A WORLD OF CITIES: THE CAUSES AND CONSEQUENCES OF URBANIZATION IN POORER COUNTRIES

Edward L. Glaeser
Harvard University

Abstract
Historically, urban growth required enough development to grow and transport significant agricultural surpluses or a government effective enough to build an empire. But there has been an explosion of poor mega-cities over the last 30 years. A simple urban model illustrates that in closed economies, agricultural prosperity leads to more urbanization, but that in an open economy, urbanization increases with agricultural desperation. The challenge of developing world mega-cities is that poverty and weak governance reduce the ability to address the negative externalities that come with density. This paper models the connection between urban size and institutional failure, and shows that urban anonymity causes institutions to break down. For large cities with weak governments, draconian policies may be the only way to curb negative externalities, suggesting a painful trade-off between dictatorship and disorder. A simple model suggests that private provision of infrastructure to reduce negative externalities is less costly when city populations are low or institutions are strong, but that public provision can cost less in bigger cities. (JEL: R00, O10, O18)

1. Introduction

Between 1950 and 2010, the world's urbanization rate increased from under 30% to over 50% (United Nations 2012). In some countries, such as China and Korea, urbanization accompanied income growth, following the familiar historic pattern. But in other places, such as Haiti and Pakistan, urbanization occurred despite persistent poverty and weak governance. What explains the emergence of poor mega-cities like Karachi and Kinshasa? How different are their urban problems from those faced by the world's wealthier cities?

Section 2 of this paper documents the rise of poor country urbanization. In 1960, the urbanization rate was under 10% in most nations where per capita annual income was below $1,000 (all dollar amounts are 2012 US dollars). No similarly poor country is that rural today. The United States became one-third urbanized in 1890, when per

---

The editor in charge of this paper was Fabrizio Zilibotti.

Acknowledgments: Yueran Ma provided great research assistance. Giacomo Ponzetto provided helpful comments on the paper. The Taubman Center for State and Local Government provided financial support.
E. Glaeser is a Research Associate at the NBER.

E-mail: eglaeser@harvard.edu

Journal of the European Economic Association October 2014 12(5):1154–1199

DOI: 10.1111/jeea.12100

capita Gross Domestic Product (GDP) was over \$5,000. Poor countries today have become one-third urbanized with per capita GDPs under \$1,200. Cities with around one million people have existed historically in polities with similarly low income levels, including classical Rome, Baghdad, and Kaifeng, but these were capitals of capably governed empires. For the first time in history, the world has numerous mega-cities in countries that are both poor and poorly governed.

Despite the increasing urbanization of poor countries, the cross-country link between urbanization and income is even stronger today than it was in 1960. The highly urbanized nations are much wealthier than they were in the past and much wealthier than moderately urbanized nations today. Since 1960, the poorer nations that have the most income growth also had the highest initial urbanization levels. These facts might make us wary about policy proposals that mandate less urban growth.

After documenting these facts, Section 3 proposes a simple explanation of these phenomena that draws on Matsuyama (1992) and Gollin, Jedwab, and Vollrath (2013). Following those papers, I argue that globalization alters the urbanization process. In an age of autarky, cities must be fed with domestic agricultural surpluses shipped via effective transportation infrastructure. In today's more globalized world, Port-Au-Prince can be fed with water-borne American rice.

The model shows that in closed economies, urbanization increases with agricultural productivity or transport improvements, but in open economies, these comparative statics reverse. A closed economy with unproductive soil and weak infrastructure will have tiny cities, but an identical open economy will become a mega-city fed by trade with the outside world.

To test the globalization hypothesis, I examine the link between agricultural productivity, country size, and urbanization in 1961 and 2010. Using country size as a proxy for openness, I find that in both years, agricultural productivity is far more positively correlated with urbanization in large countries than in small countries. I also find that the connection between local agricultural productivity and urbanization fell dramatically between 1961 and 2010. This decline is compatible with the view that a growing global food trade means that urbanization can now proceed without a domestic agricultural surplus.

This poverty is particularly problematic because the world's poorest mega-cities have particularly meager resources and particularly massive problems. Urban density generates both positive and negative spillovers, such as congestion, contagious disease, and crime. These problems cannot always be managed by rich cities, so it is unsurprising that poor and poorly governed mega-cities suffer particularly from the negative externalities associated with density.

Section 4 focuses on the appropriate responses to urban externalities, which depend on both city size and institutional capacity. People create externalities, by driving on crowded roads or carelessly disposing of human waste, and the costs of these externalities rise with city size. Limited institutional capacity is captured by assuming a maximum expected penalty that can be imposed on misbehaving law enforcement officials. This maximum is presumably determined by the level of corruption and collusion within the system. Places with better institutions can better punish police malfeasance. If private citizens can bribe with take-it-or-leave-it offers, then the maximum penalty that can be imposed on policemen also equals the maximum legal penalty that can be imposed on private citizens. Arrested citizens can always avoid higher legal penalties by paying a bribe that exceeds the policemen's expected penalty for taking the bribe.

As in Glaeser and Shleifer (2003), the government can respond to externalities with ex-post punishment, ex-ante restrictions, or no punishment (i.e., anarchy). The optimal policy depends on the combination of city size and institutional strength. In small cities and villages, the social costs of bad behavior are low, so anarchy is optimal since it avoids the fixed costs of establishing a legal system.

In large cities with strong institutions, ex-post punishment is optimal. Strong institutions enable punishments large enough to deter all malfeasance. In sufficiently large cities, eliminating negative externalities is valuable enough to cover the fixed costs of the legal system.

As in Djankov et al. (2003), there is an institutional possibilities frontier that determines whether ex-post punishment can be effective. If institutional quality falls below a threshold, so that effective penalties become too low, then ex-post punishment cannot deter malfeasance. As the probability of catching a perpetrator declines with city size (Glaeser and Sacerdote 1999), bigger cities need bigger punishments to deter bad behavior. Accordingly, the minimum institutional threshold for effective ex-post punishment is higher in larger cities. Institutions that were strong enough to deter misbehavior in a smaller city may prove inadequate as city size increases and the probability of catching a criminal falls.

In metropolises with weak institutions and large populations, including many of the world's poorest mega-cities, anarchy creates terrible costs and they do not have the option of effective ex-post punishment. In such cases, the third option—ex-ante restriction—may be the best possibility. Ex-ante restriction is meant to capture policies like curfews or stop-and-frisk that try to stop harmless activities that can lead to bad behavior. If these harmless activities bring lower expected returns than real misbehavior or if they are easier to observe, then they can be deterred with punishment levels low enough to be feasible even when institutions are weak. The downside of ex-ante prevention is that many noncriminals are stopped for perfectly benign activities, but if city sizes are large and institutions are weak, this may be the best option. Many cities from London to Singapore have accepted less liberty as the price for taming urban chaos.

If criminals cannot make take-it-or-leave-it offers to policemen, and bribes are instead determined following a Nash bargaining rule, then large formal ex-post punishments can always effectively deter crime. By setting a sufficiently large formal penalty—which will never be paid—bribe levels grow big enough, even with weak institutions, to deter crime. In this case, the primary risk is that policemen will attempt to also extort the innocent. As in Djankov et al. (2003), high penalties reduce "disorder", the social costs of private malfeasance, at the cost of more "dictatorship", the social costs of public malfeasance.

Section 5 turns to the provision of infrastructure and its relationship with city size, wealth, and institutional strength. Infrastructure provisions, such as sewers, aqueducts, and extra highways reduce the downsides of density. Infrastructure may either complement or substitute good behavior. Roads may be less valuable if they are congested (complements), but aqueducts may be more valuable if people are polluting the local well (substitutes). Institutional quality will make infrastructure more attractive when infrastructure and good behavior are complements, but less attractive when they are substitutes.

Institutions also influence the cost of providing infrastructure. Following Engel, Fisher, and Galetovic (2013), Section 5 also examines institutional arrangements for providing infrastructure such as public–private partnerships. Public–private partnerships can reduce the waste and corruption that occur with public provision when institutions are weak. The downside of private provision is that private infrastructure builders will themselves corrupt the state, especially if the infrastructure requires implicit or explicit subsidies. I focus on the land acquisition process, which has been particularly important in Chinese cities, which may be subverted by private developers. The model suggests that private infrastructure provision will be less attractive in big cities, because high land values increase the social losses when private developers subvert the land acquisition process.

Section 6 concludes. The mega-cities of the developing world have significant problems that are impossible to eradicate given the current combination of weak institutions and poverty. Yet it seems likely that the process of urbanization itself is the most likely path towards the prosperity and institutional strength that will eventually lead to more livable cities.

2. The Rise of Poor Mega-cities

In 1960, poor nations were overwhelmingly rural. South Korea was one of a handful of poor states with an urbanization rate over 25%. The majority of poor countries are now more urbanized than Korea was then. According to United Nations' data, the urbanization rate in less developed countries rose from 18% in 1950 to 47% in 2011 (United Nations 2012). In 50 years, China has grown from 16% urban to over 50% urban. Botswana has grown from 3% to 60% urban. Botswana and China fit the historical pattern where urbanization has accompanied economic growth. Real per capita incomes are 17 times higher in Botswana today than they were in 1960. China's real per capita income has increased 25-fold.

Yet rapid urbanization is also occurring in nations where incomes have remained stagnant. Over the past half century, urbanization has increased from 5% to 28% in Bangladesh and from 7% to 24% in Kenya. In both countries, per capita incomes have risen by less than 500 real dollars over the same period. Moreover, in poorer nations, urbanization has often meant the outsized growth of a single dominant metropolis.

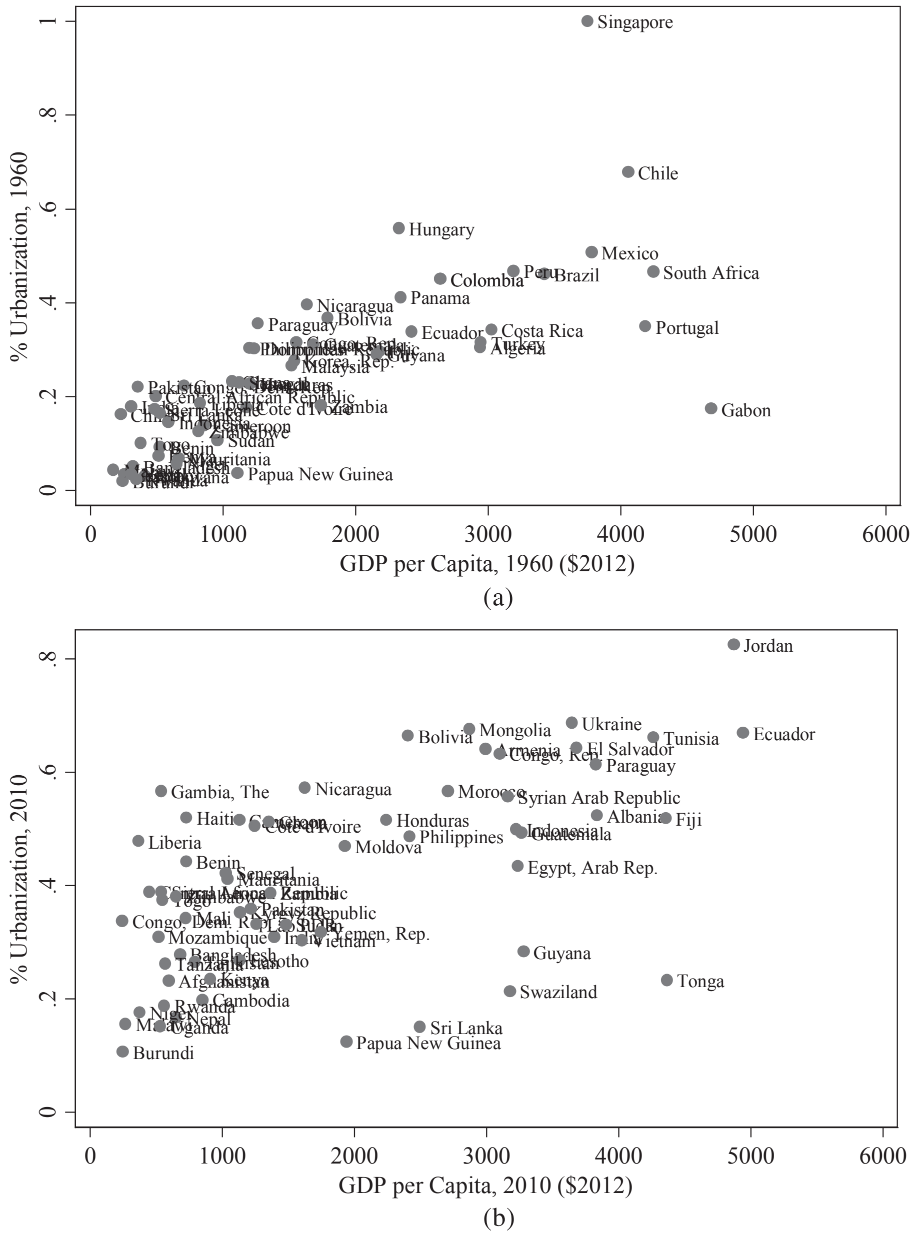

FIGURE 1. Relationship between per capita incomes (in 2012 dollars) and urbanization: (a) 1960; (b) 2010. Source: World Bank.

Three million people inhabit the Nairobi agglomeration and Dhaka is home to 15 million inhabitants.

Figures 1(a) and 1(b) both display the relationship between per capita incomes in 2012 dollars and urbanization across those countries that have per capita incomes

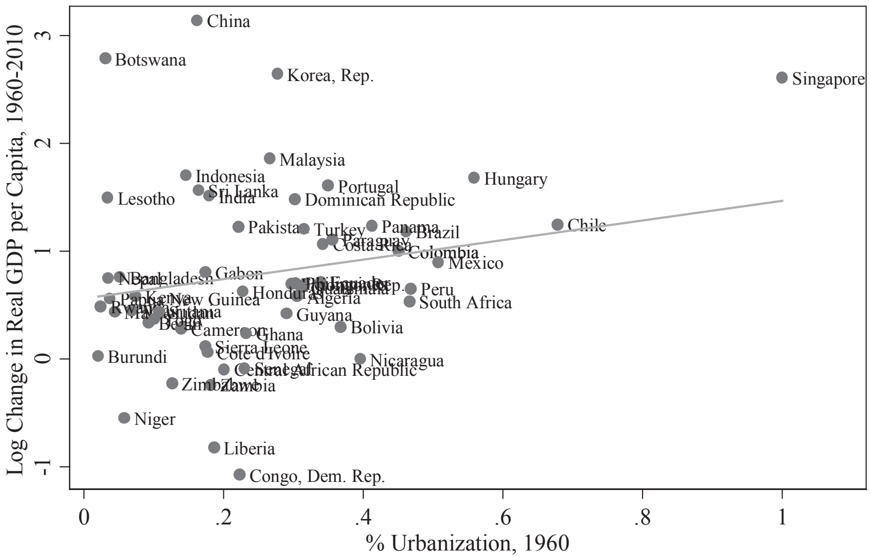

FIGURE 2. Relationship between initial urbanization and GDP growth. Source: World Bank.

below \$5,000. Figure 1(a) shows the strong positive relationship for 1960, where there are no really poor places with high levels of urbanization. Figure 1(b) shows the same relationship for 2010. Not only has the overall level of urbanization increased, but the growth has been particularly dramatic in poor countries.

