CORE DISCUSSION PAPER

2003/40

# Does Urbanization Always Foster Human Capital Accumulation?

Luisito BERTINELLI[1]

May 2003

**Abstract**

The importance of human capital has been crucial in explaining the process of economic development. In the present study, we perform cross-country estimations, measuring the relation between human capital accumulation and urbanization. Using a macro level approach we highlight a U-shaped relation, where urbanization rates below 40 per cent deter human capital accumulation. This especially holds for developing countries, raising policy concerns on issues of over-urbanization.

**JEL classification**: O18, R10

**Keywords**: urbanization, human capital accumulation, cross-country estimations

[1]FNRS fellow. CORE, Université catholique de Louvain, 34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium.

Comments by Salvador Barrios, Harris Selod, Eric Strobl and Jacques Thisse are gratefully acknowledged.

# 1 Introduction

The recent growth literature has emphasized the importance of education and human capital in the process of economic growth and development. Human capital being unequally distributed across countries, parts of the inequalites that persist worldwide may come from differential educational attainments. Although human capital accumulation and growth go hand in hand, there is evidence showing that some areas are more prone to investment in human capital. In particular, urban areas favoring interactions between people, enhance the returns to investment in education (Jacobs (1969), Lucas (1988)). This in turn may at least partly explain differences of develoment across countries. The present study's aim is to test for this idea by running cross-country estimations quantifying the impact of urban areas on human capital accumulation.

There may be numerous resons why urban areas should be more prone to higher educational attainments. Among those, human capital externalities have been most prominent. In particular, labor economists have extensively analyzed educational attainments to provide measures of returns to investment in education. However, a major issue that has mostly been dismissed relate to the *private vs. social* returns to education, i.e., whether gains from education totally accrue from individuals performing this investment or whether there may be societal benefits from individual investment in human capital, as for instance lower unemployment rates, lower crime rates among others (Moretti (2004)).

Social effects resulting from individual actions are referred to externalities in the economics literature, and a possible explanation for the lack of studying these externalities probably lies in the difficulty to tackle such impalpable phenomena. There are nonetheless two major reasons why knowledge externalities should be studied: (i) The first reason refers to recent developments in economic growth literature. Recent work on endogenous growth theory, starting from Lucas (1988), rely on the existence of human capital externalities, which play the role of the renewable energy to the engine of economic growth; (ii) The second motivation relates to welfare issues. It is a well established result in economics that the divergence between private and social returns to education lead to sub-optimal market outcomes, requiring hence public intervention to correct for this market failure.

In the present study, we intend to provide indirect evidence on the existence of human capital externalities in cities. Externalities, because they crucially rely on interactions between agents, are mostly local in nature. Densly populated locations should favor interaction among people and, hence, cross-fertilization of ideas (Jacobs(1969)): *skills feed themselves.* This in turn may increase the incentives to invest in education. Following Lucas (sect.6, 1988), we hypothesize urban areas to be such locations and measure the impact of urbanization on human capital accumulation.

Generally, human capital entails two distinct concepts which are *experience* and *education*. The distinction between these two definitions is important in empirical studies. The first one relates roughly to the vast literature on learning-by-doing, and has been used in theoretical studies by Lucas (1993) and in empirical studies by Peri (1999) among others. In the second definition, human capital is identified with educational attainment.[1] Lucas (1988) as well as the empirical labor economics literature on the determination of returns to education is based on this second definition. The economic theories of human capital state that human capital fosters productivity growth, which in turn stimulate wages (Becker (1964), Mincer (1974), Willis (1999)). Hence, a conventional but indirect mean of estimating the returns to education is to observe wages. However, doing so, one is confronted with unobservables, which might influence wages as well. This notably holds for human capital externalities.

Despite their unobservable nature, there seems to be a general agreement on the fact that human capital externalities and, more generally, knowledge externalities arise where interactions between economic agents are frequent and important (what Leamer and Storper (2001) refer to as *handshake interactions*). This is typically the case for densely populated areas, and hence, studies aiming at measuring human capital externalities have done so on regional and/or city data (see for instance Moretti (2004) for a very insightful survey).

Qualitative differences in the results of studies adopting this approach are principally due to the econometric choice of specification and technique, the definition of human capital that is used, and the geographic unit at which the externality is accounted for. For instance, whereas Rauch (1993) and Moretti (1998) find positive and significant externalities, Rudd (2000) and Ciccone and Peri (2002) find non-significant externalities when working on micro level data. Results of Acemoglu and Angrist (2000) are ambiguous according to the time period. Thus, despite a theoretical consensus on the existence of human capital externalities, empirical evidence remains inconclusive.

These previously mentioned results may depend on the choice of spatial unit of observation. As already alluded above, evidence on knowledge

[1] Unless mentionned otherwise, in what follows, we will use these two terms, *human capital* and *educational attainment/ investment* interchangeably.

externalities tend to support strong distance decay effects (Lamorgese and Ottaviano (2002)). If proximity between people matters, then we should observe that highly urbanized countries have higher returns to education and hence, higher incentives to invest in human capital. This means that capturing human capital externalities can only occur when working on *sufficiently* small geographic units. In Rudd (2000) as well as in Acemoglu and Angrist (2000) for instance, US state level data is used, which may be too rough a measure to capture knowledge spillovers.

As a first approximation one could argue that cities respectively metropolitan areas should be the unit of observation. Cities are more productive than rural areas, and this productivity is positively linked to their size (Moomaw (1981), Tabuchi (1986) among others). Explanation of this relate to the idea that dense areas favor interaction among people and, hence, cross-fertilization of ideas (Jacobs(1969)): *skills feed themselves*. Evidence on this is the city specific wage premium. Glaeser and Maré (2001) state that in the US, SMSA workers earn in real terms 36 per cent more than their counterparts working outside these metropolitan areas.

Departing from the measurement of returns to education approach, the present study's aim is to provide indirect evidence on social returns to education in cities. It however differs from work done by labor economists, in using macro rather than micro level data. Starting from the fact that human capital externalities are local in nature, it must be that dense areas favor the internalization of these externalities. Cities thus are the ideal ground for human capital accumulation to grow, and we expect countries with high levels of urbanization to highlight stronger growth of human capital.

The rest of the paper is organized as follows: section 2 provides some explanations about the data that has been used and displays some basic statistics. Section 3 introduces the empirical framework and presents the results. Section 4 concludes.

## 2 Data and basic statistics

As stated above, the present study departs from the usual micro level studies of labor economists in that it is based on macro level observations. Hence our database is based on country level data. More specifically, the data originates from different sources and was compiled by James Davis and Vernon Henderson. The data spans the period 1960-90 by 5-year periods. Our main focus will be data on urban population (stemming from the UN World Urbanization Prospects) and levels of human capital (as proxied by the average years of secondary and higher education of the population aged 25 and more; the source is from the Barro and Lee (1993) data set). Further details are provided in the data source appendix. Although our human capital variable does not adjust for the quality of education, nor the number of hours of education per week (respectively per year), it is to our knowledge the only unified source that spans about 100 countries over 30 years, and thus constitutes the best available proxy.

