
Link to published version (if available):
10.1111/j.1728-4457.2012.00493.x

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights
This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
http://www.bristol.ac.uk/pure/about/ebr-terms
Urbanization as a Global Historical Process: Theory and Evidence from sub-Saharan Africa

SEAN FOX

The process of urbanization has traditionally been understood as a natural byproduct of economic development. While there is no doubt that economic expansion in the urban sector can stimulate rural-to-urban migration, hence urbanization, a strictly economic theory of the process fails to account adequately for the phenomenon of “urbanization without growth” observed in sub-Saharan Africa in the 1980s and 1990s (Fay and Opal 2000). Inspired by this apparent anomaly, I propose an alternative, historically grounded theory of urbanization and deploy it to explain the stylized facts of Africa’s urban transition, namely the late onset of urbanization vis-à-vis other major less-developed world regions and the persistence of both urbanization and rapid urban population growth in the late twentieth century despite economic stagnation (see Table 1).

I argue that urbanization should be understood as a global historical process driven by population dynamics associated with technological and institutional change. While urban settlements emerged in many regions before the nineteenth century, the proportion of the global population residing in urban areas remained low. Historical evidence indicates that urban population growth in the preindustrial era was ultimately restricted by two factors: 1) the scarcity of surplus energy supplies (primarily food) to support nonagricultural populations, and 2) an inability to control infectious and parasitic diseases, which thrive in densely populated settlements. In other words, limitations on the availability of food supplies coupled with high disease burdens in urban settlements imposed a natural ceiling on urban population growth and hence urbanization in the preindustrial era.

A combination of technological and institutional innovations in the eighteenth and nineteenth centuries began to alleviate these constraints. While these emerged primarily in Europe, they were later diffused worldwide
through colonialism, trade, and, in the post-colonial period, international development assistance, thereby setting in motion an inexorable process of world urbanization. Crucially, many of these innovations have contributed directly to both mortality decline (hence population growth) and economic development, leading to the spurious conclusion that economic development is the motive force behind urbanization. Although there is no question that structural shifts in labor markets contribute to rural-to-urban migration, the historical record indicates that ultimately improvements in disease control and food security underpin urbanization.

Applying this thesis to the African case, the late onset of the region’s urban transition can be attributed to natural geographic endowments that rendered the local production and acquisition (through trade) of surplus food supplies, as well as disease control, especially difficult. The rapid pace of urban growth in Africa since 1960 is explained by mortality decline and improved access to surplus food supplies made possible by the application of technologies and the consolidation of institutions introduced through colonialism, trade, and international development assistance. Countries that have experienced more rapid economic and demographic growth have urbanized more quickly in the era following World War II, but the absence of economic growth has not been sufficient to arrest urban population growth wherever mortality has continued to fall and sufficient food supplies have remained available. In short, the seemingly unique characteristics of Africa’s urban transition are explicable within the framework of the historical theory of urbanization proposed here.

The article proceeds as follows. I first provide a critical review of economic and demographic theories of urbanization. Drawing on the work of economic historians and historical demographers, I sketch the stylized facts of world urbanization in the preindustrial era and articulate an integrated, his-

<table>
<thead>
<tr>
<th>Region</th>
<th>1960 Urban population (% of total)</th>
<th>1975–2005 Urban growth rate</th>
<th>Urbanization rate</th>
<th>GDP growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-Saharan Africa</td>
<td>14.8</td>
<td>4.4</td>
<td>1.7</td>
<td>–0.1</td>
</tr>
<tr>
<td>East Asia and Pacific</td>
<td>20.5</td>
<td>3.5</td>
<td>2.1</td>
<td>3.1</td>
</tr>
<tr>
<td>South Asia</td>
<td>16.7</td>
<td>3.4</td>
<td>1.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>35.1</td>
<td>3.6</td>
<td>0.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Latin America and Caribbean</td>
<td>48.9</td>
<td>2.8</td>
<td>0.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>

*a* Compound average annual rate of population growth in urban areas.

*b* Compound average annual rate of change in the percent of a country’s population residing in urban areas.

*c* Estimates based on real GDP per capita (constant 2000 US$).

Source: Calculated from World Bank World Development Indicators Database, accessed January 2012.
historically grounded explanation of urbanization based on these observations. I then apply this theory to urbanization trends in Africa. First, I offer statistical evidence that cross-country variation in colonial experience—an important historical determinant of institutional and technological change—accounts for a significant fraction of variation in patterns of early urbanization in the region. Second, I examine the dynamics of African urbanization in the post-colonial era and demonstrate empirically that “urbanization without growth” and exceptionally rapid urban population growth are largely accounted for by sub-Saharan Africa’s unique historical circumstances and population dynamics. I conclude with a brief comment on the policy implications of the theory and evidence presented here.

Economic and demographic theories of urbanization

The traditional economic theory of urbanization, which has dominated in both academic and policy circles since the 1950s, revolves around the relationship between structural economic change and the spatial dynamics of the labor market. The premise is straightforward: as the modern urban sector (i.e., manufacturing and services) expands, surplus labor from the “backward” rural economy (i.e., agriculture) is drawn to towns and cities, attracted by higher wages (Lewis 1954; Fei and Ranis 1964). In other words, this economic model suggests that urbanization is fundamentally driven by rural-to-urban migration stimulated by a wage gap between rural and urban areas that arises in the early stages of industrialization.  

As early as the 1950s, however, scholars recognized that rates of urbanization in many developing countries were incommensurate with the growth of wage-based employment opportunities in urban areas, resulting in under- and unemployment—a phenomenon dubbed “over-urbanization” (Davis and Golden 1954). To explain this deviation from the classic dual-economy model of urbanization, and the implied failure of the market to allocate labor efficiently between rural and urban areas, Harris and Todaro (1970) proposed a revised model in which migration decisions are influenced by expected as opposed to actual earnings in the urban sector. Over-urbanization is explained in the Harris–Todaro model as a consequence of wage-distorting government interventions in the labor market that inflate the wages of a few and raise the expectations of the masses. In such contexts, the model suggests, policies of wage equalization or mobility restriction will lead to net welfare improvements.

