The Urban Mortality Transition
and the Rise of Poor Mega-Cities

By REMI JEDWAB AND DIETRICH VOLLRATH*

Draft: March 11, 2016

Today the world’s largest cities lie in poor countries, unlike the historical norm. We use novel historical data to document that these poor mega-cities arose only after the post-war mortality transition, and develop a general equilibrium model of location choice with heterogeneity in congestion costs and demographics across locations to account for this. We calibrate the model to data from a sample of developing countries, and show that the urban mortality transition accounted for one-quarter of urbanization and half of slum growth post-1960. Simulations show that family planning is more effective than migration restrictions and infrastructure in slowing poor mega-city growth.

JEL: O11; O40; O18; R11; R12; R13; J10; N00
Keywords: Mega-Cities; Structural Change; Urbanization; Slums; Congestion; Urban Mortality; Demographic Regime; Growth

* Jedwab: Department of Economics, George Washington University, jedwab@gwu.edu. Vollrath: Department of Economics, University of Houston, devollra@central.uh.edu. We would like to thank Quamrul Ashraf, Hoyt Bleakley, Filipe Campante, Gilles Duranton, Edward Glaeser, Jessie Handbury, Mariaflavia Harari, Vernon Henderson, Erik Hurst, Aart Kraay, Craig McIntosh, Stelios Michalopoulos, Giacomo Ponzetto, Diego Puga, Paul Romer, Esteban Rossi-Hansberg, Rodrigo Soares, Anthony Venables, Nico Voigtlaender, Hans-Joachim Voth, David Weil and seminar audiences at George Mason, Georgetown, George Washington, FGV-EESP, Los Andes, Maryland, Michigan, New South Wales, NYU, OCDE, Oxford, CEPR Symposium on Development Economics (UPF), Pompeu Fabra-CREI, PUC-Rio, Rosario, Southern Economic Association Meetings (New Orleans), United Nations-ESCAP, Urban Institute, USAID, U.S. Department of State, Wharton, Williams, World Bank-DECRG, World Bank-George Washington University Urbanization Conference and World Economic History Congress (Kyoto) for helpful comments. We thank the Institute for International Economic Policy and the Elliott School of International Affairs (SOAR) at George Washington University for financial assistance.
Urbanization has gone hand in hand with economic growth throughout history. However, the post-war period witnessed “poor country urbanization” (Glaeser, 2014) that has generated poor mega-cities in developing nations. Dhaka, Karachi, Kinshasa, Lagos, and Manila are some of the largest cities on the planet today. In contrast, only six of the currently largest 30 cities (London, Los Angeles, New York, Osaka, Paris, and Tokyo) are in high income countries. The prevalence of poor mega-cities today runs counter to historical experience. In the past, the world’s largest cities were in the most advanced economies. Today’s mega-cities are unlike their historical peers in that their size is not indicative of prosperity.

We first document this puzzling phenomenon using the largest dataset ever built on the spatial aspects of the demographic transition from antiquity to modern times, and show that the poor mega-cities of the post-war era experienced rates of natural increase - the birth rate minus the death rate - well above levels seen in historic mega-cities. This was due to the severe drop in urban death rates following the urban mortality transition of the mid-20th century. Poor mega-cities grew in absolute terms because of both in-migration and natural increase, setting them apart from historical “killer” cities that grew only through in-migration. Beyond their absolute pace of growth, we further show that poor mega-cities today are very densely populated, with much of this density due to a high share of population in slums. These poor mega-cities also have low human capital, with high dependent population shares and low education completion rates.

To explain the rise of the poor mega-cities, we construct a general equilibrium model of labor allocation across locations, similar in spirit to work on structural change (Caselli and Coleman II, 2001; Gollin, Parente and Rogerson, 2002; Duarte and Restuccia, 2010; Buera and Kaboski, 2012; Herrendorf, Rogerson and Valentinyi, 2013; Lagakos and Waugh, 2013). We incorporate locational choice as in the literature on equilibrium city size (Henderson, 1974; Duranton and Puga, 2004; Desmet and Rossi-Hansberg, 2013) and combine that with elements of Malthusian models of population and growth (Galor and Weil, 1999,
2000; Ashraf and Galor, 2011). Compared to Michaels, Rauch and Redding (2012), we incorporate endogenous aggregate population growth, and distinguish between different types of urban locations. Our model allows population to move freely among locations that differ in their demographics, productivity growth, agglomeration effects, and congestion costs. We concentrate on three types of locations so that we can match the model to available data. In addition to rural locations, we specify two types of urban locations: formal (e.g. business districts) and informal (e.g. slums). Formal locations have high productivity growth, but are severely affected by congestion. Informal locations have low productivity growth but small congestion costs. Because of urban congestion, there is a negative effect of population growth on welfare - a feature of Malthusian models - regardless of whether the economy is heavily agricultural or not.

A key insight of the model is that the aggregate population growth rate directly influences the allocation of population across the three types of locations. In particular, an increase in population growth (whatever the spatial source) leads to an increase in the population share of low-congestion locations. The urban mortality transition had the direct effect of raising population growth, and holding the composition of the population constant, this caused absolute urban populations to rise. But in addition, the rapid population growth disproportionately increased the size of low-congestion informal urban locations. Hence the urban mortality transition accelerated urbanization, and also ensured that urbanization took place through an expansion of informal locations such as slums. The growth of slums then fed back into lower welfare by lowering average productivity growth and increasing population growth. Poor mega-cities are, in part, the result of the successes of the urban mortality transition.

To quantify the effects of the urban mortality transition, we calibrate to a sample of 41 poor countries observed from 1950 to 2005 with data on urbanization, population growth, and slum shares. We find that the urban mortality transition accounted for one-quarter of urbanization, and half of slum growth post-1960.
In addition, urban populations in our sample would be only two-thirds of their current size and welfare would be about 20% higher without the urban mortality transition.¹

Using our calibration, we then forecast urban populations and welfare forward over time. Without significant declines in population growth, poor countries will continue to see rapid expansion of urban areas, and this will occur mostly through the growth of informal urban locations. The importance of demographics for the future prospects of poor mega-cities is also apparent when we examine the effects of various policies. By far the most effective policies are family planning programs that reduce birth rates. These arrest the growth of cities, the size of the informal population, and enhance welfare. Restricting migration into cities generally, or specifically into informal urban areas (e.g. slum clearance), does little beyond lowering welfare by forcing people to remain in rural areas. Building infrastructure in existing cities or creating new large cities from scratch do not appear to be effective policies given the high costs of such policies.

This paper adds to the literature on structural change and urbanization in several ways. First, while there is some work on urbanization without growth (Fay and Opal, 2000), little attention has been given to the characteristics of poor mega-cities (see Duranton (2014), Desmet and Henderson (2014) and Jedwab and Vollrath (2015) for recent surveys which reference poor mega-cities briefly). To our knowledge, only Desmet and Rossi-Hansberg (2014) address this important issue for the development process. Second, our paper connects with the recent literature on optimal city size (Albouy, 2008; Albouy et al., 2015; Eeckhout and Guner, 2015). However, with the exceptions of Au and Henderson (2006) and Desmet and Rossi-Hansberg (2014), this literature focuses on developed countries. In addition, these papers only compare cities of different sizes, rather than considering the whole distribution of population, including the rural areas. Third, with the exception of Jedwab, Christiaensen and Gindelsky (2015), the

¹We do not attempt to put a number on the direct welfare gain from increased life expectancy.
literature has largely ignored the role of population growth in “poor country urbanization”. Yet, this explanation appears to be quantitatively important. Fourth, the literature has focused on studying agglomeration effects, and says little about congestion in poor mega-cities. A few exceptions are Desmet and Rossi-Hansberg (2013, 2014), Duranton (2015) and Hanlon and Tian (2015). We also document that these poor mega-cities have high dependency ratios and are low-skilled, in contrast to the literature that emphasizes that cities help disseminate knowledge (Glaeser et al., 1992; Moretti, 2004; Gennaioli et al., 2013).

Our work also adds to the literature on the economic effects of demography. Population growth promotes economic growth if high population densities encourage human capital accumulation or technological progress (Becker, Glaeser and Murphy, 1999; Galor and Weil, 2000; Lucas, 2004; Desmet, Greif and Parente, 2015). But the presence of a fixed factor of production (e.g. agricultural land) implies that living standards are inversely proportional to population size. Our work shows that such “Malthusian” pressures arise not only because of limits on natural resources, but also due to congestion effects in urban areas. Significantly, this implies that Malthusian forces need not disappear just because economies experience structural change and urbanize.

The poor mega-city outcome is one that arises perversely because of the success of interventions that limited urban mortality rates while urban fertility remained relatively high. In this, our work is similar to others that emphasize the

---

2 Jedwab, Christiaensen and Gindelsky (2015) study the correlations between natural increase and urbanization for a restricted sample of developing countries from 1750 to date. They focus on the growth of the urban population as a whole relative to the total population, hence *urbanization rates*, while here we look specifically at the absolute population growth of cities, hence *poor mega-cities*. Their work is also purely descriptive, while our quantitative analysis allows us to characterize the trajectory of these poor mega-cities, and identify the policies that may alleviate the problem of poor mega-city growth.

3 Our explanation does not rule out other explanations, such as urban bias (Ades and Glaeser, 1995; Davis and Henderson, 2003; Henderson, 2003), natural disasters (Barrios, Bertinelli and Strobl, 2006; Henderson, Storey, Deichmann, 2013) and resource exports (Gollin, Jedwab and Vollrath, 2015).

4 The literature on the relationship of population and economic growth is vast (see Galor (2012) for a survey). Barro and Becker (1989); Becker, Murphy and Tamura (1990) and Manuelli and Seshadri (2009) provide models of the negative relationship of income and fertility. Unified growth models depend on a rise in the demand for human capital to induce sustained growth, which is driven by acceleration in technological change (Galor and Weil, 1999, 2000; Ashraf and Galor, 2011). This demand may also be driven by gender-biased technological change (Galor and Weil, 1996), the decline in benefits of child labor (Doepke and Zilibotti, 2005), improvements in health (Cervellati and Sunde, 2005; Soares, 2005), trade (Galor and Mountford, 2006, 2008), culture (Tertilt, 2005), or structural change (Vollrath, 2011).
negative economic effect of mortality interventions and/or the positive economic impacts of mortality increases (see Young (2005) and Acemoglu and Johnson (2007) for prime examples).\(^5\) Our paper is directly related to Voigtländer and Voth (2013\(b\)), who show how the Black Death triggered a transition to a new equilibrium with higher wages in historical Europe. As urban mortality was high, increased urbanization due to income effects after the Black Death constrained aggregate population growth, allowing income to rise even further. In our analysis, population growth accelerates the growth of slums, and because of the urban mortality transition, slums have low mortality rates, high rates of natural increase, and low productivity growth, thus reducing living standards.\(^6\)

Lastly, the paper does not imply spatial misallocation, unlike the literature on the urban-rural income gap (Gollin, Lagakos and Waugh, 2013; Young, 2013) and the broader literature on misallocation and productivity (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013; Duranton et al., 2015). While we can allow for migration restrictions that generate a suboptimal wedge in wage growth between locations, in our baseline analysis low welfare comes from the fact that individuals do not internalize the negative externalities of their fertility decisions on congestion and living standards in all locations. Our simulations show that family planning is more effective than migration restrictions in slowing poor megacity growth.

In the next section we document the rise of poor mega-cities, and the underlying demographic features of these cities. We then present and calibrate our model, and calculate the impact of the urban mortality transition on the growth of poor

\(^5\)Our work is linked to studies that consider the effect of lower mortality on population and economic growth (Weil, 2007; Bleakley, 2007, 2010; Bleakley and Lange, 2009; Bloom, Canning and Fink, 2014) as well as those examining the effect of unexpected decreases in population (Voigtländer and Voth, 2013\(a\); Ashraf, Weil and Wilde, 2011). Relative to those works we consider the implications of mortality changes from the perspective of cities, and highlight how urban demographics matter for the development process.