These graphs do not mean that the overall connection between urbanization and income has declined, but rather that it has stayed roughly constant. If the logarithm of per capita GDP is regressed on urbanization, across all countries, the estimated coefficient is 4.75 (standard error of 0.32) in 1960 and 5.3 (standard error of 0.36) in 2010. Even among countries with per capita GDP levels below \$5,000 (in 2012 dollars), the estimated coefficient when the logarithm of per capita GDP is regressed on urbanization has been roughly constant, around 3.3 over the 50-year period. However, among poorer nations, the r-squared of that regression has dropped significantly, from 0.58 to 0.33, reflecting the increasing number of extremely poor, urbanized nations.

Among poorer nations, there is also a link between initial urbanization and GDP growth. Figure 2 shows that among poorer countries, a 10% higher urbanization rate in 1960 is associated with a 14% larger increase in per capita GDP over the next 50 years. I am not suggesting a causal relationship between urbanization and GDP growth, but this robust correlation should make policy makers pause before embracing strategies aimed at reducing city growth.

To provide a more concrete sense of poor but urbanized places, Table 1 lists the eight nations in my sample where incomes are below \$1,250, populations are over ten million, and urbanization is over one-third: the Democratic Republic of the Congo, Zimbabwe, Mali, Haiti, Pakistan, Senegal, the Cote D'Ivoire, and Cameroon. Haiti and Pakistan are the only two non-African countries. Every country has a one

TABLE 1. Urbanization.

Country	Largest city (population)	Percent urbanized	Percent in the million-plus agglomeration	GDP per capita (2012 $)	GDP per capita (PPP, 2012 $)
Democratic Republic of Congo	Kinshasa (8.4 million)	34%	12%	$251	$398
Zimbabwe	Harare (1.5 million)	39%	12%	$703	missing
Mali	Bamako (1.9 million)	35%	12%	$724	$1,246
Haiti	Port-au-Prince (2.1 million)	53%	21%	$760	$1,191
Pakistan	Karachi (13.5 million)	36%	8%	$1,229	$2,680
Senegal	Dakar (2.9 million)	43%	23%	$1,018	$1,899
Cote D'Ivoire	Abidjan (4.15 million)	51%	21%	$1,162	$1,875
Cameroon	Douala (2.3 million)	52%	11%	$1,144	$2,234

Source: World Bank data and United Nations (World Urbanization Prospects: The 2011 Revision).

million or more person agglomeration; three have metropolitan areas with four million people or more. Another eight smaller countries also have incomes below $1,250 and urbanization rates below one-third, including Liberia, which is almost as poor as the Democratic Republic of the Congo and the Kyrgyz Republic.

In 1960, no countries with urbanization rates over one-third had per capita incomes under $1,250 in 2012 dollars and only six had incomes under $2,500. Five out of those six, including Nicaragua and Ecuador, were in Latin America and one was in central Europe (Hungary). Notably, both regions were physically near wealthy trading partners in Europe and the United States. In 1960, urbanization in South American nations was typically about 7% higher than income levels would predict. The region's high and highly centralized urbanization was considered to be an important puzzle (Durand and Paláez 1965). Latin America has remained unusually urbanized, but since the region is much wealthier, it is no longer the epicenter of impoverished mega-cities.

To put the urbanization of today's poorer nations in perspective, Figures 3 and 4 show the time path of urbanization for the United States and then for England and Wales (based on Friedlander 1970).[1] The United States was one-third urbanized in 1890, when its income was nearly $6,000 in 2012 dollars, and one-half urbanized in the 1920s, when its income was close to $10,000. The more economically open United

1. The implied time series coefficient when urbanization is regressed on income is 0.24 in the United States and 0.25 for England and Wales, which are both relatively close to the cross-sectional coefficients discussed earlier.

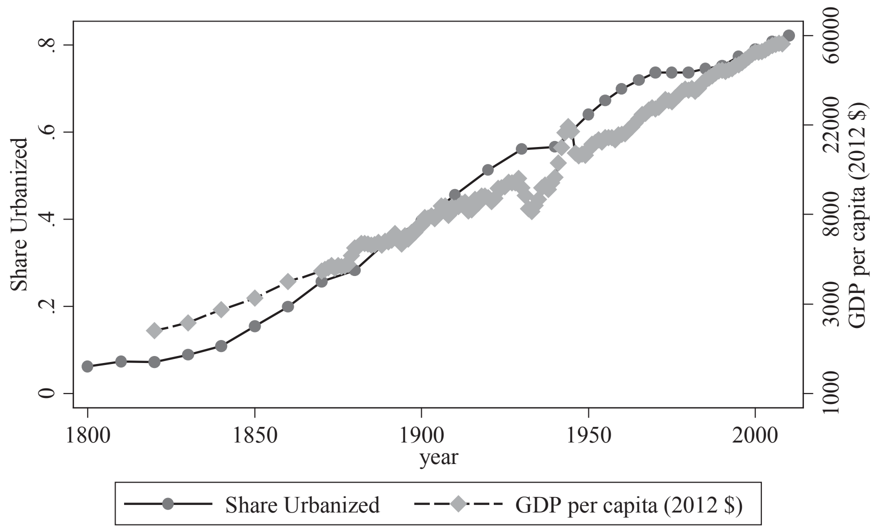

FIGURE 3. Urbanization and GDP in the USA. Sources: European Urbanization 1500–1800 by Jan de Vries, United Nation World Urbanization Prospects, Maddison Project Database.

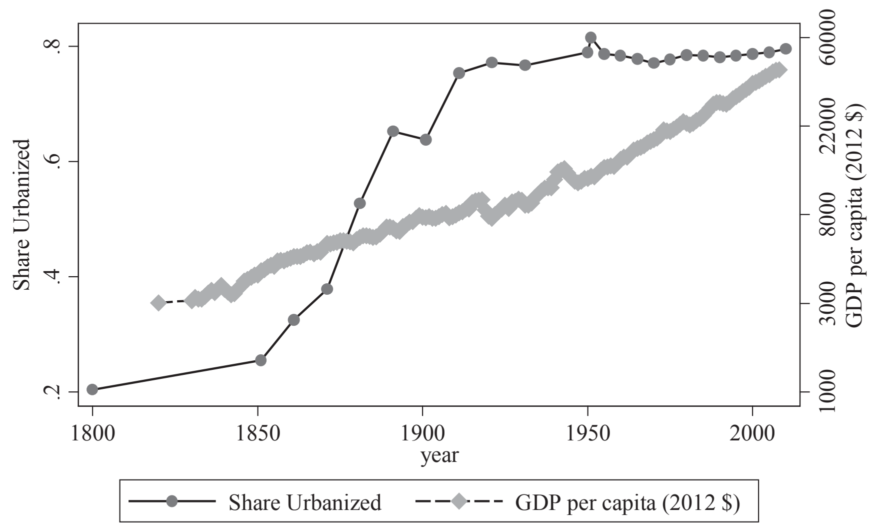

FIGURE 4. Urbanization and GDP in the UK. Sources: European Urbanization 1500–1800 by Jan de Vries, Friedlander (1970), United Nation World Urbanization Prospects, Maddison Project Database.

Kingdom became one-third urbanized in 1861 when its income was around $5,000, and one-half urbanized in 1881, when its incomes were closer to $6,000. In France, Germany, and the Netherlands, urbanization levels did not reach 50% until income levels were well over $5,000 (DeVries 1984). These incomes all significantly exceed those found in poorer, urbanized places today.

There were mega-cities in the more distant past with substantially lower income levels. Table 2 lists those cities that may have reached a population of around one

TABLE 2. World's largest cities through history.

World's largest cities throughout history	Year reached one million (world maximum until 1800)	Approximate per capita income (2012 $)	Notes
Roman Empire	1	$1,400, Italy	Fed with grain from Spain and Egypt, provided because of imperial power
		$1,000, Roman Empire	
Chang'an Kaifeng Hangzhou	800 1000 1200	$940	Political centers of a vast empire; figures highly debatable
Baghdad	900	$1,140	Highly centralized Abbasid Caliphate
Beijing	1800	$1,050	Small share of overall country
Edo/Tokyo	1725	$1,000	Metro region
London	1810	$3,000	Just predates mass industrialization
New York	1875	$4,565	First nonpolitical mega-city

Source: Tertius Chandler, *Four Thousand Years of Urban Growth: An Historical Census* (1987); Angus Maddison, Statistics on World Population, GDP and Per Capita GDP, 1–2008 AD.

million before 1875. Rome apparently had one million inhabitants around the time of the Julio-Claudian dynasty, when per capita incomes in the Empire were about $1,000 in today's dollars. Incomes in Italy were somewhat higher. The capitals of China and the Abbasid Caliphate reached that population level between 800 and 1200, at substantially lower national income levels (Bolt and van Zanden 2013). Beijing and Tokyo reached one million inhabitants between 1700 and 1800, also at about $1,000 per capita. By contrast, London and New York did not reach the one million person threshold until their national incomes reached $3,000 and $4,565, respectively.

Every one of these cities, except for New York, was a capital of a large empire. These empires existed precisely because their governments were capable of conquering and administering vast land areas. These countries did not have per capita incomes comparable to those in the United States in 1880, but they had governments strong enough to bring food to the capital and to battle the downsides of density.

Julius Caesar, for example, fought traffic congestion in Rome by forbidding carts from driving in the city during the first ten hours of the day. Public waterworks and sewer systems were renowned in Rome, Baghdad, and Kaifeng. These places were not wealthy, but they had governments competent enough to build and maintain empires. Kinshasa does not.

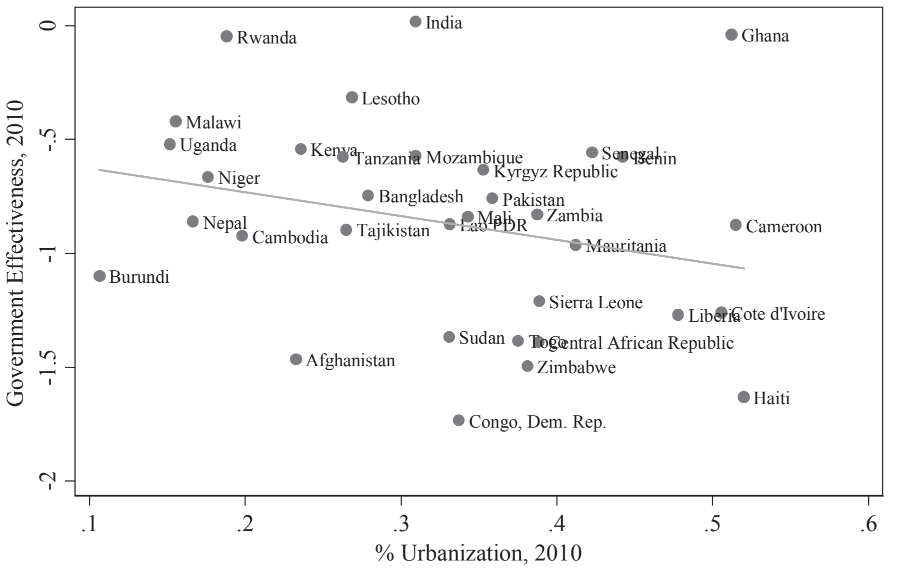

FIGURE 5. Relationship between government effectiveness and urbanization in 2010. Source: World Bank.

Today's poor but urbanized nations cannot rely on public sector competence. The overall correlation between government effectiveness and urbanization is significantly positive, but only because richer countries have better governments. Among poorer countries, any positive correlation disappears entirely. Figure 5 shows the weak negative correlation between 2010 government effectiveness and urbanization across the 34 countries with per capita incomes under $1,500 and populations over two million. The poor-but-urbanized nations of Haiti and the Democratic Republic of the Congo have among the world's lowest ratings of governmental effectiveness.

3. Understanding the Rise of Poor Mega-cities

I now turn to a positive model of urbanization, prosperity and openness. The model is in the spirit of Krugman (1991), but is particularly indebted to Matsuyama (1992) and Gollin, Jedwab, and Vollrath (2013).[2] The model's main point is that urbanization depends on rural prosperity in closed economies, but in open economies, rural desperation will push farmers to the city. These results suggest that globalization may be responsible for the unprecedented emergence of mega-cities in poor countries, which can now feed themselves with imported grains.

I consider a single city in the middle of a one-dimensional land area that is $2\bar{d}$ units long. The total population split between farms and city equals N. The city occupies no

2. Unlike Matsuyama (1992), this paper is specifically about urbanization rather than industrialization, and unlike Gollin, Jedwab, and Vollrath (2013), urbanites here produce traded goods rather than services.

land and makes manufactured goods that can be shipped at no cost. The agricultural land produces farm goods with iceberg transportation costs. If one unit of the goods is shipped "d" units of distance away, then $e^{-\tau d}$ goods arrive. All agricultural land must be sent to the city to be traded, so the distinguishing characteristic of any farm is its distance "d" from the city center, which ranges from 0 to $\bar{d}$.

The agricultural sector is characterized by a distribution of land density (which is uniform), a distribution of population density, denoted $n(d)$, which is a function of the distance from the city (d), and a distribution of agricultural output. Output per worker equals $\underline{C} + A_A L^{\gamma}$, where $\underline{C}$ and γ are constants, L represents land per worker, and A_A represents agricultural productivity. All parameters are weakly positive and $0 < \gamma < 1$. As land per worker is the inverse of density, output per worker equals $\underline{C} + A_A n(d)^{-\gamma}$, and total agricultural output at distance d from the city center will equal $\underline{C} n(d) + A_A n(d)^{1-\gamma}$.

Individual utility is $\theta_i / (\alpha^{\alpha}(1-\alpha)^{1-\alpha}) \, (C - \underline{C})^{\alpha} \, M^{1-\alpha}$, where θ_i represents local amenity levels which equal 1 in the rural sector, C represents consumption of the primary agricultural good, M represents consumption of the composite manufacturing good, and α is a constant between zero and one. For algebraic convenience, I have assumed that the minimal level of farm production ($\underline{C}$) also equals the minimal level of food needed for survival. I normalize the price of the agricultural good in the city to equal one; the price of the manufactured good is denoted by P_M.