While decades of research have consistently demonstrated a strong cross-sectional association between indicators of economic development (e.g., income per capita and structure of output) and levels of urbanization at the national level, empirical tests of the wage-differential mechanism assumed in both the classical and Harris–Todaro models have produced ambiguous
results, explaining only a small fraction of the variation in rates of urbanization across countries (see Mazumdar 1987; Weeks 1995; Becker and Morrison 1995; Fay and Opal 2000; Lall, Selod, and Shalizi 2006). This outcome may reflect the failure of these models to take account of the role of the urban informal sector (the “third sector”), which is where most migrants (especially in Africa) are found (Bhattacharya 1993). It may also be a consequence of a narrow focus on economic incentives for migration.

Qualitative and quantitative studies conducted in the 1960s and 1970s identified many noneconomic motives for migration, such as the desire of youth to escape the control of older generations; of women to escape gender discrimination, join husbands in town, or take advantage of the “thick” market for spouses in urban areas; and of others to acquire the social prestige associated with urban life or to pursue their aspirations in the “bright lights” of the big city (see Byerlee 1974; Mazumdar 1987; Becker and Morrison 1995). More recent studies have explored the impacts of ethnic conflict, war, and climatic changes in spurring migration to urban areas (Fay and Opal 2000; Barrios, Bertinelli, and Strobl 2006).

Given the many reasons people leave the countryside for the city, it is not surprising that studies have failed to confirm the primacy of the wage-differential mechanism. But few alternative explanations have been proposed. For example, in their effort to solve the paradox of urbanization without growth in sub-Saharan Africa throughout the 1980s and 1990s, Fay and Opal (2000) note that in Africa and many other regions, “urbanization continues even during periods of negative [economic] growth, carried by its own momentum” (p. 25). The only explanation the authors provide for this momentum is a vague speculation that external forces (e.g., globalization) may be at work.

A missing consideration in most studies is the potential role of population dynamics. In fact, the process of urbanization has received marginal attention in demography. Notably there has been little cross-country research on the topic (Dyson 2011). The few studies that have sought to explain variation in rates of urbanization and urban population growth emphasize the dynamics associated with the demographic transition. There are essentially two sources of urban population growth: natural increase in urban areas and rural-to-urban migration (Cohen 2003). The onset of mortality decline ahead of fertility decline in urban areas raises the rate of urban natural increase, and urban populations expand regardless of whether they are net recipients of rural migrants. Urbanization, of course, could occur in a population without rural-to-urban migration if urban natural increase exceeded rural natural increase over a sustained period, but in reality this has rarely happened. With regard to rural-to-urban migration—the proximate cause of urbanization—it has long been assumed among demographers influenced by Malthusian arguments that rapid population growth in rural areas (stimulated by mortality
decline) places strains on natural resources (e.g., land and water) resulting in declining living standards, thereby contributing to the “push” factors that drive people into cities (Preston 1979; Kelley and Williamson 1984).

There is robust empirical evidence that mortality decline and the acceleration in population growth that follows are important determinants of urban population growth. Preston (1979), for example, demonstrated a strong one-to-one correlation between total population growth and urban growth, and many studies have observed that urban natural increase generally contributes more to overall urban population growth than does rural-to-urban migration, although the relative contribution of each tends to shift as a country urbanizes, with natural increase playing an increasingly important role (e.g., Davis 1965; Preston 1979; Cohen 2003; Lall, Selod, and Shalizi 2006). By contrast, evidence on the relationship between rates of population growth and rates of urbanization is virtually non-existent.

Whether the onset of the demographic transition is a necessary and/or sufficient condition for the urban transition to occur is difficult to deduce from cross-country research. Hypothetically, both urbanization and urban growth could occur solely through rural-to-urban migration in a context of zero overall population growth and rapid economic growth. Conversely, if the Malthusian arguments are right, rapid urbanization in Africa in a context of economic stagnation might be explained by the region’s exceptionally rapid population growth—a possibility that has never been empirically tested. However, the work of historical demographers supports the view that the onset of the demographic transition is a necessary and possibly sufficient condition for urbanization.

**Toward a historical theory of world urbanization**

Abstracting from historical accounts of the emergence, growth, and decline of cities across the world over the past 6000 years, one can glean two key theoretical insights concerning the underlying factors that drive urbanization. Both insights relate to the question of why the world’s urban population remained small and static (hovering around 5 percent of the world population) for thousands of years, before experiencing a rapid and sustained expansion beginning in the late nineteenth century (see Figure 1).

The answer provided by historical demographers is that cities were deadly places to live until recently. Before the nineteenth century, urban settlements with rudimentary water and sanitation infrastructure were especially conducive to the spread of infectious and parasitic diseases. As a consequence, death rates tended to exceed birth rates, turning cities into “demographic sinks” (Graunt 1662/1964; de Vries 1984; Bairoch 1988; Lowry 1990; Dyson 2011). With negative rates of natural increase, cities depended upon a
constant inflow of rural migrants to sustain their populations. This suggests that the propensity to migrate from rural to urban settlements has been a consistent feature of human behavior for as long as cities have existed—even when migrants stood to suffer from higher rates of morbidity and mortality than they would have in the countryside. The important implication is that any realistic model of the urbanization process should assume some constant rate of rural-to-urban migration, all other things being equal.

This disease constraint on urban growth also clarifies the mechanism linking the demographic transition to urbanization by identifying mortality decline stimulated by disease control as the demographic dynamic of greatest causal significance. Where the burden of disease is eased, mortality decline contributes to urban population growth in three ways: 1) by raising the rate of urban natural increase above zero; 2) by increasing demographic pressure in rural areas, potentially spurring migration; and 3) by transforming rural migrants into a source of urban population growth instead of mere population maintenance.

