1. Introduction

Today more than half of the 7 billion inhabitants of the planet live in urban areas, with this share expected to keep rising. While urbanisation has been long recognised as a fundamental element of the process of economic development, sustainable urbanisation has become one of the main and more pressing challenges for developing countries, where millions live lacking adequate access to basic services like electricity, clean water and sanitation. Building on previous evidence on urban concentration and economic growth, differentiated effects of urban concentration on national economic performance are analysed in this paper. The paper contributes to the literature by providing empirical evidence on how different characteristics of the urban environment - in particular the quality of urban infrastructure - strongly determine whether growth-enhancing benefits of growth-deterring congestion costs prevail in the process of urban concentration, something that previous studies on urban concentration and economic growth have not considered empirically. Looking at different world regions the analysis finds that while increasing urban concentration might have been associated with growth in Asian countries, it seems that congestion diseconomies have prevailed over agglomeration benefits in Sub-Saharan African (SSA) countries due to their significant deficiencies in terms of access to basic urban services.

The focus on access to basic services lies with two major reasons. The first relates to magnitude. According to UN-Habitat Reports, today at least 1 billion people worldwide, of whom the vast majority are in the developing world, live in slums lacking access to basic services. Growing at high rates (higher than 4.5 per cent per annum in Sub-Saharan Africa) slums are expected to host 2 billion inhabitants by 2030. The second reason relates to the fact that access to basic services is expected to play a key role in the trade-off between the benefits and the costs that come with urban concentration, especially in developing countries. On the one side low coverage of basic services are likely to handicap the benefits from agglomeration (as specialisation, labour pooling and knowledge diffusion) as they hinder physical and social mobility and interaction, information flow and knowledge spillovers and trust. On the other side, deficiencies in terms of access to basic services dramatically increase congestion costs for urban inhabitants in terms of transport costs, but also in terms of disease transmission, pollution, conflict and crime (most likely reducing the capacity of cities to develop and attract talent and investment). As the World Development Report (2011) highlights, access to basic services is fundamental for the well functioning of large cities.

Slums are traditionally considered as a phasing phenomenon characteristic of fast-growing economies, and representing a temporary stage in the structural change from rural to industrial activities. However, slums have tended to grow more in poor and stagnant countries where urbanisation and urban concentration do not seem associated with economic growth (Fay and Opal 2000; Kim 2008; Bloom et al. 2008). Indeed as the WDR (2011) acknowledges the growth-enhancing benefits from urban concentration, it also warns about the risks of "rapid urbanisation" in developing countries.[1] With most of their inhabitants having been born in the slum where they live, and with their living standards hardly improving over time, slums in developing countries today are considered a form of poverty trap for a majority of their residents (Marx et al. 2013).[2] In fact, growth of large agglomerations in developing countries today is mostly given in slums, being their growth more the outcome of fast natural growth than the outcome of rural-urban migration: Jedwab et al. (2014) report a contribution of natural increase to urban growth for 10 African countries from 1950 to 2010 of 2.9%, compared to a contribution of 1.8% due to migration. Even growth driven by migration has been more associated with push rather than pull factors (Lipton 1977; Bates 1981; Bairoch 1988; Barrios et al. 2006; Swanson and Buckley 2013), with population being "expulsed" from rural areas rather than attracted to urban areas by the prospects of better living standards.[3] In this line, several authors are now referring to *Malthusian* urban dynamics, especially in SSA (Jedwab et al. 2014; Swanson and Buckley 2013).[4] With more than half of the 7 billion inhabitants of the planet living today in urban areas, it is indeed very likely that in many developing countries the Malthusian dilemma of low living standards has in some way moved from the countryside to the main urban centres, where a large proportion of urban dwellers reside under inadequate living conditions and where congestion effects of population growth are expected to dominate the positive effects from urban concentration.

In relation to the existing literature this paper is closely linked to previous empirical studies on the relationship between urban concentration and long-run economic growth at country level (Henderson 2003; Bertinelli and Strobl 2007; Brulhart and Sbergami 2009; Leitão 2103; Castells-Quintana and Royuela 2014) as well as to the growing empirical evidence of urban processes in developing countries not necessarily linked to economic development (Firebaugh 1979; Ades and Glaeser 1995; Davis and Henderson 2003; Bloom et al. 2008; Behrens and Pholo-Bala 2013; Gollin et al. 2014). The paper is also linked to the literature on efficient city size. This literature has highlighted functional characteristics of cities, beyond physical size, as further determinants of the benefits and costs form agglomeration (Richardson 1972; Capello and Camagni 2000; Royuela and

[1] The UNFPA State of World Population 2011 estimates that there are 60 million new urban inhabitants every year worldwide, most of them in the developing world. Comparing the speed of urbanisation processes in Asia and Africa between 1950 and 2010, on the one hand, and in Europe between 1800 and 1910, on the other hand, Jedwab et al. (2014) conclude that developing countries today have experienced the same growth in urbanisation in half the time.

[2] Marx et al. (2013) summarise evidence on living standards based on surveys carried-out in slums around the world. According to these surveys, the majority of slum residents were either born in the slum where they live or have been living there most of their live (or moved from a different slum).

[3] Even when driven by urban pull factors, expectations of high returns from moving to urban areas do not necessarily materialise and can lead to additional pressure from new incomers, as the well-know Todaro paradox describes (Todaro 1976). This will be especially true when both rural and urban incomes are close to subsistence levels, as it is the case in SSA.

[4] In a Malthusian equilibrium societies with greater availability of resources have higher population density but living standards remain low unless productivity is sufficiently increased. Such equilibrium was the rule for most human history (See Ashraf and Galor 2011 for a modern modelling of Malthusian equilibrium as well as for transition dynamics towards sustained growth). In its purely rural setting a Malthusian equilibrium has also been considered as a relevant possibility today for many poor countries with large rural populations and largely dependent on low-productivity agriculture and mineral exports (Weil and Wilde 2009).

Suriñach 2005; Camagni et al. 2013). The paper can additionally be related with the literature on the effects of climate change on economic performance (Miguel et al. 2004; Barrios et al. 2006; Brückner and Ciccone 2011). Finally, the paper also relates to the more scarce literature on the role of urban infrastructures on economic performance. Of high relevance for the analysis presented here is the work by Field and Kremer (2006), who find a significant role of access to basic services on improving economic performance of urban residents, as it also is the work by Lewis (2014) showing how local governments in Indonesia that invest more heavily in infrastructure are better able to cope with the detrimental effects of rapid urbanisation on local economic growth. In an analysis close to the one done here, Sekkat (2013) studies the relationship between urban concentration, poverty and infrastructure in a cross-country setting, but looking at nation-wide, rather than urban-specific, infrastructure.[5] To the best of my knowledge no paper empirically addresses in a cross-country framework the role that the urban environment plays in the relationship between urban concentration and economic growth. This paper contributes to fill that gap.

The remainder of the paper is organised as follows. Section 2 sets a simple theoretical framework for the empirical specification to be derived. In section 3 the data used is described along some basic stylised facts. Section 4 discusses estimations and results. The section includes an analysis using the complete world sample, some robustness checks, and a specific analysis for SSA. Finally, section 5 concludes and derives policy implications from the results.

2. A simple theoretical framework

The empirical analysis is based on a GDP per capita growth framework, following works as Henderson (2000) and Brülhart and Sbergami (2009).[6] To derive an empirical specification we can depart from a neoclassical framework of economic growth basis for standard cross-country growth regressions.[7] In this framework one can consider country-specific characteristics (as for instance resources, institutions, location and characteristics of the economic geography) to allow for heterogeneity in initial conditions, as well as in efficiency growth paths, that influence economic growth. Accordingly, cross-country differences in output per capita growth are expected to depend not only on initial levels of output per capita and factor accumulation, but also on differences in these country-specific characteristics:

---

[5] See Straub (2011) for a recent survey of the literature on infrastructure and development. For a focus on Africa see Ayogu (2007) and Calderón and Servén (2010).

[6] While Henderson (2000) is based on a GDP per capita growth specification, Henderson (2003) focuses on TFP growth (but also estimates a GDP per capita growth model as robustness). While both analyses are similar, a GDP per capita growth specification allows for the use of a larger dataset.

[7] From a neoclassical perspective, economic growth is related to growth due to technological progress and to the gap between the initial level of output and the steady state to which the economy converges, with the expectation that countries with lower levels grow faster. See Durlauf et al. (2005) for a more detailed explanation of how to derive cross-country growth regressions from neoclassical economic growth theory.

$$\Delta y_i = \beta\big(log y_{i,0}\big) + \psi X_{i,0} + \pi Z_{i,0} + \varepsilon_i \qquad (1)$$

where Δy_i is per capita average growth rate of country i, $y_{i,0}$ is initial output per capita, $X_{i,0}$ represents variables reflecting factor accumulation (i.e. the standard Solow determinants) plus a constant term, and $Z_{i,0}$ a vector of country-specific characteristics explaining cross-country differences in efficiency growth (the evolution of technology) or in initial conditions.