I assume that labor is free to move between farm and city and across different farm locations, and this implies a spatial equilibrium where utility levels are equalized across space. I assume that agricultural property rights are weakly defined, so that farmers are squatters. Farmers can locate anywhere and share in the total agricultural output produce of that place. A farmer at distance d will earn $A_A n(d)^{-\gamma} P_M^{\alpha-1} e^{-(1-\alpha)\tau d}$.[3] The spatial equilibrium then implies

$$n(d) = e^{-((1-\alpha)\tau/\gamma)d} n(0),$$

so population density decays exponentially with distance to the city.[4]

If the urban sector has a population level of N_U, I assume that εN_U urban residents are entrepreneurs who operate independent firms producing differentiated products, and the rest $(1-\varepsilon)N_U$ of urban residents are manufacturing workers employed by the entrepreneurs. Each manufacturing worker produces A_M units of manufactured goods. Following Ethier (1979), these products are aggregated at no cost into a nondifferentiated manufactured good, so that if each firm "i" produces $x(i)$ units of the differentiated good, the total supply of the nondifferentiated good equals $(\int (x(i)^{\sigma} di)^{1/\sigma}$. Manufactured goods cannot be shipped until they are aggregated,

3. The results are similar if farmers play rents, although I would then have to account for the distribution of those rents. One important difference is that since farmers receive the average output of a production process with diminishing returns, there is overcrowding of rural areas. Essentially, this is a squatter society.

4. If there are N_A farmers in this zone, then welfare levels in rural sector equal $A_A N_A^{-\gamma} P_M^{\alpha-1} (2 - 2e^{-\psi \bar{d}})^{\gamma} \psi^{-\gamma}$, where ψ denotes $(1-\alpha)\tau/\gamma$, which is declining in the size of the agricultural sector.

which provides the justification for agglomerating industrial activities in the city as in Ciccone and Hall (1996).

Firms will follow the usual constant markup policy, and total output in the city will equal $A_M N_U^{1/\sigma}(1-\varepsilon)\varepsilon^{(1-\sigma)/\sigma}$. Workers' wages, denominated in the agricultural good, will equal $P_M\sigma/((1-\varepsilon)N_U)$ times total output; entrepreneurial profits equal $P_M(1-\sigma)/(\varepsilon N_U)$ times total output. If there is free entry into entrepreneurship at the cost of one's time, then wages and entrepreneurial profits are equal and $1-\sigma=\varepsilon$. I will instead assume that a fixed proportion of the urban population, $\varepsilon < 1-\sigma$, become entrepreneurs, so that a proclivity towards entrepreneurship becomes an important urban asset.[5] Workers only learn if they are entrepreneurs when they come to the city, and they cannot leave the city if they learn that they are not an entrepreneur, so the spatial equilibrium then implies that expected utility in the city equals utility in the rural sector.[6]

I assume that quality of life in the city (θ_i) equals $\theta_0 e^{-\delta N_U}$, where $\theta_0 > 0$ and δ are parameters. The relationship between city size and quality of life can capture various effects of crowding including contagious disease, congestion, and even housing costs.[7] With these assumptions, Proposition 1 follows. All proofs are in the Appendix.

PROPOSITION 1. *There exists a unique spatial equilibrium with a positive level of urbanization where residents are indifferent between the agricultural sector and the urban sector. In equilibrium, the share of workers in the urban sector is independent of* A_M *and* ε*, rising with* A_A*, and falling with* N*,* δ*,* τ*, and* $\underline{C}$*.*

The equilibrium is unique because even though the productivity in the city goes to zero as the number of people in the city gets small, the price of manufactured goods also goes to infinity. Proposition 1 illustrates that within a closed economy, agricultural productivity must pave the way towards urbanization. Since individuals have a fixed amount of calories that they need to consume, agriculture must be productive enough to support a large urban population. Manufacturing productivity has no impact on the level of urbanization, because a lower price of manufactured goods perfectly offsets the rise in per worker productivity.

The transportation system critically impacts the urbanization process in a closed economy, because if food does not make it to market, then the cities cannot be fed. From Augustus' Rome to Eisenhower's America, transportation networks have disproportionately depended on the power of a capable state.

5. Alternatively, there could be a fixed number of entrepreneurs in the economy, or there could be skilled and unskilled workers and only skilled workers have a chance to become an entrepreneur. This latter assumption can provide a micro-foundation of human capital externalities (Glaeser, Ponzetto, and Tobio 2010).

6. This assumes that earnings for both entrepreneurs and urban laborers exceed $\underline{C}$. To assume only that expected earnings across both groups exceed $\underline{C}$, as I do later, it is necessary to assume risk-sharing across urbanites.

7. Since the city occupies no land, to be technically correct, congestion would have to take the form of long lines at a skyscraper, and costs of living would have to represent the cost of high-rise apartments.

Urbanization declines with the level of population in this system. As more people are born, they do not flock to the cities. They must stay on the farms to ensure that people are sufficiently well fed. A natural parallel to this in European history is that it is alleged the Black Death made urbanization easier by increasing the per capita food supply.

In the spirit of Matsuyama (1992), I now consider urbanization in an open economy, meaning that the city has access to a harbor and anyone who comes to the city can buy or sell manufactured goods at a now exogenous price, denoted P_M, again denominated in units of the agricultural good, even if no one lives in the city. I assume that $1/\sigma$ may be greater than $1 + \delta$, so that there are potentially increasing returns to urbanization, $1/\sigma < 2 + \delta$.

In this case, it is quite possible that there are multiple equilibria reflecting the increasing returns in the urban sector, as described by the following proposition.

PROPOSITION 2. *An equilibrium with a positive level of urbanization exists if and only if*

$$(1 - \varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\, A_M P_M > \underline{C},$$

and A_A is less than a cutoff value of A_A^*, which is rising with A_M, P_M, θ_0, and τ, and falling with $\bar{d}$, δ, and $\underline{C}$. The value of A_A^* is rising with ε and also with N as long as δ is sufficiently small. If an equilibrium exists with positive urbanization, then generically, there will exist two levels of urbanization that satisfy the spatial equilibrium condition. At the equilibrium with less urbanization, increases in urbanization will cause welfare in the urban sector to exceed welfare in the rural sector. At the equilibrium with higher urbanization, further increases in urbanization will cause urban welfare to fall below rural welfare.

If rural welfare exceeds urban welfare at all levels of urbanization, then everyone farms and buys manufactured goods produced abroad. The possibility of a city only exists if

$$(1 - \varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\, A_M P_M > \underline{C},$$

which ensures that urban workers will not starve, and if A_A is not too high. The no-starvation condition assumes risk-sharing between workers and entrepreneurs. If that risk sharing does not exist, then cities require that $\sigma\,(\varepsilon N)^{(1-\sigma)/\sigma} A_M P_M > \underline{C}$, a more stringent condition, so that laborers do not starve. An urban equilibrium also requires that agricultural productivity is not too high.

Proposition 2 highlights that if there is any urbanization at all, then there are multiple equilibria due to the power of agglomeration. In the closed economy, a small city means an expensive manufactured good, and that drew some people to the manufacturing sector. In the open economy, that price effect does not exist, and only agglomeration economies matter. Moreover, in open economies, zero urbanization is a decided possibility, although complete urbanization is not. If there are no farmers, then agricultural productivity becomes infinite.

The existence of multiple equilibria suggests that small changes in underlying parameters, such as urban or rural productivity, which would barely register in a closed economy, can create massive urban change in an open economy. These aspects of the model seem to fit the rapid switch of the developed world from being overwhelmingly rural to being significantly urban.

I now consider the comparative statics in the second larger urbanized equilibrium, which is "stable", in the heuristic sense that higher values of urbanization reduce the relative advantages of urbanization.

PROPOSITION 3. *If a spatial equilibrium exists with some urbanization, then at the equilibrium with a higher level of urbanization, the level of urbanization is rising with* A_M, P_M θ_0, *and* τ *and falling with* A_A, $\bar{d}$, *and* $\underline{C}$. *Urbanization will be rising with* ε *and also with* N *as long as* δ *is sufficiently small.*

Many of the comparative statics in Proposition 1 are reversed. Urbanization now depends positively on the variables that determine productivity or quality of life in the urban sector, such as A_M, P_M, and θ_0. Variables that decrease rural productivity, such as A_A, or τ, will cause urbanization to increase.

Together, Propositions 2 and 3 highlight the fact that in open economies, urbanization can be the result of misery rather than productivity. Because trade alleviates the need for agricultural productivity, cities can develop despite poor hinterlands and bad transportation. This represents an opportunity (i.e., the ability to escape terrible rural poverty) and a challenge (i.e., the difficulties of urbanization in poor and poorly governed nations).

Is there too much or too little urbanization in equilibrium? Even though the spatial equilibrium ensures that expected utility levels are equalized across space, there are three distinct market failures in both open and closed economies: adverse urban diseconomies, benefits from more land per worker in the rural sector, and agglomeration economies. The first market failure suggests that cities are too large; the second two suggest that cities are too small. At this point, we are far from being confident about the relative magnitude of these effects, especially in the developing world, and we cannot tell if the growing cities of the world's poorest countries are too big or too small.[8]

3.1. Mega-cities versus Dispersed Urbanization

What determines whether urbanizing countries develop dispersed cities, as in Europe or the United States, or a dominant mega-city, as in much of Latin America and sub-Saharan Africa? Ades and Glaeser (1995) argue that large primate cities are often political capitals of unstable or dictatorial regimes that channel rents to nearby citizens to keep the peace. Krugman and Livas (1996) suggest that the third world's mega-cities

8. Glaeser and Gottlieb (2008) discuss the estimation of these magnitudes in the United States. While there exists some clear evidence for both urban diseconomies and agglomeration economies, the imprecision of current estimates bedevils any policy advice about urban bigness.

may reflect import substitution policies that reduce international trade. The discussion that follows illustrates that the reverse can also be true, where globalization leads to more centralization.

I now consider the possible formation of a second urban center. I assume that the land segment is circular, so that it is possible to form a second city in the middle of $\bar{d}$ units of agricultural land. In the case of the open economy, both cities can trade with the outside world. In the case of the closed economy, I focus on the symmetric equilibrium where farmers will trade only with the nearest city. In both cases, the two cities split their countries into two economies identical to those discussed previously, except that they have half as much land and half as many people. To simplify matters, I also assume that $\underline{\boldsymbol{C}} = 0$.

A symmetric two-city equilibrium will always exist in the closed economy, for the same reasons that a one-city equilibrium always existed in Proposition 1. A symmetric two-city equilibrium is less likely to exist in an open economy than a one-city equilibrium, because there is a smaller range of potential city sizes that make urbanization attractive relative to rural life when there are multiple cities. Proposition 4 asks about the minimum population needed to form a second city that would attract urban pioneers, given that one city exists, and keeping the population of the first city fixed.

PROPOSITION 4. *In the open economy, a second city will only attract urban pioneers if it is larger than a threshold, which equals the size of the first city, if the first city is smaller than the welfare-maximizing city size* $((1-\sigma)/\sigma\delta)$*, or the minimal city size that exceeds agricultural welfare, if the first city is larger than welfare-maximizing city size. In a closed economy, a second city of any size can attract urbanites, as long as* $\tau\bar{d}$ *is sufficiently large and* $1 < (1+\alpha)\sigma$.

In an open economy, urban pioneers reap no advantages from being closer to farmers. Forming a second city depends only on the internal economics of agglomeration and congestion. In the open economy, urban welfare is first rising then falling with city. If the existing city is smaller than the welfare-maximizing size $((1-\sigma)/\sigma\delta)$, then any new city must be at least as large as the old city. If the existing city's population exceeds the welfare-maximizing point, then the new city must be large enough so that urban welfare exceeds rural welfare. At a minimum, a new city would need to be as large as the smaller equilibrium city in a one-city economy discussed in Proposition 2.

In a closed economy, however, a new city can start, no matter how small, as long as transport costs are sufficiently large and $1 < (1+\alpha)\sigma$. In the closed city, the urban pioneers make a market for nearby farmers, and a new city benefits from low-cost agricultural products.

In Europe today, a relatively low share of the population lives in cities with more than one million inhabitants, holding per capita incomes constant. The European difference is that these cities formed during a period of high transportation costs and limited global trade, at least in core agricultural products. Cities were dispersed

TABLE 3. Urbanization and agricultural productivity, 1961 and 2010.

	(1)	(2)	(3)	(4)
Year	1961	1961	2010	2010
Log of agricultural productivity	0.095***	0.051**	0.054***	0.00
	(0.024)	(0.021)	(0.018)	(0.019)
Log of agricultural productivity × Demeaned log of population	0.038***	0.025**	0.025***	0.021**
	(0.012)	(0.010)	(0.009)	(0.008)
Log of population	−0.205***	−0.134**	−0.157***	−0.122**
	(0.070)	(0.056)	(0.056)	(0.051)
Continent dummies	No	Yes	No	Yes
Observations	119	119	139	139
R-squared	0.189	0.531	0.085	0.304

Notes: Agricultural productivity is defined as cereal yield in kilograms per hectare times hectares per capita. The interaction between agricultural and population has demeaned the population in the given year so that the raw coefficient on agricultural productivity can be interpreted as the impact of agricultural productivity at the mean level of population. Source: Data come from the World Bank.
Standard errors in parentheses. A constant is included in the regression but not reported.
***Significant at 1%; **significant at 5%.

because each urban area was often a market town for nearby farmers. Nineteenth-century American cities also spread out to access local agricultural areas and other geological amenities, such as the coal mines near Pittsburgh. In the open, but poor, economies of the 21st century, there is little reason for urban areas to be near domestic agricultural resources, and urban populations, therefore, concentrate in mega-cities.

3.2. An Empirical Test

To test the hypothesis that globalization weakened the link between local agricultural productivity and urbanization, I examine cross-national data from 1961 and 2010. I use 1961, rather than 1960, because of greater availability of World Bank agricultural data from that year. I measure agricultural productivity as the product of cereal yield per hectare and arable land per capita in the country, which is an approximation of cereal yield per capita. As agricultural productivity itself partially reflects technological development, which in turn partially reflects urbanization, these results must be seen as being merely suggestive.

Table 3 regresses:

$$\begin{aligned}\text{Urbanization} &= a + b_1 \text{Log}(Prod_{Ag}) \\ &\quad + b_2 \text{Log}(Pop) + b_3 \text{Log}(Prod_{Ag}) * \text{Log}\,(Pop)\,,\end{aligned} \tag{1}$$

where $Prod_{Ag}$ represents the agricultural productivity measure and Pop represents population. I include population primarily because of the well-known correlation between openness and country size (Alesina and Wacziarg 1998). While actual trade flows themselves depend on the level of development and urbanization, country population is at least somewhat more independent. The coefficient on the interaction between population and agricultural productivity tests the hypothesis that the connection between urbanization and agricultural productivity is weaker in smaller, and hence more open, countries. In all regressions, I have demeaned the population variable in the interaction term, so the estimated coefficient b_1 can be interpreted as the impact of agricultural productivity at the mean level of population in the given year.