Although evidence to support the disease constraint argument comes primarily from a handful of historical case studies, it is compelling enough for Dyson (2010) to claim that “No population that has experienced a reduction in its death rate from a high level to a low level has failed to urbanize” (p. 126). In other words, mortality decline is a necessary precondition for urbanization and urban growth to occur. Moreover, if some constant net positive rate of rural-to-urban migration is assumed, mortality decline (and the rapid population growth that follows immediately in its wake) can also be interpreted as a sufficient condition.
Economic historians provide a complementary explanation for the limited scale of urbanization prior to the nineteenth century based on a compelling logical premise: cities can only exist where a surplus of energy (i.e., food and fuel) is available to support a large nonagricultural population (Lowry 1990). It follows that the size of the urban population in any given region is a function of the quantity of surplus energy it is able to acquire, which in turn is jointly determined by agricultural productivity and transportation costs (Bairoch 1988).

The limitations imposed on urban population growth by agricultural productivity and transport costs largely explain the geography of early cities, which emerged almost exclusively in areas naturally conducive to surplus food production (e.g., fertile river valleys) or locations with naturally low trade costs (i.e., on coasts and along rivers) (Childe 1950; Davis 1955; Bairoch 1988). Indeed, it is still possible to detect the profound and long-lasting influence of natural geography on patterns of urbanization using a simple ordinary least squares regression analysis in which a country’s level of urbanization and urban population size in 1960 (the earliest date for which comprehensive data are available) are modeled as a function of relatively time-invariant geographic characteristics that influence agricultural productivity and transportation costs. Table 2 presents the results of such an analysis based on a sample of 126 countries. Independent variables include soil quality (measured as

<table>
<thead>
<tr>
<th>TABLE 2 Effect of geographic characteristics on proportions urban and urban population size, 126 countries, 1960</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Percent urban</strong></td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Log of coastline (km)</td>
</tr>
<tr>
<td>(.030)</td>
</tr>
<tr>
<td>Log of waterways (km)</td>
</tr>
<tr>
<td>(.024)</td>
</tr>
<tr>
<td>Soil quality</td>
</tr>
<tr>
<td>(.110)</td>
</tr>
<tr>
<td>Log of GDP per capita</td>
</tr>
<tr>
<td>(.095)</td>
</tr>
<tr>
<td>Log of national population</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>R-squared</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

NOTES: Standard errors in parentheses. Significance at the 1, 5, and 10 percent levels is indicated by ***, **, * respectively.
SOURCES: See Table A.2.
the percentage of a country’s land area with soils that are very or moderately suitable for six key rainfed crops), length of coastline in kilometers and total length of navigable waterways in kilometers. A control for GDP per capita in 1960 is included to capture the inevitable mediating effects of technological and institutional changes prior to 1960 on natural geographical constraints. In the second specification, in which the dependent variable is the total size of the urban population (as opposed to the percentage of the national population residing in urban areas), an added control for national population size is included because of the natural correlation between total population and urban population size (Davis and Henderson 2003).

The results confirm that countries with better soil quality and more extensive “natural” transportation infrastructure were both more urbanized (column 1) and had larger overall urban populations (column 2) in 1960 than countries that were less favorably endowed, as predicted. There is some colinearity in this model: countries with longer coastlines tend to have more kilometers of navigable waterways. While this obscures accurate interpretation of effect magnitudes for these variables, it does not affect the overall fit of the model.

The case for causality in this model is strong: the time-invariant characteristics of coastline, waterways, and soil potential are clearly exogenous to urbanization and urban population size in 1960. Moreover, the inclusion of GDP per capita in 1960 as a control variable is likely to capture the influence of any omitted variables, given the strong correlation between GDP and urbanization.

The history of ancient Rome provides a useful illustration of how physical geography affected access to surplus energy supplies and hence shaped the fortunes of urban settlements in the preindustrial era. At its peak in 200 CE, Rome is estimated to have contained over one million residents (including both citizens and slaves), a population size that far exceeded the total surplus production capacity of the Italian peninsula. To satisfy its energy requirements, Rome was forced to import anywhere from 75 to 95 percent of its wheat supplies from distant territorial possessions. Given the state of transportation technology at the time, this was an extremely costly means of surplus acquisition that contributed to the eventual financial ruin of the empire. With the shift of the imperial center to Constantinople in 330 and the subsequent collapse of the publicly financed system of interregional grain distribution, Rome’s population plummeted to just 50,000 inhabitants by 700—a size more in line with the productive capacity of its hinterland (see Bairoch 1988; Reader 2004).

Low agricultural productivity and high transportation costs in the preindustrial era also explain why the proportion of the world’s urban population remained unchanged for so long. Increasing agricultural output in the premodern era was driven primarily by bringing more land under cultivation rather than by rising yields. Thus, although the global urban population may
have risen in absolute terms, it could not expand in relative terms because of very limited improvement in surplus output. Moreover, the potential for regional specialization and exchange in agricultural goods was very limited, since transportation costs remained well above the threshold that would have made such trade economically viable (Braudel 1984; Bairoch 1988).

A binding “surplus constraint” would indicate that the expansion of a food surplus is a necessary condition for urban populations to grow. The rise and fall of cities in the preindustrial era could therefore be understood as a reflection of both shifting disease burdens and fluctuations in the capacity of individual settlements to acquire surplus food supplies.

Both the disease constraint and the surplus constraint arguments trace the origins of world urbanization to a confluence of social and technological changes in Northern Europe in the eighteenth and nineteenth centuries. Innovations such as nitrogen fertilizer, crop rotation, and mechanization drove a surge in agricultural productivity (Bairoch 1988; Cameron 1997; Maddison 2007). The harnessing of inanimate sources of energy to fuel railroads, steamships, and eventually automobiles led to a dramatic reduction in transportation costs (Bairoch 1988; Crafts and Venables 2003). Improvements in hygiene, medical knowledge, maternal education, and urban planning practices and the expanded availability of health care led to a secular decline in mortality rates (Szreter 1997; Bloom and Sachs 1999; Reher 2004; Livi-Bacci 2007). Political-institutional changes such as the consolidation of private property rights, improved third-party contract enforcement, and an expansion of the role of governments in the provision of public goods (notably, health care, education, and infrastructure) reinforced and sustained these trends (Szreter 1997; Cameron 1997; Maddison 2007). Collectively, these changes catalyzed a permanent shift in Europe from a Malthusian economy characterized by stagnant per capita income growth and high mortality to a modern growth regime characterized by secular rises in factor productivity and life expectancy (Galor and Weil 1999).