\(^6\)The paper contributes to the related economic history literature. Other studies have focused on urban mortality in England or the U.S. in the 19th century (Williamson, 1990; Haines, 2008; Costa and Kahn, 2006; Voigtländer and Voth, 2013\(b\); Hanlon and Tian, 2015), whereas we track the demographic behavior of many cities, and their “surrounding” areas, for many countries and over a long period of time. For the post-war period, using the Demographic and Health Surveys and census information from IPUMS, we also provide information on measures of congestion at the city and country level.
mega-cities. The final section discusses the policy implications of the findings.

I. THE RISE OF POOR MEGA-CITIES

By “mega-cities” we mean the largest cities in the world. Table 1 shows the largest 30 cities in select years from 1700 to 2015. The mega-cities of 1700 were located in the most economically advanced countries in that period. While London and Amsterdam had wages that were relatively high then, cities such as Beijing, Istanbul, and Tokyo all had wages equivalent to those found in cities such as Paris and Naples (Özmucur and Pamuk, 2002; Allen et al., 2011).

By 1900, the nature of the list of largest cities changed along several dimensions. First, the absolute sizes were roughly ten times larger than in 1700. Second, the cities dominating this list were the leading cities from the richest countries. London, New York, and Paris are all found on the list. Further down, we see Boston, Liverpool, Manchester, and Philadelphia, all centers of industrialization. There were several large cities in poor countries: Beijing, Kolkata, and Mumbai. But none approached the size of the world leaders. Comparing 1900 to 1700, one can also see that their growth in this period was not on the same scale as the richer cities. Beijing and Istanbul increased only from 700,000 to about 1 million in the same 200 years that London went from 600,000 to 6.5 million.

In 1950 the top cities remained those in advanced nations, but we see the very beginnings of mega-city growth in poor countries. Kolkata and Shanghai both had more than four million inhabitants, putting them in the top 10 cities in the world in 1950. Beijing, Cairo, Mexico, and Mumbai were all over 2 million inhabitants.

In 2015, the nature of the largest cities has changed dramatically. First, the absolute scale of cities is 3-5 times larger than in 1950. Second, the composition of the list is now dominated by cities in developing countries. Only London, Los Angeles, New York, Osaka, Paris, and Tokyo are in rich countries. Instead we see cities such as Delhi, Mexico, Mumbai, and Sao Paulo. Low-income countries have cities present on this list, such as Dhaka (Bangladesh), Karachi (Pakistan), Kinshasa (DRC), Lagos (Nigeria), and Manila (Philippines). These cities have at
least 11 million people, making them larger in size than most mega-cities in 1950.

The shift of the world’s largest cities from rich to poor countries can be seen in Figure 1. This plots the number of cities with more than one million inhabitants for two groups, developed countries (based on their 2015 GDP per capita) and developing countries. In 1900 and 1950 nearly all million-plus sized cities were in currently developed nations. This switched between 1950 and 2015, however, and this reversal is projected to increase well into this century (2030).7

Further, there is evidence that city size in the past was a robust indicator of city-level living standards, but that relationship has broken down in more recent years. For the pre-1910 period, our data are for welfare ratios calculated using wages and price indices for minimal consumption baskets in different cities (see Web Data Appendix for details on data sources). The 111 observations are at the city-year level, so that for several cities we have multiple observations over time. We rank all these observations based on their absolute city size. We then rank all these observations based on their real wage, and plot the rank of wages against the rank of city size. Figure 2 shows there was a positive relationship historically (correlation of 0.62). This is not to say that the cities of industrializing Europe or North America had high absolute living standards. But city size indicated something regarding relative living standards at the time, as larger cities tended to have higher wages. For the modern period, we do not have real wage measures comparable to historical data. However, we do have a city product index for a sample of 111 cities of at least 300,000 inhabitants in 2010 (United Nations Habitat, 1998, 2012).8 For these observations we plot the rank of living standard against the rank of size, and find a correlation of only 0.28 (Figure 2).

To highlight the differences over time, several cities are shown in the figure. From 1700 to 1900 both Amsterdam and London shift to the top right, indicating

---

7 The patterns hold if we exclude China and India (Web Appx. Fig. 1) or focus on the proportion of population living in mega-cities and using other cut-offs than one million inhabitants (Web Appx. Fig. 2). Countries are also more and more urbanized at a constant income level (Web Appx. Fig. 3).

8 The product index is measured using such inputs as capital investment, formal employment, trade, savings, exports and imports, and household consumption. Details are available in the UN reports.
they were growing as they developed. We do not have living standard data for New York in 1700, but New York was also one the wealthiest and largest cities by 1900. From that point forward, however, these cities slipped down the rankings in size while maintaining their position in wages. In comparison are a number of currently poor mega-cities. Kolkata in 1825, Delhi in 1875, and Jakarta in 1900 are all relatively small, and relatively poor. Yet, they have all moved up to become among the largest cities of the world by 2010. However, this has not been associated with a move up in the rankings in living standards. Other poor cities such as Dacca, Kinshasa and Lagos are among the largest cities in the world.

This change in the composition of the largest cities is only going to be exacerbated in the future. The final column of Table 1 shows in parentheses the projected growth rate from 2015–2030 for each mega-city. The largest values are for poor mega-cities, such as Lagos (4.2%), Kinshasa (3.7), Dhaka (3.0), and Karachi (2.7). In comparison, rich mega-cities have growth rates close to zero.

II. THE CHARACTERISTICS OF POOR MEGA-CITIES

A. The Urban Mortality Transition and Mega-City Growth

Historically, in-migration was the dominant source of new city dwellers as the rates of natural increase were low in urban areas, typically because of high urban mortality rates. This can be seen in Figure 3, where we compare the crude birth rate (CBR) and crude death rate (CDR) of mega-cities across different historical eras. We used various sources described in Web Appendix Table 1 – population censuses, demographic surveys, sanitary reports, and historical studies – to reconstruct historical demographic data for a sample of the largest cities in the world, from antiquity to date. For the 263 city-period observations in our sample, we also obtained the same demographic data for the other cities of the same country-period, the rural areas, and the entire country.9

We first focus on the 100 largest cities as of 1900, according to Chandler (1987).

9For the oldest city-periods in our sample, we use the work of anthropometrists who studied graves and skeletons to obtain data on average fertility (i.e. birth rates) and life expectancy (i.e. death rates).
Panel A plots the birth and death rates for 49 of these cities before 1900 where we have data. The sample includes various pre-industrial cities such as ancient Rome, Teotihuacan, Renaissance Florence, and London in the 17th century. It also includes various emblematic cities of the Industrial Revolution, such as Boston, Liverpool and Manchester in the 1880s. As can be seen, birth rates were high on average (33.9 per 1,000 people), but death rates were also high (30.7). Mortality was indeed historically high in cities due to poor water quality, inadequate sewage and waste disposal, and dense areas favoring the spread of contagious diseases (Williamson, 1990; Haines, 2008; Costa and Kahn, 2006; Voigtländer and Voth, 2013b). As a result, the cities all lie near the 45-degree line, indicating that the cities experienced almost no natural increase (2.9, i.e. 0.3% per year). Their growth, which averaged 2.5% per year (Jedwab, Christiaensen and Gindelsky, 2015), could occur only through in-migration. Panel B then shows that cities remained near the 45-degree line at the turn of the 20th century, when using available data for 88 of the 100 largest cities in 1900. City natural increase was still low on average, at 6.1 per 1,000 people (0.6% per year).

In the post-war period, there was a distinct change in mega-city demographics. In Panels C and D, we study the 100 largest cities of the future (in 2030 according to the United Nations (2014)), and show their birth and death rates when available in the 1960s (N = 57) and the 2000s (N = 100) respectively. Focus first on the relatively rich cities in the lower left of Panel C, in the 1960s. Their birth rates fell along with their death rates, and so their rate of natural increase remained muted. Overall, it is apparent from Panels A to C that historically cities were “sliding down” the 45 degree line as they grew. In comparison are the nascent poor mega-cities in the upper left of Panel C, well above the 45 degree line. In the 1960s these

---

10 We were able to find data for some of the following industrializing cities in the 1800s and/or the 1820s and/or the 1850s and/or the 1880s: Amsterdam, Antwerp, Baltimore, Belfast, Berlin, Birmingham, Boston, Bristol, Brussels, Chicago, Dublin, Edinburgh, Glasgow, Hamburg, Leeds, Liege, Liverpool, London, Manchester, Munich, Newcastle, New York, Paris, Philadelphia, Rotterdam and Stockholm.

11 All the points in Panel A only represent “normal” periods, but each city was at times afflicted by very severe shocks to mortality. For example, during the Black Death in the 14th century, crude death rates of 250-750 (i.e., 25-75%) were seen across European cities (Jedwab, Johnson and Koyama, 2015).
cities differed from earlier eras in one distinct way: their death rates were very low. Dhaka, Karachi, Kinshasa and Lagos, despite being in countries with much lower levels of income per capita, all had death rates that were similar to those seen in London, New York, or Paris in the same year. Developing mega-cities in the 1960s were mainly “shifted left” compared to their historical peers. This can also be seen by comparing the respective locations of Cairo, Delhi and Mexico in Panels B and C. This led to very large rates of natural increase for emerging mega-cities. For example, in the African mega-cities in the figure, crude rates of natural increase were roughly 35, or 3.5% per year. Even absent migration, this implies that these cities doubled in size every 20 years.

This difference continued in the 2000s, as shown in Panel D. Rich mega-cities remained in roughly the same position as in the 1960s. The developing mega-cities shifted down to lower birth rates. However, the death rates in poor mega-cities were lower than the historical comparisons, for the most part falling below 10 per thousand. Thus in the 2000s poor mega-cities continued to have rapid rates of natural increase. A notable exception were Chinese cities (e.g. Beijing), which in the 1960s (Panel C) looked similar to other developing cities, but moved in the 2000s (Panel D) to a pattern of death and birth rates similar to rich mega-cities. Johannesburg is also an outlier with a high death rate due to HIV.

The deviation of the developing cities in Panels C-D from the historical norms appears due to the mortality transition. Following World War II (WW2), there was a sudden improvement in health in poor countries (Stolnitz, 1955; Davis, 1956; Preston, 1975), due to: (i) the discovery of effective techniques for mass production of antibiotics such as penicillin (1942) and streptomycin (1946), which treated diseases such as cholera and dysentery, (ii) the invention of vaccines against the yellow fever (1937), poliomyelitis (1962) and measles (1963), (iii) the creation of the World Health Organization (1948), which disseminated knowledge of the new technologies to poor countries, and (iv) disease eradication campaigns, whether against smallpox or malaria. From the perspective of the poor
countries of that period, this international epidemiological transition represented an exogenous shock to mortality (Acemoglu and Johnson, 2007).

Another consequence of the mortality transition was to raise crude rates of natural increase in the largest cities up to the rates typically seen in rural areas. As can be seen in Figure 4, the urban areas of poor countries were disproportionately affected. Indeed, rural natural increase was already high before the 20th century (Panel C), due to high rural fertility (Panel A) and low rural mortality (Panel B). Urban fertility was relatively low (Panel A), while urban mortality was high due to the urban penalty (Panel B). As a result, urban rates of natural increase remained low (Panel C). This changed post WW2, as the international epidemiological transition diffused to large cities and colonizers invested in urban public health by building water supply and sewerage systems and preventing disease outbreaks (Bairoch, 1988; Njoh and Akiwumi, 2011; Njoh, 2013; Fox, 2012, 2014; Fox and Goodfellow, 2016). This shock to mortality occurred in both the largest cities and the other cities of the same country (Panel B), which thus have similar natural growth rates (Panel C). Cities now also have rates of natural increase similar to rural areas. Therefore, due to the urban mortality transition (UMT), natural increase is now equally high across all areas.