Regression 1 shows that as agricultural productivity increased by 1 log point (approximately doubled) in 1961, urbanization increased by 9.5 percentage points. This effect is, as predicted, much stronger in more populous countries. In a country that is one log point larger than the international average, the estimated impact of one extra log point of agricultural productivity is that urbanization rises to 13.3 percentage points. Holding other variables constant, more populous countries were less urbanized.

Regression 2, like Regression 4, includes continent dummies, which cause the estimated coefficient on agricultural productivity to drop by almost 50%. One-half of the cross-national relationship between agricultural productivity and urbanization in 1961 is explained by differences across continents. The agricultural productivity effect and the interaction between that effect and population remain significant even with continent dummies.

In Regression 3, we find that, without continent dummies, the coefficient of agricultural productivity is about 43% less in 2010 than in 1961. The interaction between productivity and population is also weaker, although both variables remain significant. These results support the hypothesis that local agricultural strength has become less important to city-building. Regression 4 shows that the basic agricultural productivity coefficient drops to zero in 2010, once continent dummies are included, although their interaction between productivity and population remains statistically significant.

These results do not prove the hypothesis illustrated by the model, but they do fail to reject the model's key implications. The link between agricultural productivity and urbanization has declined over time, and it is weaker in small countries than in big countries. By 2010, there is a slightly negative correlation between agricultural productivity and urbanization in smaller countries, which is exactly what the model predicts. We now turn to the second part of the paper: urban governance in poorly governed nations.

4. Externalities and Density

Cities can create great positive externalities that go far beyond merely reducing shipping costs for intermediate manufactured goods. Over human history, cities have spread the knowledge that creates everyday human capital and the new ideas that generate economic and cultural breakthroughs. Yet, cities also create powerful negative externalities. The same urban proximity that speeds the flow of goods and knowledge also spreads infectious disease and facilitates crime. The negative externality of congestion is practically synonymous with density.

I now turn to the causes of low quality of urban life: the externalities associated with density, including traffic congestion, water-borne or other contagious diseases, and crime. High housing costs are another downside of urban living, but I will not focus on them here because high prices are not externalities.[9] These downsides of density can be mitigated through a combination of infrastructure investment and behavioral modification. With enough money and public competence, even the densest agglomerations can enjoy a virtually unlimited supply of clean water through desalination and relatively uncongested roads through effective electronic road pricing. American cities have become safer thanks to an expensive combination of large-scale imprisonment and effective policing. An essentially unlimited supply of usable space can be built on a tiny island if buyers will pay the costs of scraping the sky.

Yet many developing world mega-cities have neither the wealth nor good government that these solutions entail. I now turn to the problem of urban externalities when institutions are weak. I first discuss crime and punishment and then turn to infrastructure in Section 5.

4.1. Addressing Urban Externalities: Anarchy and Ex-Post Punishment

In Djankov et al. (2003) and Glaeser and Shleifer (2003), the ability to punish bad behavior is limited by a society's institutional strength. This paper embeds that insight into an urban setting, where urban density increases the cost of bad behavior and decreases the probability of catching perpetrators. Consequently, effective punishment may be most valuable but least feasible in big cities. While the variants of models that follow are connected, they are completely distinct from the model described in Section 3, and as such, I reuse some of the previously used notation for different purposes.

Individuals first decide whether or not to take a "harmless action", such as walking around in the city, which carries a benefit of "h". If they take this harmless action, a proportion "v" of the population has the opportunity to undertake a "harmful action", such as introducing waste into a potential water source or stealing goods from a neighbor. Individuals only learn if they have the opportunity to undertake the "harmful action" after they undertake the "harmless action".

The harmful action generates private benefits of $B + A$ or B, and those benefits are known before individuals decide whether to generate the harm, but after they have taken the harmless action. A share πv of the population will receive the higher benefit level if they undertake the harmful action, and the remainder will receive the lower benefit level.

The social cost of the harmful action is $C(N)$, where N reflects the city size, where $C'(N) > 0$ and $C(0) > A + B$. Significantly, the action does more damage if the

9. High housing prices can however reflect urban policies, like restricting land uses.

city is larger or denser. If the harmful action is careless disposal of diseased waste, then the social cost rises with city size because more people may potentially become infected. If the harmful action is driving on a crowded street, then large city sizes mean that more people may be slowed by the congestion. If the action is crime, then the connection with city population is less clear, but a crime in a big city may generate more widespread fear and create more social costs from self-protection.

I assume an additive utilitarian social welfare function so the private benefits from misbehavior can be readily weighed against its social costs, so $C(0) > A + B$ implies that it is always better if the harmful action does not occur. I will consider three public responses to the harmful activity: anarchy, ex-post punishment of the harmful action, and ex-ante prevention. In this section, 1 will consider only the first two options, and in the next section 1 consider the costs and benefits of preventing the "harmless action" as a second best means of reducing the "harmful action".

The anarchy option means that there are no public penalties of any sort. Per capita social welfare is $h + v\,(B + \pi A - C\,(N))$, which sums the benefit from the harmless action plus the probability of being able to undertake the harmful action times the expected social costs of that action. The downsides of anarchy rise with city size, suggesting the complementarity between urban growth and effective government.

An ex-post punishment system costs NK_P to establish. These costs scale with city population, but if there were also fixed costs, that would create an additional reason for bigger government in bigger cities. If the system is established, the probability of catching someone who has perpetrated the harmful action is $d(N)$, where $d'(N) < 0$. Arrests per reported crime decline steadily with city size across American cities (Glaeser and Sacerdote 1999). Arrest per actual crime seems to fall even more steeply with urban population, because people in cities are less likely to report crimes, perhaps because they know that crimes are less likely to be solved. One reason why police are less likely to solve urban crimes is that crimes are often solved by considering the range of possible suspects, and the number of possible suspects is so much larger in a big city. Anonymity is a fact of urban life and anonymity bedevils attempts to track down malefactors.

Once the ex-post punishment system is established, the system can select a punishment level "P" applied to people caught doing harm. The social costs of that punishment equals λP, where λ may be less than one (if the punishment is a fine) or potentially greater than one (if the punishment is prison). If the punishment is actually administered, then it will deter individuals who have low benefits from doing harm if $d(N)P > B$, and individuals with high benefits from doing harm if $d(N)P > A + B$.

If the punishment was actually administered, then setting a penalty of $(A + B)/d(N)$ or more would deter all harm. If the punishment was administered, then ex-post punishment would dominate anarchy if and only if $v\,(C\,(N) - B - \pi A) > K_P$. I assume that this must hold for sufficiently high levels of N, but not for N close to zero, and that implies that ex-post punishment will dominate anarchy if and only if city population exceed some threshold. Since it costs no more to deter all crime than to deter some crime, it makes sense to deter all harmful behavior if formal punishments are all actually administered.

But institutional limitations may mean that formal punishments are not fully applied. I assume that the maximum expected punishment that can be imposed on a law enforcement officer who takes a bribe is $\bar{P}$. In countries with weak institutions, corrupt law officers are protected by a web of allies. When institutions are weak, police officers refuse to inform on other police officers and judges ignore official malfeasance. When democratic institutions are weak, there may be little popular pressure to punish official corruption.

I also initially assume that law breakers, when caught, make take-it-or-leave-it offers to policemen. These offers are then accepted if and only if they exceed the maximum expected punishment faced by a policeman who takes a bribe: $\bar{P}$. The maximum ability to punish police for accepting bribes becomes the maximum effective penalty on criminals as well. No matter how high the official penalties for crime may be, the maximum effective punishment will be no more than the minimum bribe level that is acceptable to a law enforcer: $\bar{P}$.

These assumptions suggest that city size and institutional quality come together to produce an "institutional possibility frontier", which determines the possible behavior that can be effectively deterred. If $\bar{P}d\ (N) < B$, then low probability of arrest, due to large city size, and weak institutions together mean that no misbehavior can be deterred. In this case, ex-post punishment does no better than anarchy at deterring crime, but costs more to establish. If $B < \bar{P}d\ (N) < A + B$, then ex-post punishment can deter some but not all misbehavior. In that case, the penalty should be set at $B/d\ (N)$, since higher penalties generate higher social costs with no offsetting benefits. If $\bar{P}d\ (N) > A + B$, then it is possible to deter all misbehavior, and good behavior can be created with a penalty of $(A + B)/d\ (N)$.

Figure 6 shows the two institutional possibility frontiers implied by those two equations. The top curve is an example of $d^{-1}(B/\bar{P})$ and the bottom curve illustrates $d^{-1}((A + B)/\bar{P})$. These curves illustrate the maximum level of city size in which ex-post punishment deters crime for a given level of institutions. Both curves rise with institutional quality, $\bar{P}$, reflecting the fact that ex-post punishment can be effective in larger cities with better institutions. When city sizes rise beyond the level implied by $d^{-1}((A + B)/\bar{P})$, then criminals with strong incentives cannot be deterred. When city size rises beyond the level implied by $d^{-1}(B/\bar{P})$, then no crime can be deterred.

Even if crime can be deterred, anarchy may still be better than ex-post punishment, as illustrated in Proposition 5. I assume that the social cost of a bribe equals $\lambda\bar{P} + \delta(Bribe - \bar{P})$, so that social welfare is always higher when misbehavior leads to formal punishment rather than to bribe-paying of equivalent value.[10] The $\delta(Bribe - \bar{P})$ term will be irrelevant here when the accused make take-it-or-leave-it offers, but will matter if bribes are determined more flexibly. I assume that $\lambda \geq \delta$.

10. This assumption is debatable and made solely to eliminate extraneous cases from consideration.

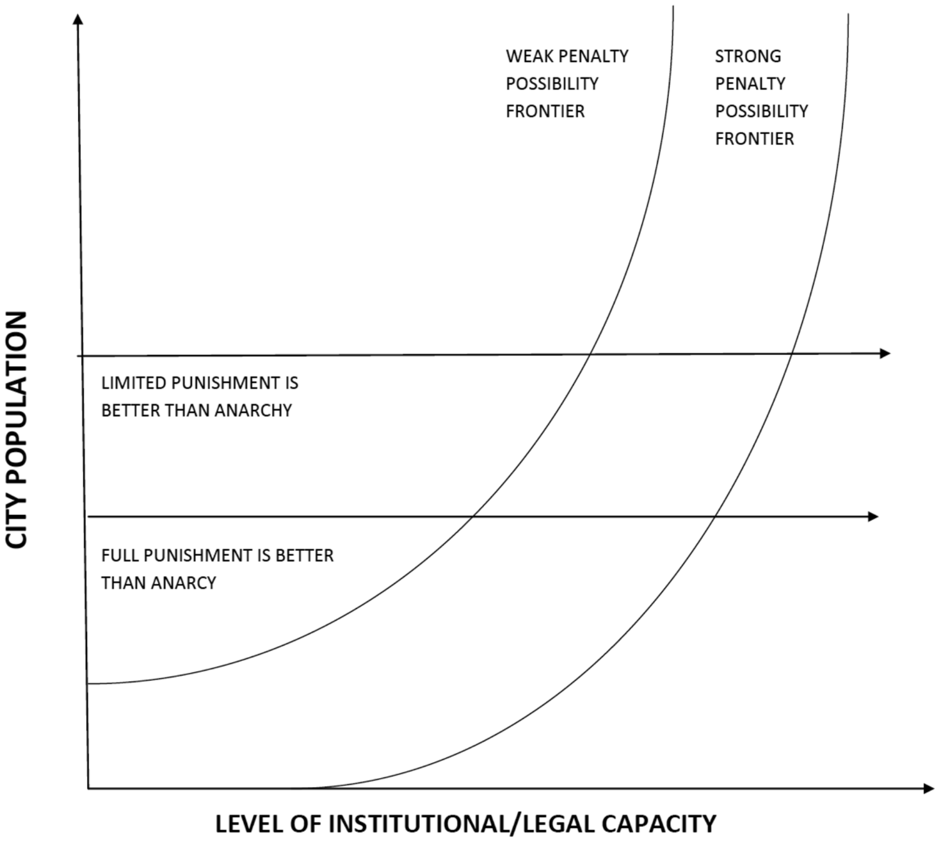

FIGURE 6. Institutional quality and city size. City size and institutional quality come together to produce an "institutional possibility frontier", that determines the possible behavior that can be effectively deterred. These curves illustrate the highest level of city size that permits penalties to be high enough to deter crime. For levels of population below the level of the bottom horizontal line, anarchy always carries the lower costs. For higher levels of city population, then full punishment carries lower costs than anarchy, but full punishment will not be possible when institutional quality is low.

PROPOSITION 5. *There exist two threshold levels of city population, denoted N^A_{LP} and N^A_{FP}, with $N^A_{LP} > N^A_{FP}$. For cities with population levels below N^A_{FP}, anarchy minimizes costs. For cities with population levels between N^A_{LP} and N^A_{FP}, if $\bar{P} > (A + B)/(d(N))$, then ex-post punishment deters all crime and minimizes costs, but if $\bar{P} < (A + B)/(d(N))$, ex-post punishment does not deter all crime and anarchy minimizes costs. For cities with population levels above N^A_{LP}, if $\bar{P} < B/(d(N))$, then ex-post punishment deters no crime and anarchy minimizes costs. If*

$$\frac{A+B}{d(N)} > \bar{P} > \frac{B}{d(N)},$$

ex-post punishment with a penalty of $B/(d(N))$ will deter some crime and minimize costs. If $(A + B)/(d(N)) < \bar{P}$, ex-post punishment can deter all crime and minimize costs. N_{LP}^{A} and N_{FP}^{A} are both rising with K_P, π, and B, and falling v. N_{FP}^{A} is also rising with A, and N_{LP}^{A} is rising with λ.

This proposition is illustrated in Figure 6. When city population lies below the bottom horizontal line, which illustrates N_{FP}^{A}, anarchy always carries the lower costs. For higher levels of city population, then full punishment—if it is effective at preventing misbehavior—creates lower costs than anarchy, but effective punishment is impossible when institutional quality is low. For moderately sized cities, anarchy dominates limited punishment but not full punishment, and so these cities remain in anarchy until institutional quality becomes high enough to support full punishment.

For the largest cities, the appropriate policy depends on institutional strength. When institutions are very weak, then anarchy is the only feasible option, which leads to sizable social losses since density implies large costs of misbehavior. When institutions are strong, then ex-post punishment can effectively eliminate all misbehavior and that policy minimizes costs. At intermediate levels of institutional strength, ex-post punishment is capable of deterring only misbehavior with low returns, but this is still preferable to anarchy.