Against a backdrop of rising surplus, intensified regional trade, and falling mortality, the stage was set for European urbanization. Between 1800 and 1900, the proportion of Europe’s population living in cities nearly tripled (growing from around 10 percent to 30 percent), and by the turn of the millennium approximately 70 percent of Europe’s population lived in urban areas (Bairoch and Goertz 1986; United Nations 2010). Through trade, colonialism, and, in the latter half of the twentieth century, international development assistance, the key technological and institutional developments that propelled Europe’s urban transition were diffused to other regions, stimulating urbanization there as well. The onset of the urban transition in any given country or region should therefore be understood as part of a global historical process linked to technological and institutional change and diffusion, not simply as a product of endogenous economic and demographic forces.
Figure 2 provides a stylized diagram of this historically grounded theory of urbanization. In brief, the underlying causes of the urban transition are the advent of technologies and institutions that facilitate disease control and surplus energy availability (i.e., productivity growth and reductions in transport costs). These factors stimulate mortality decline in both rural and urban areas. Mortality decline facilitates urban population growth directly by raising the rate of urban natural increase and indirectly by raising the rate of rural-to-urban migration. Technological and institutional changes also promote economic development, which exerts a positive effect on urbanization and urban growth by further stimulating rural-to-urban migration as demand for labor in non-agricultural sectors expands. However, economic growth is not a necessary condition for urbanization to occur. Given that noneconomic motivations for migration are ever-present, countries may experience net positive rates of urbanization as long as disease control and food security are maintained in urban areas.

The fact that many of the technological and institutional changes that drive mortality decline and facilitate surplus expansion also drive economic development is the source of the spurious conclusion that urbanization is fundamentally a byproduct of economic development. These two processes can become decoupled—as is illustrated by the case of sub-Saharan Africa, to which I now turn.

**Geography, colonialism, and early urbanization in sub-Saharan Africa**

Archeological evidence and oral histories confirm the presence of urban settlements in sub-Saharan Africa for over 2000 years (Anderson and Rathbone 2000). However, these settlements remained relatively small, few, and dispersed in comparison to other regions of the world, and most proved
ephemeral. As Figure 3 illustrates, the region’s urban transition did not begin until the middle of the twentieth century.

Drawing on the earlier discussion of constraints to urban growth, the late onset of Africa’s urban transition can largely be explained by natural geographic endowments. Africa’s climate, soils, topography, and disease ecology represent considerable obstacles to surplus agricultural production (Diamond 1997; Bloom and Sachs 1999). A high ratio of land area to coastline, few navigable rivers, and low population densities are significant natural barriers to trade, contributing to exceptionally high transportation costs in the region even today and limited scope for specialization and innovation (Bloom and Sachs 1999). Climatic and ecological characteristics render the region especially susceptible to infectious and parasitic diseases. As a result, countries in sub-Saharan Africa have consistently exhibited some of the highest mortality rates in the world since comparable records became available in the 1950s (Bloom and Sachs 1999; Acemoglu, Johnson, and Robinson 2001; Iliffe 2007).

The alleviation of geographical constraints on urbanization began in the colonial era. In the early colonial period the slave trade, violent conflicts, the introduction of foreign pathogens, and the disruption of traditional systems of production and networks of trade contributed to a shrinking of the region’s population (Iliffe 2007). After World War I, however, colonial governments began to invest more heavily in primary commodity production, launched health campaigns to combat epidemic diseases, expanded transport infrastructure, and introduced new agricultural technologies and cultigens such as cassava, which is drought resistant and has become an important anti-famine crop across Africa (Iliffe 2007; Clapham 2006). While these changes

---

**FIGURE 3** Levels of urbanization by major world regions, 1850–2050: estimates and projections

![Graph showing levels of urbanization by major world regions, 1850–2050](image)

collectively improved surplus availability and stimulated a secular decline in mortality rates, urbanization remained limited because of colonial restrictions on African mobility and residence in urban areas, poor urban living conditions, and limited wage employment opportunities.

After World War II the colonial powers, especially Britain and France, changed tack and launched a “modernization” drive designed to prepare colonies for eventual independence. This involved a significant expansion of public education and health services, further infrastructure development, and limited industrial investments (Cooper 2002; Iliffe 2007). Vaccination schemes led to sharp reductions in mortality associated with epidemic diseases; child mortality rates began to fall because of better treatments for polio, measles, diarrhea, and malnutrition; and improved road and rail transport contributed to reductions in famine-related mortality by rendering affected areas more accessible to emergency aid (Iliffe 2007). As mortality rates fell, Africa’s population began to grow rapidly. The gradual relaxation of government restrictions on African mobility, coupled with higher demand for labor in urban areas during and immediately after World War II, accelerated rural-to-urban migration. Rapid urban expansion exacerbated poor housing conditions and led to consumer price inflation and unemployment—factors that proved instrumental in catalyzing the growth of labor unions, which played a pivotal role in securing independence in the region (Cooper 2002; Iliffe 2007).

In sum, Africa’s urban transition was set in motion by technologies and institutions introduced in the late colonial period that facilitated mortality decline (and a subsequent population boom in the region) and increased the availability of surplus food supplies. However, the nature and impact of colonialism varied widely across countries, and this variation provides a means of assessing the relative impact of colonialism on urbanization. A plausible hypothesis is that those countries in which colonial powers were more economically and politically assertive experienced higher degrees of technological and institutional transfer and diffusion, thereby creating more favorable conditions for urbanization and urban growth. To test this hypothesis, I use capital investment during the colonial period as an indicator of colonial influence.

Figure 4 presents evidence that variation in levels of colonial capital investment in sub-Saharan African territories is correlated with variation in early urbanization in the region. The X-axis represents log-transformed values of the total amount of publicly listed capital invested in individual European colonial territories (in thousands of pounds sterling) between 1870 and 1936, as catalogued by Frankel (1969). The Y-axis represents log-transformed values of the total urban population of each corresponding territory in 1950. Although only 20 observations are available, the figure shows a clear correlation between early colonial investment and early urbanization.\(^\text{12}\)

Additional evidence of colonial influence on patterns of early urbanization in the region can be drawn from two other indicators: the relative degree
of “indirect rule,” based on legal records, and a measure of administrative depth, proxied by proportions of civil servants in the population. An analysis of their effects is set out in the Appendix.