Panel C also shows for the 263 city-period observations the mean population growth rate of the entire country. Total population growth greatly accelerated in the 20th century. When crudely decomposing the change in total natural increase between the 1900s and the 1960s into its two main sources, urban natural increase and rural natural increase, we find that about three fourths of it was explained by urban natural increase. Indeed, the urban rate of natural increase dramatically increased during the same period. In addition, these countries also urbanized

---

12 For each mega-city, we use its share in the total urban population, its rate of natural increase and the urban rate of natural increase, to obtain the mean rate of natural increase of the other cities.

13 The urban mortality transition observed in Panel B is starker if we focus on developing countries only (based on their 2015 GDP per capita, i.e. the countries whose income level is below the income of Slovakia, the last country to have become a developed country according to the International Monetary Fund). Their urban death rate was stable around 35, but 40 for the poorest, before the urban mortality transition ("pre-1900s" and "1900s"), after which it dropped to 10 and 15 respectively ("1960s"). 40 was the urban death rate of developed countries before they started developing with the Industrial Revolution.
during that period. Had the urban rate of natural increase remained low, this increased urbanization would have slowed population growth. But since urban natural increase was high, urbanization increased total population growth.

How much of urban natural increase comes from the informal areas versus the formal areas of cities will prove important when simulating the effects of the UMT in the model section. To answer this question, we used similar sources as for the largest cities to obtain the crude birth rate and death rate of the slum areas of four industrializing cities in the late 19th century and eight poor mega-cities today.\textsuperscript{14} Knowing the city population share of the slum areas, we reconstruct the demographic rates of the non-slum areas. In the 19th century, both slum and non-slum areas had low rates of natural increase, at 0.5\% per year. Slum fertility was high, but slum mortality was also high. In the poor mega-cities of today, the rates of natural increase are higher, at 1.0\% per year in the non-slum areas and 2.6\% per year in the slum areas. Natural increase is thus disproportionately higher in slums, which suggests that poor mega-cities are likely to remain informal in the future. Simple decompositions show that three-fourths of the change in urban natural increase between the late 19th century and today can be explained by the higher rate of natural increase in slums.\textsuperscript{15}

Lastly, a consequence of the decline in urban mortality was that the absolute growth of cities expanded to previously unseen levels. For each of the 30 largest cities in 2015 (Table 1), we collected information on the largest annual change in population (in thousands) experienced by the city, as well as the decade this occurred (data not shown but available upon request). We find that poor mega-cities such as Delhi (620 thousand during the 2000s), Dhaka (440; 2000s), Karachi (410; 2000s) and Lagos (350; 2000s) added population at rates well above those seen in historical mega-cities. London (90; 1890s), New York (220; 1920s) and

\textsuperscript{14}We consider London’s slums, Manchester’s townships, New York’s tenement-house wards and Paris’ “quartiers pauvres” in the 1880s, and Manshiet in Cairo, Korail in Dhaka, Orangi in Karachi, Tondo in Manila, Neza in Mexico, Dharavi in Mumbai, Kibera in Nairobi and Rocinha in Rio in the 2000s.

\textsuperscript{15}When comparing the poor mega-cities of the 19th century and the 20th century respectively, we find that slum fertility has remained high, at 35 per 1,000 people. But while slum mortality was also high in the 19th century, at 30 per 1,000 people, it is now as low as in the non-slum areas, at 7 per 1,000 people.
Paris (110; 1950s) never received inflows of people at rates like those seen in the poor mega-cities of today. For poor countries in the 20th century, the UMT created a positive shock to the absolute growth of cities, and given congestion effects, this created the potential for poor mega-cities to emerge and prevail.

B. Congestion in Poor Mega-cities

We now document how measures of congestion relate to growth in the 100 largest cities of the world today. We examine city-level correlations with respect to natural increase, to show that mega-cities growing more quickly tend to be both more congested and less productive than slower-growing peers.

Figure 5.A plots the projected growth rate from 2015-2030 of the 100 projected largest cities in 2030 against their rates of natural increase in the 2000s.\(^\text{16}\) The relationship is strongly positive, which indicates that natural increase will likely be a meaningful driver of city population growth (R-squared of 0.51). The Chinese cities (e.g. Beijing) are outliers to this relationship, with low rates of natural increase, but high expected growth rates due to in-migration (excluding them raises the R-squared to 0.69). Cities that are expected to stagnate in size (e.g. Tokyo) all have rates of natural increase of roughly zero. In Panel 5.B we find a similar correlation between past city natural increase (1960s) and past city growth (1960-2010) (R-squared of 0.50), but for 57 mega-cities only.

In both Panels A and B, the unconditional effect of natural increase (when defined per 100 people) is about 1.\(^\text{17}\) Each newborn in a city thus raises the population of that city by one in the long run. This holds at the country level even though migration costs within countries are relatively low and population can reallocate across locations. Duranton (2015) finds that the effect of the log birth rate (per 100 people) in 1993 on the change in log population between 1993 and 2010 for about 1,000 Colombian municipalities is 0.24, not one. Likewise, if

\(^{16}\)The projected growth rates are based on non-linear extrapolation given the rates of growth pre-2015, estimated independently of the rates of natural increase.

\(^{17}\)Regressing total population growth (%) on aggregate natural increase (per 100 people) for a panel of 201 countries every 5 years in 1960-2010, we find a coefficient of 1 (not shown).
we include country fixed effects in the regression in Panel A, to compare mega-cities within the same country, the effect of natural increase is almost nil (not shown). However, finding a coefficient of one across countries is not surprising given that (i) our analysis is cross-country, and people cannot easily migrate to different countries; and (ii) secondary cities and the countryside also grow at high rates and are similarly congested, thus shutting down a potential “safety valve” for residents of mega-cities.

Poor mega-cities that grow through high rates of natural increase are different from historical cities that grew mainly through in-migration. Panels C through H of Figure 5 plot several characteristics of cities against their rate of natural increase in the 2000s as we have then data for the 100 top cities (see Web Data Appendix for full details on sources). Please note that the correlations in the figures do not show causal relationships. Our goal is simply to characterize the conditions that reinforce conditions in poor mega-cities.

First, in panel C is log density, which shows a positive relationship. Perhaps surprisingly, rich mega-cities such as New York (2,000 people per sq km) or Paris (4,000) are much less densely populated than poor mega-cities such as Dhaka (44,000), Karachi (23,000) or Kinshasa (17,000). Poor mega-cities have both high rates of natural increase and are already dense, meaning they are likely to become even denser in the future. While the literature considers density as a measure of economic development when comparing the cities of a same country, it measures underdevelopment when comparing the mega-cities of different countries.18

This density is potentially indicative of congestion rather than strong agglomeration effects. Panel D shows that slum shares (measured at the national level) are much higher in countries with poor mega-cities. Roughly two-thirds

18 Densities are defined using both the population and the area of the whole agglomeration. Focusing on the central place or slums only should not affect the results. Manhattan’s density is 28,000 people per sq km today. However, mega-cities in poor countries contain areas with much higher densities. The slums of Mumbai, Nairobi, Dhaka and Cairo have densities of 350,000, 300,000, 200,000 and 110,000 respectively (see Web Data Appendix for sources). The Lower East Side in New York City was historically the densest slum of the developed world, even reaching 140,000 in 1910. Other slums in currently developed cities were less dense: Les Halles in Paris (100,000) and the East End in London (90,000).
of the urban residents live in slums in countries containing Dhaka, Karachi, Kinshasa, or Lagos. In Panel E, we also show the strong negative correlation between the “city infrastructure index” designed by United Nations Habitat (1998, 2012) and city natural increase in the 2000s (R-squared of 0.66). This is not surprising considering that poor mega-cities had annual growth rates of 3-7% in 1960-2010 (Panel B), thus doubling in population size every 10-25 years.\textsuperscript{19}

The high density of poor mega-cities does not indicate a large supply of productive workers. Panel F shows the child dependency ratio (the share of those under 14 to population aged 15-64) across our sample of cities. In poor mega-cities with high rates of natural increase, this ratio reaches more than 50 percent (R-squared of 0.66). Further, the labor force that does exist in poor mega-cities is low skilled. In the poorest mega-cities, the share of college-educated workers is close to zero (Panel G). Lastly, in Panel H of figure 5, we show that city natural increase correlates negatively with our measure of city income, the “city product index” of United Nations Habitat (1998, 2012), for the 2000s (R-squared of 0.52). It could be that mega-cities that grow through natural increase are poor because they are highly congested, whether in terms of housing, infrastructure, children or unskilled workers. Or birth rates could remain high in poor mega-cities, because their residents do not have the incentives to have and invest in fewer children, especially if the returns to education are low in these cities.\textsuperscript{20}

Overall, mega-cities in poor countries appear to be developing as “giant villages” rather than as high-productivity agglomerations. Obviously, we would find similar relationships if we were to study the countries as a whole. What is puzzling here is that we find these strong relationships at the mega-city level as well. Mega-cities constitute the modern sector of their countries, and they are

\textsuperscript{19}The city infrastructure index combines information on access to water, sanitation, electricity, roads, and housing (details available in the UN-Habitat reports). The results are similar if we use data on the slum share at the city level for fewer observations, the housing price-to-income ratio, access to water, waste management, or measures of commuting and pollution (Web Appx. Fig. 5 Panels A-H).

\textsuperscript{20}These results hold if we alternatively look at total dependency ratios, primary and secondary completion rates, informal employment, unemployment, national income, city poverty rates and the ratio of the city product index to the log of city size (Web Appx. Fig. 5 Panels I-P).
often the ports of entry for innovations from the rest of the world. Yet, poor mega-cities do not appear that “modern”, suggesting that we should not automatically expect structural change and urbanization to have transformative effects.

III. A MODEL OF HETEROGENOUS LOCATIONS AND POPULATION GROWTH

To explain the rise of poor mega-cities, and the role that the urban mortality transition (UMT) played, we build a general equilibrium model of location choice with heterogeneity in congestion costs and demographics. The presence of congestion effects in all locations means that population growth puts a drag on real wage growth. In this sense, the model is “Malthusian” because any kind of congestion effects - rural or urban - imply that faster population growth lowers real wage growth. Beyond that, heterogeneity in congestion effects means that aggregate population growth influences the distribution of population across locations. Higher population growth rates - like the ones created by the UMT - push people into low-congestion locations. This creates a drag on productivity growth while keeping fertility high, leading to poor mega-cities.

A. Location-Specific Production and Income

We allow for \( J \) total locations in the economy. For each specific location \( j \), the real wage net of rents is determined by

\[
 w_j = A_j F_j(N_j)
\]

where \( A_j \) is productivity in that location, and \( N_j \) is the total number of residents in that location. \( w_j \) is the real wage net of rents.\(^{21}\) The function \( F() \) captures agglomeration and congestion effects in that location. We are not attempting to produce a specific model of these effects, but take them as given and examine

\(^{21}\)Rents within each location are distributed evenly across all residents of the location. This is a typical assumption within the urban literature when the focus is on the spatial distribution of population, and it avoids having to track the distribution of ownership.
the impact of endogenous population growth when they are present. For the function \( F() \), we assume that it is quasi-concave, and that

\[ F_j'(N_j^*) = 0 \text{ for some } N_j^* > 0. \]  

This implies that each \( F_j() \) is single-peaked, reaching a maximum at \( N_j^* \). Because of this, for any population size \( N_j > N_j^* \), it must be that \( F'(N_j) < 0 \), or that wages are decreasing with population size at some point.  