Introducing urban concentration as a determinant of growth:

The degree of urban concentration represents one variable that could be considered within the vector $Z_{i,0}$. The degree of urban concentration is a relevant characteristic affecting growth in efficiency (Henderson 2003), as it reflects agglomeration economies that remain unexploited, and therefore offering possibilities for growth, or that become exhausted and subject to congestion:[8]

$$\Delta y_i = \beta\big(log y_{i,0}\big) + \psi X_{i,0} + \lambda UC_{i,0} + \pi Z_{1i,0} + \varepsilon_i \qquad (2)$$

where UC_{i0} is the degree of urban concentration and $Z_{1i,0}$ other relevant country-specific factors. However, as suggested, the way urban concentration affects growth in efficiency depends on specificities of the urban process. In particular, urban infrastructures define the urban environment, leading to different capacities for cities to benefit from agglomeration economies and to control congestion diseconomies. As Henderson (2005) notes, "public infrastructure affects not just the resources devoted to urban living such as commuting and congestion costs, but also affects production efficiency - the extent to which knowledge spillovers are fully realized and exploited." Bertinelli and Black's (2004) stylized urban economics model indeed suggests an empirically testable prediction; that the growth-enhancing benefits from concentration are significantly affected by the quality of urban infrastructure affecting the urban production technology.[9] Hence, taking this prediction into account, equation (2) extends to:

$$\Delta y_i = \beta\big(log y_{i,0}\big) + \psi X_{i,0} + \lambda_1 UC_{i,0} + \lambda_2 G_{i,0} UC_{i,0} + \pi Z_{1i,0} + \varepsilon_i \qquad (3)$$

where $G_{i,0}$ captures specificities of the urban process as the quality of urban infrastructure. Equation (3) is our main equation of analysis.

---

[8] According to Henderson (2003), "urbanisation represents sectoral shifts within an economy as development proceeds, but is not a growth stimulus per se. However, the form that urbanisation takes, or the degree of urban concentration, strongly affects productivity growth" (Henderson 2003, pp. 67).

[9] Bertinelli and Black (2004) introduce dynamic human capital externalities, along traditional congestion externalities in the urban sector, to study how urbanisation influences economic growth at country level. In this framework urbanisation enhances growth by the structural change given by the reallocation of resources, and through higher human capital accumulation that increases productivity. Thus, "to the extent that urbanisation encourages human capital accumulation, cities become the engines of economic growth."

3. Data and Stylised Facts

3.1. Data:

To study the relationship between urban concentration and growth I rely on panel data for as many countries in the world as possible depending on data availably between 1960 and 2010, covering more countries and a longer time span than previous studies on urban concentration and growth. The dependent variable is national economic *growth*, for which data from the Penn World Tables are used. For UC_{i0} the proportion of urban population living in the primate city (*primacy*), as the most standard measure in the literature on urban concentration, is considered.[10] Data for primacy comes from the World Bank. For the quality of urban infrastructure several measures are considered. Following the World Development Report (2011), I focus on three key indicators: access to improved sanitation, improved water source, and electricity. As data for all these variables is scarce, when I introduce them in the analysis the panel only considers the 1990-2010 period.[11] Finally, as control variables (X_i and Z_{1i} in equation 3) I begin by considering investment, as share of GDP, fertility rates, and average years of secondary and higher education of the adult population, following Henderson (2000) specification. For urban infrastructure variables, as well as for control variables, I rely on a variety of sources. Appendix A lists all variables' names, definitions and sources. For robustness a wide variety of other control variables are also considered, following Brülhart and Sbergami (2009) and the literature on cross-country economic growth. For the analysis done for SSA, data on rainfall is used to instrument for economic growth (as explained below). Rainfall data comes from the National Aeronautics and Space Administration (NASA) Global Precipitation Climatology Projects (GPCP), and has been used in previous papers as Brückner and Ciccone (2011). Additionally, also for SSA, data on light density at night is used as robustness for measurement errors in income per capita. This data comes from the Defence Meteorological Satellite Program's Operational Linescan System (DMSP-OLS) and archived by the National Oceanic and Atmospheric Administration (NOAA), and has recently been used as proxy for income by several authors (Henderson et al. 2012; Mveyange 2014; Lowe 2014; Henderson et al. 2014).

[10] Primacy measures consider main metropolitan areas (including core city and satellite cities). As a ratio, primacy is more easily comparable across countries than measures of urbanisation. Moreover, it has been shown that primacy correlates very highly with other measures of concentration (as the Hirschman-Herfindahl index for which there is very limited coverage) and reflects fairly well parameters behind Zipf's law curves (the fact that when we rank cities from largest to smallest, rank times population size is approximately the same constant for all cities). The largest city in the country, therefore, delineates all other city sizes and is sufficient information to calculate any comparative index of national urban concentration (Henderson 2003).

[11] Main results and discussion focus on access to improved sanitation. According to the World Bank, sanitation remains one of the most off-track Millennium Development Goals (MDG) globally. Access to improved sanitation not only lies at the heart of many other development challenges but the lack of it is also currently holding back economic growth in many less-developed countries. In the robustness section, I discuss results using improved water source and electricity. I further consider infant mortality rates, as a common and basic indicator of health, and access to mass urban transport systems.

3.2. Some basic stylised facts:

Before performing econometric analysis, an initial look at urban concentration patterns and economic performance worldwide during the previous decades allows us to highlight some basic but interesting stylised facts. The first of these is that while the proportion of urban population living in the primate city (primacy) has stayed relatively constant over time at around 40 per cent of total world urban population, there are important differences between developed and developing countries and across world regions. While the average is about 35 per cent for developed countries, it is higher than 43 for developing countries. Figure 1 shows primacy levels around the world while Table 1 presents descriptive statistics for primacy and economic growth, the correlation between the two, and basic figures related to the urban environment, all for different world regions. Higher values of primacy tend to be concentrated in poorer regions of the globe (as Latin America and the Caribbean -LAC- with average close to 50 per cent, and SSA with an average above 42). The second fact relates to the fast pace of urbanisation processes in developing countries, and especially the current growth of large agglomeration in these countries. While in 1970 large primate cities in developing countries had on average a similar size of those in developed countries (around 1.2 million inhabitants), in 2010 primate cities in developing countries had on average almost one million inhabitants more (with an average of 3.4 million) than their counterparts in developed countries.[12] The third fact relates to the heterogeneity in the correlation between urban concentration and subsequent economic growth. While there is a negative, although insignificant, correlation for the world sample (-0.03), the picture changes if we consider the correlation by level of development; urban concentration is positively correlated with growth in developed but not in developing countries. By regions the correlation is positive in Europe, Asia, LAC and North Africa, and negative in North America, Oceania and SSA. The final stylised fact relevant for our analysis refers to urban infrastructure and the urban environment, where we also find important heterogeneities across countries. In particular, urbanisation in many developing countries indeed appears as characterised by a large proportion of urban inhabitants living under inadequate conditions. While access to basic services was already virtually universal in developed countries in 1990, it was not in developing countries, with important differences among them and particularly significant deficiencies in SSA. These deficiencies in SSA appear as remarkably severe in terms of access to improved sanitation and electricity and remain quite persistent (sanitation increasing since 1990 on average less than 5 percentage points and electricity around 10). Figures 2a and 2b display maps of access to improved sanitation and electricity worldwide. For access to improved sanitation,

[12] I calculate these world averages using World Bank data for the largest agglomeration in 193 countries worldwide and considering only countries with a total population of at least 1 million inhabitants. 150 out of these 193 agglomerations are in developing countries. Also note that averages hide high variability in size. Jakarta, Shanghai and Bombay in Asia, Lagos and Cairo in Africa, Mexico City and Sao Paulo in Latin America, are all above or close to 20 million inhabitants in their respective metropolitan areas with a population still growing at a fast pace.

while the average for Asia was close to 85 per cent, it was 44 for SSA (taking average values between 1990 and 2005). In terms of electricity the average coverage in SSA reached only half of the urban population. Similarly, in terms of infant mortality - reflecting access to health services - the average was 11 children per 1000 live births in the developed world, higher than 62 in developing countries, and exceeding 93 in SSA. In terms of transport, none of the primate cities in SSA had a massive transport system by 2000.[13] In general, looking at data on urban population living in slums, we find an average of 57 per cent of urban population in developing countries, the figure reaching 77.9 for SSA. These dramatic deficiencies in SSA do not seem just the consequence of low-income levels. As Figure 3 shows for access to improved sanitation, even controlling for income levels SSA countries present significantly lower levels of urban infrastructure.[14] Such deficiencies are in all probability hampering agglomeration benefits while raising congestion costs in Sub-Saharan African cities.

If we take into account the high heterogeneity in the quality of urban infrastructures, we can see a clearly distinguishable correlation between urban concentration and long-run economic growth; positive in countries with relatively high quality of urban infrastructures and negative in countries with relatively low quality (Figure 4).

[Insert Figure 1: Population living in largest city (percentage of urban population)]

[Insert Table 1: Some basic figures]

[Insert Figure 2: Access to improved sanitation and electricity]

[Insert Figure 3: Access to improved sanitation by income levels]

[Insert Figure 4: Access to improved sanitation by income levels]

4. Estimations and results

4.1. Urban concentration and economic growth in a panel of countries:

Following the literature on urban concentration and economic growth, I begin by estimating the specified growth equation based on cross-country panel data (for 137 countries) and without considering differentiated urban patterns across countries. I split 1960-2010 into 5-year periods.[15] Equations of this type using panel data represent dynamic models. Estimation of these growth

[13] Lagos inaugurated a bus rapid transit system in 2008, and Accra has now planned a metro monorail project.

[14] A simple regression analysis yields highly significant lower levels of urban infrastructure for SSA countries (16 percentage points on average for sanitation) compared to other developing countries of same income per capita levels. Ghana presents a gap of almost 50 percentage points in terms of access to sanitation.

[15] I also experimented with 10-year periods in order to reduce any short-term noise from the business cycle, but at the expense of losing observations. Results using 10-year periods are very similar to those presented throughout the paper using 5-year periods.

models raises some concerns: reverse causality, unobserved time-invariant country-specific characteristics, and the presence of initial income as a regressor. I therefore complement more standard panel estimation techniques (as pool-OLS and Fixed Effects -FE) with System-GMM estimations, in order to deal with some of these concerns and as common in empirical studies estimating dynamic models.[16] For SSA (section 4.4) I extend the empirical analysis with panel Fixed Effects Instrumental Variables (FE-IV) estimations taking advantage of the exogenous variability given by rainfall data.