While no single piece of this statistical mosaic provides definitive confirmation that urbanization in Africa was historically inhibited by unfavorable geographic endowments and was ultimately set in motion by technologies and institutions introduced by European colonizers, collectively the evidence provides significant support for these arguments.

Urbanization and urban growth in the postcolonial era

The growth in urban populations that began in the late colonial period accelerated in the independence era owing to a confluence of demographic, political, and economic factors. The mortality decline that began in the late colonial era continued, while fertility rates remained exceptionally high, resulting in population increase of historically unprecedented scale. Many countries experienced a surge in rural-to-urban migration in the early independence period due to the elimination of residence restrictions on Africans in urban areas and a sharp increase in urban employment opportunities at-
tributable to the expansion and Africanization of civil service administrations and investments in urban public works (Miner 1967; Stren and Halfani 2001; Iliffe 2007). I refer to this phenomenon in the statistical analysis below as a “postcolonial adjustment effect.” And economic growth rates in the region reached historic highs, fueled by a boom in public and private investment (much of it provided by international actors) and strong growth in commodity exports (ibid.). As a result, rates of urbanization and urban population growth reached exceptionally high levels between 1960 and 1975 (see Table 3).

Subsequently, unsustainable fiscal expansion, poor macroeconomic management, deteriorating terms of trade, and a global recession following the 1973 oil price shock resulted in a region-wide economic crisis. By the early 1980s sub-Saharan Africa was experiencing negative per capita income growth and fiscal retrenchment in the form of donor-imposed structural adjustment programs. The consequences in urban areas were severe. Public- and private-sector employment contracted sharply, real wages declined, investments in housing and urban infrastructure came to a virtual standstill, and the rural/urban wage gap that arose in the early independence era essentially vanished (Potts 1995; Weeks 1995; Becker and Morrison 1995). Yet urbanization and urban population growth rates remained generally high in Africa, with a few notable exceptions. This can be explained by continued mortality decline and by steady surplus expansion sustained by imports and aid. Figure 5 shows that surplus food supplies (measured in tons of cereals and starchy roots) generally kept pace with urban population growth, even in the crisis years of the 1980s and early 1990s. It also shows that this growth was made possible by imports more than by productivity growth.

Perhaps the most well-known exception to this narrative is Zambia, which experienced de-urbanization in the 1980s and 1990s. This anomaly is generally attributed to a severe economic downturn and the effects of structural adjustment on urban livelihoods (Potts 1995), yet many other countries experienced similar crises without de-urbanization. However, Zambia also experienced a sharp decline in food supplies and a reversal in the trend of declining mortality beginning in the late 1970s as a result of a crisis in the public health sector, deteriorating nutrition, and increases in mortality related to these factors. The table below provides a summary of demographic and economic trends in Africa from 1960 to 2005:

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Demographic and economic trends in Africa, 1960–2005 (percent per annum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urban population growth</td>
<td>5.1</td>
</tr>
<tr>
<td>Rate of urbanization</td>
<td>2.5</td>
</tr>
<tr>
<td>GDP per capita growth*</td>
<td>2.1</td>
</tr>
<tr>
<td>Population growth</td>
<td>2.5</td>
</tr>
</tbody>
</table>

*Estimates based on real GDP per capita (constant 2000 US$).
Source: Calculated from World Bank World Development Indicators Database, accessed January 2012.
to malaria and HIV/AIDS in particular (Dyson 2003; Garenne and Gakusi 2006). Moreover, under-five mortality rates rose much faster in urban than in rural areas between the mid-1970s and early 1990s and fell faster (from the mid-1990s) in rural areas than in cities (Garenne and Gakusi 2006). As Figure 6 illustrates, de-urbanization in Zambia was preceded by an abrupt and sustained contraction in per capita food supplies and was accompanied

**FIGURE 5** Urban population growth and food supply in sub-Saharan Africa, 1961–2000

![Graph of urban population growth and food supply](image)

**NOTE:** Food supply data include both domestic production and imports of cereals and starchy roots.

**SOURCES:** Food supply data from FAOSTAT online database, accessed June 2010. Population figures from the United Nations.

**FIGURE 6** Food supply, mortality, and percent urban in Zambia, 1965–2000: A case of de-urbanization

![Graph of food supply, mortality, and percent urban](image)

**NOTE:** See Figure 5.

**SOURCES:** See Figure 5.
by rising mortality rates. While some of the factors that contributed to rising mortality rates were a direct consequence of Zambia’s economic failures, others (such as rises in malaria and HIV/AIDS mortality) were independent of it (ibid.). In other words, Zambia experienced a unique combination of misfortunes that led to the resurgence of the surplus and disease constraints that inhibit urbanization—and in this case resulted in a rare episode of deurbanization.

To summarize, Africa’s urban transition in the postcolonial period was driven by a combination of rapid population growth set in motion in the late colonial period, a postcolonial adjustment involving the Africanization and expansion of employment opportunities in urban areas, and early international aid and investment. Through the recessionary years of 1975–1990 and the slow-growth recovery of the 1990s, the transition was sustained by persistent demographic expansion. In other words, both urbanization without economic growth and exceptionally high urban growth rates in the late twentieth century are explicable once Africa’s unique postwar political and population dynamics are taken into account.

As a test of this argument, and of the broader theory of urbanization outlined above, Table 4 presents the results of an OLS regression analysis in which average annual rates of urbanization and urban growth are modeled as a function of 1) average annual rates of population growth, 2) average annual rates of per capita income growth, and 3) the sectoral composition of output, measured as the average percentage contribution of agriculture to GDP over the relevant period. According to the theory outlined above, population growth and economic growth rates should both be positively correlated with rates of urbanization and urban growth, while agriculture as a percentage of GDP should be negatively correlated (as a result of labor being retained in the rural sector). The model also incorporates a dummy variable for sub-Saharan African countries to determine whether they share some unobserved characteristics that account for persistent urbanization in the absence of economic growth and for exceptionally high urban population growth rates.