The graph suggests that larger cities may see some form of externality-preventing law before smaller cities, but it may be imperfect. It also suggests that for some levels of institutional quality, the connection between order and city size might be nonmonotonic. A small city that grows may initially adopt ex-post punishment and become more law abiding, but as the city grows further, this policy becomes ineffective and anarchy again rules until institutions are improved.

While order may have a nonmonotonic relationship with city size, Proposition 5 and Figure 6 suggest that law and order monotonically increase with institutional quality, but this might not be the case. Many cities with low institutional quality have adopted quite draconian methods to address crime and other bad behavior, such as Julius Caesar's banning of all wheeled vehicles from Rome during the first ten hours of daylight. Singapore's early adoption of relatively tough penalties for minor abuses, such as spitting, might be another example. These policies, and New York's stop-and-frisk, can be seen as an alternative approach that veers towards dictatorship in an attempt to promote urban livability.

4.2. Ex-Ante Prevention: The Institutional Appeal of Stop-and-Frisk

To consider a wider range of policies, I now allow ex-ante prevention. In the model, this means that the government will impose penalties on individuals who take the harmless action, which always precedes the harmful action. The harmless action might just be walking after dark. The harmless action creates private benefits and no social costs, so there is always a social loss from ex-ante prevention.

There are many historic examples of governments adopting such an approach. Sabine (1937) reports the attempts to control waste in medieval London that included fines levied on individuals who left waste in public spaces (ex-post punishment) and locked gates preventing access to waterways (ex-ante prevention). Traffic congestion in classical Rome was reduced by Julius Caesar's ban on driving during certain hours. A somewhat more strained interpretation of that division is that it distinguishes between the standard ex-post punishment for crimes and stop-and-frisk policies that involve searching ordinary pedestrians who appear suspicious.

Stopping the harmless action may be easier than punishing the harmful action because the harmless action may be easier to detect and easier to deter. By assumption, people who take the harmless action do not know if they will also have the opportunity to commit a crime. If ex-ante prevention is effective, therefore, it deters everyone from the harmless action and consequently deters everyone from the harmful action as well. It is therefore unnecessary to impose added penalties on the harmful action if the harmless action is effectively banned.

The net expected private benefits of engaging in the harmless action equal "*h*" plus the net private benefits to committing the crime. I assume that the probability of catching someone engaged in the harmless action is $d_0(N) > d(N)$, whether or not the individual also engages in the harmful act, where $\lim_{N\to\infty} d_0(N) = 0$ and

$$\lim_{N\to\infty} \frac{d_0(N)}{d(N)} = 1.$$

The cost of establishing this system again equals NK_P. Otherwise ex-ante prevention would never be attractive. The maximum penalty is again $\bar{P}$, and the expected benefits of the harmless action are $h + vB + v\pi A$. I assume that

$$\frac{h + vB + v\pi A}{d_0(N)} < \frac{B}{d(N)},$$

so that it is always possible to engage prevention of harmless actions at a lower level of institutional quality than is needed to engage ex-post punishment with light penalties. This leads to the following proposition.

PROPOSITION 6. *Ex-ante prevention can be optimal only if*

$$\frac{b + vB + v\pi A}{d_0(N)} < \bar{P} < \frac{A + B}{d(N)}.$$

Ex-ante prevention dominates anarchy if and only if city size is greater than a threshold, N^A_{PR}, that is increasing with K_P, A, B, h, and π and decreasing with v. If

$$h < \pi\left(\frac{K_P + v\lambda B}{1-\pi} - vA\right),$$

then effective ex-ante prevention always dominates ex-post punishment with weak penalties, but if

$$h > \pi\left(\frac{K_P + v\lambda B}{1-\pi} - vA\right),$$

then effective ex-post punishment with weak penalties dominates ex-ante prevention and will be optimal and feasible if $\bar{P} > B/d(N)$, and city size exceeds N^A_{LP}.

Ex-ante prevention is only optimal when institutions occupy a middle ground. They must be weak enough so that effective complete ex-post prevention is not an option—that regime always creates lower costs. They must be strong enough so that ex-ante prevention actually deters the harmless action, which is why

$$\frac{b + vB + v\pi A}{d_0(N)} < \bar{P} < \frac{A + B}{d(N)}.$$

As in all cases involving public intervention, there is a minimum population threshold, denoted N^A_{PR}, below which anarchy dominates.

Yet even if the city population exceeds that threshold, and if institutional quality occupies the relevant middle ground, ex-ante prevention may still be worse than limited punishment. The choice between the two policies depends on "*h*", the returns to the harmless action. If the harmless action carries large private social costs, then imperfect ex-post punishment dominates ex-ante prevention whenever it is institutionally possible. If the harmless action is less valuable, then ex-ante prevention is more attractive.

This proposition suggests that draconian rather than light policies can be the cost-minimizing strategy when institutional quality is fairly low. Targeting the harmless action does create significant social costs—individuals are prevented from undertaking a benign action—but it is easier to target the harmless action because the probability of detection is higher and the penalties needed to prevent the harmless action are lower. Ex-post punishment requires stronger institutional strength because the penalties must be more severe to be effective, and as a result, they are more likely to lead to subversion.

One case where this strategy seems to have been followed is the stop-and-frisk tactics used by the New York Police Department. Crime had risen significantly in New York City between 1960 and 1975, and crime remained high in 1990, partially because of the crack epidemic. The New York Police Department followed somewhat more draconian tactics under Mayor Rudolph Giuliani and Bill Bratton, his chief of police, which were continued under Mayor Bloomberg. One particularly contentious approach has been stopping and frisking people who appeared "dangerous".

This policy can be seen as targeting anyone who appeared to be a threat. They could be searched and arrested if they carried something illegal, like drugs or a weapon. This greatly increased the probability of catching someone, although it was not obvious that the person was really causing much harm. The social costs of the policy were at least temporarily accepted because it was seen as bringing order to a city where the punishments and probability of arrest for serious crimes did not seem to deter crime. The standard logic of Becker (1968) suggests that the city could have raised the penalties, but it did not have the power to do that, not because of bribery risk (although that was also present during earlier years), but because of a lack of legal authority.

4.3. Flexible Bribery

We now drop the assumption that arrestees make take-it-or-leave it offers. We now assume that there is Nash bargaining between the criminal and the policeman. If the official penalty for a crime is P, and the policeman's expected penalty from taking a bribe is $\bar{P}$, then the bargained bribe will equal $\sigma \bar{P} + (1-\sigma)\, P$ if $P > \bar{P}$. The parameter σ reflects the bargaining power of the arrestee. If $P < \bar{P}$, then there will be no bribery and the arrested individual will pay the official penalty.

While the assumption of flexible bargaining may seem realistic relative to the take-it-or-leave-it assumption discussed earlier, I also assume that there are no credit constraints that limit the ability to extract large bribes. If those credit constraints bind, then there will be limits on punishment and the results will be similar to those discussed in the previous section.

For the simplicity of expression, here I drop the previous assumption that there are two levels of private benefits from taking harmful actions ($A + B$ and B) and instead assume that the private benefit is always equal to B. All misbehavior can be deterred if P is high enough so that $\sigma \bar{P} + (1-\sigma)\, P > B$, but the downside of very high penalties is that this creates the possibility of police extortion through false arrests (Friedman 1999). To capture this possibility, I assume that the police can accuse those they find engaging in the harmless action (which occurs again with probability $d_0\,(N)$) of perpetrating a misdeed, and unless a bribe is paid, this will cause ψP in harm to those people who are accused. They will therefore be willing to pay a bribe of $\sigma \bar{P} + (1-\sigma)\,\psi P$ as long as $\psi P > \bar{P}$. As before, I assume that bribery leads to a social loss of λ times the bribe. For simplicity, I also assume here that $\pi = 0$, and

$$vC'(N) - d_0'(N)\left(\frac{\delta\psi B}{d(N)} + (\lambda - \delta(1-\sigma\,(1-\psi)))\,\bar{P}\right) + \frac{\delta\psi B d_0(N) d'(N)}{(d(N))^2} > 0,$$

which essentially limits the slope of $d(N)$ relative to $d_0\,(N)$ and $C(N)$.

When bribery is flexible, sufficiently high punishments can always deter the harmless action and all crime.

PROPOSITION 7a. *If $\psi > 0$, then a sufficiently high level of formal ex-post punishment will deter the harmless action and all crime, and this will reduce costs relative to anarchy as long as N is greater than a threshold which is rising with K_P, h, and B and falling with v.*

The intuition of this proposition is straightforward. If penalties are sufficiently high, then the police will attempt to extort anyone who is out on the street. This will create a sufficient deterrent against the harmless action so that ex-ante prevention occurs. This policy of high formal penalties deters so much that it leads to no bribery in equilibrium, and this dominates anarchy when city sizes are sufficiently large. In this case, the ability to extort essentially eases the institutional possibilities frontier—ex-ante prevention becomes possible at all levels of institutional quality, because people are sufficiently afraid of police extortion.

It is somewhat more difficult not to deter the harmless action, but still to deter the harmful action, and it is impossible to do that without some social losses due to bribery.

PROPOSITION 7b. *If $B/d(N) < \bar{P}$, then it is possible to have effective ex-post punishment without bribery or extortion, and if*

$$\frac{\psi B}{(1-\sigma+\sigma\psi)\,d\,(N)} < \bar{P} < \frac{B}{d\,(N)},$$

then ex-post punishment can be effective (with bribery) but with no extortion of the innocent. If $(\psi B/(1-\sigma+\sigma\psi)d(N)) > \bar{P}$, then effective ex-post punishment will always involve extortion of the innocent, but this will not deter the harmless action if $h/(d_0(N)) > \psi B/d(N)$ and $\bar{P}$ is sufficiently low. Only ex-ante prevention is possible if

$$\frac{h}{d_0\,(N)} < \frac{\psi B}{(1-\sigma+\sigma\psi)\,d\,(N)}$$

and

$$\frac{\psi B}{(1-\sigma+\sigma\psi)\,d\,(N)} > \bar{P} > \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0\,(N)} - \frac{\psi B}{d\,(N)}\right).$$

Ex-post punishment with extortion, when it is possible, is always preferable to ex-ante prevention and preferable to anarchy if N is greater than a threshold level that is increasing K_P, B, λ, ψ, σ, and $\bar{P}$.

Proposition 7b describes the possible outcomes if a regime attempts to follow a policy of ex-post punishment. If institutions are strong, relative to city size, ex-post punishment can deter the misbehavior without either bribery or extortion. With somewhat weaker institutions, flexible bribery actually eases the institutional possibilities frontier and makes it possible to deter the misbehavior without leading to extortion. When institutions are weaker still, then extortion will also occur. Somewhat paradoxically, extortion is more likely to deter the harmless action when institutions are strong than when they are weak, as long as $h/(d_0(N)) > \psi B/d(N)$, because law enforcement will demand higher bribes.[11] If institutions are really weak, then with flexible bribery it is always possible to deter the harmful action without deterring the harmless action, again as long as $h/(d_0(N)) > \psi B/d(N)$.

There exists an institutional possibilities frontier defined by

$$\frac{\psi B}{(1-\sigma+\sigma\psi)\,d\,(N)} = \bar{P},$$

11. Naturally, lower values of ψ can also be seen as reflecting stronger institutions and these will make it less likely that extortion will deter the harmless behavior.

which determines whether ex-post punishment can coexist with no extortion of the innocent. If $\bar{P}$ is sufficiently high, then ex-post punishment without extortion is possible and always preferable to ex-ante prevention. Ex-post prevention is also preferable to anarchy if city size is sufficiently large.

If $\bar{P}$ is lower than this threshold, then penalties that are high enough to effectively deter misbehavior will always lead to extortion of the innocent. If the extortion is severe, then it will deter the harmless action as well as the harmful action and ex-post punishment essentially becomes ex-ante prevention, because the threat of being held up by the police will deter individuals from venturing forth. If the extortion is less severe, then it becomes an added social cost, but in this case, ex-post punishment is always preferable to ex-ante prevention.

If ex-post punishment is compatible with individuals undertaking the harmless action, then the per capita costs of extortion are increasing with λ, ψ, B, σ, and $\bar{P}$, which explains why all of those parameters increase the population threshold at which ex-post punishment (with extortion) is preferable to anarchy. The impact of city size is uncertain, because while higher values of city size also reduce the probability that an individual will be caught and extorted by the police, higher values of N also increase the size of the minimal penalty needed to prevent the harmful action, which also increases the size of the bribe that is paid contingent upon getting caught.

Figure 7 illustrates the linear case where $d_0(N) = (d_0^1)/N$ and $d(N) = d^1/N$, where

$$\left(1 - \frac{(1-\delta)\sigma(1-\psi)}{\lambda - \delta}\right) d^1 b > d_0^1 \psi B$$

so there exist multiple regions. The two horizontal lines illustrate the population cutoffs determining whether anarchy is preferable to either ex-post punishment without extortion or ex-ante prevention. The minimum population needed for ex-ante punishment to dominate anarchy is higher than the minimum threshold needed for ex-post punishment without extortion to dominate anarchy. There is also an upward sloping line between the two, for low levels of N, that illustrates the minimum cutoff needed for punishment with extortion to dominate anarchy. This line meets the line indicating the threshold for prevention to dominate anarchy at the values of N and $\bar{P}$ where prevention and ex-post punishment with extortion carry the same social costs.

There are also three upward sloping but largely vertical lines. The leftmost line indicates the threshold at which ex-post punishment with fines dominates prevention. For lower levels of $\bar{P}$, ex-post punishment dominates, because the bribes that are paid are lower and the social costs of punishing the bribe-taking policemen are lower. For higher levels of $\bar{P}$, prevention dominates, and for still higher levels of $\bar{P}$, as indicated by the middle essentially vertical line, effective ex-post punishment will also deter the harmless action, so prevention is the only real possibility. The rightmost line indicates the institutional possibilities frontier at which extortion-free punishment becomes possible, and for levels of institutional quality above that point, there is no benefit from prevention.