[Table 1]

Table 1 provides some summary statistics. Accordingly, urbanization has increased by about 14 percentage points (36 per cent), whereas human capital as well as GDP/capita have more than doubled (166 per cent respectively 122 per cent) during the 1960-90 period.[2] However as can be deduced from the standard deviations, there are large variations in our sample. In particular, as can be expected, countries with low initial levels of urbanization (i.e. less than 30 per cent), increased most (47 per cent), whereas intermediate (between 30 and 60 per cent) and high (more than 60 per cent) urbanized countries have seen their urbanization rates increase less (9 respectively 3 per cent).

A closer look at the data reveals that countries with high urbanization rates are also those entailing high levels of human capital (see Fig.1). Relating both urbanization and human capital, we obtain correlation coefficients between 0.66 and 0.70 according to the period. A result worth mentioning appears also when correlating human capital to *present and past* values of urbanization. The same correlations have been computed by replacing human capital levels by their variations (see Tables 2a and 2b).[3]

[Table 2a]

[Table 2b]

This very rough measure of the relation between our two variables of interest hints to some non negligible comovement. If we consider our human capital variable as a stock measure and its 5-year variation as a flow measure, then it seems that past values of urbanization matter at least as much as present values for the stock of human capital, whereas there is a time decay effect when considering flows of human capital.

However, so far relations rather than functions have been presented, and figures would also be consistent with results supporting the idea that human capital accumulation induces forward looking agents to migrate to

[2] Remember that the urbanization rate is bounded between 0 and 1, contrary to the other two variables which are open-ended.

[3] These results are all significant at 1 per cent. Similar results were obtained when weighting by a measure of GDP/capita.

cities. Hence, in the next section, we impose a parametric structure on our data to determine the contribution of cities in the development of human capital.

# 3 Empirics

## 3.1 Specification

We are interested in determining the contributing factors of human capital accumulation. An issue we are confronted with in running cross-country regressions is country-specific heterogeneity. To address this issue, we introduce fixed effects in our estimations, hence allowing for a different constant per country, i.e., allowing omitted variables to have differential impacts across different countries. Our basic specification is given by:

$$\begin{aligned}\underbrace{\Delta\left(human\ capital\right)_{it}}_{NT\times 1} &= \alpha+\beta\underbrace{\left(human\ capital\right)_{i,t-1}}_{NT\times 1}+\delta\underbrace{\left(urbanization\right)_{i,t-1}}_{NT\times 1}\\ &+\underbrace{\left(controls\right)_{i,t-1}}_{NT\times K}\underbrace{\gamma}_{K\times 1}+\mu_i+\eta_t+\epsilon_{it} \qquad (1)\end{aligned}$$

where $\alpha$, $\beta$ and $\delta$ are scalars, $\mu_i$ and $\eta_t$ are country and period-specific effects respectively, and $\epsilon_{it}$ is an iid random term, verifying the following properties:

$$E\left(\epsilon_{it}\right)=0,\text{ and }E\left(\epsilon_{it}\epsilon_{js}\right)=\begin{cases}\sigma_t^2 & \text{if } i=j\\ \sigma_i^2 & \text{if } t=s\\ 0\text{ otherwise}\end{cases}.$$

Although fixed effects results are reported in our results, the coefficients are plagued by endogeneity as $\left(human\ capital\right)_{i,t-1}-\overline{\left(human\ capital\right)_i}$ is by construction correlated with $\epsilon_{it}-\bar{\epsilon}_i$. Following Arellano and Bond (1991), we thus wipe out fixed effects by first differencing specification (1) and then instrumenting the RHS endogenous variables:

$$\Delta\left(human\ capital\right)_{it}-\Delta\left(human\ capital\right)_{i,t-1}=\left(X_{i,t-1}-X_{i,t-2}\right)\varphi+\left(\epsilon_{it}-\epsilon_{i,t-1}\right) \qquad (2)$$

The first RHS expression can be instrumented by $X_{i,t-2}$, which is highly correlated with $(X_{i,t-1}-X_{i,t-2})$ but uncorrelated with $(\epsilon_{it}-\epsilon_{i,t-1})$ as long as $\epsilon_{it}$ are serially uncorrelated. For the following period, $X_{i,t-1}$ is a valid instrument for $(X_{i,t}-X_{i,t-1})$, but so is $X_{i,t-2}$, and so forth for subsequent periods. Hence typically for dynamic panels, one has a different number of instruments for different periods, leading to an overidentification issue. The literature suggests two ways out: either discard some instruments (Anderson and Hsiao, 1981), leading to a loss of some information, or estimating the overidentified system via GMM (Arellano and Bond, 1991), exploiting a different number of instruments in each time period. The latter solution has been adopted here. Estimations are run on differenced series and the number of instruments are generally limited to two periods in order to avoid overfitting biases. Except for time dummies, all explanatory variables are instrumented.

### 3.2 Basic results

In what follows, we should stress that we are not interested in measuring differences in returns to human capital in cities and in more remote areas for example. The point here is to detect, whether the distribution of people, and more particularly, their agglomeration in space is neutral or not in terms of incentives to invest in human capital. Doing so, we avoid the problem of choosing the optimal spatial unit

In the sequel, interpretations will be based on the two-step heteroskadasticity consistent GMM estimates. Results for fixed effects are also provided for the interested reader.[4] The assumptions on no serial correlation in the error term are tested and hold for all our estimations: there is significant first order correlation and no evidence of second order serial correlation, thus the disturbances are serially uncorrelated.

In the first and the fifth column of Table 3, basic specification results are reproduced, where only the base period dependent variable as well as time dummies have been added as RHS explanatory variables. In a growth type interpretation, we can note that there is convergence between the average levels of human capital. Actually, the bulk of the 30 least educated countries had an *annual* human capital growth rate of about 4.9 per cent between 1960 and 1990, whereas for the 30 most educated, this rate has been about 2.6 per cent. Moreover, GMM estimates point to slightly smaller speeds of convergence. This is a common result when working with dynamic panels. Whereas OLS levels estimates of $\beta$ tend to be upward biased, Within estimates are downward biased (Bond *et al.*, 2001).[5] The introduction of

[4] All estimations were run with DPD98 package for unbalanced panels and on STATA 7.0.

[5] This is exactly what we find here. GMM coefficients are in between the OLS and Within $\beta$. ($\beta$s for OLS estimations are given by $0.059^{***}$,0.001, $-0.032$ $^{**}$ and $-0.040^{**}$ for first, second, third and fourth columns respectively).

the urbanization variable accelerates the rate of convergence. In any case, the presence of cities seems to favor human capital accumulation: coefficients are always positive and highly significant, and even more so when accounting for possible endogeneity, supporting Glaeser and Maré (2001)'s statement: "The evidence suggest that cities speed the accumulation of human capital" (p.316). This is true, even after controlling for general levels of development, as is done in columns (3) and (7) respectively (4) and (8). Including GDP/capita, life expectancy at birth and fertility, which should account for the general level of development of countries, the coefficient of urbanization, though decreasing, remains positive and highly significant.[6] *A five percentage point deviation of urbanization fosters a supplementary five-year human capital acceleration of 0.07 and 0.085 per cent, according to columns (7) respectively (8).*

[Table 3]

### 3.3 Non-linear impact of urbanization

In Table 3, we have made the assumption that the marginal gains of urbanization are independent of the absolute levels of urbanization. Hence, this amounts to stating that the United States, with a 75 per cent urbanization rate in 1990, and Albania, with 37 per cent, do benefit in the same fashion from a one percentage point increase in city dwellers, everything else being constant. The fact that the possible number of contacts between pairs of persons increases exponentially when urbanization increases points to the fact that marginal gains from urbanization should actually be increasing. Least, from an econometric point of view, the fact that the urbanization variable is bounded below and above strongly suggests that in between these two bounds, effects are non-linear.