Define \( \epsilon_j \) as the negative of the elasticity of real wages in location \( j \) with respect to population size,

\[ \epsilon_j = -\frac{F_j'(N_j)N_j}{F_j(N_j)}. \]

We write this as a negative so that \( \epsilon_j \) can be interpreted as a measure of congestion effects. If \( \epsilon_j > 0 \), then greater population in location \( j \) implies lower wages, which holds when \( N_j > N_j^* \), and congestion effects outweigh agglomeration effects. This definition allows us to write wage growth as

\[ \frac{\dot{w}_j}{w_j} = \frac{\dot{A}_j}{A_j} - \epsilon_j \frac{\dot{N}_j}{N_j}. \]

With \( \epsilon_j > 0 \), greater population growth lowers wage growth, capturing the “Malthusian” effect we mentioned before.

---

\(^{22}\) We assume all locations produce a homogenous good, but the model could easily be adapted to allow for differentiated goods by location (see Web Theory Appendix 1). Further, we do not explicitly allow for physical capital accumulation. If we take the economy to be open to capital flows, then the rate of return, and the capital/output ratio, would be pinned down at all times in all locations.

\(^{23}\) Duranton and Puga (2004) provide a review of several theories of the origin of agglomeration and congestion, but each features this property of \( F() \).
B. The Distribution of Individuals Across Locations

Individuals have utility over both the real wage, \( w_j \), and the amenities, \( Q_j \), of the location they live in

\[
V_j = \ln w_j + \alpha \ln Q_j.
\]

This formulation of utility matches that used by Desmet and Rossi-Hansberg (2013) in their work on systems of cities. \( Q_j \) does not explicitly depend on \( N_j \), and is simply a location-specific term, weighted by \( \alpha \). Any effects of congestion are presumed to work through the net wage.\(^{24}\) We then assume that the spatial distribution of individuals, meaning the size of each \( N_j \), is an equilibrium outcome based on three conditions:

1) **Labor mobility.** Individuals move costlessly between locations, implying that \( V_j = \bar{V} \) for all locations \( j \), where \( \bar{V} \) is the equilibrium level of utility.

2) **Stability.** \( \epsilon_j > 0 \) for all locations \( j \), which ensures that the equilibrium is robust to small perturbations.

3) **Adding up.** \( \sum_{j=1}^{J} N_j = N \). This closes the model and determines \( \bar{V} \).

The assumption regarding stability bears some further explanation. Given our assumptions about \( F_j(N_j) \) being single peaked, utility is single-peaked with respect to \( N_j \) as well. Hence for any given outside option, \( \bar{V} \), there are two possible equilibrium allocations of population to location \( j \), one with \( N_j < N_j^* \), and one with \( N_j > N_j^* \). The condition says that each location’s equilibrium allocation of population, \( N_j > N_j^* \), and hence the real wage is declining with further population, ensuring stability.\(^{25}\) The stability assumption

\(^{24}\)Having congestion effects work through amenities as well would not change our quantitative exercise. Indeed, faster population growth would then decrease the stock of amenities per capita, and thus welfare. Whether welfare decreases because real wages are lower and/or due to fewer amenities per capita should thus give somewhat similar results in our context.

\(^{25}\)Stability follows from a perturbation argument. If \( N_j > N_j^* \), and \( \epsilon_j > 0 \), then if we moved people into location \( j \), the wage would fall. This would incent people to move to other locations, and offset the
reproduces the standard prediction that locations will be “too big” because individuals do not internalize their effect on agglomeration and congestion within locations (Henderson, 1974; Duranton and Puga, 2004). This doesn’t imply that agglomeration effects are absent, simply that the congestion these individuals produce lowers the wage by more than the agglomeration effects raise it.\textsuperscript{26}

Finally, given labor mobility, it must be the case that the following holds,

\begin{equation}
\frac{\dot{w}_j}{w_j} = \frac{\dot{w}_i}{w_i}
\end{equation}

for any pair $i, j$. This assumption does not imply that the \textit{level} of wages is equal across locations, only that the growth rate is equal. Locations can have different levels because of different amenities, $Q_j$. This assumption also does not imply that location populations grow at the same rate, as they can grow or shrink in size relative to each other because of differences in congestion costs, $\epsilon$.

\textbf{C. The Dynamics of Location Size and Net Wage}

Before proceeding further, we define several terms. First, let the growth rate of any variable $X$ be denoted with a “hat”, so that $\dot{X} = \dot{X}/X$. Second, define the share of population in location $j$ as $s_j = N_j/N$. By definition, the growth rate of aggregate population, $N$, can be written as the weighted sum

\begin{equation}
\dot{N} = \sum_{j=1}^{J} s_j \dot{N}_j.
\end{equation}

From the prior section, we know that wages grow at the same rate in all locations. Therefore, from (4) it must be that the population growth of any location $j$ with initial movement. On the other hand, an equilibrium with $N_j < N_j^*$ is unstable to small perturbations such as this. In this case the wage would rise if one person moved in, and this would incent more people to move into location $j$. The equilibrium with $N_j < N_j^*$ is thus unstable.

\textsuperscript{26}We could, however, imagine that agglomeration effects dominate congestion effects in low-population countries, thus promoting long-run growth (Becker, Glaeser and Murphy, 1999; Desmet, Greif and Parente, 2015). While this hypothesis could apply to prehistorical, antique and medieval societies, we believe that they do not characterize well the developing countries in the 20th century, as they were already densely populated by the time of the mortality transition.
respect to any location \( i \) satisfies

\[
\hat{N}_j = \frac{1}{\epsilon_j} \left[ \hat{A}_j - \hat{A}_i + \epsilon_i \hat{N}_i \right].
\]

Plugging this into the weighted sum in (7) and re-arranging yields

\[
\hat{N}_i = \frac{\bar{\epsilon}}{\epsilon_i} \left[ \hat{N} + \sum_{j=1}^{J} \frac{s_j}{\epsilon_j} (\hat{A}_i - \hat{A}_j) \right],
\]

where

\[
\bar{\epsilon} = \frac{1}{\sum_{j=1}^{J} \frac{s_j}{\epsilon_j}}
\]

is the harmonic mean of the \( \epsilon_j \) values across all locations. From (9) we see that population growth in any location \( i \) depends on several factors. First, higher aggregate population growth, \( \hat{N} \), tends to push up population growth in location \( i \). Regardless of where new people originate, the additional population is spread across existing locations to ensure that utility is equalized.

Second, differentials in productivity growth rates between location \( i \) and the other locations \( j \) will influence population growth in \( i \). When \( \hat{A}_i > \hat{A}_j \), workers are drawn to location \( i \) away from location \( j \). The strength of this effect depends inversely on the congestion elasticities, \( \epsilon_j \). If \( \epsilon_j \) is large, then if workers move out of location \( j \), the wage will rise rapidly as congestion is alleviated, slowing the flow of workers to location \( i \). Finally, the ratio of \( \bar{\epsilon}/\epsilon_i \) captures the relative congestion of location \( i \). The smaller is \( \epsilon_i \), the more elastic is \( \hat{N}_i \) with respect to \( \hat{N} \).

Knowing the growth rate of population in location \( i \), we can solve for the growth
rate of wages by using (9) in (4),

\[
\hat{w} = \tau \sum_{j=1}^{J} \hat{A}_j \frac{\hat{s}_j}{\hat{c}_j} - \tau \hat{N}.
\]

This equation shows that wage growth in any location depends on the weighted sum of productivity growth across all locations relative to the aggregate population growth rate, because of the free movement of labor between locations. Finally, given the utility function in (5), it is simple to show that welfare grows proportionally to wages, or \( \hat{V} = \hat{w} \).

With all these mechanics in place, we can establish some general results related to aggregate population growth – and hence to the UMT and the rise of poor mega-cities.

**The negative effect of population growth on welfare.** From (11) it is clear that wage growth (and hence welfare growth) is negatively related to population growth. The mean value \( \tau \) dictates how powerful this effect is, and captures the “Malthusian” nature of production. Because all locations have some degree of congestion, an increase in population growth lowers wage growth. This effect holds even though population is free to move between locations costlessly, and does not depend on whether one defines locations as “rural” or “urban”.

**The effect of population growth on the distribution of population.**
The share of population in each location, \( s_i \), evolves according to the following simple relationship

\[
\hat{s}_i = \hat{N}_i - \hat{N}
\]

and if location \( i \)'s population is growing faster than the aggregate growth rate of population, then the share of population in location \( i \) is rising over time. The derivative of the growth rate of population share, \( \hat{s}_i \), with respect to aggregate
population growth, $\hat{N}$, can be found from (9),

\[
\frac{\partial \hat{s}_i}{\partial \hat{N}} = \begin{cases} 
> 0, & \text{if } \epsilon_i < \bar{\epsilon} \\
< 0, & \text{if } \epsilon_i > \bar{\epsilon} \\
= 0, & \text{if } \epsilon_i = \bar{\epsilon}
\end{cases}
\]

(13)

In the first case, a location $i$ has a low elasticity relative to the rest of the economy and hence population disproportionately moves into that location, as it can absorb population growth more easily. This is driven by the fact that wage growth is equalized across locations. From (11), we see that an increase in population growth lowers wages across all locations. But locations with relatively low values of $\epsilon_i$ can absorb more workers before their wages fall to the equilibrium level. Hence as population growth rises, low-$\epsilon$ locations see their population share rising (or falling more slowly) compared to the average location. Heterogeneity in congestion across locations implies that the level of aggregate population growth influences the distribution of population; it is not the case that all locations grow proportionally when aggregate population growth increases.

**Dynamic consequences of population growth.** The full consequences of population growth depend on how the changing distribution of population across locations influences wage growth and population growth. For wage growth, (11) shows that as the $s_j$ shares change, the weights on the $\hat{A}_j/\epsilon_j$ terms change. If $\hat{A}_j$ is particularly low in locations where $\epsilon_j$ is low, then wage growth may well decelerate with rapid population growth.

Further, the distribution of population may influence the future population growth rate. Consider the following general form, with

\[
\hat{N} = \sum_{j=1}^{J} s_j CRNI_j,
\]

(14)

where $CRNI_j$ is the crude rate of natural increase in location $j$. As population
shares change, this may lower or raise $\dot{N}$ itself depending on the pattern of $CRNI_j$ across locations. The exact evolution of population distribution ($s_j$), wages/welfare ($\hat{w}$), and population growth ($\dot{N}$) thus depends on the precise assumptions made regarding the values of $\epsilon_j$, $\dot{A}_j$, and $CRNI_j$ across the different locations. In the next section we specify these and derive the implications for developing countries after the UMT.

IV. THE RISE OF POOR MEGA-CITIES

To account for the rise of the poor mega-cities we documented earlier in the paper we will be more concrete about the types of locations and the values of the relevant parameters. Those assumptions will allow us to explain how the distinct shock of the UMT put in motion forces that led to the growth of poor mega-cities.

A. Formal, Informal, and Rural Locations

We divide up the J generic locations into three categories: rural locations, urban formal locations, and urban informal locations. Rural locations, of which there are $R$, operate a technology that is meant to capture not only agriculture, but also the service activities of rural residents. Formal locations, of which there are $F$ total, are locations that operate a formal sector technology. Manufacturing and professional services would be examples of what we have in mind for the formal locations.

Informal locations, of which there are $L$, are located in urban areas and are meant to capture the combination of physical locations that lack urban infrastructure, as well the group of workers that provide low-level personal services. Slums are an example of our concept of informal locations, although we think informal locations may encompass more than just those slums.