Table 2 presents the result for the first set of estimations of the basic growth model. Columns 1 to 4 present results for different estimation techniques.[17] Control variables have the expected sign reflecting conditional convergence, a positive effect of higher investment and educational levels and a negative effect of higher fertility rates.[18] In column 5 I introduce primacy. Results yield a positive and significant effect (although just at the 10%). But there are reasons to expect that the relationship between urban concentration and growth will vary according to the level of development.[19] Following Henderson (2000), column 6 considers a more flexible functional form for the effect of primacy on growth; I introduce not just a linear effect of primacy but also an interaction term with initial income per capita (in logs) and another interaction term with the square of this initial income per capita. Results support the Williamson hypothesis - with a negative coefficient for primacy, a positive for its interaction with income and a negative for the interaction with the square of income (all coefficients significant at the 1%). In Figure 5 this quadratic effect of primacy on growth, depending on income levels, is plotted. At very low levels of development the effect of primacy is negative. It then becomes positive and increases as income raises up to levels around $9500 per capita (in PPP converted, at 2005 constant prices) to then start declining.[20] Finally, I take into account the possibility of significant differences across world regions. As column

[16] Both Henderson (2003), using first-differences GMM, and Brülhart and Sbegami (2009), using system-GMM, rely on GMM estimations and provide a good explanation on the suitability of these methods for cross-country data on urban concentration and economic growth. In particular, system-GMM (Blundell and Bond, 1998) estimates are expected to be more efficient than any other dynamic GMM estimators, especially when the coefficient of the lagged dependent variable is close to one and the between sample variance is large compared to the within sample variance (as is the case here). For GMM estimations I present standard AR(1), AR(2) and Hansen tests for validity of internal instruments. As Bazzi and Clemens (2013) note, there is yet no reliable and straightforward test for the strength of the instrument set in Sys-GMM estimations. Correlation analysis of our key variables, nevertheless, reveals substantial explanatory power for lagged differences to explain levels and for lagged levels to explain first differences.

[17] OLS, FE, GMM and System GMM -SysGMM- results are presented but I focus throughout the paper on SysGMM results (and panel FE-IV estimations for SSA).

[18] I also calculate the annual speed of convergence to ease comparability of results with previous papers. The values found are within the range of what is commonly found in the literature, although differing depending on the estimation technique considered.

[19] While increasing urban concentration is desirable and expected in early stages of development, de-concentration eventually occurs as development proceeds. The optimal degree of urban concentration declines as development proceeds as knowledge gets accumulated, lowering the scope from agglomeration economies, and as better infrastructure allows efficient de-concentration to avoid congestion costs (Henderson 2003). Furthermore, the optimal level of urban concentration is expected to decline with the level of development also as institutional environments improve (Henderson 2003; Barca et al. 2012).

[20] Semiparametric estimations confirm this nonlinear relationship between primacy and growth. Results available upon request.

7 shows, while there seems to be a positive relationship between primacy and growth for the world sample, there is a significantly different relationship for LAC and SSA.[21]

[Insert Table 2: Urban concentration and growth in a panel of countries]

[Insert Figure 5: The Williamson hypothesis]

4.2. Positive and negative effects of urban concentration depending on the urban process:

Results in Table 2 suggest that the sign and the form of the relationship between urban concentration and growth are not uniform. The relationship is likely to vary not only with the level of development but also with other country's characteristics. In particular, as noted before, the quality of urban infrastructure might be fundamental to unleash positive synergies from agglomeration economies or to increase congestion costs, in both cases affecting national productivity. In Table 3 I present results for estimates of equation (3), letting the effect of urban concentration to depend on the quality of urban infrastructure. Results are presented using access to improved urban sanitation facilities (*sanitation*) as a proxy for the quality of urban infrastructure.[22] The coefficients for both the direct effect of urban concentration and for its interaction with sanitation are highly significant under OLS (column 1), being negative the first and positive the second. Results are less significant when I estimate by FE (column 2) or SysGMM (column 3). However, as noted in the descriptive analysis, the quality of urban infrastructure substantially differs between developed and developing countries. Accordingly, in columns 4 and 5 I split the sample into developed and developing countries. SysGMM results are now non-significant for developed countries but they are highly significant for developing countries. The absence of enough variability between developed countries in the variables considered for urban infrastructure could explain their non-significance. As we have seen, access to basic services is very high and quite homogenous among developed countries. However, there is much higher heterogeneity among developing countries, with some of them reaching developed world figures but other lagging behind and with less than half of urban population having access to these services. In the case of developing countries results suggest that urban concentration is negative associated with economic growth for countries with low levels of sanitation. By contrast, the association becomes positive as access to sanitation increases.[23] Hence, the growth-enhancing benefits of urban concentration

---

[21] In fact, when I analyse urban concentration by the different world regions, its effect on growth seems to have been positive and significant only in Asia and Europe. When distinguishing between developed and developing countries, rather than between world regions, while linear effects of primacy are only positive and significant in the former countries, it is in developing countries where the evidence of the Williamson hypothesis is clearer (in line with Bertinelli and Strobl 2007).

[22] Below I discus some results (presented in the appendix) using other proxies for the quality of urban infrastructure.

[23] I also obtain similar results when considering *growth* in urban concentration and *growth* in sanitation rather than their levels.

prevail over congestion costs only when basic services spread to the majority of the urban population.

[Insert Table 3: Urban concentration depending on the urban process]

4.3. Robustness checks:

We can check the robustness of the results found in several additional ways. In first place one could worry that the positive effect of the interaction between primacy and sanitation is due to the fact that higher sanitation is correlated with higher income levels (where urban concentration could have more beneficial effects). Nevertheless, as column 6 of Table 3 shows, main results for developing countries hold when we introduce an interaction between urban concentration and income levels. Likewise, results hold as we control for the Williamson hypothesis, introducing interactions with income levels and their square (column 7 of Table 3). Results are also robust to other regional differences in the relationship between urban concentration and economic growth beyond differences in urban infrastructure (in column 8 the effect of urban concentration is allowed to vary across world regions). While access to sanitation is a good proxy and very pertinent for the analysis, there could be different contexts in which the role of other urban infrastructures might be more relevant, for example transport infrastructure (mobility and transport costs being a central issue of congestion analysis in the urban economics literature). In this line, and to expand the analysis, I replicate some of the estimations using other variables for the quality of urban infrastructure. On one side Appendix B presents panel results for access to improved water source (*water*) and access to electricity (*electricity*). Results are non-significant for access to water, but are for access to electricity. On the other side Appendix C presents some cross-section results. Cross-section analysis is more common in the long-run economic growth literature and, as discussed before, allows us to consider other variables, as transport systems for which there is not enough time variation. Cross-section results for sanitation are in line with panel results. Results also hold when other variables are considered, as *electricity* or *transport_systems*, although the significance is reduced and depends on the controls used.[24] When a composite measure for urban infrastructure is considered, rather than just one indicator, estimations yield highly significant results (and robust to all the considered controls).[25]

---

[24] Following Brülhart and Sbergami (2009), our cross-section controls include 18 variables found to be robustly associated with long-run growth by Sala-i-Martin et al. (2004) along population growth rates, higher education, fertility, investment share, and population density - to further capture agglomeration between countries. As in the panel analysis, when I analyse by world regions cross-section estimations yield a positive relationship between urban concentration and long-run growth (1990-2010) for Asia, while negative and highly significant for SSA (being robust to all considered controls).

[25] I simply create a composite measure standardizing *sanitation*, *water*, *electricity* and *transport_systems*, and aggregating them with equal weight.

4.4. Sub-Saharan Africa:

A focus on SSA has two main motivations. One resides in the particular deficiencies in urban infrastructure, which could be behind results from previous papers reporting negative effects on growth of growing urbanisation despite increasing returns from agglomeration (i.e. Brückner 2012).[26] The second motivation is methodological. In particular, there might still be concerns about reverse causality from growth to primacy and to the quality of urban infrastructure in the results of Tables 2 and 3. As noted, SysGMM estimations are expected to address endogeneity concerns. However, SysGMM estimations rely on internal instruments (i.e. variables' transformations and lags). Valid external instruments for primacy and for the quality of the urban infrastructure are hard to find. Yet, we can find reliable external instruments for economic growth, at least for Sub-Sahara African countries. Being still relatively dependent on agriculture and agricultural-dependent activities, economic growth in SSA countries is significantly determined by rainfall.[27] Following Brückner (2012; 2013), I exploit this exogenous variation to construct instrumental variables that allows us to purge the possible effect that economic growth might have on our key variables, urban concentration and sanitation (reverse causality). The use of exogenous instruments allows us to control for simultaneity bias concerns in a more direct way, alternative to SysGMM and without having to rely on internal instruments. The strategy is based on a two-steps procedure. In a first step primacy and sanitation are regressed on economic growth by using a panel Fixed Effects Instrumental Variables (FE-IV) approach using rainfall and rainfall squared as instruments for growth:

$$UC_{it} = \rho_1(\Delta y_{it-1,t}) + a_i + b_t + \varepsilon_{it} \quad (4)$$

$$G_{it} = \rho_2(\Delta y_{it-1,t}) + a_i + b_t + \varepsilon_{it} \quad (5)$$

where a_i are country fixed effects and b_t are year fixed effects. The introduction of country fixed effects allows us to control for time-invariant country-specific omitted variables, while the introduction of year fixed effects allows us to control for global shocks. Appendix D presents first-stage OLS estimation for growth on rainfall and rainfall squared, and FE-IV estimations of equations (4) and (5). Rainfall (and its square) appears as highly significant to explain

[26] According to Brückner (2012), high ethnic fractionalization, very low economic development and excessive size of primate cities drive negative effects of growing urbanisation in Africa. Brückner suggests that the negative role of the excessive size of primate cities relates to their large squatter settlements with inadequate access to transport, water, sanitation, electricity, and health services, but he does not explicitly examines the role of these services.