In Table 4, we add a non-linear impact on the variable of interest, i.e., urbanization. Columns (1) and (4) of Table 4 are similar to columns (3) and (7) of Table 3, except that we additionally test for the non-linear returns of urbanization on human capital accumulation. Speeds of convergence tend to remain stable. More interestingly however, results in columns (1) and (4) depict a U-shaped relation between urbanization and human capital accumulation, hence increasing marginal impact of urbanization (Figure 2). According to column (4) results, urbanization highlights a deterrent effect on educational attainments for urbanization levels below 40 per cent.[7] In

[6] Other controls have been included (road density, telephone main lines per 1000 inhabitants, trade relative to GDP), but left the results qualitatively unchanged.

[7] Note that this results should not be taken at face value. Changing the control variables

1960, about $\frac{2}{3}$ of the countries were in this situation, and in 1990, still $\frac{1}{3}$. Hence this points to the existence of an under-urbanization trap. This idea echoes with the idea originally developed by Harris and Todaro (1970) according to which cities are too big from an optimal point of view. Of course, our approach being macro rather than micro based, we cannot identify urban concentration issues, which is the prime issue in Harris and Todaro. Nonetheless, urban migration seems to act counterproductively against educational investments in the first stages of development.

[Table 4]

A further observation is the unbounded increasing effect urbanization has on human capital (beyond the under-urbanization trap). The conclusions to such a process lead to a dead end, where countries with urbanization rates equal to 100 per cent have the highest human capital accumulation rates. First, one has to note that this result is dictated by our particular specification choice, i.e., introduce urbanization as a polynomial of degree 2. Running again the same specification as in Table 4 column (4), with a degree 3 polynomial instead, the coefficient of $(urbanization)^3$ is negative and statistically significant ($-2.916^*$). Hence, we have a maximum beyond which, congestion effects eventually dominate positive knowledge externalities.

The idea above is that urbanization can favor human capital *accumulation*, provided there is a significant *level* of human capital. Stated differently, urbanization and human capital levels work hand in hand. To test for this idea, in columns (2) and (3) respectively (5) and (6) of Table 4 we run basic specifications by adding the following interaction terms, measuring the marginal impact of urbanization, according to the base period level of human capital:

$$urbanization_{it}[\delta + \lambda_1 human\ capital_{it}] \qquad (3)$$

$$urbanization_{it}[\delta + \lambda_1' human\ capital_{it} + \lambda_2'(human\ capital_{it})^2] \qquad (4)$$

The $\delta$ remains always positive, thus urbanization per se is non negligeable. Of more interest are the results on the interaction terms: $\lambda_1$ and $\lambda_1'$ are positive, supporting the idea that the marginal impact of urbanization increases with the absolute level of educational attainments of a population. The higher the average level of education, the higher the potential of externalities and hence the larger the role of cities as means of diffusion of

induces slight changes in the *urbanization* coefficients, and hence, in the minimum of the parabola. The important point is that additional city dwellers seem to deter human capital accumulation for *low* levels of urbanization.

knowledge. This result is consistent with Glaeser (1994)'s claim that *schooling has a very powerful effect on increasing schooling growth.* However, according to the sign of $\lambda_2'$, the marginal impact of urbanization decreases with the level of human capital above a certain threshold level of education. Note however that this level lies at about 4 years, which is by far above the 1990 value of average secondary and higher education in our sample and only concerns countries in the upper 5 percentile of the distribution leaving room for the remaining 95 percentiles to benefit from positive marginal impact of urbanization to be fostered by the level of human capital.

### 3.4 Some further extensions

Results so far support the idea that urbanization is a catalyst to human capital accumulation. People may benefit more from their human capital by living in/ migrating to cities, and the higher their level of human capital, the higher their benefits from being in cities. However these results were obtained in a well-defined scheme: (i) a set of about 100 countries basically chosen for their data availability; however, there are large differences among countries in our data set; in particular, we mix up developed and developing countries (ii) moreover, urbanization is supposed to foster secondary *and* higher education attainments both together; but we do not provide any insights on alternative definitions of human capital; (iii) *urbanization* is hypothesized to be the ideal ground of *handshake interactions*, but definitions of urbanization may be quite different across countries, depending on the definition of cities, and may actually only be a proxy for agglomeration of people and economic activity, which may as well be proxied by population density!

To partly tackle these issues, we present in Tables 5, 6 and 7 some extensions of the estimations proposed in Tables 3 and 4. More particularly, three sets of estimations are presented hereafter: estimations with three different definitions of human capital are rerun: averages of secondary, higher and total education (Section 3.4.1); rather than suppose that urbanization is the best proxy for agglomeration of people, we replace our agglomeration variable by a population density variable (Section 3.4.2); finally, estimations on country groups are run, so as to ease the heterogeneity between countries, and allow for different slope coefficients between country groups (Section 3.4.3).

### 3.4.1 Alternative definitions of human capital

In Tables 3 and 4, human capital has been defined as the average educational attainments of the population aged 25 and more. Educational attainments have been chosen as the secondary and higher years of education. Primary education was excluded because in many countries, primary education is compulsory. The age limit of 25 years is taken in order to account for the working population and not the student population.

Columns (1) and (2) respectively (7) and (8) of Table 5 are based on secondary education variables whereas columns (3) and (4) respectively (9) and (10) are based on higher education variables, the aim being to disentangle the aggregate effects of secondary and higher education described in Tables 3 and 4. Finally, columns (5) and (6) respectively (11) and (12) of Table 5 account for average years of total education (primary, secondary and higher). The point is that, although in most developed countries, primary education has been compulsory for a long time, this is not the case in many developing countries. Moreover, one has to take account of population momentums: a person aged 60 may be living in a country where primary education is compulsory but has not been so 50 years ago.

[Table 5]

Except for higher education (column (9)), coefficients highlight the expected signs, but are mostly statistically non significant at our chosen significance level. Results in column (7) are in line with column (7) of Table 3, although the impact of urbanization is much weaker. The same holds for column (8), but the minimum of the U-shaped curve is reached for a 49 per cent rate of urbanization (compared to 40 per cent above). As we have mentioned before, these values have to be taken with caution: they support the idea of an under-urbanization trap, though the threshold value is uncertain.[8] For the higher education regressions in column (9), the coefficient of urbanization has exactly the opposite sign as expected, although for column (10), urbanization impacts positively on human capital accumulation for a range of urbanization rates above 36 per cent.

Besides p-values above 0.1, coefficients are very low in absolute value for columns (9) and (10), pointing hence to non significant results from a statistical as well as economic perspective. This may stem from the fact that less people reach higher education, thus their supplementary months/years of higher education due to urbanization are *diluted* when computing averages.

[8] Further support to this idea of under-urbanization trap is provided in Figure 3 in the Appendix, where the evolution of urbanization and human capital is plotted for three developing countries.

Furthermore, it results that the marginal impact of urbanization on human capital accumulation is much flatter for higher education than secondary education. Finally, columns (11) and (12) point to more cumbersome results: marginal impact of urbanization are highest in this case. An explanation to this may come from taking into account primary education, which may bias our results.