The data consist of an unbalanced panel dataset with 353 observations from over 150 countries spanning three 15-year intervals that roughly correspond to global economic trends (1960–75, 1975–90, and 1990–2005) in order to limit the influence of short-term fluctuations in economic and demographic conditions. This arrangement also permits the inclusion of two more control variables: interactive regional dummies to determine whether there was a significant postcolonial adjustment effect (see Fay and Opal 2000). These are AFRICA*P1 and AFRICA*P2, where P1 and P2 represent dummy variables for the 1960–75 and 1975–90 periods respectively. Finally, the initial level of urbanization in each period is controlled for.

As expected, population and per capita GDP growth rates are both positively and significantly correlated with urbanization and urban growth rates,
while the contribution of agriculture to GDP is negatively and significantly correlated. The Africa dummy is insignificant; however, the period interaction dummies are both positive and significant. The fact that the coefficient on the first interactive dummy (AFRICA*P1) is larger than that of the second (AFRICA*P2) is indicative of the postcolonial adjustment effect noted in the historical narrative presented above. The models have near-identical coefficients because of the naturally strong correlation between rates of urban population growth and rates of urbanization.

The design of this model precludes definitive statements about causation: it merely confirms contemporaneous correlation between the variables of interest. It is possible that omitted variables or reverse causality is driving the results. However, there is nothing in the quantitative literature to suggest that significant determinants of urbanization or urban growth are omitted from these models, and there is little reason to suspect reverse causality. Higher levels of urbanization are generally associated with lower fertility rates, hence lower population growth rates. Consequently, if rates of urbanization or urban growth were having an effect on population growth rates, it would most likely

### Table 4: Determinants of urbanization and urban growth rates, 1960–75, 1975–90, and 1990–2005

<table>
<thead>
<tr>
<th></th>
<th>Urbanization rate</th>
<th>Urban growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population growth</td>
<td>.139***</td>
<td>1.139***</td>
</tr>
<tr>
<td>Per capita GDP growth</td>
<td>.126***</td>
<td>.126***</td>
</tr>
<tr>
<td>Agriculture (% of GDP)</td>
<td>−.015***</td>
<td>−.015***</td>
</tr>
<tr>
<td>AFRICA</td>
<td>.058</td>
<td>.056</td>
</tr>
<tr>
<td>AFRICA*P1</td>
<td>1.509***</td>
<td>1.510***</td>
</tr>
<tr>
<td>AFRICA*P2</td>
<td>.689***</td>
<td>.690***</td>
</tr>
<tr>
<td>Urbanization₁</td>
<td>−.469***</td>
<td>−.469***</td>
</tr>
<tr>
<td>R-squared</td>
<td>.586</td>
<td>.821</td>
</tr>
<tr>
<td>Observations</td>
<td>353</td>
<td>353</td>
</tr>
</tbody>
</table>

**Notes:** Standard errors in parentheses. Significance at the 1, 5, and 10 percent levels is indicated by ***, **, * respectively. Urbanization₁ refers to the level of urbanization in a given country at the beginning of the 15-year period examined (i.e., 1960, 1975, or 1990).

**Sources:** See Table A.2.
be a negative effect. Similarly, all of the existing empirical evidence suggests that GDP growth drives urbanization, and there is no evidence that levels of urbanization affect GDP growth rates (Bloom, Canning, and Fink 2008).

Nevertheless, a second test examining the determinants of changes in levels of urbanization and urban population size serves as a robustness check. In this case, the model is designed to test whether lagged values of the independent variables of interest add predictive power to a model that includes lagged values of the dependent variables as independent variables. I examine whether population and GDP growth rates between time $t_1$ and time $t_2$ help to predict levels of urbanization and urban population size at time $t_2$ when levels of urbanization and urban population size at time $t_1$ are controlled for. As in the previous model, the data are divided into three 15-year periods. Given that past levels of urbanization and urban population size explain over

| TABLE 5 Determinants of changes in relative and absolute size of urban populations, 1960–2005 |
|-----------------------------------------------|-----------------------------------------------|
| Population growth$_{t_1 \rightarrow t_2}$ | .082*** (.017) |
| Per capita GDP growth$_{t_1 \rightarrow t_2}$ | .049*** (.010) |
| Agriculture (% of GDP)$_{t_1}$ | -.008*** (.002) |
| AFRICA | .057 (.077) |
| AFRICA*P1 | .254** (.101) |
| AFRICA*P2 | .197** (.093) |
| Urbanization$_{t_1}$ | .881*** (.020) |
| Log of urban population$_{t_1}$ | .804*** (.019) |
| Log of national population$_{t_2}$ | .188*** (.020) |
| R-squared | .953 |
| Observations | 353 |

NOTES: Standard errors in parentheses. Significance at the 1, 5 and 10 percent levels is indicated by ***, **, * respectively. Urbanization$_{t_1}$ refers to the level of urbanization in a given country at the beginning of the 15-year period examined (i.e., 1960, 1975, or 1990). The same principle applies to the variables “Log of urban population” and “Log of national population.”

SOURCES: See Table A.2.
90 percent of variation in contemporary levels, the independent variables of interest must be robustly correlated with the dependent variables to exhibit statistical significance. The results of this test are presented in Table 5.

The results again are consistent with the hypotheses: population growth and economic growth continue to exhibit positive and significant effects on levels of urbanization and urban population size, while the share of agriculture in GDP remains negative. There is also no evidence of an “Africa” effect, but there is significant evidence of a postcolonial adjustment effect as captured by the two interaction dummies. The fact that these results echo those of the previous model lends further support to the theory outlined above. Moreover, given that levels of urbanization and urban population size at time \( t_2 \) cannot logically have influenced rates of urbanization or GDP growth in the previous 15-year period, this model offers further evidence of a causal relationship between population growth and GDP growth (on the one hand) and urbanization and urban population growth (on the other).