[27] Higher levels of rainfall are expected to increase agricultural productivity and therefore economic growth in these countries. One should also consider rainfall squared, as too much rainfall can lead to floods detrimental for agriculture. See Miguel et al. (2004), Brückner and Ciccone (2011) and Brückner (2012) for more on the significance of rainfall as an exogenous variable determining economic growth in SSA countries. There is also a relatively recent and increasing literature on the effects of decreasing long-term trends of rainfall, associated with climate change, in Sub-Saharan Africa (see for instance Barrios et al. 2006).

variation in economic growth in SSA, as previously found in the literature (see footnote 27). By construction the residual variation on primacy and sanitation from our FE-IV estimations of equations (4) and (5) capture any variation in these variables that is not due to economic growth. In a second step I use these residual variations in primacy and sanitation as instruments for actual primacy and sanitation to estimate, again by FE-IV, our economic growth equation (3) for SSA.[28] Appendix E provides formal proof for why this FE-IV strategy using residual variation can properly address simultaneity bias as long as one has good instruments for the dependent variable (in our case rainfall, and its square, as instruments for economic growth). Table 4 presents the results.[29]

[Insert Table 4: Urban concentration and growth in SSA]

Similar to results in column 7 of Table 2, FE-IV results yield a negative and highly significant coefficient for primacy in SSA (column 1 of Table 4) and in line with Brückner (2012). Regarding the role of urban infrastructure, FE-IV results - column 2 of Table 4 - are also similar to those in Table 3 (with a negative coefficient for primacy and a positive for its interaction with sanitation, both being highly significant). Coefficients are robust to the considered controls as to the introduction of an interaction term between urban concentration and income levels (column 3).[30] They are also highly significant if access to improved water source (column 4) or access to electricity (column 5), rather than sanitation, are considered.

Finally, recent literature has highlighted potential measurement error of income per capita in poor countries, especially sub-Saharan African ones. As a final robustness check I use data on light density at night to measure income (as proposed by Henderson et al. 2012). In Table 5 FE and FE-IV estimations of equation (3) for SSA are replicated using light density at night (as aggregated at the national level by Henderson et al. 2012) and divided per population to proxy for income per capita. Results for primacy and for its interaction with urban infrastructure remain significant. Interestingly, the effect of our interaction term is even larger, while the coefficients for primacy and sanitation remain almost exactly of the same magnitude as those in Table 4.

---

[28] As the instruments used in the growth equation are generated regressors, standard errors on the slope coefficients are usually incorrect for hypothesis testing. However, as shown by Wooldridge (2010, p. 125) and noted by Brückner (2013), in the special case of testing that slope coefficients are equal to zero these standard errors are correct.

[29] Standard tests confirm, on one hand, the relevance and validity of rainfall and its square as instruments for growth in our regressions for primacy and sanitation. Angrist-Pischke F tests and Hansen J tests are reported respectively in Appendix 4. On the other hand, tests also confirm the relevance of the residual variation in primacy and sanitation, once the reverse causality from growth has been removed, as instruments for actual primacy and sanitation in the growth equation. Kleibergen-Paap F and LM tests are reported in Table 4.

[30] We obtain similar results regardless of the estimation technique: OLS, FE or SysGMM. Results also hold if we consider a role for ethnic polarisation. As suggested by Brückner (2012), important ethnic divisions increase the severity of negative externalities in urbanised areas. Result available upon request.

[Insert Table 5: Estimates for SSA using light density at night data]

In sum, results confirm - in this case for SSA - the role of urban infrastructure when it comes to analysing the relationship between urban concentration and economic growth. According to estimates, for urban concentration to have a positive impact on growth access basic services must at least cover 70 per cent of urban population. But access to basic services is still very deficient in SSA, as we have seen. Only 3 countries out of 34 reached that 70 per cent threshold of urban population with access to sanitation in 1990 (Djibouti, Mauritius and South Africa), three more countries in 2005 (Angola, Botswana and Seychelles).

5. Conclusions and policy implications

Urban concentration plays an important role in the process of economic development, as the panel results presented confirm. But there are wide heterogeneities across countries in terms of urban processes and urban environments. One aspect of the urban environment that is critical when analysing the relationship between urban concentration and economic growth is the quality of urban infrastructure. The data analysed in this paper indeed reflects important differences across countries in terms of access to basic public services, especially in the developing world. The econometric results provide evidence on the relevance of these differences to explain diverse results found in the literature in what refers to the effect of urban concentration in different regions of the world. The role of access to basic services seems robust to a long list of controls and econometric techniques. In this regard, it has been analysed how urban concentration can be negatively associated with national economic growth under urban environments with deficient urban infrastructure. This situation seems common in Sub-Saharan Africa, where access to improved sanitation and electricity appear as especially deficient and currently hampering structural change as well as the net benefits from urban concentration.

In this line, for large agglomerations in developing countries today, a *Malthusian trap* might be a relevant reality, as population growth in these agglomerations exceeds the supply of resources (understood here as urban infrastructure), leading to congestion costs exceeding the benefits from agglomeration. Regarding policy implications previous works have suggested that when urban congestion is due to natural increase rather than due to migration, as seems to be the case for large agglomerations in Sub-Saharan Africa, investments in urban infrastructure are fundamental (Jedwab et al. 2014). Access to basic services, in particular, is not just desirable per se in terms of quality of life for urban residents but also in terms of capital accumulation as well as in terms of economic efficiency at national level, as they allow for the realisation of agglomeration economies and the control of congestion costs. Consequently, guaranteeing that adequate urban infrastructure in these large cities (as in all urban areas) keeps pace with their rapid increase in population not only improves living conditions but can also induce a transition away from Malthusian dynamics.

Results suggest that the net benefits from agglomeration can arise in places where that is not the case today if efforts are made to improve the quality of the urban environment, and it should not be different in Sub-Saharan Africa. According to the results provided, the negative effects of urban concentration that the literature has implied in this region can be associated precisely with its severe lack of adequate basic infrastructure. But, as in other regions, improvements in urban infrastructure, leading to upgraded urban environment, can also unleash agglomeration economies while helping control congestion costs in Sub-Sahara African countries. In other words, the lower economic performance of Sub-Saharan Africa can be in part explained by hampered agglomeration economies due to deficient urban infrastructures. Clearly further research on urban patterns could be of great value to better understand the relationship between urban concentration and national economic performance, an issue of major relevance for developing countries today.

References:

Ades, A. and Glaeser, E. (1995). 'Trade and circuses: Explaining urban giants,' *The Quarterly Journal of Economics* 110(1): 195-227.

Ashraf, Q. and Galor, O. (2011). 'Dynamics and stagnation in the Malthusian epoch,' *American Economic Review* 101(5): 2003-41.

Ayogu, M. (2007). 'Infrastructure and economic development in Africa: A review,' *Journal of African Economies* 16(10): 75-126.

Bairoch, P. (1988). *Cities and Economic Development: From the Dawn of History to the Present*. Chicago: The university of Chicago Press.

Barca, F., McCann, P., and Rodríguez-Pose, A. (2012). 'The case for regional development intervention: Place-based versus place-neutral approaches,' *Journal of Regional Science* 52(1): 134-152.

Barrios, S. Bertinelli, L., and Strobl, E. (2006). 'Climate change and rural-urban migration: The case of Sub-Saharan Africa,' *Journal of Urban Economics* 60: 357-371.

Bates, R. (1981). *Market and States in Tropical Africa: The Political Basis of Agricultural Policies*. Berkeley: University of California Press.

Bazzi S, Clemens M A (2013). Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth. *American Economic Journal: Macroeconomics* 5(2): 152-186.

Behrens, K. and Pholo Bala, A. (2013). 'Do rent-seeking and interregional transfers contribute to urban primacy in Sub-Saharan Africa?' *Papers in Regional Science* 92: 163-196.

Bertinelli, L., and Black, D. (2004). 'Urbanization and growth,' *Journal of Urban Economics* 56: 80-96.

Bertinelli, L., and Strobl, E. (2007). 'Urbanisation, Urban Concentration and Economic Development,' *Urban Studies*, 44(13): 2499-2510.

Bloom, D. E., Canning, D., and Fink, G. (2008). 'Urbanization and the wealth of nations,' *Science (New York, N.Y.)*, 319(5864): 772-5.

Blundell, R., and Bond, S. (1998). 'Initial conditions and moment restrictions in dynamic panel data models,' *Journal of Econometrics* 87(1): 115-143.

Brückner, M. (2013). 'On the simultaneity problem in the aid and growth debate,' *Journal of Applied Econometrics* 28: 126-150.

Brückner, M. (2012). 'Economic growth, size of the agricultural sector, and urbanization in Africa,' *Journal of Urban Economics* 71: 26-36.

Brückner , M., and Ciccone, A. (2011). 'Rain and the democratic window of opportunity,' *Econometrica* 79: 923-947.

Brülhart, M., and Sbergami, F. (2009). 'Agglomeration and growth: Cross-country evidence,' *Journal of Urban Economics* 65: 48-63.