### 3.4.2 Urbanization vs. population density

The claim of this paper is that *proximity* of people favors absorption of the external effects of human capital. Hence, densities of population rather than urbanization may as well be a proxy for the proximity of people. The more people per unit area, the smaller the average distance between each member of the community. In other studies, there have been a certain number of studies working with densities, the most prominent of which has been Ciccone and Hall (1996).

In order to disentangle the differential effect of urbanization and population density to assess for their (dis)similarities, Table 6 reports results where we have run specifications (7) and (4) of Table 3 respectively Table 4. The most striking result is that none of our variables of interest appear to be statistically significant (and by far not as one can deduce from the standard deviations in parenthesis). Second, columns (1)-(3) highlight the expected signs, but column (4) refers to an *inverted* U-shaped curve. Two interpretations are at hand, which may explain this lack of statistical significance in each specification and inconsistency across specifications: first, we have used a very crude measure of density, including river, lac, mountains, agricultural land in our area measure. As the presence of these natural obstacles to population settlements have huge variances between countries, this may discredit our results.[9] Second, if human capital externalities exist, they may be very local in nature. This idea of close proximity is indeed present in the urbanization variable, but less so in the population density measure, partly due to the mentioned issue above here. A conclusion which can be drawn from these results relates directly to the population distribution across space: an average measure of distribution of people across space is not enough, its skewness is also very important when dealing with human capital accumulation. A more straightforward way to acknowledge the fact that urbanization entails more information than population density is to recognize that urbanization variations capture variations in urban and total

[9] Note that the correlation coefficient between the population density and the urbanization variables is less than 0.3.

population variations, whereas population density changes eventually only entail changes in total population, the areas of countries being fixed.

[Table 6]

### 3.4.3 Country groups

In the basic regressions, we have pooled all countries together in one regression. Heterogeneity was accounted for, through country fixed effects. Nonetheless, marginal impact of urbanization is forced to be identical across countries, be it large or small countries, rich or developing countries, Asian or Latin-American countries. The econometrics on cross-country studies have been sharply criticized by Maddala (1999) for instance, who states that "choosing the countries used in the panel study is as crucial as (if not more than) the choice of the estimation method" (p.434).

Our aim in what follows is not to make a country-by-country analysis, or make a full-fledged analysis on *reasonable* country groups. On a more modest scale, we provide some results of our basic specification on *high-income* and *low-income countries* (the country composition of every group is detailed in the appendix). This approach is at best a tentative attempt to account for the heterogeneity in our sample.

Interpretations of the results should be done with great caution, as samples are not very large, above all when it comes to run GMM estimations. Except for column (8), signs of the variables of interest come out as expected, i.e., positive impact of urbanization on human capital accumulation with a lower threshold below which urbanization has a deterrent effect. This is true for fixed effect as well as for first-difference GMM estimators. Furthermore, the urbanization estimators comes out to be statistically non-significant three times out of twelve, each time in the quadratic regressions. As the linear urbanization term rather than the quadratic one is not significant two times out of three, this leads us to believe that a degree 2 polynomial is a quite poor approximation of the real-world non-linear effect of urbanization. The minimum of the U-shaped curve of column (6) is about 25%, which is in line with what we have found for the whole sample.

Looking at the marginal impact of urbanization, one can recognize that it is consistently higher for high income countries than for low income countries for the linear as well as for the quadratic case. A plausible reason why this is so may come from the fact that our urbanization variable might capture different realities when it comes to compare developed and developing countries. For instance, patterns of urbanization are very in developing and developed countries are very different. Whereas urbanization rates are roughly twice as high as urban concentration rates in high income countries, they are almost identical in the low income ones. As shown in Henderson (2003), the impact of primacy on GDP/capita growth is not neutral. The same may be true for human capital accumulation, leading to the difference we find in our results. But more importantly from a policy perspective is the fact that the threshold effect only appears in the case of developing countries. This may be one reason why urbanization in the developing world raises so much concern (see for instance the World bank's *World Development Report* 1999/2000)

A further consideration may be devoted to the primary aim of this subsection, namely to take account of heterogeneity among countries. One possible proxy to determine whether this goal has somewhat been achieved is to give a look at the speeds of convergence of the different groups. When comparing the human capital coefficient of columns (7) and (8) of Table 7, with columns (7) of Table 3 and (4) of Table 4, we can indeed see that the speeds of convergence have substantially increased for our high income sub-sample. The reverse however holds for the least developing country groups, pointing to large heterogeneity in this group. Further breakdowns where operated on unreported results, separating for instance Latin American countries and Sub-Saharan African countries, and led to considerably higher speeds of convergence (and hence, higher homogeneity across countries).[10] Essentially, results do support the role of cities as engines of human capital accumulation, but at the cost of loss of statistical significance due to too small sample sizes.

[Table 7]

### 3.5 Long-run effects

So far, we have supposed that urbanization impacts on human capital accumulation within a 5-year period delay. However, urbanization may have persistent effects on human capital, be it only due to population momentums in the age pyramid. Hence, our hypothesis may be too restrictive. To check this potentially persistent effect of urbanization, we introduce further lags of this variable in our regressions. Lags were restricted up to 3 in order to maintain a reasonable sample size. Table 8 only reports results for the variables of interest.

Interestingly, results mostly highlight the expected sign as above, however, only the first 5-year lag variables are statistically significant, pointing

[10] These results are available upon request.

to a limited persistence of urbanization effect through time.

[Table 8]

## 4 Conclusion

This paper has empirically addressed the analysis of the link between urbanization and human capital accumulation, using a rich cross-country data set. We have shown that even after controlling for the level of development of countries, urbanization plays a non-negligeable role in speeding human capital accumulation. Cities act as an ether, where barriers to knowledge flows shrink. Closeness between people favors interactions, which may be at the root of spillovers from human capital. In return, incentives to invest in education are reinforced, leading hence to higher levels of education. The sole fact of distributing unevenly population across space may hence act as a catalyst to foster investment in human capital.

A further important feature noticed throughout the analysis is the existence of an under-urbanization trap, below which urbanization goes against accumulation of human capital. A small excursion in the urban history may enlighten this observation: after the Industrial Revolution, most European cities were industry-cities. Besides benefiting from economies of scale due to investment in physical capital, the aim of gathering workers in city plants was to reduce cost of monitoring them, in opposition to Proto-industrial periods, where labor force mainly worked at home. Obviously, this is a very crude representation of the much more complex industrial city formation and growth. The core idea, however is that cities did not entail high levels of human capital. Most of the workers at that time were actually unskilled, and more importantly, had no incentive to invest in education. This may be at the root of the under-urbanization trap that has been highlighted, especially for developing countries.

So why is it that contemporaneous urbanization is so closely associated with human capital formation? The fact is that development dragged in its wake more service-oriented activities, requiring higher levels of interactions among people. In cities, manufacturing-based activities were pushed at the outskirts and replaced by more service oriented and specialized activities, with, as a corollary, higher skill requirements. In short, during early stages of development, educational and skill levels are very low in (industrial) cities, and more important, there is no incentive to increase it. Only more skill-oriented activities foster educational attainments in cities, and hence human capital externalities. Cities make up a more fertile ground for skills to blossom, but only from a certain level of development upwards!