This figure essentially suggests that there are three stages of development for larger cities. At very low levels of institutional quality, it is cost-minimizing to have ex-post

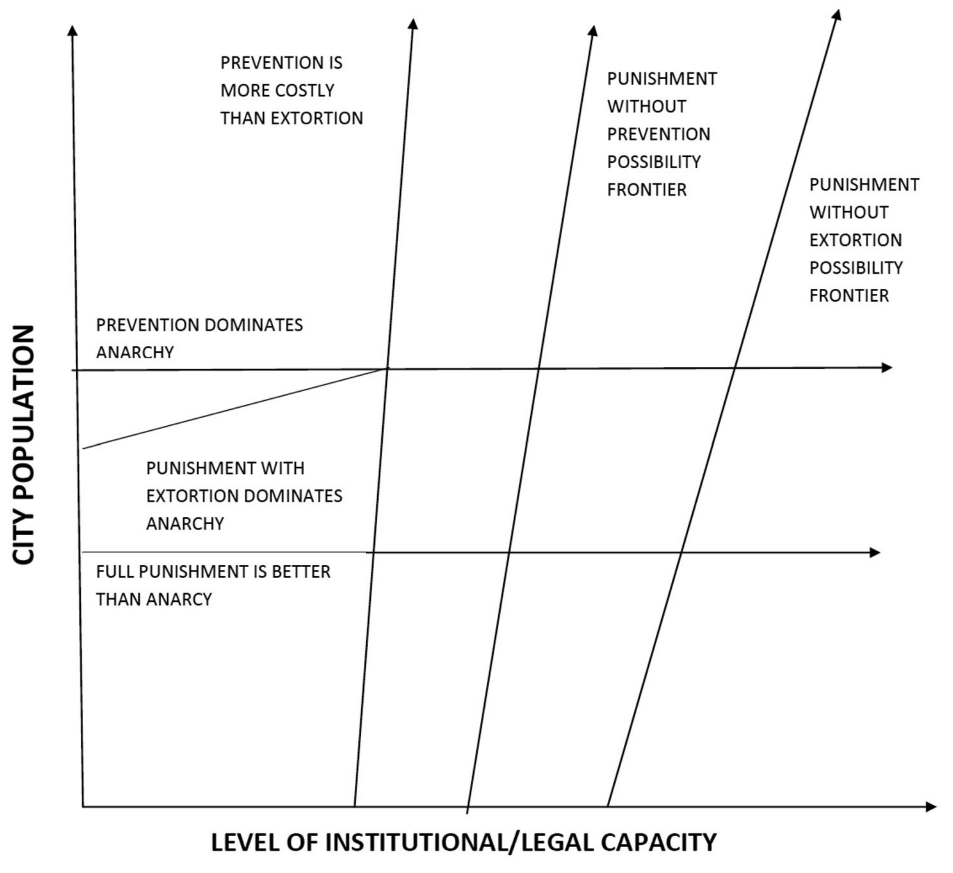

FIGURE 7. Prevention, punishment and city size. The two horizontal lines illustrate the population cutoffs determining whether anarchy is preferable to either ex-post punishment without extortion or ex-ante prevention. The minimum population needed for ex-ante punishment to dominate anarchy is higher than the minimum threshold needed for ex-post punishment without extortion to dominate anarchy. There is also an upward sloping line between the two that illustrates the minimum cutoff needed for punishment with extortion to dominate anarchy. The leftmost mostly vertical line indicates the threshold at which ex-post punishment with fines dominates prevention. The middle essentially vertical line shows effective ex-post punishment will also deter the harmless action, so prevention is the only real possibility. The rightmost line indicates the institutional possibilities frontier at which extortion-free punishment becomes possible, and for levels of institutional quality above that point, there is no benefit from prevention.

punishment, but it must be expected that this will lead to extortion and bribery. At intermediate levels of development, ex-ante prevention may be the only real option. At the highest levels of institutional development, ex-post punishment is again optimal.

For smaller cities, either anarchy is always appropriate, because the social costs of the harmful action are sufficiently low, or it is optimal to engage in ex-post punishment at high enough levels of institutional quality so that extortion and bribery do not occur. For intermediate levels of city size, it is possible that ex-post punishment with bribery is optimal at very low levels of city size, but then anarchy becomes optimal at intermediate institutional quality ranges, and then ex-post punishment with bribery becomes optimal again at high levels of institutional quality. For cities in this range, prevention is never optimal.

5. Urban Infrastructure

Many of the largest urban policy decisions concern the provision of infrastructure. America's cities and towns were spending as much on water at the start of the 20th century as the Federal government was spending on everything except for the Post Office and the Army (Cutler and Miller 2006). Sewers, highways, and housing all interact with urban externalities. Good sewers deal with the waste that can make cities pestilential; highways become less valuable when they are congested. In a sense, aqueducts decrease the effective density by importing water from other less dense areas.

In Section 5.1, I focus on the connection between infrastructure provision and the ability to address externalities using ex-post punishment, as discussed in the previous section. The model is the same, except that I have added an ability to improve the quality of life by investing in infrastructure. I simply assume that infrastructure is provided at a cost and do not model the challenges of actually delivering infrastructure. In Section 5.2, I turn to the problems that poor governance creates for the provision of infrastructure.

5.1. Infrastructure and Externalities

I now turn to urban infrastructure, and its relationship with externalities. If the total quantity of infrastructure is denoted "*I*", and if the proportion of the population that take the harmful action is denoted $\hat{v}$, then the per capita benefit created by infrastructure equals $b(I(1-\sigma\hat{v}N)/N^{\gamma})$, where $I(1-\sigma\hat{v}N)/N^{\gamma}$ represents the effective infrastructure and $b(I)$ is increasing, weakly concave and equal to zero if $I(1-\alpha\hat{v}N) = 0$. I assume that $1 \geq \alpha\hat{v}N$ for all feasible values of $\alpha\hat{v}N$, so infrastructure always carries a positive return. The harmful action effectively reduces the amount of available infrastructure that can be used. Polluting the water reduces the value of a reservoir or congestion reduces the value of city streets.

The function N^{γ} reflects the congestion of the infrastructure due to city population. Holding infrastructure per capita constant, effective infrastructure is declining with city size $\gamma > 1$, and rising with city size if $\gamma < 1$. This connects with the previous model, if $b(.)$ is a linear function $b_0 I(1-\sigma\hat{v}N)/N^{\gamma}$, and then $b_0\alpha N^{1-\gamma} = C(N)$.

If the cost of providing *I* units of infrastructure is $K_I I$, then the optimal level of infrastructure satisfies

$$N^{1-\gamma}(1-\alpha\hat{v}N)b'\left(\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}\right) = K_I.$$

This implies that the preferred level of infrastructure is rising with $\hat{v}$ if and only if

$$-\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}\frac{b''\left(\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}\right)}{b'\left(\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}\right)} > 1.$$

Social harm reduces the return to infrastructure (a price effect) but increases the need for infrastructure (essentially an income effect) and if the latter effect is stronger, which requires the function $b'(.)$ to be sufficiently elastic, then the optimal level of infrastructure increases with the level of social harm.

To complete the model, I return to the level of harmful behavior. I again assume that the private benefit of the harmful action is B and that $\pi = 0$. I also assume no extortion and no ex-ante prevention, and that those who are accused of a harmful action make take-it-or-leave-it offers to policemen, which ensures that the maximum penalty is $\bar{P}$. This implies that $\hat{v}$ either equals v, if there is anarchy, or 0, if there is full prevention. Full prevention is only possible if $\bar{P}/d(N) > B$. I assume that K_P is sufficiently modest so that at some high enough level of population, welfare is greater than vB given optimal expenditure and effective punishment.

PROPOSITION 8. *Anarchy always has lower costs than ex-post punishment when N is low and higher costs when N is close to* $1/\alpha v$. *If* γ *is close to one, or if* $1 > \gamma$ *and* $1 > x(-b''(x)/b'(x))$ *for all x or if* $1 < \gamma$ *and* $1 < x(-b''(x)/b'(x))$ *for all x, then there will exist a unique population threshold, which determines whether anarchy carries lower costs than ex-post punishment. If* $1 > x(-b''(x)/b'(x))$, *then infrastructure is higher under punishment than under anarchy and higher levels of* K_I *cause the punishment threshold to rise. If* $1 < x(-b''(x)/b'(x))$, *then optimal infrastructure investment is higher under anarchy than with ex-post punishment and increases in* K_I *cause the punishment threshold to rise.*

The proposition suggests that the connection between infrastructure and rule-of-law depends on the shape of the function $b(\cdot)$. If $b(\cdot)$ is extremely concave, then anarchy leads to more, not less, infrastructure. In that case, higher costs for infrastructure reduce the population threshold for preferring rule of law as rule of law is essentially a substitute for infrastructure. Conversely, if $b(\cdot)$ is less concave, for example $b(x) = x^{\sigma}$, with $1 > \sigma > 0$, then rule of law is a complement with infrastructure, meaning that optimal infrastructure increases if society imposes punishments sufficient to prevent harmful behavior. Similarly, higher costs of infrastructure increase the minimal threshold needed for rule-of-law to reduce costs.

The comparative statics connecting per capita infrastructure and population depend also on the shape of the function b and the value of γ. For example if $1 > x(-b''(x))/b'(x)$, then per capita infrastructure is rising with N almost everywhere if $(1-\gamma)/(2-\gamma) > \alpha vN > 0$ and decreasing with N almost everywhere if $(1-\gamma)/(2-\gamma) < 0 < \alpha vN$. If $(1-\gamma)/(2-\gamma) > 0$, but for some value of N below the threshold at which rule of law becomes desirable, $\alpha vN > (1-\gamma)/(2-\gamma)$, so optimal infrastructure is first rising and then falling with city population. If

$1 > x(-b''(x))/b'(x)$, then there is always a discontinuous increase in infrastructure when rule of law is adopted. Similarly, if $\bar{P}$ rises, making rule of law possible, there will be a similar increase in the level of optimal infrastructure.

If $1 < x(-b''(x))/b'(x)$ for all x, then the conditions are largely reversed and there is a discontinuous drop in infrastructure if rule of law is adopted. The importance of $x(-b''(x))/b'(x)$ reflects the fact that this term determines whether a price effect or an income effect dominates in the first order condition. When rule of law is adopted, then the stock of effective infrastructure rises, and this reduces the return to infrastructure through an income effect. Conversely, rule of law also ensures that each dollar invested produces more effective infrastructure and this increases the return to infrastructure through a price effect. For the second effect to dominate, $x(-b''(x))/b'(x)$ must be less than one.

5.2. *Public versus Private Provision of Infrastructure and Rule of Law*

We now turn from the benefits of infrastructure to its costs, and consider two potential tools for providing infrastructure. I assume here that there is one piece of infrastructure that can be built and its benefits are fixed, but the provision of infrastructure can be enormously wasteful, and that waste occurs through two primary channels.

The first, simplest channel is that the public pays far too much for the infrastructure either because of incompetence or corruption. New York's Tweed Courthouse, for example, was a particularly famous example of massively overpaying for construction inputs in exchange for kickbacks to city officials. At a less pernicious level, whenever public workers are paid more than the prevailing wage rate, the costs of public provision rise. Private providers have a track record of subverting the political process to increase their rents.

The second primary channel of waste is the construction of bad projects. Bridges-to-nowhere are a staple of American political discourse. Many economists have alleged that high-speed rail projects in Spain and China cost far more than the benefits they deliver. It is certainly possible, even with competent provision and scrupulous book-keeping, to waste billions by providing the wrong pieces of infrastructure.

I focus on the first channel and on the comparison between private and public provision. I assume that two ingredients are needed for the provision of public infrastructure—labor and land—and that there is a trade-off between the two. Providing the infrastructure on less land requires more labor, specifically if A units of land are acquired then $L(A)$ units of effective labor are needed. The value of $L(0)$ is finite.

The opportunity cost of labor is denoted by w, and private providers pay exactly that amount. Public providers are unable to perfectly monitor their workers, and I assume that if public workers spend a fraction "s" of a unit of time shirking, there is a probability $\phi(s)$, which is an increasing convex function, that they will be caught. If caught, they face a punishment of $\bar{P}$. If the benefit of shirking is equal to the opportunity cost of time w times a multiplier $z < 1$, then they will maximize $zws - \phi(s)\bar{P}$, which leads to an optimal level of shirking for the public workers determined by $zw = \phi'(s^*)\bar{P}$. The labor cost of the project is therefore $(wL(A))/(1 - s^*)$, because to get one effective unit of labor, the public must pay for $1/(1-s^*)$ units of labor. The total cost of the project, therefore, holding land use constant, is strictly decreasing in the value of $\bar{P}$.

The total social cost of the project however is

$$\frac{wL\,(A)\,(1-zs^*)}{1-s^*}+\frac{\phi(s^*)\lambda\bar{P}}{1-s^*},$$

reflecting both the loss of value of time from shirking and the costs of punishment.

One way to embed the loss from private provision is to assume that the private entity was able to wheedle more cash out of the public entity by bribery. This certainly has occurred, upon numerous occasions, but instead, I assume that the primary cost of private rather than public provision occurs through land acquisition. The land must be acquired from private entities, and I assume that the true opportunity cost of the land is $P_L(N)$, which is increasing without bound in city size. To eliminate the hold-up problem, the public allows eminent domain, but there is a public adjudicator, who ensures that land acquisition is supposed to occur at a fair price, and in the cost of the public provision it does. The strength and weakness of public provision is weak incentives (Hart, Shleifer, and Vishny 1997). Weak incentives make it difficult for the public entity to eliminate shirking, but they also mean that the public entity will not try to subvert the land adjudication process.

The public manager then chooses the amount of land to minimize financial costs $wL(A)/(1-s^*)+P_L(N)A$, which leads to a first-order condition of $-[wL'(A)/(1-s^*)]=P_L(N)$. I am assuming that the public manager is choosing land to minimize total costs to the public sector instead of total social costs. The first best would set $-wL'(A)=P_L(N)$, so because the public entity faces artificially high costs of labor, it ends up acquiring more land than it would if there was no shirking.

The private entity is assumed to face a fixed price contract, and as such it has an incentive to reduce land acquisition costs if it can. There is a minimum price that the adjudicator can set of $\underline{\boldsymbol{P}}_L$, so there is a potential gain of $P_L\,(N)-\underline{\boldsymbol{P}}_L$ that can occur if the private provider can bribe the adjudicator to set a lower price. If the adjudicator takes a bribe, he again faces a maximum penalty of $\bar{P}$ and has a probability of being caught of d per unit of land. As such, bribery will occur when $d(P_L(N)-\underline{\boldsymbol{P}}_L)>\bar{P}$. As such, bribery becomes more likely if the city size is larger. Again, there is an institutional possibilities frontier based on city size. The larger the city, the more likely the frontier is to be breached because the returns from expropriating land are higher. If the private entity makes a take-it-or-leave-it offer to the public adjudicator, then the effective cost of land becomes $\underline{\boldsymbol{P}}_L+d\bar{P}$, and the private entity will also use more land, relative to labor, in its infrastructure provision.

Proposition 9 assumes that $\lambda=0$.