**Conclusion**

Urbanization should be viewed as a global historical process propelled by technological and institutional changes that alleviated the surplus and disease constraints that limited urban population growth in the preindustrial era. These changes initially emerged in Europe and subsequently spread, albeit unevenly, through conquest and trade. This historically grounded view of urbanization stands in contrast to the traditional view that urbanization is a byproduct of industrialization. While it is true that mortality decline and expanded access to surplus food supplies—prerequisites for urbanization—often go hand-in-hand with economic development, they do not always do so.

In the case of sub-Saharan Africa, colonizers introduced key technological and institutional innovations that alleviated geographically determined surplus and disease constraints, which had previously inhibited urban population growth in the region. Colonial influence varied significantly across countries within the region, and this variation accounts for a significant fraction of variation in patterns of early urbanization across countries.

In the years after World War II, gains in life expectancy and increased access to surplus food supplies occurred more rapidly than economic development in sub-Saharan Africa. As a result, many countries in the region experienced urbanization without economic growth. Moreover, rapid mortality decline coupled with minimal fertility decline led to population increases of historically unprecedented proportions—increases that largely account for the extraordinary urban growth rates in the region.

There is, in short, nothing particularly unusual about Africa’s urban transition when viewed through the historical lens outlined here. The implication of this theory, from a practical policy perspective, is that the process of urbanization cannot be restrained. For governments interested in alleviat-
ing demographic pressure in urban areas, the only humane policy option is one targeted at encouraging fertility decline in order to reduce population growth rates. While fertility decline is a natural consequence of urbanization, targeted interventions such as family planning initiatives may serve to accelerate the process and ease the social strains associated with rapid urban population growth.

Appendix

This appendix offers further evidence of colonial influence on urbanization. Table A.1 presents bivariate correlations between two indicators of colonial influence (the independent variables) on the one hand and variables related to disease control, food security, and early urbanization (the dependent variables) on the other. In the upper half of the table, the independent variable is a measure of “indirect rule” in 33 former British colonies worldwide developed by Lange (2004). The indirect rule index reflects the percentage of legal cases settled by traditional authorities as opposed to formal courts in 1955. The higher the percentage, the greater the extent to which British authorities relied on local powerbrokers to maintain order in their territories. Using this index, Lange (2004) demonstrated that higher degrees of colonial indirect rule resulted in less effective public institutions in the postcolonial period. In the lower half of the table, the independent variable is the number of colonial civil servants per capita in British and French territories in Africa circa 1936, as calculated by Richens (2009). This serves as a proxy for administrative depth. As with capital investment, it is reasonable to suppose that higher degrees of direct political rule and administrative depth resulted in greater technological and institutional transfer and diffusion, hence more favorable conditions for early urbanization.

The dependent variables in Table A.1 include the infant mortality rate in 1960, food supply in 1960 (measured in calories per capita per day), the number of registered physicians per thousand population around 1960, the average annual rate of change in the infant mortality rate in the 15 years after 1960, the level of urbanization in 1960, and total urban population in 1960. Controls (apart from the inclusion of national population in column 6) are omitted because of the small sample sizes of the colonial data.

The results indicate that higher degrees of indirect rule are associated with higher infant mortality, smaller food surplus, and fewer physicians per capita in 1960, (columns 1a–3a), while greater administrative depth is associated with lower infant mortality rates, greater food surplus, and more physicians per capita in 1960 (columns 1b–3b). The direction of causality in these correlations is not necessarily clear. As Acemoglu, Johnson, and Robinson (2001) have argued, patterns of colonial settlement may have been influenced by the region’s disease environment and agricultural potential, so these correlations could be interpreted as indicative of reverse causality. In other words, colonial rulers may have invested economically and politically in territories with lower disease burdens and greater agricultural potential.

On the other hand, there is no theoretical reason to believe that mortality decline following independence was driven by anything other than the further diffusion of technologies and the consolidation of institutions introduced during the colonial
period that affect public health. As column 4 shows, African countries that had more robust colonial legal institutions and greater colonial administrative capacity experienced more rapid declines in infant mortality in the early postcolonial period. In this case the direction of causality is clear: postcolonial changes in mortality cannot logically have driven colonial settlement patterns. Finally, columns 5 and 6 confirm the association demonstrated above between colonial capital investment and early urbanization: the indicators of indirect rule and administrative depth are both significantly correlated with the size and percentage of a country’s urban population in 1960.