Furthermore, when running our regressions, we have implicitly supposed that the link goes from urbanization to human capital accumulation, hence setting a precise causal link. Though this may be a reasonable assumption, the reverse causality could also be supported. This is for instance the case in Stark and Wang (2002) where agents form expectations on their returns to capital if they migrate. Hence, investing in human capital is a *prior* to migration. Two remarks can be made at that point: first, this causality problem is intimately linked to the endogeneity problem, and in our estimations, we have corrected for endogeneity. Second, our setting is *urban* versus *rural* area, in contrast to Stark and Wang who hypothesize migration from *poor countries* to *rich* ones. In the present context, it would not be reasonable to suppose that human capital investment in rural areas and subsequent migration to cities for the simple reason that education can mostly only be achieved in cities, at least for higher education! Thus, our causal link is much less restrictive than one may a priori suppose.

Finally, possible directions for further research should disentangle the effect of *urbanization* versus *urban concentration*. Henderson (2003) has shown that too much urban concentration deters growth. Does this result hold for human capital accumulation too? Moreover, separating the effect of urbanization and urban concentration would allow us to better tackle the issue of *developed* (characterized by high levels of *urbanization*) versus *developing countries* (characterized by high levels of *urban primacy*).

# References


[1] Acemoglu, D. and J. Angrist (2000), How Large Are Human Capital Externalities? Evidence from Compulsory Schooling Laws, NBER Macroannual 2000, 9-59

[2] Anderson, T.W. and C. Hsiao (1981), Estimation of Dynamic Models with Error Components, Journal of the American Statistical Association, 76(375), 598-606

[3] Arellano, M. and S. Bond (1991), Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, Review of Economic Studies, 58(2), 277-97

[4] Barro, R.J. and J.W. Lee (1993), International Comparisons of Educational Attainment, Journal of Monetary Economics, 32(3), 363-94

[5] Becker, G.S. (1964), Human Capital. New-York, Columbia University Press and NBER.

[6] Bond, S.R., A. Hoeffler and J. Temple (2001), GMM Estimation of Empirical Growth Models, CEPR DP No3048

[7] Ciccone, A. and G. Peri (2002), Identifying Human Capital Externalities: Theory with an Application to Cities, IZA WP No488

[8] Glaeser, E.L. (1994), Why does Schooling generate Economic Growth, Economics Letters, 44(3), 333-7

[9] Glaeser, E.L. and D.C. Maré (2001), Cities and Skills, Journal of Labor Economics, 19(2), 316-42

[10] Harris, J.R. and M.P. Todaro (1970), Migration, Unemployment & Development: A Two-Sector Analysis, American Economic Review, 60(1), 126-42

[11] Henderson, V. (2003), The Urbanization Process and Economic Growth: The So-What Question, forthcoming Journal of Economic Growth

[12] Jacobs, J. (1969), The economy of cities, Random House, New York

[13] Lamorgese, A. and GIP. Ottaviano (2002), Space, Factors and Spillovers, mimeo

[14] Leamer, E.E. and M. Storper (2001), The Economic Geography of the Internet Age, Journal of International Business Studies, 32(4), 641-65

[15] Lucas, R.E. Jr. (1993), Making a Miracle, Econometrica, 61(2), 251-72

[16] Lucas, R.E. Jr. (1988), On the Mechanics of Economic Development, Journal of Monetary Economics, 22(1), 3-42

[17] Maddala, G.S. (1999), On the Use of Panel Data Methods with Cross-Country Data, Annales d'Economie et de Statistique; 0(55-56), 429-48

[18] Marshall, A. (1890), Principles of Economics, London: Macmillan

[19] Mincer, J. (1974), Schooling, experience, and earnings, New York (N.Y.): Columbia University Press

[20] Moomaw, R.M. (1981), Productivity and City Size? A Critique of the Evidence, Quarterly Journal of Economics, 96(4), 675-88

[21] Moretti, E. (2004), Human Capital Externalities in Cities, in JF. Thisse and JV. Henderson, Handbook of Urban And Regional Economics, Volume 4.

[22] Peri, G. (1999), Local Human Capital Externalities: An Overlapping Generation Model and Some Evidence on Experience Premia, CESifo WP No219

[23] Rauch, J.E. (1993), Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities, Journal of Urban Economics, 34(3), 380-400

[24] Rudd, J.B. (2000), Empirical Evidence on Human Capital Spillovers, Board of Governors of the Federal Reserve System, Finance and Economics DP 2000/46

[25] Stark, O. and Y. Wang (2002), Inducing Human Capital Formation: Migration as a Substitute for Subsidies, Journal of Public Economics, 86(1), 29-46

[26] Summers, R. and A. Heston (1988), A new set of international comparisons of real product and price levels estimates for 130 countries, 1950-1985, Review of Income and Wealth, 34(1), 1-25.

[27] Tabuchi, T. (1986), Urban Agglomeration, Capital Augmenting Technology, and Labor Market Equilibrium, Journal of Urban Economics, 20(2), 211-28

[28] Willis, R.J. (1999), Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions, In: Ashenfelter, O. and D. Card (Eds.), Handbook of Labor Economics, Vol.1, Ch.10 Elsevier

# 5 Appendix

## 5.1 Tables

| | urbanization rates | average education (years) | GDP/capita (constant $, 1985) |
|---|---|---|---|
| 1960 | 0.39 | 0.62 | 2477 |
| 1975 | 0.47 | 0.99 | 3985 |
| 1990 | 0.53 | 1.65 | 5490 |
| Min | 0.017 | 0.01 | 313 |
| Max | 1 | 6.22 | 20018 |
| n° obs | 770 | 771 | 716 |

Table 1: Summary statistics

| | human capital | | | |
|---|---|---|---|---|
| | **1960** | **1970** | **1980** | **1990** |
| **urbanization(t)** | *0.67* | *0.66* | *0.67* | *0.71* |
| **urbanization(t-5)** | | *0.67* | *0.69* | *0.72* |
| **urbanization(t-10)** | | *0.68* | *0.70* | *0.73* |
| **urbanization(t-15)** | | | *0.70* | *0.74* |

Table 2a: Correlation between urbanization and human capital

| | human capital variation (5-year) | | | |
|---|---|---|---|---|
| | 1960 | 1970 | 1980 | 1985 |
| **urbanization(t)** | *0.17* | *0.38* | *0.25* | *0.39* |
| **urbanization(t-5)** | | *0.37* | *0.23* | *0.38* |
| **urbanization(t-10)** | | *0.35* | *0.21* | *0.37* |
| **urbanization(t-15)** | | | *0.2* | *0.35* |