Calderón, C., and Servén, L. (2010). 'Infrastructure and economic development in Sub-Saharan Africa,' *Journal of African Economies* 19(1): 13-87.

Camagni, R., Capello, R., and Caragliu, A. (2013). 'One or infinite optimal city sizes? In search of an equilibrium size for cities,' *Annals of Regional Science* 51: 309-341.

Capello, R., and Camagni, R. (2000). 'Beyond optimal city size: An evaluation of alternative urban growth patterns,' *Urban Studies* 37: 1479.
Castells-Quintana, D., and Royuela, V. (2014). 'Agglomeration, inequality and economic growth,' *Annals of Regional Science* 52(2): 343-366.
Davis, J., and Henderson, V. (2003). 'Evidence on the political economy of the urbanization process,' *Journal of Urban Economics* 53: 98-125.
Durlauf, S., Johnson, P. and Temple, J. 2005. 'Growth Econometrics,' in Philippe Aghion and Steven Durlauf (Eds.), *Handbook of Economic Growth*, Elsevier: 255-677.
Easterly, W., and Levine, R. (1997). 'Africa's growth tragedy: Policies and ethnic divisions,' *Quarterly Journal of Economics* 112(4): 1203-50.
Fay, M. and Opal, C. 2000. Urbanization without growth: a not-so-uncommon phenomenon. World Bank Policy Research Working Paper 2412.
Field, E. and Kremer, M. (2006). 'Impact evaluation for slum upgrading interventions.' *Doing impact evaluation series #3*, World Bank, July 2006.
Firebaugh, G. (1979). 'Structural determinants of urbanization in Asia and Latin America, 1950-1970,' *American Sociological Review* Vol. 44, No. 2: 199-215.
Gallup, J., Sachs, J., and Mellinger, A. (2001). 'Geography Datasets,' *Center for International Development at Harvard University (CID).*
Gollin, D., Jedwab, R., and Vollrath, D. (2014). 'Urbanization with and without industrialization,' Institute for International Economic Policy Working Papers 2014-01.
Henderson, J.V. (2000). 'The effects of urban concentration on economic growth,' *NBER Working Papers Series* No. 7503.
Henderson, J.V. (2003). 'The urbanization process and economic growth: The so-what question,' *Journal of Economic Growth* 8: 47-71.
Henderson, J.V., (2005). 'Urbanization and Growth,' in Philippe Aghion and Steven Durlauf (Eds.), Handbook of Economic Growth, Vol. 1B: 1543-1591.
Henderson, J.V., Storeygard, A., and Weil, D. (2012). 'Measuring economic growth from outer space,' *American Economic Review* 102(2): 994-1028.
Henderson, J.V., Storeygard, A., and Deichmann, U. (2014). '50 years of urbanization in Africa,' *World Bank Policy Research Working Paper* 6925.
Jedwab, R. Christiansen, L., and Gindelsky, M. (2014). Demography, urbanization and development: Rural push, urban pull and… urban push?. Unpublished manuscript. George Washington University.
Kim, S. (2008). 'Spatial Inequality and Economic Development: Theories, Facts and Policies.' Working Paper No. 16, *Commission on Growth and Development.*
Leitão, N. (2013). 'A panel data modelling of agglomeration and growth: cross-country evidence,' *Theoretical and Empirical Researchers in Urban Management* 8(1): 67-77
Lewis, B. (2014). 'Urbanization and economic growth in Indonesia: good news, bad news and (possible) local government mitigation,' *Regional Studies* 48(1): 192-207.
Lipton, M. (1977). *Why Poor People Stay Poor: Urban Bias in World Development.* Cambridge: Harvard University Press.
Lowe, M. (2014). Rail revival in Africa? The impact of privatization. Paper presented at the joint RES-SPR Conference on "Macroeconomic Challenges Facing Low-Income Countries."
Marx, B., Stoker, T., and Tavneet, S. (2013). 'The economics of slums in the developing world,' *Journal of Economic Perspectives* 27(4): 187-210.
Miguel, E., Satyanath, S., and Sergenti, E. (2004). 'Economic shocks and civil conflict: and instrumental variables approach,' *Journal of Political Economy* 112(41): 725-753.
Mveyange, A. (2014). Night lights and regional income inequality in Africa. Unpublished work.
Richardson, H.W. (1972). 'Optimality in city size, systems of cities and urban policy: a sceptic's view,' *Urban Studies*: 29-48.
Royuela, V., and Suriñach, J. (2005). 'Constituents of quality of life and urban size, ' *Social Indicators Research* 74: 549-572.
Sala-i-Martin, X., Doppelhofer, G., and Miller, R.I. (2004). 'Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach,' *American Economic Review* 3: 813-835.
Sekkat, K. (2013). 'Does infrastructure mitigate the effect of urban concentration on poverty in developing countries?' *Economic Research Forum Working Paper Series*, WP No. 800.
Straub, S. (2011). 'Infrastructure and development: A critical appraisal of the macro-level literature,' *Journal of Development Studies* 47(5): 683-708.
Swanson, H. and Buckley, R. (2013). Are we creating Malthusian cities in Africa? Conference poster.
Todaro, M. (1969). 'A model of labor migration and rural unemployment in less developed countries,' *American Economic Review* 59(1): 138-148.
United Nations Department of Economic and Social Affairs. (2010). *World Population Prospects.* UN DESA Press.

Weil, D., and Wilde, J. (2009). 'How relevant is Malthus for economic development today?' *The American Economic Review* Vol. 99, No. 2: 255-260.
Williamson, J. (1965). 'Regional inequality and the process on national development,' *Economic Development and Cultural Change* 4: 3-47.
Wooldridge, J. (2010). '*Econometric Analysis of Cross Section and Panel Data*: *Second Edition*. MIT Press: Cambridge. MA.
World Bank. (2011). *World Development Report 2011*.

Appendix A: Variables' names, definitions and sources:

Basic growth model	Description	Source
growth	Cumulative annual average per capita GDP growth rate	Constructed with data from PWT 7.1 (Summers and Heston), using real GDP chain data (rgdpch)
primacy	Population living in largest city (percentage of urban population)	World Bank - World Development Indicators
ln(rgdpch)	Per capita GDP (in logs)	Constructed with data from PWT 7.1 (Summers and Heston), using real GDP chain data (rgdpch)
ki	Investment share (percentage of GDP)	PWT 7.1. (Summers and Heston)
fertility	Fertility rates	World Bank - World Development Indicators
schooling23	Average years of secondary and tertiary schooling of adult population	Barro and Lee dataset
Further controls		
primary_edu	Percentage of primary schooling attended in total population	Barro and Lee dataset
higher_edu	Percentage of higher schooling attended in total population	"
pi	Price level of investment	PWT 7.1. (Summers and Heston)
kg	Government consumption (percentage of GDP)	"
openk	Openness	"
life_exp	Life expectancy at birth	World Bank - World Development Indicators
dens65c	Density in coastal regions. 1965	Gallup et al. (2001)
tropicar	Proportion of population living in tropical areas	"
malfal66	Malaria	"
elf60	Ethno linguistic fractionalization	Easterly and Levine (1997)
buddha	Fraction of Buddhist	Sala-i-Martin et al. (2004). (BACE dataset)
confuc	Fraction of Confucian	"
east	Dummy for East Asian countries	"
laam	Dummy for Latin American countries	"
mining	Percentage of GDP in mining	"
muslim00	Fraction of Muslim	"
safrica	Dummy for Sub-Sahara African countries	"
spain	Dummy for Spanish colony	"
pop_density	Population density	World Bank - World Development Indicators
pop_growth	Population growth rate	Constructed with data from PWT 7.1 (Summers and Heston), using data on population
Urban infrastructure		
sanitation	Population with access to improved sanitation facilities (percentage of urban population)	World Bank - World Development Indicators
water	Population with access to improved water source (percentage of urban population)	"
electricity	Access to electricity (percentage of urban population)	World Bank - Sustainable Energy for All database
transport_systems	Dummy variable indicating if primate city has a massive transport system (metro, tram or rapid bus)	Constructed by the authors
telephones	Telephone lines (per 1000 inhabitants)	World Bank - World Development Indicators
infant mortality	Infant mortality rates (per 1000 births)	"
slums	Population living in slums (percentage of urban population)	UN-Habitat
rainfall	Annual rainfall aggregated at the country level	Global Precipitation Climatology Projects (GPCP)
growthavsd	Per capita growth of light density at night	Constructed with data from Henderson et al. (2012)

Appendix B: System GMM results with *water* and *electricity*:

	(1) G=*water*	(2) G=*water*	(4) G=*electricity*	(4) G=*electricity*
	World	Developing	World	Developing
Dependent variable:	growth	growth	growth	growth
UC	0.0256	-0.0536	-0.0224**	-0.0183**
	(0.0455)	(0.0519)	(0.0097)	(0.0081)
G	0.0136	-0.0209	-0.0144**	-0.0057
	(0.0228)	(0.0120)	(0.0072)	(0.0054)
UC*G	-0.0003	0.0005	0.0003**	0.0002**
	(0.0005)	0.0005)	(0.0001)	(0.0001)
Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Observations	497	347	540	374
Number of countries	129	91	137	95
AR(1) p-value	0.071	0.087	0.029	0.050
AR(2) p-value	0.203	0.276	0.187	0.179
Hansen test p-value	0.180	0.271	0.118	0.068