PROPOSITION 9. *If $d(P_L(N)-\underline{\boldsymbol{P}}_L)<\bar{P}$, then private provision generates the first best, and if $d(P_L(N)-\underline{\boldsymbol{P}}_L)>\bar{P}$, there exists a unique population threshold above which public provision reduces costs and below which private provision reduces costs. Public land acquisitions are greater than private acquisition at low levels of population, but at and above the population threshold, private land acquisitions are greater than* *public land acquisitions. The threshold is increasing with d, the probability of catching a corrupt land adjudicator.*

The intuition of Proposition 9 is straightforward. Private provision of infrastructure reduces costs when cities are small but not when cities are large. The fundamental loss from private provision is the subversion of the land acquisition process, and that becomes more and more problematic as city size rises, because the value of the expropriated land gets arbitrarily large. Since the private providers pay a fixed amount, they ignore this value and continue to use the same amount of land, no matter how valuable the city's space becomes.

The results of the model are meant to highlight this issue rather than provide any definitive answer. It may be that public expropriation is just as big a problem as private expropriation, and in that case, private provision would always dominate. Alternatively, the private provider may have other means of expropriating rents and if those scale up with city size, they will also push against private provision if cities are sufficiently big.

6. Conclusion

This paper has argued that the recent rise of poor world mega-cities is a historical anomaly. Typically wealth, or at least political competence, accompanied the growth of large urban agglomerations. Yet for 50 years, there has been explosive urban growth in some of the world's poorest and most poorly governed countries.

In the first modeling exercise of the paper, I argued that this phenomenon is best seen as a by-product of globalization. The cities of the West developed in essentially closed economies, where local agricultural productivity and transportation technology were needed to feed urbanites. The cities of the developing world today can feed themselves with imports paid for with mineral exports or foreign aid. As in Matsuyama (1992), the switch from closed to open economies reverses key comparative statics. In a closed economy, agricultural prosperity increases urbanization, while in an open economy, agricultural desperation causes cities to grow.

In the second modeling exercise of the paper, I examined the governance of big cities with weak institutions. An institutional possibilities frontier emerged that reflected the combination of city population and institutional quality. Larger cities are more likely to need governmental interventions to alleviate externalities, but urban anonymity makes it more difficult to enforce good outcomes, especially when weak institutions make bribery common. When I allowed for both ex-ante prevention and ex-post punishment, I found the more dictatorial approach of ex-ante restrictions to be more effective than ex-post punishment when institutions are weak and city sizes are large.

Infrastructure can alleviate the downsides of density, but it is unclear whether infrastructure becomes more or less attractive as institutional quality increases. The key condition depends on whether restrictions on harmful behavior increase the stock of effective infrastructure, thereby alleviating the need to spend more, or increase the marginal gains from incremental infrastructure. I also discussed private and public provision. If the cost of public provision is that public employees provide low effort, but the cost of private provision is subversion of the land acquisition process, then public provision would dominate when city sizes are sufficiently large while private provision of infrastructure would dominate with smaller cities.

I have spent little time on slums, land tenure, and housing costs, which are also major public issues in poor mega-cities. Purchasing-power-adjusted Class A office rents in Mumbai are among the highest in the world (Gomez-Ibanez and Masood 2006). In a previous draft (Glaeser 2013), I argued that new construction in poor mega-cities faces the double threat of public regulation and private expropriation. The western experience is that regulation of land followed property rights protection, but poorer mega-cities regulate but do not protect.

This phenomenon is an understandable result of weak institutions. Regulations that limit large structures are cheap to enforce. Tall structures are large, obvious, and expensive. Construction can be deterred by even a small probability of demolition due to zoning violations. By contrast, enforcing property rights by pushing large numbers of poor people off undeveloped land is difficult politically and physically, and squatters have little to lose. Limited enforcement helps explain why property rights are not protected, but not why new development is so regulated. In some cases, regulations may facilitate rent extraction, and occasionally, restrictions may correct real externalities. But in many cases, stringent land use controls are yet another example of inappropriate institutions, borrowed from Western cities, and brought to places where they do far more harm. A primary theme of this paper is that rules must reflect local conditions, particularly the combination of urban size and institutional capacity.

Given the terrible costs when density is combined with limited resources and poor government, it may be tempting to take the view that mega-city growth should be constrained. Yet, even if a static calculation suggested that such limitations would be beneficial, the dynamics effects of urbanization push in the opposite direction. I documented the robust correlation between urbanization and income growth between 1960 and 2010 among poorer countries. While city-building is no guarantee of income growth, it has led to more prosperity more reliably than rural living in areas with poor soil quality.

Just as importantly, history suggests that urbanization helps build the institutional strength that can ultimately lead to better government. The connections and social movements that form readily in the dense confines of urban areas can ultimately force democracy and discipline government. The rural past of the poorer world has shown little evidence of positive institutional development. The cities of the developed world have been wellsprings for the growth of political organization. Indeed, it is possible that an urban contribution to the growth of institutional quality may be the most important benefit of urban growth.

Appendix

Proof of Proposition 1. To solve for P_M, I equate the agricultural surplus sold in the city with the value of manufacturing goods that are sought to be sold so

$$P_M = \frac{(1-\alpha)\left(A_A N_A^{1-\gamma} g(\tau, \bar{d}) - N_U \underline{C}\right)}{\alpha A_M N_U^{1/\sigma}\,(1-\varepsilon)\,\varepsilon^{(1-\sigma)/\sigma}},$$

where

$$g(\tau, \bar{d}) = \left(\frac{2\gamma}{\tau}\right)^{\gamma} \frac{(1-\alpha)^{1-\gamma}}{(1-\alpha+\alpha\gamma)} \frac{1 - e^{-(((1-\alpha+\alpha\gamma)\tau)/\gamma)\bar{d}}}{\left(1 - e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)^{1-\gamma}}.$$

Welfare in the city equals

$$\theta_0 e^{-\delta u N}\left((1-\varepsilon)(\varepsilon N_U)^{(1-\sigma)/\sigma} A_M P_M - \underline{C}\right) P_M^{\alpha-1} \text{ or}$$

$$\frac{\theta_0 e^{-\delta u N}}{\alpha}\left((1-\alpha)\, A_A g(\tau, \bar{d}) \frac{(1-u)^{1-\gamma}}{u N^{\gamma}} - \underline{C}\right) P_M^{\alpha-1}$$

while welfare in the countryside equals

$$A_A\,((1-u)\,N\psi)^{-\gamma}\,(2-2e^{-\psi\bar{d}})^{\gamma}\,P_M^{\alpha-1}.$$

The value of

$$(1-\alpha)\, A_A g(\tau, \bar{d}) \frac{(1-u)^{1-\gamma}}{u N^{\gamma}} - \underline{C}$$

goes to infinity as u goes to zero, and is negative at sufficiently high levels of u. Moreover, as long as

$$(1-\alpha)\, A_A g(\tau, \bar{d}) \frac{(1-u)^{1-\gamma}}{u N^{\gamma}} > \underline{C},$$

then

$$\frac{\theta_0 e^{-\delta u N}}{\alpha}\left((1-\alpha)\, A_A g(\tau, \bar{d}) \frac{(1-u)^{1-\gamma}}{u N^{\gamma}} - \underline{C}\right)$$

is strictly decreasing in u, while $A_A\,((1-u)\,N\psi)^{-\gamma}\,(2-2e^{-\psi\bar{d}})^{\gamma}$ is strictly increasing in u. It follows that there must exist a unique value of u at which the welfare levels are the same.

At this point,

$$e^{-\delta u N}\left(\frac{1-u}{u}\,(1-\alpha)\,g(\tau, \bar{d}) - (1-u)^{\gamma}\frac{N^{\gamma}\underline{C}}{A_A}\right) = \alpha(2-2e^{-\psi\bar{d}})^{\gamma}\,\frac{\psi^{-\gamma}}{\theta_0}. \quad \text{(A.1)}$$

The derivative of the left-hand side with respect to u is negative whenever the left-hand side is positive. Equation (A.1) can be rewritten as

$$h\,(u\,(z)\,, z) = \alpha(2-2e^{-\psi\bar{d}})^{\gamma}\,\frac{\psi^{-\gamma}}{\theta_0}, \quad \text{(A.2)}$$

where z refers to a vector of parameters. Since the derivative of h with respect to u is negative in the equilibrium, we know that u is rising with θ_0 and rising with any parameter that increases h, without impacting the other side. From this logic it follows that u is rising with A_A, falling with N, δ, and $\underline{\mathbf{C}}$. Urbanization is independent of A_M and ε.

To see the impact of τ, it is helpful to rearrange the equation to be

$$e^{-\delta u N}\left(\frac{(1-u)(2\gamma)^{\gamma}(1-\alpha)^2}{u\alpha(1-\alpha+\alpha\gamma)}\left(\frac{1-e^{-(((1-\alpha+\alpha\gamma)\tau)/\gamma)\bar{d}}}{1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}}\right)\right.$$
$$\left.-\left(\frac{(1-\alpha)\tau(1-u)}{1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}}\right)^{\gamma}\frac{N^{\gamma}\underline{\mathbf{C}}}{A_A\alpha}\right)=\frac{(2\gamma)^{\gamma}}{\theta_0}.$$

The left-hand side must continue to be declining in u around an equilibrium. The derivative with respect to τ then equals

$$e^{-\delta u N}\times\frac{1-u}{u}\frac{2^{\gamma}\gamma^{\gamma-1}(1-\alpha)^3}{\alpha\bar{d}(1-\alpha+\alpha\gamma)}\frac{e^{-(((1-\alpha+\alpha\gamma)\tau)/\gamma)\bar{d}}}{\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)^2}$$
$$\times\left(\frac{\alpha\gamma}{1-\alpha}\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)-e^{\alpha\tau\bar{d}}+1\right)$$
$$-\left(1-\left(1+\frac{(1-\alpha)\bar{d}\tau}{\gamma}\right)e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)$$
$$\times\frac{\gamma\tau^{\gamma-1}}{\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)^{\gamma+1}}(1-\alpha)^{\gamma}\frac{((1-u)N)^{\gamma}\underline{\mathbf{C}}}{A_A\alpha}.$$

The first term is always negative and the second term is always positive. Hence urbanization is always falling in transportation costs. □

Proof of Proposition 2. In this case, the spatial equilibrium requires that

$$\left((1-\varepsilon)(\varepsilon N)^{(1-\sigma)/\sigma}u^{(1-\sigma)/\sigma}A_M P_M-\underline{\mathbf{C}}\right)(1-u)^{\gamma}e^{-\delta u N}$$
$$=\frac{N^{-\gamma}A_A}{\theta_0}\left(\frac{2\gamma\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}. \qquad \text{(A.3)}$$

The left-hand side is negative as u goes to zero and zero when u goes to one. As such, the left-hand side must be lower than the right at either extreme. The first derivative of the left-hand side with respect to u is

$$(1-u)^{\gamma}e^{-\delta uN}$$
$$\left[\left((1-\varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\,u^{(1-\sigma)/\sigma}A_M P_M\right)\left(\frac{1-\sigma}{\sigma u}-\frac{\gamma}{1-u}-\delta N\right)-\left(\frac{\gamma}{1-u}+\delta N\right)\underline{\mathbf{C}}\right].$$

This assumes that there is risk-sharing between laborers and entrepreneurs so that everyone receives the expected output. If there was no risk sharing, I would need a further condition that $\sigma\,(\varepsilon N)^{(1-\sigma)/\sigma}\,u^{(1-\sigma)/\sigma}A_M P_M > \underline{\mathbf{C}}$.

The second derivative is always negative if the first-order condition is satisfied, and as such, it can reach only one critical value and that must be a maximum. Since the function begins below its ending value, it must have such one unique maximum between zero and one. If there exists any value of u for which

$$(1-u)^{\gamma}e^{-\delta uN}\left((1-\varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\,u^{(1-\sigma)/\sigma}A_M P_M-\underline{\mathbf{C}}\right)$$
$$>\frac{N^{-\gamma}A_A}{\theta_0}\left(\frac{2\gamma\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}$$

then generically (i.e., except when that value equals the maximum itself), there must be two values of u that satisfy the equality (since the left-hand side begins and ends below the right). At the lower crossing point, the left-hand side is rising with u, and hence the welfare in the urban sector is rising more quickly than the welfare in the rural sector. But at the higher crossing point, the left-hand side is rising more slowly with u, and hence welfare in the urban sector is decreasing relative.

The condition for urbanization to occur at all is that the maximized value of

$$(1-u)^{\gamma}e^{-\delta uN}N^{\gamma}\left((1-\varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\,u^{(1-\sigma)/\sigma}A_M P_M-\underline{\mathbf{C}}\right)$$

is greater than

$$\frac{A_A}{\theta_0}\left(\frac{2\gamma\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}.$$

If $A_A=0$, this will occur, as long as $(1-\varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\,A_A P_M > \underline{\mathbf{C}}$, but as A_A goes to infinity, the condition will never hold. As the left-hand side and its maximum are independent of A_A, there must exist a unique value of A_A above which urbanization does not occur and below which urbanization occurs. This value, denoted A_A^*, satisfies

$$(1-\hat{u})^{\gamma}e^{-\delta\hat{u}N}N^{\gamma}\left((1-\varepsilon)\,(\varepsilon N)^{(1-\sigma)/\sigma}\,\hat{u}^{(1-\sigma)/\sigma}A_M P_M-\underline{\mathbf{C}}\right)$$
$$=\frac{A_A^*}{\theta_0}\left(\frac{2\gamma\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma},$$

where $\hat{u}$ denotes the maximized value of u. As the right-hand side is increasing with A_A^*, its value is declining with any value that causes the right-hand side to rise and increasing with any variable that causes the left-hand side to rise. As such, A_A^* is rising with A_M, P_M θ_0, and τ and falling with δ, $\bar{d}$, and $\underline{C}$. The A_A^* is rising with ε as long as $1-\sigma > \varepsilon$, and N if and only if $\gamma > \delta\hat{u}N$. □

Proof of Proposition 3. At any spatial equilibrium (A.3) applies, but at a stable equilibrium, the left-hand side must be decreasing with u. As such, any variable that causes the left-hand side to increase will cause the equilibrium level of urbanization to rise, and any variable that causes the right-hand side to increase will reduce the equilibrium level of urbanization. This implies that urbanization will be rising with A_M, P_M θ_0, and τ and falling with A_A, $\bar{d}$, δ and $\underline{C}$. Urbanization will be rising with ε as long as $1-\sigma > \varepsilon$ and N if and only if $\gamma > \delta\hat{u}N$. □

Proof of Proposition 4. Open economy: Welfare in the city equals $\theta_0 e^{-\delta N_U^*}(1-\varepsilon)(\varepsilon N_U^*)^{(1-\sigma)/\sigma} A_M P_M^\alpha$, where N_U^* reflects the equilibrium level of population discussed in Proposition 2, which must satisfy

$$\left(N_U^*\right)^{(1-\sigma)/\sigma}\left(N-N_U^*\right)^{\gamma} e^{-\delta N_U^*} = \frac{A_A\left(2\gamma\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)\right)^{\gamma}}{\theta_0(1-\varepsilon)\,(\varepsilon)^{(1-\sigma)/\sigma}\,A_M P_M\,((1-\alpha)\tau)^{\gamma}}. \quad \text{(A.3}'\text{)}$$

To attract urban pioneers, the new city must offer the same utility level.