Table 2b: Correlation between urbanization and human capital variation

| | Fixed effects | | | | First-difference GMM | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **human capital** | *-0.311****<br>*(0.032)* | *-0.311****<br>*(0.031)* | *-0.409****<br>*(0.035)* | *-0.461****<br>*(0.036)* | *-0.030****<br>*(0.004)* | *-0.226****<br>*(0.019)* | *-0.265****<br>*(0.026)* | *-0.418****<br>*(0.043)* |
| **urbanization** | | *0.912****<br>*(0.304)* | *0.967****<br>*(0.379)* | *1.066****<br>*(0.373)* | | *2.738****<br>*(0.204)* | *1.393****<br>*(0.317)* | *1.680****<br>*(0.561)* |
| **GDP/capita** | | | *0.222****<br>*(0.055)* | *0.166****<br>*(0.056)* | | | *0.406****<br>*(0.089)* | *0.319****<br>*(0.083)* |
| **life expectancy** | | | *-1.614****<br>*(0.384)* | *-1.474****<br>*(0.379)* | | | *-0.240*<br>*(0.440)* | *-0.960**<br>*(0.552)* |
| **Fertility** | | | | *-0.369****<br>*(0.086)* | | | | *-0.679****<br>*(0.139)* |
| **$R^2$-within** | *0.23* | *0.25* | *0.29* | *0.32* | | | | |
| **Sargan's overidentification test** | | | | | *25.69* | *28.25* | *19.45* | *17.92* |
| **Wald test: joint significance** | | | | | *55.14*** | *211.2*** | *214.2*** | *163.1*** |
| **Wald test: time dummies** | | | | | *3166*** | *50.20*** | *23.44*** | *15.57*** |
| **AR(1) test** | | | | | *-3.927*** | *-4.498*** | *-3.754*** | *-3.566*** |
| **AR(2) test** | | | | | *0.627* | *0.402* | *0.199* | *0.747* |
| **#observations** | *659* | *659* | *615* | *615* | *548* | *548* | *507* | *507* |
| **#countries** | *111* | *111* | *108* | *108* | *111* | *111* | *108* | *108* |

Table 3: Impact of urbanization on human capital accumulation

| | Fixed effects | | | First-difference GMM | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **human capital** | -0.452*** | -0.679*** | -0.711*** | -0.233*** | -0.380** | -0.262* |
| | (0.035) | (0.084) | (0.084) | (0.047) | (0.171) | (0.147) |
| **urbanization** | -1.655** | 0.792** | 0.431 | -2.793** | 1.124** | 0.782* |
| | (0.675) | (0.378) | (0.400) | (1.347) | (0.484) | (0.463) |
| **(urbanization)$^2$** | 2.775*** | | | 3.476*** | | |
| | (0.596) | | | (0.882) | | |
| **urbanization*human capital** | | 0.405*** | 0.676*** | | 0.196 | 0.403*** |
| | | (0.115) | (0.154) | | (0.214) | (0.141) |
| **urbanization*(human capital)$^2$** | | | -0.044*** | | | -0.051*** |
| | | | (0.017) | | | (0.011) |
| **GDP/capita** | 0.193*** | 0.212*** | 0.194*** | 0.221** | 0.361*** | 0.178*** |
| | (0.054) | (0.054) | (0.055) | (0.093) | (0.084) | (0.067) |
| **life expectancy** | -0.817** | -1.155*** | -0.956** | 1.374* | 0.221 | 0.070 |
| | (0.413) | (0.401) | (0.406) | (0.802) | (0.541) | (0.430) |
| **$R^2$-within** | 0.32 | 0.31 | 0.32 | | | |
| **Sargan's overidentification** | | | | 17.27 | 18.89 | 20.06 |
| **Wald test: joint significance** | | | | 158.1** | 281.9 | 281.8** |
| **Wald test: time dummies** | | | | 28.92** | 30.97 | 37.21** |
| **AR(1) test** | | | | -3.975** | -3.730** | -3.757** |
| **AR(2) test** | | | | 0.555 | 0.156 | 0.440 |
| **#observations** | 615 | 615 | 615 | 507 | 507 | 507 |
| **#countries** | 108 | 108 | 108 | 108 | 108 | 108 |

Table 4: Non-linear impact of urbanization

| | Fixed effects | | | | | | First-difference GMM | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Secondary education | | Higher education | | Total education | | Secondary education | | Higher education | | Total education | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| human capital | -0.441*** | -0.469*** | -0.206*** | -0.267*** | -0.511*** | -0.530*** | -0.251*** | -0.316*** | 0.137*** | 0.110*** | -0.355*** | -0.339*** |
| | (0.036) | (0.037) | (0.034) | (0.036) | (0.040) | (0.040) | (0.046) | (0.051) | (0.029) | (0.030) | (0.066) | (0.061) |
| urbanization | 0.780** | -1.083* | 0.158** | -0.431*** | 1.856*** | -0.741 | 0.479 | -3.142** | -0.065 | -0.024 | 3.983* | -0.228 |
| | (0.345) | (0.613) | (0.082) | (0.145) | (0.625) | (1.102) | (0.547) | (1.565) | (0.103) | (0.251) | (2.164) | (3.038) |
| (urbanization)$^2$ | | 1.971*** | | 0.643*** | | 2.752*** | | 3.182*** | | 0.033 | | 4.330 |
| | | (0.540) | | (0.132) | | (0.964) | | (1.098) | | (0.182) | | (3.203) |
| GDP/capita | 0.198*** | 0.174*** | 0.025** | 0.017 | 0.326*** | 0.295*** | 0.255*** | 0.300*** | 0.023* | 0.020 | 0.643*** | 0.554*** |
| | (0.050) | (0.050) | (0.012) | (0.012) | (0.090) | (0.090) | (0.081) | (0.091) | (0.013) | (0.017) | (0.160) | (0.191) |
| life expectancy | -1.343*** | -0.743** | -0.190** | -0.039 | -0.314 | 0.606 | 0.565 | 1.599** | 0.007 | 0.012 | 0.359 | 0.999 |
| | (0.345) | (0.378) | (0.075) | (0.080) | (0.598) | (0.676) | (0.623) | (0.825) | (0.073) | (0.126) | (2.407) | (2.189) |
| $R^2$-within | 0.30 | 0.31 | 0.14 | 0.18 | 0.35 | 0.36 | | | | | | |
| Sargan's overidentification test | | | | | | | 21.35 | 17.06 | 19.33 | 19.14 | 19.80 | 18.74 |
| Wald test: joint significance | | | | | | | 69.07** | 93.79** | 146.9** | 82.36** | 178.0** | 123.6** |
| Wald test: time dummies | | | | | | | 15.04* | 21.49** | 34.79** | 20.95** | 34.69** | 25.27** |
| AR(1) test | | | | | | | -3.420** | -3.541** | -2.290* | -2.311* | -4.867** | -4.794** |
| AR(2) test | | | | | | | 0.411 | 0.625 | -0.762 | -0.806 | 1.222 | 1.267 |
| #observations | 609 | 609 | 615 | 615 | 599 | 599 | 502 | 502 | 508 | 508 | 492 | 492 |
| #countries | 107 | 107 | 107 | 107 | 107 | 107 | 107 | 107 | 107 | 107 | 107 | 107 |

Table 5: Various definitions of human capital

| | Fixed effects | | First-difference GMM | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **human capital** | -0.407*** | -0.409*** | -0.240*** | -0.232*** |
| | (0.035) | (0.035) | (0.028) | (0.035) |
| **population density** | 0.121 | -0.341 | 0.288 | 0.510 |
| | (0.096) | (0.475) | (0.859) | (1.357) |
| **(population density)$^2$** | | 0.058 | | -0.109 |
| | | (0.058) | | (0.296) |
| **GDP/capita** | 0.224*** | 0.226*** | 0.303** | 0.273** |
| | (0.057) | (0.057) | (0.121) | (0.130) |
| **life expectancy** | -1.376*** | -1.371*** | 0.929*** | 1.207*** |
| | (0.374) | (0.374) | (0.180) | (0.267) |
| **$R^2$-within** | 0.29 | 0.29 | | |
| **Sargan's overidentification test** | | | 24.56 | 24.63 |
| **Wald test: joint significance** | | | 106.0** | 70.59** |
| **Wald test: time dummies** | | | 9.955 | 9.204 |
| **AR(1) test** | | | -3.772** | -3.752** |
| **AR(2) test** | | | 0.174 | -0.077 |
| **#observations** | 615 | 615 | 507 | 507 |
| **#countries** | 108 | 108 | 108 | 108 |