### TABLE A.2 Data sources

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of urbanization</td>
<td>Percentage of national population residing in urban areas</td>
<td>United Nations (2010)</td>
</tr>
<tr>
<td>Log of urban pop</td>
<td>Log of the absolute size of a country’s urban population in 1000’s</td>
<td>United Nations (2010)</td>
</tr>
<tr>
<td>Log of national pop</td>
<td>Log of the absolute size of a country’s population in 1000’s</td>
<td>United Nations (2010)</td>
</tr>
<tr>
<td>World population (Fig. 1)</td>
<td>Population in millions</td>
<td>Maddison (2009)</td>
</tr>
<tr>
<td>Historical urbanization estimates (Fig. 1 and 3)</td>
<td>Percentage of world population residing in urban areas</td>
<td>Grauman (1977)</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
<td>Source</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>--------</td>
</tr>
<tr>
<td>Log of coastline (km)</td>
<td>Log of kilometers of coastline</td>
<td>CIA World Factbook (various years)</td>
</tr>
<tr>
<td>Log of waterways (km)</td>
<td>Log of navigable waterways</td>
<td>CIA World Factbook (various years)</td>
</tr>
<tr>
<td>Soil quality</td>
<td>Percentage of land area with soil very or moderately suitable for 6 key rainfed crops</td>
<td>Gallup, John L. and Jeffrey D. Sachs, with Andrew Mellinger, “Geography and economic development” (CID Working Paper No. 1, March 1999). Available at: «<a href="http://www.cid.harvard.edu/ciddata/ciddata.html%C2%BB">http://www.cid.harvard.edu/ciddata/ciddata.html»</a></td>
</tr>
<tr>
<td>GDP per capita</td>
<td>Real GDP per capita in 1990 international Geary–Khamis dollars</td>
<td>Maddison (2009)</td>
</tr>
<tr>
<td>Colonial investment, 1870–1936</td>
<td>Sum of publicly listed capital invested in African colonial territories between 1870 and 1936, calculated in pounds sterling</td>
<td>Frankel (1969)</td>
</tr>
<tr>
<td>Indirect rule</td>
<td>Percentage of legal cases adjudicated by “traditional” authorities in British colonies, 1955</td>
<td>Lange (2004)</td>
</tr>
<tr>
<td>Calories per capita</td>
<td>Average available calories per capita per day in 1960</td>
<td>Food and Agriculture Organization, available at «<a href="http://faostat.fao.org%C2%BB">http://faostat.fao.org»</a></td>
</tr>
<tr>
<td>Physicians per 1000</td>
<td>Physicians per 1000 population around 1960. Missing values were replaced by nearest available year</td>
<td>World Bank World Development Indicators online, accessed June 2011. Available at: «<a href="http://data.worldbank.org/data-catalog/world-development-indicators%C2%BB">http://data.worldbank.org/data-catalog/world-development-indicators»</a></td>
</tr>
<tr>
<td>Food supply (Fig. 5)</td>
<td>Production and import volume of cereals and starchy roots in tonnes</td>
<td>Food and Agriculture Organization, available at «<a href="http://faostat.fao.org%C2%BB">http://faostat.fao.org»</a></td>
</tr>
<tr>
<td>Food supply (Fig. 6)</td>
<td>Available tonnes per capita of cereals and starchy roots</td>
<td>Food and Agriculture Organization, available at «<a href="http://faostat.fao.org%C2%BB">http://faostat.fao.org»</a></td>
</tr>
<tr>
<td>Crude death rate (Fig. 6)</td>
<td>Deaths per 1000 population</td>
<td>United Nations, World Population Prospects: the 2008 Revision (2009)</td>
</tr>
</tbody>
</table>
Notes

Funding for this research was provided by the UK Department for International Development and the Crisis States Research Centre. I thank Jo Beall, Tim Dyson, Jean-Paul Faguet, John Flynn-York, Tom Goodfellow, Elliott Green, and Su Lin Lewis for useful comments on earlier drafts.

1 The theory assumes that higher urban wages reflect rising demand for labor, as well as higher marginal returns to labor, in an expanding modern sector. The wage gap stimulates migration among rational individuals seeking to maximize their incomes until the labor market clears.

2 The lack of comparable data on informal-sector wages precludes the possibility of accurately determining the influence of wage differentials between rural areas and the urban informal economy.

3 The reclassification of rural areas as urban also contributes to urbanization in a statistical sense, but reclassification is logically a byproduct of natural population increase and migration, so the focus here is on these factors.

4 Dyson (2011) provides evidence that urban natural increase outpaced rural natural increase around the turn of the twentieth century in Sweden and again in the immediate aftermath of World War II, but these are exceptional cases of relatively brief periods of urbanization without migration.

5 All population, urban population, and GDP estimates used in this article were log transformed to normalize sample distributions. Similarly, all urbanization estimates were square root transformed. In this model, coastline and waterways data were also log transformed to correct for skewness. For details of all variables and sources see Table A.2.

6 Angus Maddison’s historical estimates of world population and GDP indicate that per capita output increased from $US412 in 1 CE to just $606 in 1700 (expressed in 1990 International Geary–Khamis dollars). From this one can infer that factor productivity growth was very limited over this period. In other words, there was very little increase in energy surplus per capita.

7 The bulk of world trade in the preindustrial era involved relatively lightweight, nonperishable, and high-value items such as spices and luxury textiles (see Braudel 1984; Bairoch 1988).

8 Whether growth in surplus energy supplies is a sufficient condition for urbanization is difficult to answer definitively. History indicates that wherever surplus energy becomes available, cities form, suggesting that the very presence of the surplus is sufficient to spur agglomeration. To verify this empirically is virtually impossible.

9 In a cross-country statistical study, Bairoch and Goertz (1986) demonstrated that the pace of urbanization in nineteenth-century Europe was driven primarily by changes in agricultural productivity, by the pace of industrial growth, and by the expansion of trade. However, they found evidence that the most important factors driving urbanization varied over time within and between countries. In particular, their results concerning the role of agricultural productivity were ambiguous. In some models the coefficient was positive, suggesting that rising output facilitated urbanization, while in others the coefficient was negative. They speculate that agricultural success in certain regions resulted in the retention rather than release of rural labor.

10 Surplus food expansion contributes to mortality decline through improved nutrition, which is an important determinant of variation in disease-related morbidity and mortality rates.

11 Urbanization is a finite process. As a result, a country’s rate of urbanization naturally decreases as its level of urbanization increases. By contrast, there is technically no upper bound on urban growth. This explains continued urban growth in some fully urbanized countries such as many in South America, which continue to experience fertility rates that exceed the replacement rate of 2.1.

12 The direction of causality assumed in Figure 4 could be challenged by the argument
that colonial rulers invested more where African populations were larger (hence offered a larger pool of labor to exploit). Given that total population and urban population are highly correlated, this is a potentially valid criticism. However, given the paucity of pre-colonial cities in the region, it is unlikely that Europeans invested in areas with existing urban populations, so a causal link between colonial investment and urban population size is reasonable. Moreover, total population and level of urbanization are not at all correlated. In results not reported here, Frankel’s colonial investment figures were also found to be negatively correlated with crude death rates in 1950 and positively correlated with income per capita in 1950, consistent with the hypothesis that colonial investment had the dual effect of reducing mortality and improving access to surplus food supplies.

13 This lag between mortality decline and fertility decline is usually explained as a consequence of historical factors that have made high birth rates culturally desirable in the region (see Iliffe 2007 and Clapham 2009).

14 This approach follows the logic (although not the exact form) of a “Granger causality test” (see Granger 1969; Bloom, Canning, and Fink 2008). A variable X can be said to “Granger-cause” Y if X at time t−1 (Xt−1) explains variation in Y at time t (Yt) when Y at time t−1 (Yt−1) is included as a control variable on the right-hand side of the equation.

References