Note: Controls include *ln(rgdoch)*, *ki*, *fertility* and *schooling23*. All controls are calculated as averages over 5 years except *ln(rgdoch)* and *schooling23*, which are measured at the beginning of each period. Estimation done by SysGMM. *ln(rgdoch)*, *ki*, *fertility*, *schooling23*, *UC*, *G* and *UC*G* are treated as endogenous using lagged values between 2 and 4 periods as instruments for first differences and variables in first differences lagged between 2 and 4 periods as instruments for variables in levels. Estimations are done with small sample correction. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix C: Cross-section results:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	G=*sanitation*		G=*electricity*		G=*transport_systems*		G=composite	
Dependent variable:	growth	growth	growth	growth	growth	growth	growth	growth
UC	-0.0070***	-0.0070***	-0.0025*	-0.0031**	-0.0004	-0.0033***	-0.0067***	-0.0053***
	(0.0026)	(0.0020)	(0.0014)	(0.0013)	(0.0009)	(0.0011)	(0.0019)	(0.0018)
G	-0.0020	-0.0006	0.0004	0.0009	-0.0532	-0.0880	-0.0019	0.0016
	(0.0015)	(0.0011)	(0.0009)	(0.0012)	(0.0572)	(0.0714)	(0.0012)	(0.0018)
UC*G	0.0001***	0.0001**	0.0001**	0.0000	0.0013	0.0039*	0.0001***	0.0001**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0014)	(0.0020)	(0.0000)	(0.0000)
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Further controls	NO	YES	NO	YES	NO	YES	NO	YES
adj R square	0.294	0.637	0.291	0.609	0.231	0.611	0.316	0.674
Obs.	112	87	129	93	129	93	107	84

Note: *growth* is here calculated as cumulative annual average per capita GDP growth rate between 1990 and 2010. In column 7 and 8 composite is calculated combining *sanitation*, *water*, *electricity* and *transport_systems*. Controls include *ln(rgdoch)*, *ki*, *fertility* and *schooling23*. Further Controls include: *primary_edu, higjher_edu, pi, kg, yrsopen, life_exp, dens65c, tropicar, malfal66, elf60, buddha, confuc, east, laam, mining, muslim, safrica, spain, pop_dens, ki, fertility* and *pop_growth*. All right-hand variables are measured at the beginning of the period or closest year. Estimations are done by OLS. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix D: First step estimations for SSA:

	(1) OLS	(2) FE-IV	(3) FE-IV
Dependent variable:	growth	primacy	sanitation
rainfall	0.0028**		
	(0.0013)		
rainfall squared	-0.0001***		
	(0.0000)		
growth		-4.1597	-0.4570
		(2.9154)	(1.3064)
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Observations	143	178	143
Number of countries	38	38	38
First-stage F-stat p-value		0.000	0.005
Angrist-Pischke F stat p-value		0.053	0.093
Hansen J stat p-value		0.730	0.944

Note: Columns 2 and 3 use *rainfall* and *rainfall_squared* as instruments for *growth*. IV estimations are done with small sample correction. Angrist-Pischke F tests the significance of excluded instruments. Hansen J tests the null hypothesis of valid instruments. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix E: Adjusting for simultaneity bias, formal proof:

Building on Bruckner's (2013), this appendix briefly formalizes how simultaneity bias can be properly addressed by IV estimations using residual variation in urban concentration and in urban infrastructure that is not driven by economic growth. We start by assuming a possible simultaneous equation model:[31]

$$\Delta y = \lambda(UC) + u \quad \text{(E.1)}$$
$$UC = \gamma(\Delta y) + e \quad \text{(E.2)}$$

where Δy is growth and UC is urban concentration. We are interested in the coefficient λ. However, if $\gamma \neq 0$ then OLS estimates of λ from equation (E.1) will be biased. Nevertheless, if we can consistently estimate γ in equation (E.2) we can construct a series for UC that is adjusted for the endogenous response (i.e. $res(UC) = UC - \hat{\gamma}(\Delta y)$) and use $res(UC)$ as an instrument for actual UC in equation (E.1) to estimate λ. The instrumental variables (IV) estimate of λ will not suffer from simultaneity bias:

$$\lambda^{IV} = \frac{cov(res(UC),\Delta y)}{cov(res(UC),UC)} = \lambda + \frac{cov(res(UC),u)}{cov(res(UC),UC)} = \lambda + \frac{cov(e,u)}{cov(e,UC)} \quad \text{(E.3)}$$

Consistent estimate of γ can only be obtained if one has a valid instrument for Δy in equation (E.2) (OLS can not yield a consistent estimate of γ if $\lambda \neq 0$ in equation (E.1)). In our case rainfall, and its square, provide these valid instruments for growth.[32]

In fact, we can identify the adjustment in λ when addressing for simultaneity bias in our growth equation. The first stage estimation, in which actual UC is regressed on $res(UC)$, is:

$$UC = ð\big(res(UC)\big) + v = ð(UC - \hat{\gamma}\Delta y) + v \quad \text{(E.4)}$$

the residuals from this stage being:

$$\hat{v} = UC - \hat{ð}(UC - \hat{\gamma}\Delta y) \quad \text{(E.5)}$$

We can introduce $\hat{v}$ as an additional control in our growth equation and estimate by OLS - control function approach. The estimate for λ will be the same than λ^{IV} (see Wooldridge 2010 for the equivalence between IV and control function approach estimates in linear models):

$$\Delta y_{it,t+1} = \beta(logy_{it}) + \psi X_{it} + \lambda UC_{i,t} + \theta\hat{v} + \pi Z_{1it} + \varepsilon_{it} \quad \text{(E.6)}$$

which equals to:

$$\Delta y_{it,t+1} = \beta(logy_{it}) + \psi X_{it} + (\lambda + \theta(1 - \hat{ð}))UC_{it} + \theta\hat{\gamma}\hat{ð}\gamma_{it-1,t} + \pi Z_{1it} + \varepsilon_{it} \quad \text{(E.7)}$$

and where $\theta(1 - \hat{ð})$ will be the adjustment for simultaneity bias done to an estimate of λ in which simultaneity bias was not addressed (i.e. direct OLS without $\hat{v}$). As it can be seen, the adjustment depends on θ, which indicates the role of past growth in explaining current growth, and also on $(1 - \hat{ð})$, which captures the share on the variation of UC that is due to economic growth. If any of the two components, θ or $(1 - \hat{ð})$, is cero then the estimate reduces to the direct OLS estimate (no simultaneity bias).

---

[31] We formalize here the procedure to adjust for simultaneity bias between primacy and growth. An equivalent procedure is followed to adjust for simultaneity bias between urban infrastructure and growth.

[32] Note that there will still be omitted variables bias in our IV estimate of λ if $cov(e,u) \neq 0$. This bias will, of course, diminish as further controls are taken into account (as well as fixed effects in panel data estimations are included).

Tables and figures:

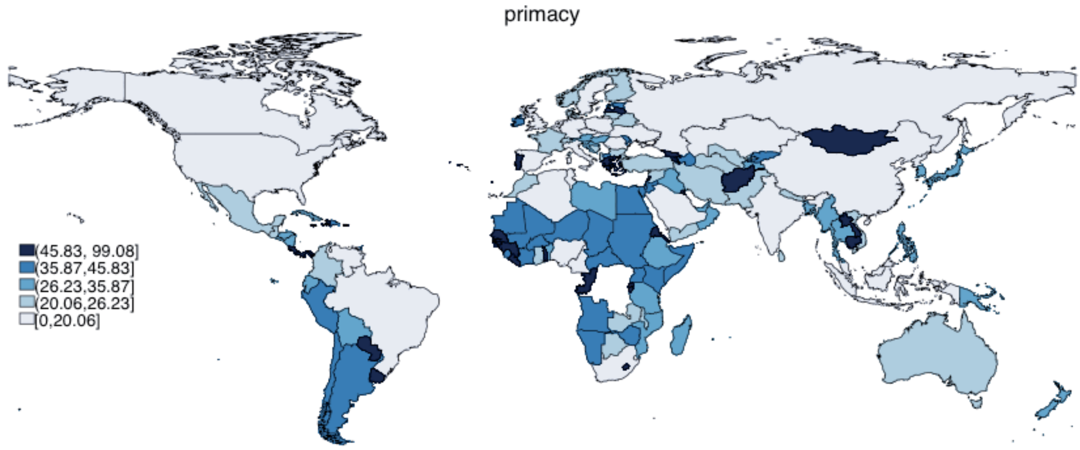
Table 1: Some basic figures

Panel A:	growth			primacy			Correlation	slums		
Region	mean	std dev	sample	mean	std dev	sample	(growth-primacy)	mean	std dev	sample
Developed	1.8	1.1	42	35.0	24.8	44	0.17			
Developing	1.9	2.3	128	43.3	24.2	149	-0.08	57.0	28.9	102
North America	1.5	0.1	3	18.0	6.7	4	-0.10	18.0		1
Europe	2.0	1.5	31	28.3	19.0	38	0.10			
Asia	2.9	2.9	39	38.5	24.1	47	0.07	52.2	24.9	26
Oceania	1.1	0.9	12	72.8	30.8	16	-0.07			
North Africa	2.1	1.2	6	25.8	9.6	6	0.20	39.5	29.0	6
LAC	1.9	1.3	34	49.3	23.7	36	0.08	33.7	23.8	28
SSA	1.3	3.0	45	42.1	16.3	46	-0.13	77.1	19.7	41
World	1.9	2.0	170	41.4	25.5	193	-0.03	57.0	28.9	102

Panel B:	sanitation			Other urban infrastructure measures				
Region	mean	std dev	sample	water mean	electricity mean	inf. mort. mean	tel. lines mean	transport mean
Developed	98.6	3.59	40	99.6	96.6	11.0	40.9	69%
Developing	70.6	25.71	142	89.4	79.8	62.4	10.8	20%
North America	100.0	0	2	100.0	100.0	8.1	59.9	100%
Europe	97.5	4.6	35	99.7	99.7	13.1	38.4	74%
Asia	83.8	18.25	45	93.3	92.3	50.2	14.1	33%
Oceania	83.3	18.39	14	93.0	74.3	33.6	16.2	6%
North Africa	87.7	15.15	6	85.9	90.5	56.6	5.8	67%
LAC	84.2	15.68	34	94.8	94.9	35.4	17.6	26%
SSA	44.1	20.26	46	81.1	52.7	93.3	2.2	0%
World	76.7	25.55	182	91.7	83.7	50.7	17.8	32%

Note: *growth* is calculated over 1970-2010. "sample" indicates the number of countries considered (for which we have data for the respective region and variable). *primacy*, *sanitation*, *water*, *electricity* and *tel. lines* are calculated as averages over 1990-2005. *transport* indicates the percentage of countries in the region for which their primate city has a massive transport system (metro, tram or rapid bus).