Table 6: Urbanization vs population density

Table 7: Country Groups

| | Fixed effects | | | | First-difference GMM | | | |
|---|---|---|---|---|---|---|---|---|
| | Low income countries | | High income countries | | Low income countries | | High income countries | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **human capital** | *-0.227**** | *-0.278**** | *-0.377**** | *-0.424**** | *-0.189**** | *-0.299**** | *-0.573**** | *-0.542**** |
| | *(0.043)* | *(0.045)* | *(0.049)* | *(0.050)* | *(0.031)* | *(0.019)* | *(0.015)* | *(0.012)* |
| **urbanization** | *1.055**** | *-0.305* | *2.027**** | *-2.463* | *0.982**** | *-1.201**** | *1.612*** | *1.669*** |
| | *(0.285)* | *(0.485)* | *(0.778)* | *(1.670)* | *(0.175)* | *(0.191)* | *(0.723)* | *(0.859)* |
| **(urbanization)$^2$** | | *1.848**** | | *3.841**** | | *2.371**** | | *0.638* |
| | | *(0.539)* | | *(1.271)* | | *(0.163)* | | *(0.904)* |
| **GDP/capita** | *0.162**** | *0.149**** | *0.306*** | *0.311*** | *0.175**** | *0.094**** | *0.115** | *-0.008* |
| | *(0.041)* | *(0.040)* | *(0.129)* | *(0.126)* | *(0.0.019)* | *(0.019)* | *(0.064)* | *(0.079)* |
| **life expectancy** | *0.075* | *0.335* | *0.664* | *1.506* | *-0.677**** | *-0.288**** | *-3.130**** | *-2.599**** |
| | *(0.200)* | *(0.210)* | *(1.146)* | *(1.159)* | *(0.0.104)* | *(0.101)* | *(0.379)* | *(0.521)* |
| **$R^2$-within** | *0.22* | *0.25* | *0.25* | *0.28* | | | | |
| **Sargan's overidentification test** | | | | | *54.74* | *53.00* | *32.56* | *33.35* |
| **Wald test: joint significance** | | | | | *191.7*** | *738.7*** | *3114*** | *3026*** |
| **Wald test: time dummies** | | | | | *427.5*** | *164.4*** | *421.8*** | *388.1*** |
| **AR(1) test** | | | | | *-3.598*** | *-3.512*** | *-2.592*** | *-2.560** |
| **AR(2) test** | | | | | *0.410* | *0.400* | *-1.032* | *-0.958* |
| **#observations** | *342* | *342* | *262* | *262* | *282* | *282* | *216* | *216* |
| **#countries** | *60* | *60* | *46* | *46* | *60* | *60* | *46* | *46* |

*NB: High income countries include high income and upper middle income as defined by the World Bank whereas Low income countries include low income and lower middle income (as in July 2002)*

| | Fixed effects | | First-difference GMM | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| urbanization$_{(t)}$ | 2.062* | -0.734 | 0.261 | -4.150*** |
| | (1.216) | (3.065) | (1.100) | (1.455) |
| urbanization$_{(t-1)}$ | 1.896 | 1.999 | 0.328 | -2.365 |
| | (1.883) | (4.379) | (1.229) | (2.319) |
| urbanization$_{(t-2)}$ | -1.546 | -3.183 | 1.694 | 0.122 |
| | (1.694) | (3.940) | (1.268) | (2.336) |
| urbanization$_{(t-3)}$ | 2.650** | 3.089 | -0.678 | 1.378 |
| | (1.176) | (2.670) | (1.112) | (1.732) |
| (urbanization)$^2_{(t)}$ | | 4.715 | | 6.119*** |
| | | (3.092) | | (2.049) |
| (urbanization)$^2_{(t-1)}$ | | -1.306 | | 2.135 |
| | | (4.195) | | (2.941) |
| (urbanization)$^2_{(t-2)}$ | | 2.022 | | 1.228 |
| | | (3.732) | | (2.352) |
| (urbanization)$^2_{(t-3)}$ | | -0.929 | | -3.022* |
| | | (2.615) | | (1.879) |
| #observations | 332 | 332 | 211 | 211 |
| #countries | 111 | 111 | 108 | 108 |

Table 8: Persistence of the impact of urbanization

## 5.2 Figures

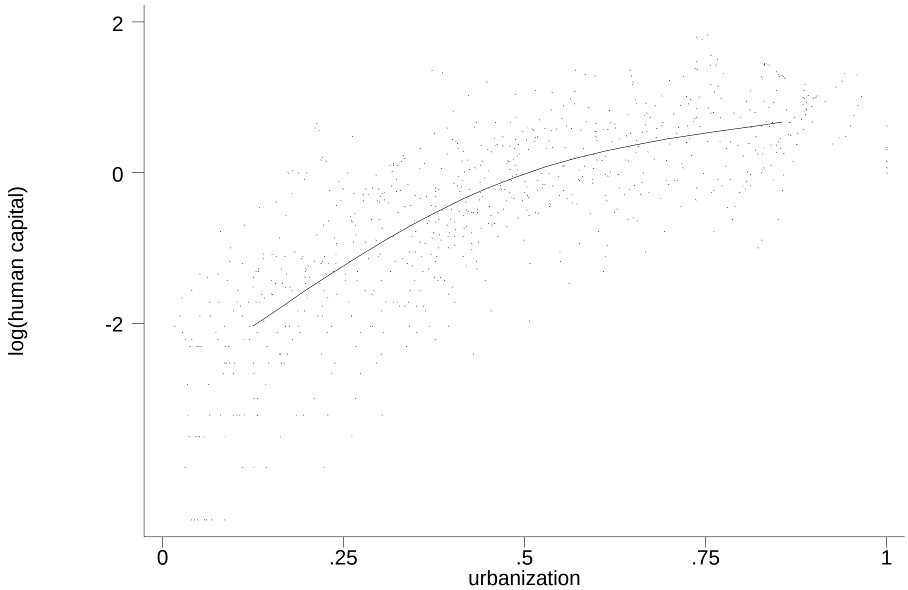


Figure 1: Urbanization vs. $log$(Human Capital)