For simplicity, we assume that all formal areas have a common growth rate of productivity, $\dot{A}_f$, and the same elasticity of wages with respect to population, $\epsilon_f$. Similarly, informal locations have a common productivity growth rate of $\dot{A}_i$, but...
and an elasticity of $\epsilon_l$, and rural areas have productivity growth of $\hat{\lambda}_r$ and an elasticity of $\epsilon_r$.

Because the parameter values are identical for each type, it is straightforward to see from (9) and (11) that we can track the total population in each of the three types of locations, as opposed to each of the $J$ locations individually.\textsuperscript{27} Hence let

\begin{equation}
N_f = \sum_{j \in F} N_j \quad N_l = \sum_{j \in L} N_j \quad N_r = \sum_{j \in R} N_j
\end{equation}

be the population living in each type of location, and the associated shares are $s_f = N_f/N$, $s_l = N_l/N$, and $s_r = N_r/N$. Wage growth with the three types of locations is

\begin{equation}
\hat{\bar{w}} = \bar{\tau} \left( \hat{\lambda}_f \frac{s_f}{\epsilon_f} + \hat{\lambda}_l \frac{s_l}{\epsilon_l} + \hat{\lambda}_r \frac{s_r}{\epsilon_r} \right) - \bar{\tau} \hat{N}.
\end{equation}

Each location of a given type is also assumed to be identical in demographic behavior, but that behavior varies across the types. Aggregate population growth is given by

\begin{equation}
\hat{N} = s_f CRNI_f + s_l CRNI_l + s_r CRNI_r.
\end{equation}

Aggregate population growth is thus endogenous, in that it changes with the distribution of population across locations.\textsuperscript{28}

\textsuperscript{27}Note that while we denote certain locations as urban or rural, we do not specify that particular urban locations are part of any particular “city”. Our model is robust to any arbitrary classification of urban locations to cities, allowing us to match any desired distribution of city sizes. If we wanted to discipline the distribution of city sizes to meet an empirical regularity such as Zipf’s law (Gabaix, 1999), that could be accommodated by assumptions regarding the underlying values of amenities, $Q_i$. But those assumptions are not necessary for us to determine the implications of population growth on the distribution of population across locations.

\textsuperscript{28}We do not provide explicit micro-foundations for fertility behavior here, but this could be incorporated without altering the structure of the model. In short, we are assuming that the time cost of children varies by type of location. See Web Theory Appendix 2 for an explanation.
B. Consequences of the Urban Mortality Transition

While simplified, we cannot characterize the effect of the UMT without specifying the parameters governing productivity growth and congestion in each type of location. We first make general assumptions so we can discuss the theoretical implications, and in the following section document more fully that these assumptions are warranted given the available data. The first assumption concerns the severity of congestion effects,

ASSUMPTION 1: \( \epsilon_l < \epsilon_f < \epsilon_r \)

This assumption implies that \( \epsilon_l \leq \epsilon \) no matter the population shares in the different locations. This in turn implies that faster population growth will act to raise the growth rate of the informal population share.

This assumption does not say that congestion is less of an absolute problem in informal than in formal or rural areas. It says that informal areas can absorb further population with a relatively small impact on net wages. Informal areas are thus not inherently “bad” locations, since they provide a valuable outlet for the economy, reducing the welfare cost of rapid population growth. The economy would be worse off without them for this reason. As will be highlighted below in the calibrations, it is not the presence of informal locations that is the fundamental issue, it is the rapid population growth that pushes people into these locations.

The second assumption concerns productivity growth rates.

ASSUMPTION 2: \( \hat{A}_l < \hat{A}_r < \hat{A}_f \)

This will influence the effect of changes in population shares on wage growth. Anything that pushes people into informal or rural areas will slow down aggregate productivity growth.

Given Assumptions 1 and 2, we can provide some general findings for the effects of the UMT before we turn to a simulation of the dynamics of population growth, wage growth, and the distribution of population. In response to the UMT, which raised \( \hat{N} \):
1. The economy urbanized more quickly. Assumption 1 indicates that \( \epsilon_r > \bar{\epsilon} \), and hence the growth rate of the rural population share, \( \hat{s}_r \), fell. As \( \hat{s}_r \) was generally negative to begin with, the UMT accelerated the movement of population out of rural areas and hastened urbanization.

2. Urbanization occurred primarily in informal locations. Assumption 1 implies that \( \epsilon_l < \bar{\epsilon} \), and so the growth rate of the informal share, \( \hat{s}_l \), rose in response to the UMT. For a sufficiently large shock to population growth, the value of \( \hat{s}_t \) is sure to be positive. The UMT not only urbanized the economy, but did so by pushing people into the locations with the lowest congestion elasticity - informal areas.

3. Population growth sped up. The immediate effect of the UMT is to raise \( \hat{N} \) directly. But in addition, by urbanizing the economy primarily through informal locations, the UMT led to higher aggregate \( CRNI \) as population moved into the relatively high-growth informal locations. The UMT thus eliminated the natural limitation on population growth that historical economies experienced as they urbanized.

4. Welfare growth slowed down. By raising \( \hat{N} \), wage growth was pushed down directly, as can be seen in equation (16). This is the typical “Malthusian” effect. But in addition, because informal locations have relatively low productivity growth the increased share of population in informal locations meant that aggregate wage growth was slower.

It is important to note that the shift into informal location is the welfare-maximizing response to the UMT, and there is no misallocation across locations. There is a negative externality, as individuals do not take into account their effect on congestion within any given location, but individuals are not constrained from moving between locations. Ultimately the decline in wage (and welfare) growth is a direct effect of increased population growth on an economy where there are congestion costs in all locations.
V. CALIBRATION AND EVALUATION

To evaluate the impact of the UMT, we calibrate our model to a sample of poor countries. After describing the data and the parameters we use, we discuss several experiments we run using the calibrated model. We first explore how urbanization would have proceeded if we eliminated the UMT from these countries. Second, we look at long-run predictions. Finally, we study the effects of policy interventions.

A. Calibration, 1950-2005

Table 2 contains a summary of the parameter and initial values used.

Demographic rates. We use information from the location-level demographic data presented earlier. For 1950, we set the rural crude birth rate (CBR) to 45 (per thousand), the informal CBR to 45, the formal CBR to 40, and the rural crude death rate (CDR) to 20. The CDR for both urban locations will be varied between 15 (the observed average) and 40 (a pre-UMT average).29

Productivity growth rates. We set productivity growth in formal locations, $\hat{A}_f$, to a 3% annual rate. Given the values of other parameters, this will result in a 2% annual growth rate in net wages for an economy that is fully urbanized in formal locations, a rough approximation of developed countries of today.30 For informal locations, there is no direct source available. Informal locations are dominated by workers in personal services and small-scale retail trade. These sectors tend to have low productivity levels and growth rates (Duarte and Restuccia, 2010; McMillan, Rodrik and Verduzco-Gallo, 2014). Duarte and Restuccia indicate that productivity growth in services among the poorest

---

29We calibrate our model to a sample of 41 poor countries. Among these 41 countries, location-level demographic data is available for 19 countries. We obtain the demographic rates by using the average rates for these 19 countries. For this restricted sample, the mean urban CBR is 42.5. Since we will assume a slum share of 50%, and since the informal CBR is probably not strictly higher than the rural CBR, we assume CBRs of 45 for the informal areas and 40 for the formal areas. We then assume similar CDRs for the informal and formal areas, in line with the descriptive results of Section II.A.

30Productivity growth in our model refers to the growth rate of wages net of rents, so there is a disconnect in using growth rates derived for aggregate output. If we assume that labor and land are used in a Cobb-Douglas production function within each location, then wages net of rents should be a constant fraction of output in each location, and hence wage growth should match aggregate growth. This doesn’t require us to assume that the wage share of output is identical across locations (i.e. the wage share in agriculture might be lower than in urban areas), only that it is constant within a location.
countries in their sample was around 1% per year, and we set $\hat{A}_t$ equal to 1%. Finally, rural sector productivity growth is set to 1.5% per year.  

**Congestion elasticities.** For the rural location congestion elasticity $\epsilon_r$, sources point to a high aggregate congestion elasticity in predominantly rural countries. Lee (1987) finds an elasticity of -1.6 and Weir (1991) finds -1.2, both using pre-Industrial Revolution data. Lee (1997) updated his own estimate to -1.0. Acemoglu and Johnson (2007) have work on the effect of the international epidemiological transition that implies an elasticity of -1.2 in a sample of low and middle-income countries in 1940-1980, all of which had urbanization rates about 10-15% in 1950. We thus use -1.2 as our preferred estimate of $\epsilon_r$.

The urban elasticities are set by targeting specific moments from a sample of 41 poor countries at the onset of the UMT. Formally, we selected countries with (a) least 1 million inhabitants in 1950, (b) an urbanization rate below 20% in 1950, (c) that had data available on the slum share in 2005, and (d) had slum shares that reached at least 30% in 2005. The average urbanization rate in 1950 for this sample is 8.9%. We do not have explicit information on the informal share in 1950, or a proxy such as the slum share. We thus assumed a value of 50%.

Given the initial urbanization rate and slum share, and a normalization of the initial size of urban population to 1.0, we find $\epsilon_f$ and $\epsilon_l$ by targeting the average urbanization rate (31.0%) and informal share (64.2%) in our sample in 2005. The

---

31 Fuglie (2010) shows that agricultural TFP growth was below 1% for nearly all developing regions in the 1960s-1970s. However, since that time TFP growth rates have increased across all regions, reaching 2-3% per year in some cases. Overall, he reports developing countries with agricultural TFP growth of 1.4% from 1961-2007. Block (2010) reports agricultural TFP growth rates for Africa over different time periods as well. Growth rates are negative through much of the 1960s and 1970s, but have been as high as 2.8% in the most recent decade, and were approximately 2% from the 1980s into the 1990s.

32 -1.2 comes from dividing their estimate of the effect of life expectancy on GDP per capita, -2.43, by the effect of life expectancy on population, 2.04 (see column (3) of their Tables 8 and 9).

33 Angola, Bangladesh, Benin, Burkina-Faso, Burundi, Cambodia, Cameroon, the C.A.R., Chad, China, Ivory Coast, the D.R.C., Ethiopia, Ghana, Guinea, Haiti, Honduras, India, Indonesia, Kenya, Laos, Madagascar, Malawi, Mali, Mozambique, Myanmar, Nepal, Niger, Nigeria, Pakistan, Rwanda, Senegal, Sierra Leone, Somalia, Sudan, Togo, Uganda, Tanzania, Vietnam, Yemen, and Zambia.

34 Historical evidence suggests that cities in the developing world were relatively more formal during the colonial era than today (Bairoch, 1988; Njoh and Akiwumi, 2011; Njoh, 2013; Fox, 2012, 2014; Fox and Goodfellow, 2016). The average slum share for the 41 countries is 64.2% in 2005, so 50% implies an increase from 1950 to 2005. Results are similar (see Web Theory Appendix 3) if we assume that cities were mostly formal (i.e. a slum share of 40%) or informal (60%) in 1950.
informal share is set based on the slum share in urban areas (United Nations, 2014). Numerically, solving for $\epsilon_f$ and $\epsilon_l$ is not computationally intense, as the model is a series of static allocation problems. We use a discrete-time approximation of the model, setting one period to equal one year. The solution is found by using a non-linear least squares routine to solve for the values of $\epsilon_f$ and $\epsilon_l$ that force the model’s predicted urbanization rate and informal share to match the targeted moments precisely. This results in an informal location elasticity, $\epsilon_l$, of 0.36. The formal location elasticity, $\epsilon_f$, is 0.92.

Table 3 shows observed data in row one. Our baseline calibration in row two uses an urban crude death rate of 15, matching the average found in the data following the UMT. In our sample, the absolute size of the urban (informal) population increased by a factor of 15.2 (18.6). Our calibration generates a similar absolute increase, predicting a factor of 13.7 (17.5).