Figure 1: Population living in largest city (percentage of urban population)



Note: values for *primacy* calculated as averages between 1970 and 2010.

Figure 2a: Access to improved sanitation (percentage of urban population)

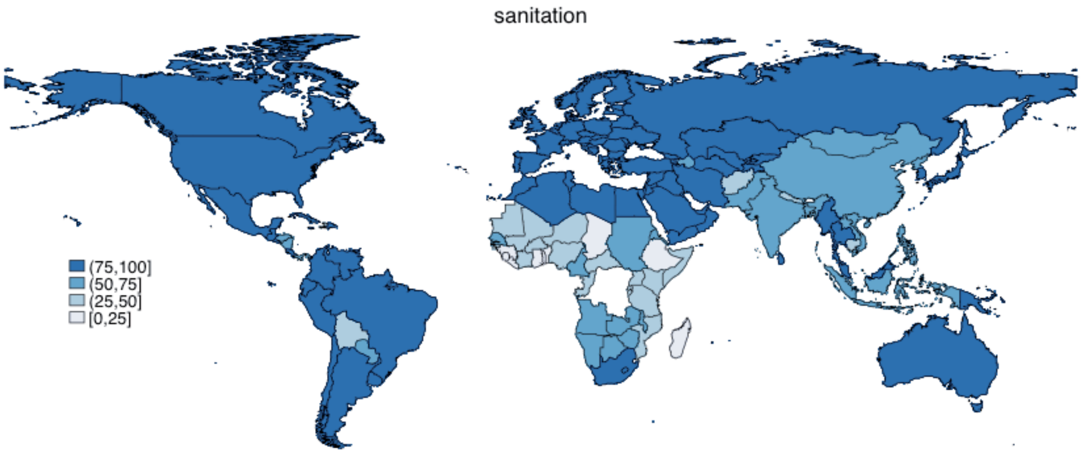
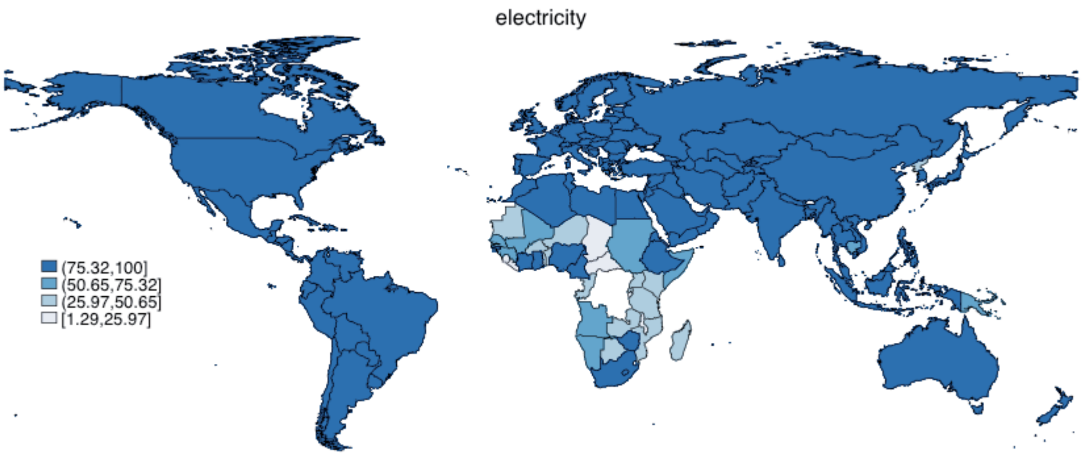


Figure 2b: Access to electricity (percentage of urban population)



Note: values for *sanitation* and *electricity* calculated as averages between 1990 and 2005.

Figure 3: Access to improved sanitation by income levels

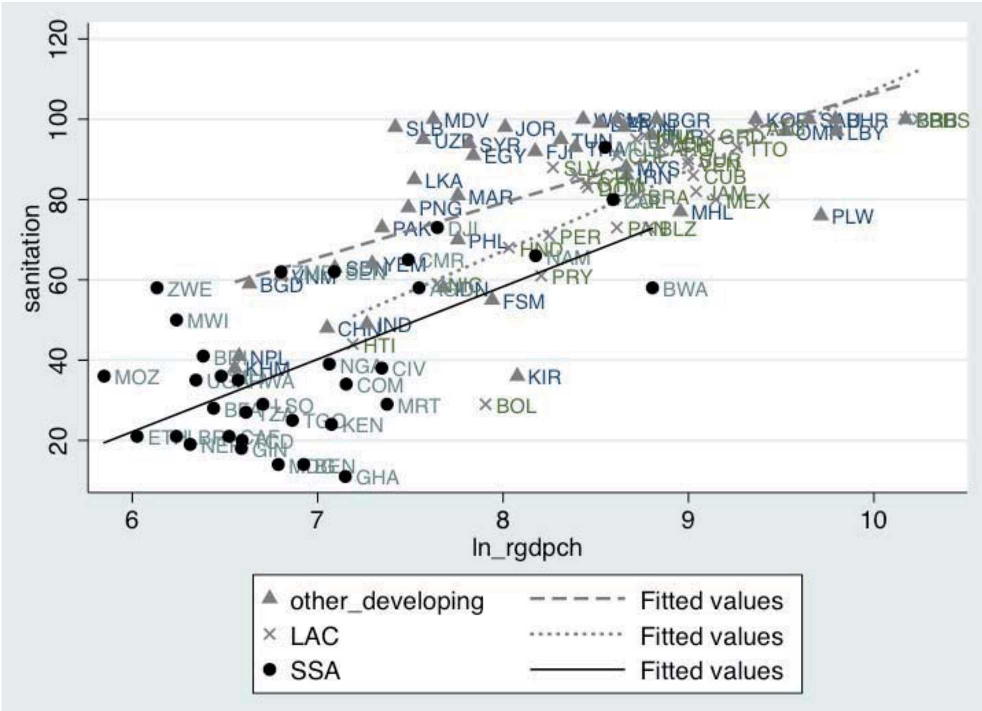
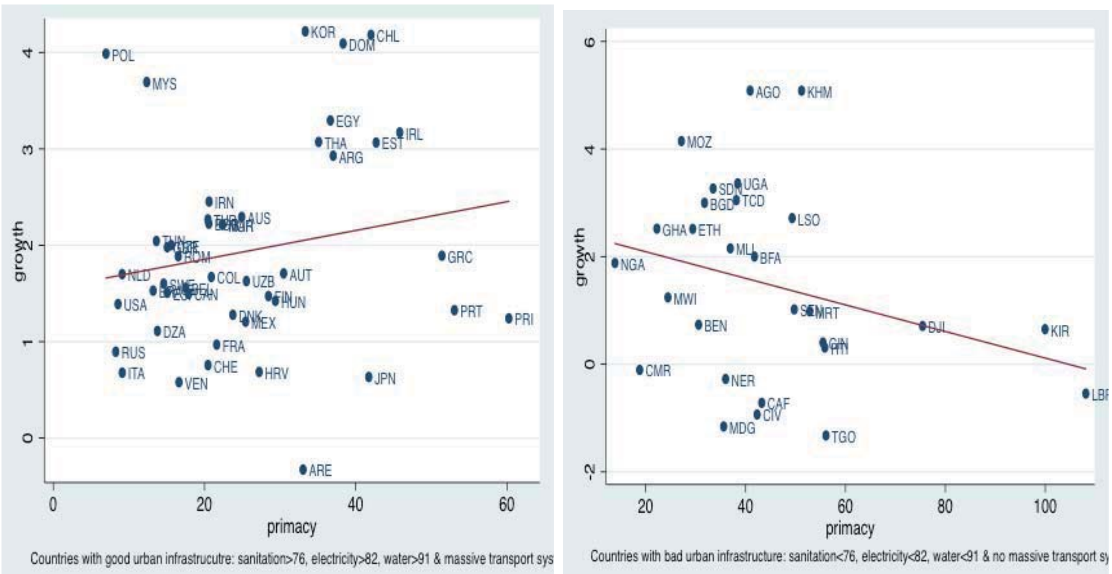


Figure 4: Correlation between primacy and growth by quality of the urban infrastructures