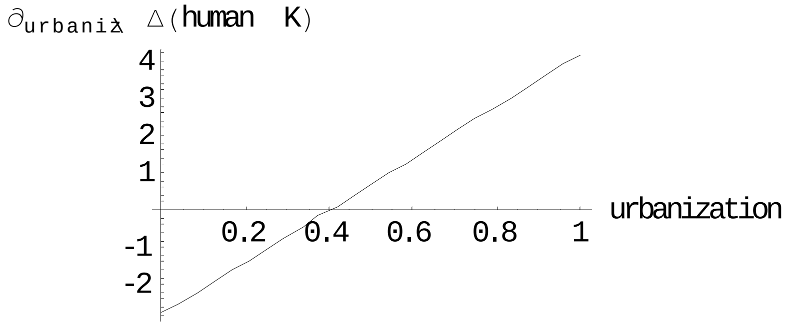


Figure 2: Marginal impact of urbanization on human capital accumulation

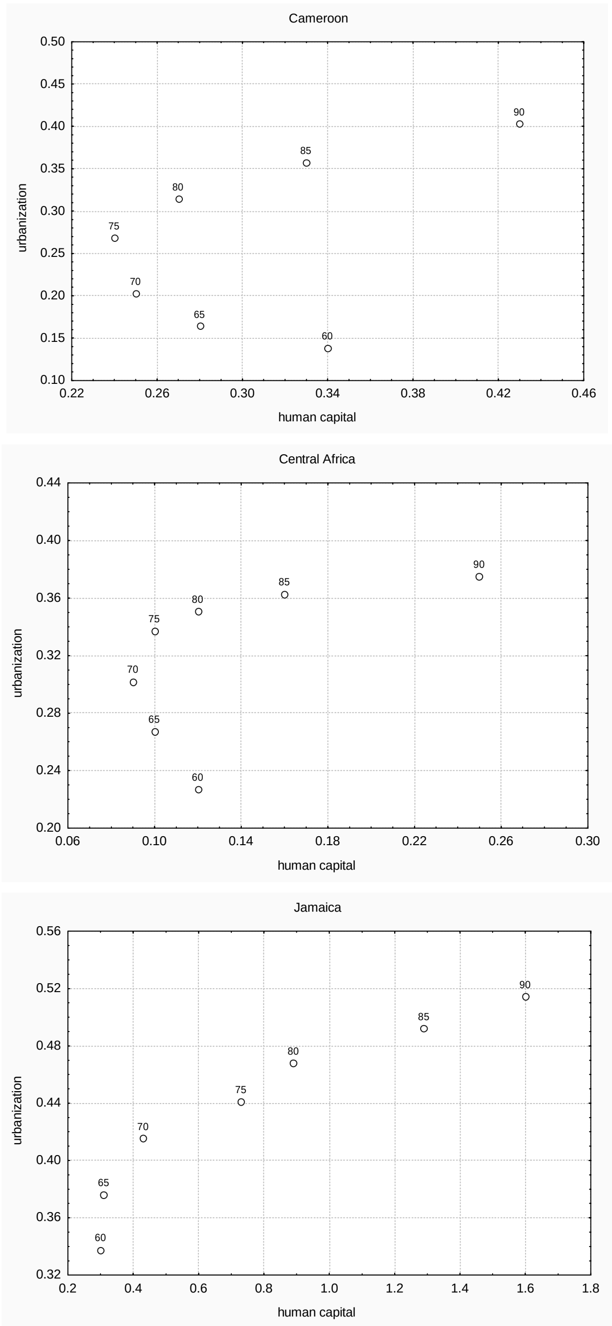


Figure 3: U-shaped relation between urbanization and human capital for some developing countries

### 5.3 Data sources

Human capital

*Secondary education*: Average years of secondary schooling in the total population aged 25 and over, Barro-Lee. Census and survey figures primarily from UNESCO Statistical Yearbooks and UN Demographic Yearbooks fill two fifths of the observations. The remaining values are estimated using UNESCO school enrollment data and a perpetual inventory method. The data are not adjusted for quality of education or length of school day or year.

*Higher education*: Average years of higher schooling in the total population aged 25 and over, Barro-Lee. Census and survey figures primarily from UNESCO Statistical Yearbooks and UN Demographic Yearbooks fill two fifths of the observations. The remaining values are estimated using UNESCO school enrollment data and a perpetual inventory method. The data are not adjusted for quality of education or length of school day or year.

*Total education*: Average schooling years in the total population aged 25 and over, Barro-Lee. Census and survey figures primarily from UNESCO Statistical Yearbooks and UN Demographic Yearbooks fill two fifths of the observations. The remaining values are estimated using UNESCO school enrollment data and a perpetual inventory method. The data are not adjusted for quality of education or length of school day or year.

**Urbanization** This variable is the ratio of the *total urban population* variable and the *total population* variable.

*National urban population*: UN World Urbanization Prospects Table A.3 Urban population. Values for Czechoslovakia, Yugoslavia and USSR are summed over former constituent republics. Values for West Germany are calculated using the World Bank WDI urban per cent of the Penn5.6 population figure.

*National population*: UN World Urbanization Prospects Table A.5 Total population. Values for West Germany, Czechoslovakia, Yugoslavia and USSR are taken from the Penn World tables 5.6 (1995 are missing observations).

**GDP/capita** *log* of real GDP per capita in constant dollars, chain index, expressed in international prices, base year 1985. Penn World Tables Mark 5.6. 1995 values are estimated using World Bank WDI yearly real GDP growth figures on 1990 rgdpcPN values.

**Life expectancy** *log* of life expectancy at birth in years, five year average from base year (e.g. average over 80-84). Czechoslovakia, Yugoslavia and USSR are averages over the constituent republics weighted by population share. The values for united Germany are substituted for West Germany. World Bank WDI.

**Fertility** *log* of total fertility rate, births per woman, five year average from base year (e.g. average over 80-84). Czechoslovakia, Yugoslavia and USSR are averages over the constituent republics weighted by population share. The values for united Germany are substituted for West Germany. World Bank WDI.

**Population density** Ratio of *total population* as defined above and *area*, defined as the land area in square kilometers, World Bank WDI. Values for West Germany, Czechoslovakia, Yugoslavia and USSR are taken from the Columbia Encyclopedia.

### 5.4 Groups of countries[11]

**High and Higher Middle Income Countries** Argentina, Australia, Austria, Bahrain, Barbados, Belgium, Botswana, Brazil, Canada, Chile, Costa Rica, Cyprus, Czechoslovakia, Denmark, Finland, France, West Germany, Greece, Hong Kong, Hungary, Iceland, Ireland, Israel, Italy, Japan, South Korea, Kuwait, Malaysia, Malta, Mauritius, Mexico, Netherlands, New Zealand, Norway, Panama, Poland, Portugal, Singapore, Spain, Sweden, Switzerland, Trinidad and Tobago, United Kingdom, United States of America, Uruguay, Venezuela.

**Low and Lower Middle Income Countries** Algeria, Bangladesh, Benin, Bolivia, Bulgaria, Cameroon, Central Africa, China, Colombia, Congo, Dominican Republic, Ecuador, Egypt, El Salvador, Fiji Islands, Gambia, Ghana, Guatemala, Guyana, Haiti, Honduras, India, Indonesia, Iran, Iraq, Jamaica, Jordan, Kenya, Lesotho, Liberia, Malawi, Mali, Mozambique, Myanmar, Nepal, Nicaragua, Niger, Pakistan, Papua New Guinea, Paraguay, Peru, Philippines, Romania, Rwanda, Senegal, Sierra Leone, South Africa, Sri Lanka, Sudan, Swaziland, Syria, Thailand, Togo, Tunisia, Turkey, USSR, Uganda, Yugoslavia, Zambia, Zimbabwe.

[11] This grouping has been made according to the World Bank definition GNI per capita. The groups are: low income and lower middle income: <\$2,975 per year; high income and upper middle income: >\$2,976 per year.