B. Eliminating the UMT, 1950-2005

We simulate the model again, this time setting the urban crude death rate to 40 in row 3 of Table 3. We are thus asking what would have happened if poor countries had experienced urban death rates similar to those experienced by the developing world prior to WW2, or Europe and the U.S. in the early 19th century. Population growth would have been slower, which would have resulted in less urbanization by 2005. The urbanization rate would have been 26.1%, implying that about one-quarter ($\approx (31.0-26.1)/(31.0-8.9)$) of the urbanization occurring after 1950 was due to the UMT. The composition of urban areas would have been different as well, with 56.7% in informal locations. Here, half of the increase in the informal share post-1950 is attributable to the UMT. The urban sector would have been 9.2 times larger in 2005 than in 1950, compared to a

---

35We are targeting the sample means of the urbanization rate and informal share, but do not use the variation in these measures across the sample. In Web Theory Appendix 4, we show that estimating $\epsilon_f$ and $\epsilon_l$ individually for each of the 41 countries, and then taking the median or mean of these estimates, gives us essentially identical results.

36Acemoglu and Johnson (2007) indicate an overall elasticity of -0.81 of GDP per capita with respect to population growth, which includes all countries, including developed countries which are primarily formal and heavily urbanized. Our targeted value of -0.92 appears reasonable given the overall sample.
factor of 13.7 in the baseline. Developing cities would thus be two-thirds as large today in these countries without the UMT. Moreover, the absolute size of the informal sector would have been only 60% of their current size (10.4 versus 17.5 times larger than in 1950). Finally, welfare in 2005 is 21% higher without the UMT. This does not account for the direct welfare effect of lower mortality, solely the effect of the UMT on wage growth.

C. Long-run Effects of the UMT, 1950-2100

No demographic changes. Table 4 shows the results when simulating outcomes another 50 and 100 years into the future (i.e. to 2050 and 2100). With an urban crude death rate of 15, the urbanization rate in 100 years is predicted to be 56.1%, with 67.4% of those urban residents in informal locations (row 1). The size of the urban (informal) sector would be about 80 (108) times larger than in 1950. If we extend this to 150 years, urbanization would be almost 78%, and the informal share would remain high, at 65.8%. In contrast, had the UMT never occurred (row 2), in 150 years the urbanization rate would be just over half, at 57.7%, and the informal share would fall to 31.8%. If we let this run out further, urbanization would eventually become a movement into formal locations and welfare would be 135% higher than with the UMT.

The long, slow, urbanization process with an urban CDR of 40 is consistent with the path followed by rich countries. If our initial period were the early 19th century instead of 1950, then these predictions are consistent with what was observed in Europe and the U.S. over the long-run. In 1800, they both had an urbanization rate of about 10%, similar to the starting position of our sample. At the eve of WW2, their urbanization rate was about 55%, close to our simulated value of 57.7%. The dominance of formal locations, and the associated improvement in welfare, are qualitatively consistent with the development of these countries over this period. The model predicts an informal share of 31.8%, which, we believe, is a plausible estimate of the slum share of their urban areas around
Endogenous CRNI. We introduce a crude endogenous effect on crude rates of natural increase in all locations, assuming CRNIs have an elasticity of -0.2 with respect to the wage. Jones, Schoonbroodt and Tertilt (2010) provide elasticities of fertility with respect to income for the U.S. from 1826-1960, and find that they decline from between -0.30 and -0.40 in the 1800’s to about -0.20 in the 1950’s. Young (2005) estimates an elasticity of fertility with respect to wages of -0.35 using household survey data from South Africa. We use a relatively small value of this elasticity to ensure that we do not overstate the effect, as our model only provides information on net wages (not total wages), and to ensure we are not overstating the effect of falling net wages due to population growth.

Introducing endogenous demographic behavior has a modest effect on urbanization and informal rates, but accelerates the pace of absolute growth (see the second panel). The urbanization rate is higher after 100 years at 62.1% (compared to 56.1%), and the informal share is 73.4% (as opposed to 67.4%). Moreover, welfare is only 84% of the baseline level, due to the fact that the growing population with the UMT leads to lower net wages, and thus given the endogenous demography, higher population growth rate. Endogenous demographic behavior following the UMT generates a kind of “trap” where a country finds itself in a spiral of fast population growth, expanding informal areas, and low wage growth.

Exogenous CRNI decline. In the third panel, we instead impose a “demographic transition”, letting the CRNI in all locations fall by 1% per year starting in 1950. With this exogenous CRNI decline, countries that had the UMT do not urbanize nearly as rapidly after 150 years, at only 49.2% versus 77.2%. The informal share also falls to 16.0%, as opposed to 65.8%. Urban areas are

---

37 Definitions of slums take into account four components: access to improved water, access to improved sanitation facilities, durability of housing, and having sufficient living area. The 1950 U.S. housing census reports that 21.2% of the U.S. urban population lived in “dilapidated” housing units and/or units without running water. Also taking into account sanitation facilities and living area should thus lead to an even higher slum share in 1950, and a slum share close to 31.8% around the time of World War II.

38 Linking population growth to incomes follows Becker (1960), but the correlation of income and population growth may instead be driven by factors such as the returns to education (Galor, 2012). Our simulation is a reduced-form method of incorporating endogenous population growth into our analysis.
only 39.7 times larger, and informal areas only 12.7 times larger, significantly smaller than the baseline. Welfare is quadruple that of the baseline. The results thus suggest that modifying the population growth rate, perhaps through policy interventions, could change the long-run trajectory of poor mega-cities.

D. Simulating Policy Effects, 2005-2055

Table 5 shows the effects of various policy experiments that might be used to alleviate the conditions in, or slow the growth of, poor mega-cities. The first row shows our starting point in 2005, with urbanization of 31.0% and an informal share of 64.2%. We normalize the size of the urban sector and of the informal sector to one in 2005. The second panel presents the results of the simulations, for which we use the parameters of Table 2, but the demographic rates of the new initial year, 2005. The first row shows our baseline outcomes for 2055, i.e. an urbanization rate of 56.5%, and a total urban size nearly 5.6 times that in 2005. As a comparison, United Nations (2014) projects an urbanization rate of 54%, and a relative size of 3.2 for the countries in our sample. Our baseline thus captures the urbanization rate well, but overstates the size, due to the fact that our baseline holds demographic rates constant over the entire period. The informal share remains high, at 65.6%.

Productivity growth. One strategy for alleviating the conditions in poor mega-cities is to foster rapid productivity growth, either through industrialization in formal locations or “green revolutions” in rural ones. Our simulations, though, indicate that these are likely to be less effective than boosting productivity growth in informal locations themselves. To see this, we separately raise productivity growth by one percentage point in each individual sector. Raising productivity growth to 4% in formal locations increases urbanization rates (row 2), but also lowers the informal share relative to our baseline. Welfare is 11% higher in this

39 We use the sample averages in the location-level data for the 19 poor countries (minus China, due to the one-child policy) in the 2000s. Birth rates in informal locations are set to 35, formal locations to 20, and rural locations to 35. Death rates are set to 10 in rural locations, and 7.5 in urban locations.

40 While we do not explicitly model human capital, we show in Web Theory Appendix 5 that if human
scenario. Raising rural productivity growth to 2.5% reduces urbanization to 41.8% and welfare is 16% higher in this case (row 3). Compared to this, raising informal productivity growth to 2% increases welfare by 40%, even though cities and informal areas are much larger than in the baseline. The effect is larger because welfare depends on the difference between productivity growth and congestion, and higher $\hat{A}_t$ generates the largest difference of the three scenarios.\footnote{We find that formal productivity growth would have to be at least 6% – what has been achieved by China’s modern industrial sector in 1993-2004 (Bosworth and Collins, 2008, Table 3) – for the economy to become predominantly formal and for welfare gains to match those found from increasing informal productivity growth. See Web Theory Appendix 6 for results.}

Lower formal congestion. An alternative policy is building additional transportation and/or housing infrastructure in existing cities, or creating well-planned cities ex nihilo. In row 5, we examine the effect of lowering the formal sector elasticity from 0.92 to 0.73, a value we estimate using an alternate sample of 24 rich countries.\footnote{These 24 countries are chosen because they had urbanization rates higher than 50% in 1950 (see Web Theory Appendix 6 for full details of this calibration). They are assumed to have informal shares of 25% in 1950, and if they are missing data in 2005, an informal share of 5% in that year. Crude birth rates and crude death rates are set at 25 and 10 in all locations respectively, matching data from the location-level demographic data set. The lower value of $\epsilon_f$ indicates slightly lower congestion effects in rich countries. The functionality of rich mega-cities is entirely consistent with a relatively large value of $\epsilon_f$, which only determines the marginal effect of a new resident. These cities benefit from relatively high productivity levels and low aggregate population growth rates compared to poor mega-cities.}

As can be seen, there is a negligible change to the overall urbanization rate and urban size relative to the baseline, while welfare is 9% higher. In row 6 we consider an extreme case where policy is able to lower formal sector congestion to the informal elasticity of 0.36. Urbanization would be higher, at 70%, and urban sizes larger as overall congestion is less severe. The informal share would drop to 11.6%, as people moved into the relatively high-productivity growth formal locations. Welfare would then be 82% higher.

Higher informal congestion. Instead of improving formal locations, policies may be aimed explicitly at reducing the attraction of informal locations. This may take the form of enforcing property rights within these locations, or physically destroying these locations, i.e. clearing slums as Haussmann did in Paris in 1853-1870. In row 7 we simulate these policies by raising informal congestion to the
level of formal congestion, $\epsilon_l = 0.92$. This lowers the informal share significantly to 40.1%, and the size of the informal population is much smaller, at only 2.6 times the 2005 starting value. However, welfare is lower, at only 77% of the baseline. Lower welfare is a result of raising the average congestion costs in the economy. Informal areas, while they may have low productivity growth, provide an outlet for population growth that alleviates congestion effects in the aggregate.

**Migration Restrictions.** Another policy approach, such as with China’s hukou system, is to limit the growth of poor mega-cities by limiting in-migration. We can evaluate the effect of such policies by adding a migration restriction to our model that generates a wedge in wage growth between locations. Let the value $\phi_{ij}$ measure the wedge between locations $i$ and $j$. We then solve the model with this wedge in place. Web Theory Appendix 7 shows how this enters into the equations. For our simulation, we set $\phi_{fr} = 0.01$ and $\phi_{fr} = 0.01$, implying that wage growth in all urban locations is higher than in rural ones by one percentage point due to a migration restriction. These kinds of migration restrictions limit urbanization to 41.8% (row 8), while city size and informal size are also smaller. However, welfare (which is a weighted average of location-specific welfare) is only 85% of the baseline. The migration restrictions lower welfare by forcing a large portion of the population to stay in relatively high-congestion rural locations.

In row 9, we flip the migration restrictions, imposing limits on the movement of urban people into rural areas. This may reflect either an explicit policy, a preference by urban dwellers to remain in cities, or a way of accounting for a shift in preferences towards urban locations after the UMT eliminated the urban penalty in life expectancy.\footnote{Europe’s killer cities during the Industrial Revolution had to offer relatively high wages to urban residents in order to compensate for the fact that they had relatively higher mortality rates than the countryside (Williamson, 1990). Conversely, urban mortality rates are slightly lower than rural mortality rates in today’s developing world (see Figure 4.B), and access to public social services is broader, which gives a direct incentive for rural residents to migrate to, and urban newborns to stay in, the cities.} Regardless of the origin, these “restrictions” raise urbanization rates to 68.9%, well above the baseline, and informal areas would be larger. Welfare is 16% higher than in the baseline, however. Rural workers...
have relatively high wage growth and are the majority of workers for most of the period between 2005 and 2055 in our simulation.

**Cross-location congestion.** Here we do not consider an explicit policy, but rather extend the model to see how different policies impact poor mega-city growth when there are cross-location congestion effects. In short, informal areas or slums in cities may create congestion effects in the formal locations of the same city. In Web Theory Appendix 8, we present a modification of the model that allows for these cross-location congestion effects. The key parameter here is $\theta_f$, the elasticity of formal sector wages with respect to the total population of a city (which is assumed to consist of one informal and one formal location).

We set the formal location’s own congestion parameter to be equal to the informal value ($\epsilon_l = \epsilon_f = 0.29$), but we add a cross-location congestion elasticity for formal areas by setting $\theta_f = 1.0$. Urbanization comes out to 54%, but the informal share in 2005 is 94% (row 10). These numbers are not strictly comparable to the baseline, as we have now altered the nature of congestion costs. They are most interesting when compared to lines 11 and 12, which combine cross-location congestion with various policies. The hypothesis is that, with this specific form of congestion effects, constraining the informal sector should allow the formal sector to develop. That is not what we find, however.

Row 11 assumes there are rural-to-urban migration restrictions. As can be seen, this does limit the urbanization rate, urban size, and informal share relative to row 10. However, there is no improvement in welfare. The restrictions lower rural welfare, and this overcomes any positive effect on wage growth from formal locations that suffer less congestion from informal locations. In row 12 we increase the congestion costs in informal locations (e.g. clearing slums). This limits informal sector growth, which has positive effects for the formal sector. Again, urbanization is lower in this scenario than in row 10, and the informal share falls

---

44 Varying the value of $\theta_f$ between 0.50 and 1.50 has little effect on the results (see Web Theory Appendix 6).
to only 50.5%. However, welfare is also lower, as the cost of higher congestion in informal areas overcomes the benefits accruing to formal workers.

**Lower population growth.** Finally, we consider policies regarding population growth, which are not generally thought of as urban policies, but which we will show can have a significant impact on poor mega-city growth. In row 13 we impose zero population growth by setting the CBR and CDR to be equal in each location. This results in essentially no additional urbanization from 2005 to 2055, as the urbanization rate ends up at only 33.6%. Urban size is unchanged, but the informal share of urban areas drops dramatically, to only 29.7%, and the size of these informal areas falls to half the size in 2005. Welfare is 145% higher than in the baseline. This is an extreme policy, and in row 14 we explore a policy of “family planning”, making all crude birth rates equal to 20, similar to those achieved on average in China, South Korea, Hong Kong, or Singapore during the 1970’s. Urbanization would only be 43.3% in 2055, and urban size only about 2.3, while the informal share would fall below half, and informal areas would only be 1.7 times as large. Welfare would be 56% higher than the baseline.

**E. Discussion**

From Table 5, we find that the three most effective policies for increasing welfare and limiting the size of the informal locations are lower formal congestion costs (row 6), zero population growth (row 13), and family planning (row 14).

Lowering formal congestion costs is promising, but may not be feasible or cost-effective. Improving city infrastructure by “building back better” in existing cities, or creating new urban locations by “better building elsewhere” are two strategies for lowering formal congestion costs. Building back better is costly, and likely requires clearing slums or other forced population disruptions. Better building elsewhere may not require moving populations from existing locations, but has massive costs of its own. In Web Theory Appendix 9, we document that the cost of creating new large cities is between 3 and 25% of GDP (mean of 12.5%), based on information from Brasilia in Brazil (1956-1960; 11-13%), Islamabad in
Pakistan (1960s; 6%), Abuja in Nigeria (1978-1991; 23-25%), 50 ghost cities in China (2000s-2010s; 7-24%), Kilamba outside Luanda in Angola (2008-2012; 3%), and the future potential new capital city of Egypt (2020s; 13%). Further, in Web Appendix Figures 6 and 7, we show that countries that have created new megacities in our sample period do not show any significant reduction in slum shares or infrastructure congestion in their other megacities.\textsuperscript{45}

Improving urban institutions (e.g. forward-looking urban planning, easing up urban regulations, documenting property rights, lowering the costs of land transactions, reducing corruption in local government) are likely less costly than either “building back better” or “better building elsewhere”. However, regardless of the cost, to have a large welfare impact they would have to reduce formal sector congestion costs to levels below those in currently rich countries. As noted above, for a sample of rich countries we estimate a formal congestion elasticity of $\epsilon_f = 0.73$, twice the size of the parameter assumed in row 6 ($\epsilon_f = \epsilon_l = 0.36$). It seems more likely that the results infrastructure spending or institutional reform could achieve are like those in row 5, where welfare increases by only 9% and there is a modest reduction in the size of informal locations.\textsuperscript{46}

Compared to lowering formal congestion costs, restraining population growth appears to be far more effective. Today, most African countries and the poorest American and Asian countries still have crude birth rates of 40 in the 2000s. Zero population growth, i.e. crude birth rates of about 10, may be politically infeasible, as it would require authoritarian measures similar to the one-child policy in China (1978-2015). But even family planning programs aimed at lowering the CBR to 20 severely limit the size of informal locations and raise welfare by 56%.

\textsuperscript{45} Ten of the cities in the figures are in countries that have massively invested in the creation of a new capital city: Karachi and Lahore in Pakistan (Islamabad became the capital city in 1960), Belo Horizonte, Rio and Sao Paulo in Brazil (Brasilia; 1960), Dar es Salaam in Tanzania (Dodoma; 1974), Abidjan in Ivory Coast (Yamoussoukro; 1983), and Ibadan, Kano and Lagos in Nigeria (Abuja; 1991).

\textsuperscript{46} One issue with forward-looking urban planning, i.e. devoting some land to urban expansion and laying out a street grid on that land so that the city can then self-organize itself, is that it would have to solve a spatial coordination failure. No single individual has an incentive to move to a new city (or part of a city), because they would enjoy no agglomeration benefits and their own wage would be lower than in the poor mega-city. A new large city would require a successful, centralized, coordinated effort. But a centralized effort would then be expensive, as highlighted above.
Hong Kong, Singapore, South Korea and China all implemented similar “two is enough” campaigns in the 1960’s and 1970’s, bringing down their crude birth rates dramatically. The WHO estimates costs for family planning programs that would lower crude birth rates to 20 of about 1.2% of one year of GDP in Sub-Saharan Africa (see Web Theory Appendix 9 for details). These policies are thus much less expensive than building new cities (cost of 12.5% of GDP).

Relative to lower population growth, no other policies we examine are able to limit poor mega-city growth without severely limiting welfare. Similar to the conclusions of Au and Henderson (2006) and Desmet and Rossi-Hansberg (2014), we find that migration restrictions are likely to be counter-productive. They limit both the size and informal share of cities, but at the cost of decreasing aggregate welfare as they do not allow population to flow into the low congestion locations.

Even if those informal areas cause cross-location congestion in formal locations, this does not alter the welfare consequences, and hence the net result is lower welfare on average. Policies meant to eliminate slums or informal locations lower welfare because they eliminate a valuable outlet for rapid population growth. Ultimately, our quantitative analysis shows that it was a surge in population growth that helped create these poor mega-cities, and reversing that surge is thus the most direct way to improve living standards and limit their size.

---

47 The World Bank (2007) documents that family planning policy is not based on income levels. China, India, Indonesia, South Africa and South Korea all adopted strong family planning policies at very low income levels, while many African, Central and South American, and Western Asian countries had weak family planning policies given their income level for most of the post-WWII period.

48 Our simulation shows the effects of lowering the CBR to 20 in all locations. Web Theory Appendix 6 shows the effect of limiting CBR’s by location, and rural family planning is the most effective, followed by family planning in slums. However, rural family planning may be more costly than urban family planning given that rural settlements are often scattered over vast areas. Determining the optimal family planning policy would require more research on the cost of location-specific policies.

49 Au and Henderson (2006) find that migration restrictions have resulted in many undersized cities, with net output per worker 17% lower at the 50th percentile of cities – although this may have changed more recently as the Hukou system has been relaxed (Li and Gibson, 2015) – suggesting welfare costs to restrictions even higher than those we estimate (a 15% welfare loss) if agglomeration effects are this strong in China. Desmet and Rossi-Hansberg (2014) find that reducing spatial differences in efficiency and amenities in China to the levels seen in the U.S. would improve welfare by 17.7% and 22.6% respectively. They also attribute these spatial differences to “existing mobility restrictions under the Hukou system”.


VI. CONCLUSION

Poor mega-cities are a relatively new phenomena, appearing in the post-war era across the developing world. Using a novel dataset of city-level demographics over time, we document that poor mega-cities arose after the urban mortality transition (UMT) following World War II, which lowered their urban mortality rates to rich-country levels. To account for this, we develop a general equilibrium model of location choice that has heterogeneity in congestion costs and demographics across locations. The model shows that the aggregate population growth rate has a direct effect on the distribution of population across locations. Calibrating our model to a sample of developing countries, we find that the UMT accounted for one-quarter of urban growth, and half of slum growth, from 1960-2005. The UMT lowered welfare from net wages by about 20%, although that must be set against the gains to welfare from lower mortality. Using our calibrated model, we showed that the most effective way of limiting poor mega-city growth and raising welfare is through limiting population growth, as opposed to imposing migration restrictions or building infrastructure.
REFERENCES


Figure 1. : **NUMBER OF MILLION-PLUS MEGACITIES, 1700-2030**

Notes: This figure shows the number of urban agglomerations of at least 1 million inhabitants in developed countries and developing countries in 1700, 1900, 1950, 2015 and 2030. Developed countries are countries whose GDP per capita (PPP, constant 2011 international $) is above $12,476 (the income level of Slovakia) in 2015. See Web Appendix for data sources.

Figure 2. : **CITY LIVING STANDARD VERSUS CITY SIZE RANK, HISTORICALLY (PRE-1910) AND IN 2010**

Notes: This graph displays the relationship between city living standards and city size for 111 cities of more than 300,000 inhabitants in 2010 and 111 city-year observations of more than 100,000 inhabitants pre-1910 (multiple observations for a same city). For each period, we rank the cities by living standards and city size and show the correlation between the two (linear fit estimated using as weights the population of each city-year). City living standards are proxied by the city product indexes of United Nations Habitat in 2000-2010 and welfare ratios (estimated for a “bare bones” consumption basket) before 1910. See Web Appendix for data sources.
Figure 3. MEGA-CITY CRUDE BIRTH AND DEATH RATES, FROM ANTIQUITY TO DATE

Notes: This figure shows the crude birth and death rates for 49 cities before 1900, 88 cities in the 1900s (the cities in the 1900s and before were selected because they were among the 100 top cities in 1900), 57 cities in the 1960s, and 100 cities in the 2000s (the cities in the 1960s-2000s were selected because they will be among the 100 top cities in 2030 according to United Nations (2014)). The mean crude birth rates (CBR), death rates (CDR) and rates of natural increase (CRNI) are shown. See Web Appendix for data sources.
Notes: This figure shows the mean crude rates of birth, death and natural increase for the mega-cities as well as for the other cities, the rural areas and the whole area of the same countries as the mega-cities before 1900 (N = 49), in the 1900s (88), in the 1960s (57), and in the 2000s (100). The sample is the same as in Figure 3. See Web Appendix for data sources.
Figure 5: MEGA-CITY NATURAL INCREASE AND CHARACTERISTICS

(a) Future Mega-City Growth

(b) Past Mega-City Growth

(c) Mega-City Population Density

(d) National Slum Share

(e) Mega-City Infrastructure Index

(f) Mega-City Child Dependency Ratio

(g) Mega-City Tertiary Education

(h) Mega-City Product Index

Notes: This figure describes the 100 largest cities in 2030 according to United Nations (2014). (a) shows the correlation between projected growth in 2015-2030 and natural increase in the 2000s. (b) shows the correlation between growth in 1960-2015 and natural increase in the 1960s. (c)-(h) show the correlations between congestion and natural increase (2000s): the city density (c), the national slum share (d), the city infrastructure index (e), the city child dependency ratio (f), the city tertiary education completion rate (g), and the city product index (h). See Web Appendix for data sources.
### TABLE 1: WORLD'S LARGEST MEGACITIES (MILLIONS), 1700-2015

<table>
<thead>
<tr>
<th></th>
<th>1700</th>
<th>1900</th>
<th>1950</th>
<th>2015 (△% 2015-2030)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Istanbul</td>
<td>0.7</td>
<td>London</td>
<td>6.5</td>
</tr>
<tr>
<td>2</td>
<td>Tokyo</td>
<td>0.7</td>
<td>New York</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>Beijing</td>
<td>0.7</td>
<td>Paris</td>
<td>3.3</td>
</tr>
<tr>
<td>4</td>
<td>London</td>
<td>0.6</td>
<td>Berlin</td>
<td>2.7</td>
</tr>
<tr>
<td>5</td>
<td>Paris</td>
<td>0.5</td>
<td>Chicago</td>
<td>1.7</td>
</tr>
<tr>
<td>6</td>
<td>Ahmedabad</td>
<td>0.4</td>
<td>Vienna</td>
<td>1.7</td>
</tr>
<tr>
<td>7</td>
<td>Osaka</td>
<td>0.4</td>
<td>Tokyo</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>Isfahan</td>
<td>0.4</td>
<td>St Petersburg</td>
<td>1.4</td>
</tr>
<tr>
<td>9</td>
<td>Kyoto</td>
<td>0.4</td>
<td>Manchester</td>
<td>1.4</td>
</tr>
<tr>
<td>10</td>
<td>Hangzhou</td>
<td>0.3</td>
<td>Philadelphia</td>
<td>1.4</td>
</tr>
<tr>
<td>11</td>
<td>Amsterdam</td>
<td>0.2</td>
<td>Birmingham</td>
<td>1.2</td>
</tr>
<tr>
<td>12</td>
<td>Naples</td>
<td>0.2</td>
<td>Moscow</td>
<td>1.1</td>
</tr>
<tr>
<td>13</td>
<td>Guangzhou</td>
<td>0.2</td>
<td>Beijing</td>
<td>1.1</td>
</tr>
<tr>
<td>14</td>
<td>Aurangabad</td>
<td>0.2</td>
<td>Kolkata</td>
<td>1.1</td>
</tr>
<tr>
<td>15</td>
<td>Lisbon</td>
<td>0.2</td>
<td>Boston</td>
<td>1.1</td>
</tr>
<tr>
<td>16</td>
<td>Cairo</td>
<td>0.2</td>
<td>Glasgow</td>
<td>1.0</td>
</tr>
<tr>
<td>17</td>
<td>Xian</td>
<td>0.2</td>
<td>Osaka</td>
<td>1.0</td>
</tr>
<tr>
<td>18</td>
<td>Seoul</td>
<td>0.2</td>
<td>Liverpool</td>
<td>0.9</td>
</tr>
<tr>
<td>19</td>
<td>Dacca</td>
<td>0.2</td>
<td>Islamabad</td>
<td>0.9</td>
</tr>
<tr>
<td>20</td>
<td>Ayutthaya</td>
<td>0.2</td>
<td>Hamburg</td>
<td>0.9</td>
</tr>
<tr>
<td>21</td>
<td>Venice</td>
<td>0.1</td>
<td>Buenos Aires</td>
<td>0.8</td>
</tr>
<tr>
<td>22</td>
<td>Suzhou</td>
<td>0.1</td>
<td>Budapest</td>
<td>0.8</td>
</tr>
<tr>
<td>23</td>
<td>Nanking</td>
<td>0.1</td>
<td>Mumbai</td>
<td>0.8</td>
</tr>
<tr>
<td>24</td>
<td>Rome</td>
<td>0.1</td>
<td>Ruhr</td>
<td>0.8</td>
</tr>
<tr>
<td>25</td>
<td>Smyrna</td>
<td>0.1</td>
<td>Rio</td>
<td>0.7</td>
</tr>
<tr>
<td>26</td>
<td>Srinagar</td>
<td>0.1</td>
<td>Warsaw</td>
<td>0.7</td>
</tr>
<tr>
<td>27</td>
<td>Palermo</td>
<td>0.1</td>
<td>Tientsin</td>
<td>0.7</td>
</tr>
<tr>
<td>28</td>
<td>Moscow</td>
<td>0.1</td>
<td>Shanghai</td>
<td>0.6</td>
</tr>
<tr>
<td>29</td>
<td>Milan</td>
<td>0.1</td>
<td>Newcastle</td>
<td>0.6</td>
</tr>
<tr>
<td>30</td>
<td>Madrid</td>
<td>0.1</td>
<td>St Louis</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Notes: The table shows the population of the world’s largest urban agglomerations in 1700, 1900, 1950 and 2015. An urban agglomeration comprises the city proper and also the suburban fringe or thickly settled territory lying outside. (△% 2015-30) is the projected annual growth rate (%) of the city between 2015 and 2030 according to United Nations (2014). These growth rates are based on non-linear extrapolation given the rates of growth pre-2015. The main sources of the data are Chandler (1987) and United Nations (2014). See Web Appendix for more details on data sources.

### TABLE 2: CALIBRATION PARAMETER VALUES

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set externally:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural crude birth rate (per 000)</td>
<td>45</td>
<td>Location-level data, 1960s</td>
</tr>
<tr>
<td>Informal crude birth rate (per 000)</td>
<td>45</td>
<td>Location-level data, 1960s</td>
</tr>
<tr>
<td>Formal crude birth rate (per 000)</td>
<td>40</td>
<td>Location-level data, 1960s</td>
</tr>
<tr>
<td>Rural crude death rate (per 000)</td>
<td>20</td>
<td>Location-level data, 1960s</td>
</tr>
<tr>
<td>Formal productivity growth, ( \hat{A}_f )</td>
<td>0.030</td>
<td>Developed country trend growth</td>
</tr>
<tr>
<td>Informal productivity growth, ( \hat{A}_l )</td>
<td>0.010</td>
<td>Various, see text for details</td>
</tr>
<tr>
<td>Rural productivity growth, ( \hat{A}_r )</td>
<td>0.015</td>
<td>Various, see text for details</td>
</tr>
<tr>
<td>Rural congestion elasticity, ( \epsilon_r )</td>
<td>1.2</td>
<td>Various, see text for details</td>
</tr>
<tr>
<td>Initial urban size</td>
<td>1.0</td>
<td>Normalization</td>
</tr>
<tr>
<td>Urbanization rate (%) in 1950</td>
<td>8.9</td>
<td>Sample average</td>
</tr>
<tr>
<td>Informal share in urban areas (%) in 1950</td>
<td>50.0</td>
<td>Various, see text for details</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Targeted:</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal congestion elasticity, ( \epsilon_f )</td>
<td>0.92</td>
<td>Urbanization rate, 2005 (mean 31%)</td>
</tr>
<tr>
<td>Informal congestion elasticity, ( \epsilon_l )</td>
<td>0.36</td>
<td>Informal % of urban pop, 2005 (mean 64%)</td>
</tr>
</tbody>
</table>

Notes: The table shows the parameter values used in the baseline simulation of the model. Location-level data refers to the data presented in section 2. Sample average refers to the sample of 41 developing countries. The two targeted parameters are matched to the 1950 and 2005 sample means (in parentheses) from the 41 developing countries. See text for a complete description of the method.
### TABLE 3: CALIBRATED OUTCOMES AT DIFFERENT DEATH RATES

<table>
<thead>
<tr>
<th>Rate (%)</th>
<th>Size</th>
<th>Share (%)</th>
<th>Size</th>
<th>Welfare</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>8.9</td>
<td>31.0</td>
<td>1.0</td>
<td>15.2</td>
</tr>
<tr>
<td>2005</td>
<td>8.9</td>
<td>26.1</td>
<td>1.0</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Calibration:**
- Urban CDR = 15:
  - Urban: 8.9 31.0 1.0 13.7 50.0 64.2 1.0 17.5 1.00
  - Informal: 8.9 26.1 1.0 9.2 50.0 56.7 1.0 10.4 1.21

**Notes:** The table shows the simulated outcomes for different urban crude death rates (CDR). The results for “Urban CDR=15” are our baseline, used to match to the observed data. The row for “Urban CDR=40” is simulated changing only the CDR. Urban (informal) size is the geometric mean across our sample of urban (informal) population in 2005 relative to 1950. Welfare is the equivalent variation in net wages necessary for an economy with CDR=15 to match welfare with CDR=40.

### TABLE 4: LONG-RUN OUTCOMES USING CALIBRATED MODEL

<table>
<thead>
<tr>
<th>From 1950 forward 100 years (in 2050):</th>
<th>150 years (in 2100):</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urb. CDR = 15</td>
<td>56.1</td>
</tr>
<tr>
<td>Urb. CDR = 40</td>
<td>42.5</td>
</tr>
</tbody>
</table>

**Notes:** The first panel shows the simulated results using the parameters of Table 2, and different urban crude death rates. In the second panel, the CRNI in all locations is assumed to be endogenous with respect to the net wage, \( w \) (assuming an elasticity of CRNI with respect to wage of -0.2). In the third panel, the CRNI in each location declines at the rate of 1% per year. Welfare is the equivalent variation in net wage necessary for the baseline (row 1) to match welfare in a subsequent row.

### TABLE 5: LONG-RUN OUTCOMES UNDER DIFFERENT POLICIES

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>31.0</td>
<td>1.0</td>
<td>64.2</td>
<td>1.0</td>
<td>n/a</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>56.5</td>
<td>5.6</td>
<td>65.6</td>
<td>5.6</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

**Productivity changes:**
- 2. Higher formal productivity growth (\( \hat{A}_f = 0.04 \))
- 3. Higher rural productivity growth (\( \hat{A}_r = 0.025 \))
- 4. Higher informal productivity growth (\( \hat{A}_l = 0.02 \))

**Congestion and migration changes:**
- 5. Lower formal congestion (\( \epsilon_f = \epsilon_{f,rich} = 0.73 \))
- 6. Lower formal congestion (\( \epsilon_f = \epsilon = 0.36 \))
- 7. Higher informal congestion (\( \epsilon_l = \epsilon_f = 0.92 \))
- 8. Rural-to-urban restrictions (\( \phi_{fr} = \phi_{lr} = 0.01 \))
- 9. Urban-to-rural restrictions (\( \phi_{fr} = \phi_{lr} = -0.01 \))
- 10. Cross-location congestion (\( \theta_f = 1.00 \))
- 11. Combined (\( \theta_f = 1.00, \phi_{fr} = \phi_{lr} = 0.01 \))
- 12. Combined (\( \theta_f = 1.00, \epsilon_f = \epsilon = 0.92 \))

**Population growth changes:**
- 13. Zero population growth (\( CBR = CDR \))
- 14. Family planning (\( CBR = 20 \))

**Notes:** The table shows the simulated outcomes in 2055 using different policy interventions. Each simulation uses the parameters of Table 2, and demographic rates in 2005 (\( CBR_r = CBR_l = 35; CBR_f = 20; CDR_r = 10; CDR_l = CDR_f = 7.5 \)), unless otherwise noted. Welfare is the equivalent variation in net wage necessary for the baseline economy (row 1) to match welfare in the given scenario.