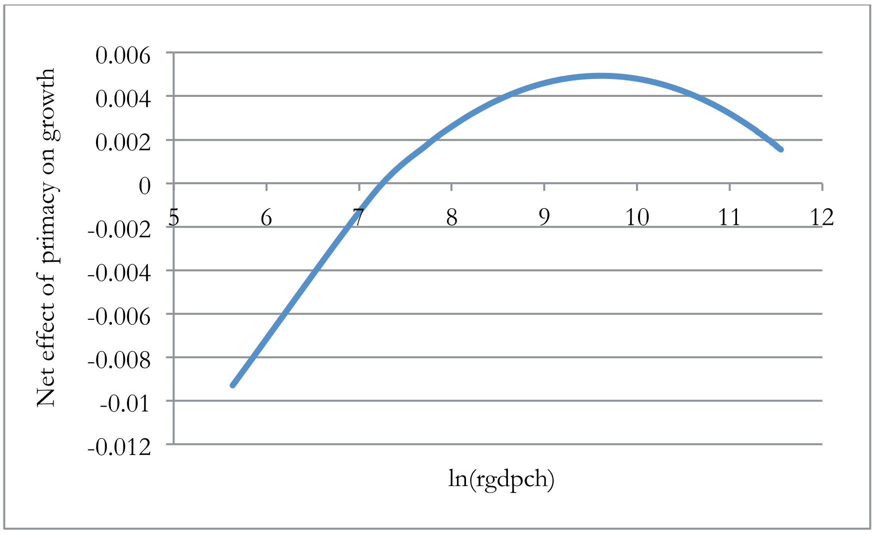
Note: Countries have been classified respect to sample mean

Table 2: Urban concentration and growth in a panel of countries

	(1) OLS	(2) FE	(3) GMM	(4)SysGMM	(5)SysGMM	(6)SysGMM	(7)SysGMM
Dependent variable: Average cumulative annual growth rates of per-capita GDP							
ln(rgdpch)	-0.1031***	-0.4280***	-0.3697***	-0.0362***	-0.0814***	-0.1252***	-0.0715***
	(0.0151)	(0.0686)	(0.1695)	(0.0779)	(0.0507)	(0.0974)	(0.0539)
ki	0.0073***	0.0078***	-0.0019	0.0039*	0.0015	-0.0015	-0.0014
	(0.0012)	(0.0021)	(0.0031)	(0.0021)	(0.0035)	(0.0032)	(0.0036)
fertility	-0.0870***	-0.0546***	0.0315	-0.0487***	-0.0629***	-0.0362**	-0.0448***
	(0.0094)	(0.0183)	(0.0237)	(0.0170)	(0.0170)	(0.0156)	(0.0131)
schooling23	0.0044	0.0116	0.1466**	-0.0186	0.0206	0.0113	-0.0141
	(0.0109)	(0.0280)	(0.0694)	(0.0515)	(0.0388)	(0.0565)	(0.0387)
UC					0.0054*	-0.0782***	0.0049*
					(0.0032)	(0.0269)	(0.0027)
UC*ln(rgdpch)						0.0173***	
						(0.0062)	
UC*(ln(rgdpch))^2						-0.0009***	
						(0.0003)	
UC*LAC							-0.0040***
							(0.0012)
UC*SSA							-0.0070**
							(0.0030)
Year FE	YES	YES	NO	YES	YES	YES	YES
Annual speed of convergence	2.10%	11.27%	9.12%	0.59%	1.70%	2.68%	1.48%
adj R square	0.196	0.216					
Observations	1204	1204	1033	1170	1204	1204	1204
No. of countries	137	137	137	137	137	137	137
AR1 test p-value			0.030	0.001	0.004	0.000	0.002
AR2 test p-value			0.352	0.412	0.437	0.582	0.552
Hansen test p-value			0.032	0.011	0.047	0.166	0.338

Note: *ki*, *fertility* are calculated as averages over 5 years. The time span goes from 1960 to 2010. All remaining variables are measured at the beginning of the period. For GMM and SysGMM estimations variables in levels lagged between 2 and 4 periods are used as instruments for first differences, and variables in first differences lagged between 2 and 4 periods are used as instruments for levels. GMM and SysGMM estimations are done with small sample correction. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Figure 5: The Williamson hypothesis



Note: Plot using SysGMM estimation coefficients (column 6 of Table 2).

Table 3: Urban concentration depending on the urban process

	World sample			Developed	Developing	Developing	Developing	Developing
	(1) OLS	(2) FE	(3)SysGMM	(4)SysGMM	(5)SysGMM	(6)SysGMM	(7)SysGMM	(8)SysGMM
Dependent variable:	growth	growth	growth	growth	growth	growth	growth	growth
UC	-0.0171***	-0.0474***	-0.0331	0.0711	-0.0462***	-0.0031	0.1152	
	(0.005)	(0.016)	(0.020)	(0.061)	(0.011)	(0.0211)	(0.0729)	
sanitation	-0.0035	-0.0057	-0.0197	0.0310	-0.0139	-0.0159	-0.0080	-0.0137
	(0.002)	(0.011)	(0.012)	(0.028)	(0.010)	(0.0109)	(0.0089)	(0.0112)
UC*sanitation	0.0002***	0.0004*	0.0004*	-0.0007	0.0005***	0.0005**	0.0004**	0.0005**
	(0.000)	(0.000)	(0.000)	(0.0006)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
UC*ln(rgdpch)						-0.0053	-0.0360**	-0.0354**
						-0.0037	-0.0177	-0.0134
UC*ln(rgdpch)^2							0.0021*	0.0018**
							-0.0011	-0.0009
UC*region								YES
Country FE	NO	YES						
Year FE	YES	YES	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Observations	500	500	500	144	356	356	356	356
No. of countries	131	131	131	37	94	94	94	94
AR(1) p-value			0.082	0.192	0.071	0.059	0.029	0.043
AR(2) p-value			0.280	0.371	0.569	0.505	0.863	0.711
Hansen test p-value			0.172	0.529	0.424	0.305	0.325	0.272

Note: *UC*region* are interaction terms between *UC* and each world region. Controls include *ln(rgdoch)*, *ki*, *fertility* and *schooling23*. All controls are calculated as averages over 5 years except *ln(rgdoch)* and *schooling23*, which are measured at the beginning of each period. The time span goes from 1990 to 2010. In SysGMM estimations *ki*, *fertility*, *schooling23*, *UC*, *sanitation* and *UC*sanitation* are treated as endogenous using lagged values between 2 and 4 periods as instruments for first differences and variables in first differences lagged between 2 and 4 periods as instruments for variables in levels. SysGMM estimations are done with small sample correction. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Urban concentration and growth in SSA

	G = *sanitation*			G = *water*	G = *electricity*
	(1) FE-IV	(2) FE-IV	(3) FE-IV	(4) FE-IV	(5) FE-IV
Dependent variable:	growth	growth	growth	growth	growth
UC	-0.0287**	-0.0874***	-0.0200	-0.3371***	-0.1754**
	(0.0141)	(0.0146)	(0.0697)	(0.0767)	(0.0752)
G		-0.0638***	-0.0725***	-0.1083***	-0.0361***
		(0.0214)	(0.0229)	(0.0301)	(0.0094)
UC*G		0.0013***	0.0015***	0.0024***	0.0007***
		(0.0005)	(0.0005)	(0.0007)	(0.0002)
UC*ln(rgdpch)			-0.0111	0.0125	0.0170
			(0.0113)	(0.0124)	(0.0118)
Country FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES
Observations	135	103	103	95	108
Number of countries	28	28	28	26	28
Angrist-Pischke F tests p-values	0.000	0.000; 0.000; 0.000	0.007; 0.000; 0.000; 0.009	0.002; 0.000; 0.001; 0.003	0.001; 0.000; 0.000; 0.002
Kleibergen-Paap F-stat	1253.02	40.15	28.19	18.37	70.07
Kleibergen-Paap LM-stat	6.63**	24.26***	23.38***	12.39***	17.10***

Note: Controls include *ln(rgdoch) ki*, *fertility* and *schooling23,* but also *rainfall* and *rainfall squared*. All controls are calculated as averages over 5 years except *ln(rgdoch)* and *schooling23*, which are measured at the beginning of each period. The time span goes from 1990 to 2010. For IV estimations, *UC*, *G* and *UC*G* series adjusted for the effect that *growth* has on them are used as instruments. Kleibergen-Paap stats test the null hypothesis that the equation is underidentified. Estimations are done with small sample correction. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Estimates for SSA using light density at night data

	(1) FE	(2) FE-IV	(3) FE	(4) FE-IV
Dependent variable:	growthavsd	growthavsd	growthavsd	growthavsd
UC	-0.0175*	-0.0127	-0.0624***	-0.0866***
	(0.0094)	(0.0103)	(0.0212)	(0.0198)
sanitation			-0.0258	-0.0601*
			(0.0304)	(0.0304)
UC*sanitation			0.0019**	0.0029***
			(0.0009)	(0.0008)
Country FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Observations	108	108	103	103
Number of countries	28	28	28	28
Angrist-Pischke F tests p-values		0.000		0.000
Kleibergen-Paap F-stat		2745.44		33.02
Kleibergen-Paap LM-stat		3.93**		21.41***

Note: Controls include *ln(avsd) ki*, *fertility* and *schooling23,* but also *rainfall* and *rainfall squared*. All controls are calculated as averages over 5 years except *ln(avsd)* and *schooling23*, which are measured at the beginning of each period. The time span goes from 1990 to 2010. For IV estimations, *UC*, *sanitation* and *UC*sanitation* series adjusted for the effect that *growth* has on them are used as instruments. IV estimations are done with small sample correction. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Institut de Recerca en Economia Aplicada Regional i Públic
Research Institute of Applied Economics

WEBSITE: www.ub-irea.com • **CONTACT:** irea@ub.edu

Grup de Recerca Anàlisi Quantitativa Regional
Regional Quantitative Analysis Research Group

WEBSITE: www.ub.edu/aqr/ • **CONTACT:** aqr@ub.edu

Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona