln A_{i0} \quad (4)$$

where v_{it} is the idiosyncratic error term with mean zero. Note that the Solow model puts the restriction that $\phi_1 = -\phi_2$. In the first step, we use (3) and (4) to derive the estimates of A_{i0}.

In the second step, we analyze the determinants of cross-country TFP differences. For this analysis, we estimate the following regression model:

$$\ln A_{i0} = \pi U + \Pi_1 H + \Pi_2 S + \Pi_3 Z + u_i \quad (5)$$

where U is the indicator for urbanization and π is the associated coefficient, H is the matrix of variables measuring health capital, Π_1 is the associated vector of coefficients, S is the matrix of indicators of education, Π_2 is the associated vector of coefficients, Z is the matrix of other explanatory variables including constant term, Π_3 is the associated vector of coefficients, and u_i is the idiosyncratic term with mean zero. The description of independent variables used to estimate equation (5) and their data sources are given in the Appendix.

3 Growth Regression

Our main sources of data are Penn World Tables 6.3 and World Development Indicators. The income data are real GDP per-worker adjusted for purchasing power parity. For the savings rate we use the ratio of real investment to real GDP. Following Mankiw et al. (1992), we set $(g+\delta)$ to be equal to 0.05. We proxy health capital by $LLE = -\ln(90 - LE)$, where $(90 - LE)$ is the shortfall of the average life expectancy (LE) at birth from 90 years. In the sample, we include only those countries for which data is available for the entire period, 1960-2005.[2]

For a panel analysis, similar to Islam (1995) we divide the total period into several shorter time spans, each consisting of five years. Thus, we have ten data points for each country: 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005. When $t = 1965$ for example, then $t - 1 = 1960$, and the savings rate, $s^K_{i\tau}$, the labor force growth rate, $n_{i\tau}$, and health capital, $\hat{L}LE^*_{i\tau}$ are averages over $1961 - 1965$. The real income per worker for time t, y_{it}, is the real income per worker for the year 1965 and y_{it-1} is the real income per worker for the year 1960.

Table 1 presents the results of the first stage regression. Breusch-Pagan (BP) and Hausman (H) tests suggest that a fixed-effect model is the appropriate model to estimate (4). We use the Arellano and Bond (1991) GMM method (AB method), a widely used method, to estimate the parameters of (4). We only report estimates of the restricted version of the Solow model as the Wald test shows that the implied restriction by the Solow model ($\phi_1 = -\phi_2$) is not rejected for any of the estimated equations. The AB method provides for one-step and two-step estimators. In what follows, we take two-step specification as our preferred model as the Sargan test results support a two-step specification in place of one-step estimates.

4 Cross-Country Productivity Differentials

Note that the estimate of TFP, $\ln A_{i0}$, though derived using data from 1960-2005 pertains to the initial productivity level i.e. the productivity level in 1960. Due to this we use 1960 values of independent variables. We proxy urbanization level by the logarithm of the proportion of the urban population (LUR) in 1960. We proxy education by the logarithm of the fraction of the population aged 15 years and above completing secondary education (LSC)

[2] We also used infant mortality rate as a proxy for health capital. Results are similar (Kumar and Kober 2012).

in 1960.[3]

We use three indicators of health to proxy the mortality rate and the disease burden. For mortality rates we use $LLE_{60} = -\ln(90 - LE_{60})$ and $LMR_{60} = \ln IMR_{60}$ where LE_{60} and IMR_{60} are the life-expectancy and the infant mortality rate for 1960. We proxy disease burden by the risk of malaria. Since malaria risk data is not available for 1960, we use Malaria Ecology (ME) as an indicator for malaria risk.

We control for other determinants of TFP such as trade openness, ethno-linguistic fractionalization, and legal origin. We proxy trade openness by the log of the ratio of export and import to GDP in current prices in 1960 (LOPEN). To capture the effects of legal origin and ethic heterogeneity, we use dummies for socialist countries (SOC) and the index of ethno-linguistic fractionalization (ETH). We also use a dummy for African countries (AFRICA) and oil exporting countries (OIL).

Estimated results are given in Table 2. In Model 1, we include all other independent variables except LUR. The results show that the coefficients of LLE_{60} and LSC are highly significant. The coefficient of ME is significant, but only at a 10% level of significance. In Model 2, we incorporate LUR. The result shows that LUR has a positive and highly significant effect on TFP, but the coefficient of LSC becomes insignificant. Similar is the result when we use LMR_{60} as a proxy for the mortality rate rather than LLE_{60} (Model 3).

Health, education, and urbanization are potentially endogenous. To control for endoegeniety we use an IV approach. We use a proportion of area of a country under tropics (TROPIC), the distance from the equator (LATI), the proportion of the population within 100 kilometers of coastal areas (LT100), a dummy for land-locked countries (LAND), a dummy for countries where parental gender bias in favor of sons is prevalent (SON), and the percentage of Muslim (MUSLIM) and Catholic (CATH) population in a country as instruments.

These variables have been chosen as instruments as literature suggests that they are important determinants of human capital and urbanization. We use SON as an instrument as empirical evidence suggests that a son-preference is wide-spread in many regions of the world and it disadvantages female children particularly with respect to access to health facilities, nutrition, and education. This in itself is likely to lead to a lower human capital. In addition, literature suggests that a mother's education and health status positively and significantly affect the health and education of children. Due to this we expect human capital to be lower in countries with a strong

[3]For results with other indicators of education see Kumar and Kober (2012).

son-preference. The list of countries in which a son-preference is prevalent is given in Table 3.

We find that these instrument variables explain significant proportion of cross-country variations in human capital and urbanization and are significant at the 1% level of significance. Models (4) and (5) present 2SLS results. Model (4) shows that the coefficients of both LLE_{60} and LUR remain significant at the 5% level of significance. However, the coefficient of ME is significant at only a 10% level of significance. In Model (5), we use LMR_{60} as a proxy for the mortality rate. Results are similar. The estimated p-values for the Sargan tests imply that the over-identifying restrictions are not rejected.

5 Conclusion

In this paper, we studied the effects of health, education, and urbanization on the TFP of a large number countries. We find that health capital affects both per-capita income and TFP. Urbanization has a positive and highly significant effect on TFP. Education has an insignificant effect on TFP. The IV estimates suggest that the effects of health and urbanization on TFP are causal.

References

[1] Barro, R. and J. Lee (2001), "International Data on Educational Attainment: Update and Implications," *Oxford Economic Papers*, 53(3), 541-563.

[2] Gallup, J., Sachs J., and Mellinger, A. D. (1999), "Geography and Economic Development", *International Regional Science Review*, 22, 179-232.

[3] Fuse, K. (2010), "Variations in Attitudinal Gender Preferences for Children Across 50 Less-Developed Countries", *Demographic Research*, 123, 1031-1048.

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

[5] Kumar, Alok and B. Kobber (2012), "Urbanization, Human Cpital, and Cros-Country Productivity Differences" available at http://web.uvic.ca/ kumara/productivitytotal.pdf

[6] La Porta, R, Lopez-de Silanes, F., Schleifer, A. and Vishny, R. W. (1999), "The Quality of Government", *Journal of Law, Economics, and Organization*, 15, 222-282.

[7] Mankiw, N. G., D. Romer, and D.N. Weil (1992), "A Contribution to the Empirics of Growth", *Quarterly Journal of Economics*, 107, 407-437.

[8] Scahs, J. (2003), "Institutions' Don't Rule: Direct Effects of Geography on Per-Capita Income", *NBER Working Paper 9490.*

[9] Williamson, N. E. (1976), "*Sons or Daughters: A Cross-Cultural Survey of Parental Preferences*", Sage Publications.

Table 1
Growth Regression Results

Explanatory Variables	AB 1-Step (1)	AB 2-Step (2)
$y_{i,t-1}$	$0.6415(0.057)^{*}$	$0.652(0.03)^{*}$
$\ln(s_{i\tau}^{K}) - \ln(n_{i\tau} + g + \delta)$	$0.0715(0.038)^{**}$	$0.0785(0.018)^{*}$
$\hat{L}LE_{i\tau}^{*}$	$0.1968\ (0.074)^{*}$	$0.2151\ (0.037)^{*}$
Implied α	0.17	0.18
Implied β	0.46	0.50
Implied λ	0.09	0.086
p Values:		
Wald Test	0.15	0.47
Sargan Test	0.003	0.13
H(0): AR(2) is absent	0.16	0.19
No. of Observations	796	796
No. of Countries	100	100

Note:

1. $*$, $**$, and $***$ indicate significance levels of 1%, 5%, 10% respectively against two-sided alternatives for the t-tests.
2. Number in brackets are standard errors.
3. All specifications included time specific effects (not reported).

Table 2
TFP , Human Capital, and Urbanization

Variable	(1)	(2)	(3)	(4)/2SLS	(5)/2SLS
LLE_{60}	1.0682*	0.6315*		0.8945**	
	(0.17)	(0.15)		(0.38)	
LMR_{60}			−0.3143*		−0.2097***
			(0.05)		(0.12)
ME	−0.0235***	−0.0254**	−0.0337*	−0.0158***	−0.0371*
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
LSC	0.1208**	0.0545	0.0527	0.1244	0.1279
	(0.05)	(0.04)	(0.03)	(0.23)	(0.13)
LOPEN	0.0087	-0.0166	0.0085	0.0879	−0.0026
	(0.06)	(0.05)	(0.06)	(0.20)	(0.14)
LUR		0.3877*	0.3562*	0.5093**	0.3896**
		(0.08)	(0.07)	(0.21)	(0.16)
SOC	−1.07*	−1.3633*		−1.1365*	
	(0.13)	(0.14)		(0.43)	
OIL	0.3002**	0.2099*	0.2172*	0.1888**	0.1941**
	(0.12)	(0.08)	(0.07)	(0.10)	(0.08)
AFRICA	−0.1417	−0.2557***	−0.1994	−0.2236	−0.2595
	(0.18)	(0.14)	(0.15)	(0.19)	(0.15)
ETH	−0.1238	−0.0638	−0.2503	−0.0702	−0.1704
	(0.26)	(0.19)	(0.17)	(0.29)	(0.24)
R^2	0.74	0.83	0.87	0.75	0.85
N	74	74	63	74	63
Sargan (p-value)				0.60	0.64

Note: 1. $*$, $**$, and $***$ indicate significance levels of 1%, 5% and 10% respectively against one-sided alternatives for the t-tests.

2. Numbers in parentheses are White Heteroskedasticity-Consistent standard errors.

Table 3
List of Countries Where a Son-Preference is Prevalent

Algeria, Bangladesh, Benin, Burkina Faso, Chad, China, Egypt, Ethiopia, Guinea, Hong Kong, India, Iran, Jordan, South Korea, Mali, Morocco, Nepal, Niger, Nigeria, Pakistan, Senegal, Syria, Tunisia

Appendix
Variables and Data Sources

- y: Real income per worker in 2005 constant price PWT 6.3
- s^K: Investment share of real GDP per capita *PWT 6.3*
- n: Calculated using LFPR and population *PWT 6.3*
- LE: Life Expectancy *World Development Indicators*
- MR: Infant Mortality Rate *World Development Indicators*
- ME: Malaria Ecology *Sachs 2003*
- LAV: The average years of school for adults *Barro and Lee (2001)*
- LPC: The fraction of population aged 15 years and above completing primary schooling *Barro and Lee (2001)*
- LSC: The fraction of population aged 15 years and above completing secondary schooling *Barro and Lee (2001)*
- LTC: The fraction of population aged 15 years and above completing tertiary schooling *Barro and Lee (2001)*
- LUR: Fraction of urban population in 1960
- LATI: The absolute distance from equator *La Porta et. al. (1999)*
- TROPIC: The percentage of area of a country under tropics *Gallup et. al. (1999)*
- LT100: The proportion of population within 100 k.m. of coast *Gallup et. al. (1999)*
- LAND: Dummy for land-locked countries

- SOC: Countries with socialist legal system *La Porta et. al. (1999)*
- ELH: Index of ethno-linguistic fractionalization *La Porta et. al. (1999)*
- MUSLIM: Proportion of muslim population *La Porta et. al. (1999)*
- CATH: Proportion of catholic population *La Porta et. al. (1999)*
- SON: Dummy for countries where son-preference is prevalent *Williamson 1976, Fuse 2010*
- OIL: Dummy for oil exporting countries