The derivative of urban welfare with respect to city size is

$$\left(\frac{1-\sigma}{\sigma N_U}-\delta\right)(1-\varepsilon)\left(\varepsilon N_U\right)^{(1-\sigma)/\sigma} A_M \theta_0 e^{-\delta N_U} P_M^{\alpha},$$

which is positive when N_U is small and the second derivative is negative as long as the first derivative is positive, or close to zero. Moreover, as N_U goes to infinity the derivative must be negative. As such, there exists a unique value of N_U denoted N_U^{Max} equal to $(1-\sigma)/\sigma\delta$ which maximizes well-being in the city.

If $N_U^* < (1-\sigma)/\sigma\delta$, which requires that

$$\left(\frac{1-\sigma}{\sigma\delta}\right)^{(1-\sigma)/\sigma}\left(N-\frac{1-\sigma}{\sigma\delta}\right)^{\gamma} e^{-((1-\sigma)/\sigma)} < \frac{A_A\left(2\gamma\left(1-e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)\right)^{\gamma}}{\theta_0(1-\varepsilon)\,(\varepsilon)^{(1-\sigma)/\sigma}\,A_M P_M\,((1-\alpha)\tau)^{\gamma}},$$

then the second city must be as large as the first city to attract urban pioneers.

If $N_U^* > (1-\sigma)/\sigma\delta$, then since welfare at the lower equilibrium level of urbanization yields urban welfare that is below welfare at the higher level of urbanization (since agricultural land per capita is less and agricultural welfare must equal urban welfare), the second city must always be larger than the lower equilibrium city size in the one-city case. If

$$N_U^* > \frac{1-\sigma}{\sigma\delta}, \text{ or } \left(\frac{1-\sigma}{\sigma\delta}\right)^{(1-\sigma)/\sigma} \left(N - \frac{1-\sigma}{\sigma\delta}\right)^{\gamma} e^{-((1-\sigma)/\sigma)}$$

$$> \frac{A_A \left(2\gamma \left(1 - e^{-(((1-\alpha)\tau)/\gamma)\bar{d}}\right)\right)^{\gamma}}{\theta_0 (1-\varepsilon)\,(\varepsilon)^{(1-\sigma)/\sigma}\, A_M P_M\, ((1-\alpha)\tau)^{\gamma}}$$

then the minimum size for the new city must be a fraction of the size of the old city, φ, that satisfies $-\mathrm{Ln}\,\varphi/(1-\varphi) = \sigma\delta/\,(1-\sigma)$, so this fraction is decreasing in σ and δ.

Closed economy: In the closed economy, when $\underline{C} = 0$, then the welfare of urban residents in any second location equals

$$e^{-\delta(N_{U2}-N_{U1})} \left(\frac{N_{U2}}{N_{U1}}\right)^{(1-\sigma)/\sigma} \left(\frac{P_{M2}}{P_{M1}}\right)^{\alpha} V_1,$$

where V_1 represents welfare in the first location, N_{Ui} is the population of city i, and P_{Mi} represents the price of manufactured goods in city i. A fraction of agricultural land, denoted Δ, will start shipping its good to the new city. The equilibrium condition that determines this fraction is that $e^{-\tau\bar{d}(1-\Delta)} P_{M2} = e^{-\tau\bar{d}\Delta} P_{M1}$, so farmers on the margin receive the same return in industrial goods for shipping to either city. The ratio of the price of the manufactured goods in city 1 to city 2 or P_{M1}/P_{M2} will equal

$$\left(\frac{N_{U2}}{N_{U1}}\right)^{1/\sigma} \left(\frac{1 - e^{-((1-\alpha+\alpha\gamma)/\gamma)\,(1-\Delta)\,\tau\bar{d}}}{1 - e^{-((1-\alpha+\alpha\gamma)/\gamma)\Delta\tau\bar{d}}}\right),$$

which must equal $e^{-\tau\bar{d}(1-2\Delta)}$.

For any value of $k = \tau\bar{d}\,(1-2\Delta)$, where

$$\tau\bar{d} > k > \frac{1}{\sigma} \ln\left(\frac{N_{U1}}{N_{U2}}\right),$$

there will exist a value of "$\tau\bar{d}$", at which

$$\left(\frac{N_{U2}}{N_{U1}}\right)^{1/\sigma} \left(\frac{1 - e^{-((1-\alpha+\alpha\gamma)/2\gamma)}\left(\tau\bar{d}+k\right)}{1 - e^{-((1-\alpha+\alpha\gamma)/2\gamma)}\left(\tau\bar{d}-k\right)}\right) = e^{-k},$$

because the left-hand side is increasing in $\tau\bar{d}$ from zero to one. As such, for any value of $\theta < 1$, there will exist a value of $\tau\bar{d}$ such that for higher values of $\tau\bar{d}$, welfare is greater than $e^{-\delta(N_{U2}-N_{U1})}(N_{U2}/N_{U1})^{(1-(1+\alpha)\sigma)/\sigma}\theta V_1$.

This implies that for any second city population, for sufficiently large values of $\tau\bar{d}$, welfare in the second city will be higher than welfare in the first city, as long as $1 < (1+\alpha)\sigma$. □

Proof of Proposition 5. The social costs of deterring all crime equals $K_P N$. The social costs of deterring the low-incentive criminals but not the high-incentive criminals equals $K_P N + N(\pi v(C(N) - A - (1-\lambda)B))$. The social costs of deterring no crime are $N(v(C(N) - \pi A - B))$. Deterring a limited amount of crime is never preferable to deterring all crime, but it is preferable to deterring no crime if and only if $K_P + vB(1 - \pi(1-\lambda)) < v(1-\pi)C(N)$. As such, there exists a city size level denoted N_{LP}^A at which the social costs from anarchy are equal to the social costs of deterring a limited amount of crime, which satisfies $K_P + vB(1 - \pi(1-\lambda)) = v(1-\pi)C(N_{LP}^A)$. For all higher levels of city size, limited punishment is preferred to anarchy and for all lower levels of city size, anarchy is preferred to limited punishment. N_{LP}^A is rising with K_P, π, and B, and falling with λ, and v.

Deterring all crime is preferable to anarchy if and only if $vC(N) > K_P + v(\pi A + B)$, and there exists a city size level denoted N_{FP}^A at which the social costs from anarchy are equal to the social costs of fully deterring crime, which satisfies $K_P + v(\pi A + B) = vC(N_{FP}^A)$. For all higher levels of city size, full punishment is preferred to anarchy and for all lower levels of city size, anarchy is preferred to full punishment. The value of N_{FP}^A is rising with K_P, π, A, and B and falling with v. If $vC(N) - v(\pi A + B) = K_P$, then $v(1-\pi)C(N) - vB(1 - \pi(1-\lambda)) < K_P$, because $C(N) > A + B(1-\lambda)$, so N_{FP}^A is always less than N_{LP}^A.

As such, for all values of N greater than N_{LP}^A, there exists a value of $\bar{P} = B/(d(N))$ (which is increasing with B and N) at which limited punishment is possible and desirable and a second higher value of

$$\bar{P} = \frac{A+B}{d(N)}$$

(which is increasing with A, B, and N), at which full punishment is possible and desirable. For all values of N less than N_{LP}^A but greater than N_{FP}^A, there exists a value of

$$\bar{P} = \frac{A+B}{d(N)}$$

at which full punishment is possible and desirable, but there is no level of institutional quality at which limited punishment is desirable. □

Proof of Proposition 6. The social costs of ex-ante prevention are always higher than the costs of high-penalty punishment. The social costs associated with ex-ante prevention are less than the costs of anarchy if and only if $K_P + h + v(B + \pi A) < vC(N)$ and we denote the value of N at which $K_P + h + v(B + \pi A) = vC(N_{PR}^A)$ as N_{PR}^A. The value of N_{PR}^A is increasing with K_P, A, B, h, and π and decreasing with v. The value of N_{PR}^A is greater than N_{LP}^A if and only if

$$h > \pi \left(\frac{K_P + v\lambda B}{1-\pi} - vA \right).$$

As such, if h is low, then low penalty is never optimal. If h is high, then it is always cost minimizing to engage in low-penalty ex-post punishment. If h is low, then for cities with population levels above N^A_{PR}, then it is cost-minimizing to impose ex-ante prevention when

$$\frac{h + vB + v\pi A}{d_0(N)} < \bar{P}$$

and then engage in ex-post punishment with full penalties when

$$\frac{A + B}{d(N)} < \bar{P}.$$

If h is high, then for cities with population levels above N^A_{LP}, it is cost-minimizing to engage in ex-ante prevention when

$$\frac{h + vB + v\pi A}{d_0(N)} < \bar{P} < \frac{B}{d(N)},$$

and then ex-post punishment with low penalties when

$$\frac{B}{d(N)} < \bar{P} < \frac{A + B}{d(N)}$$

and then ex-post punishment with high penalties when

$$\bar{P} > \frac{A + B}{d(N)}.$$

□

Proof of Proposition 7a. If the policy is equivalent to ex-ante prevention, then there are no false accusations, since individuals need to be engaging in harmless action (i.e., be out on the street) in order to be accused. In this case, ex-ante prevention is always possible as long as

$$h + vMax[0, B - d(N)(\sigma\bar{P} + (1-\sigma)\psi P)] < d_0(N)(\sigma\bar{P} + (1-\sigma)\psi P))$$

and there are no costs from false arrests. A value of P equal to

$$\frac{h + vB}{(1-\sigma)d_0(N)} - \frac{\sigma\bar{P}}{1-\sigma}$$

will achieve ex-ante prevention. The social cost is $N(K_P + h)$ relative to the first best. The social cost of anarchy is $Nv(C(N) - B)$ relative to the first best. The social costs of ex-ante prevention (through extortion) is less than anarchy if and only if $C(N) > (K_P + h - vB)/v$, which implies that anarchy is preferable if and only if N is below some threshold, then is increasing with K_P and h and falling with B and v. □

Proof of Proposition 7b. If $\bar{P} > B/d(N)$, then misbehavior can be prevented with a fixed punishment below $\bar{P}$, which implies no bribery and no extortion at a flat cost of NK_P. Ex-post punishment with full enforcement dominates anarchy if and only if $K_P < v(C(N) - B)$.

If $\bar{P} < B/d(N)$, then ex-post punishment with bribery can deter crime if $B \leq d(N)(\sigma\bar{P} + (1-\sigma)P)$, which is possible if

$$\frac{B}{(1-\sigma)\,d(N)} - \frac{\sigma\bar{P}}{1-\sigma} < P.$$

For ex-post punishment not to generate extortion of the innocent, it must be that $\psi P < \bar{P}$. Hence if

$$\frac{\psi B}{(1-\sigma)\,d(N)} - \frac{\psi\sigma\bar{P}}{1-\sigma} < \bar{P} \text{ or } \frac{\psi B}{(1-\sigma+\sigma\psi)\,d(N)} < \bar{P}$$

then ex-post punishment with bribery but without extortion is possible.

If $\psi B/((1-\sigma+\sigma\psi)d(N)) > \bar{P}$, then extortion will also occur, at a level of $\sigma\bar{P} + (1-\sigma)\psi P$, which means that at the minimum punishment level that will deter misbehavior, $B/((1-\sigma)\,d(N)) - \sigma\bar{P}/(1-\sigma)$, the level of extortion will be $\sigma(1-\psi)\bar{P} + \psi B/(d(N))$. This will not deter the harmless action if

$$\frac{h}{d_0(N)} > \sigma(1-\psi)\bar{P} + \frac{\psi B}{d(N)} \text{ or } \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right) > \bar{P}.$$

Hence as long as

$$\frac{h}{d_0(N)} > \frac{\psi B}{d(N)},$$

then it will always be possible to deter the harmful behavior if $\bar{P}$ is sufficiently high. If $\psi B/((1-\sigma+\sigma\psi)d(N)) > h/(d_0(N))$,

then when

$$\frac{\psi B}{(1-\sigma+\sigma\psi)\,d(N)} > \bar{P} > \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right),$$

only ex-ante prevention is possible. When

$$\frac{\psi B}{(1-\sigma+\sigma\psi)\,d(N)} < \frac{h}{d_0(N)},$$

then it is always possible to have only ex-ante prevention.

If

$$d_0(N)\left(\frac{\psi B}{d(N)} + \sigma(1-\psi)\,\bar{P}\right) < h,$$

then ex-post punishment carries social costs $N(K_P + \lambda d_0(N)\,\bar{P} + \delta d_0(N)\,(\psi B/d(N)) - (1-\sigma(1-\psi))\bar{P}))$, which is less than ex-ante prevention as

$$(\lambda - \delta(1 - \sigma\,(1 - \psi)))\,\bar{P} + \frac{\delta\psi B}{d\,(N)} < \frac{h}{d_0\,(N)}.$$

Ex-post punishment with extortion will reduce costs relative to anarchy if and only if $v\,(C\,(N) - B) - K_P - d_0\,(N)\,(\delta\psi B)/d(N) + (\lambda - \delta(1 - \sigma\,(1 - \psi)))\,\bar{P}) > 0$.

As long as

$$vC'(N) - d_0'\,(N)\left(\frac{\delta\psi B}{d\,(N)} + (\lambda - \delta(1 - \sigma\,(1 - \psi)))\,\bar{P}\right) + \frac{\delta\psi B d_0\,(N)\,d'\,(N)}{(d\,(N))^2} > 0,$$

which we have assumed, then the left-hand side of the equation is monotonically increasing with *N*. As long as $v\,(C\,(N) - B) - K_P > 0$ is greater than zero for *N* sufficiently large and as $v\,(C\,(0) - B) < K_P$, there must exist a value of *N* at which ex-post punishment with extortion provides equal costs to anarchy. This value is increasing with K_P, B, λ, ψ, σ, and $\bar{P}$. □

Proof of Proposition 8. I ignore the welfare related to the harmless action as it is unchanged under this scenario. Total welfare if there is anarchy equals $b(N^{1-\gamma} i_{AN}^*(1 - \alpha v N)) + vB - K_I i_{AN}^*$, where i_{AN}^* is per capita spending which satisfies

$$\frac{(1 - \alpha v N)}{N^{\gamma - 1}} b'\left(N^{1-\gamma} i_{AN}^*(1 - \alpha v N)\right) = K_I.$$

Total welfare if there is full prevention equals $b(N^{1-\gamma} i_{PR}^*) - K_P - K_I i_{PR}^*$, where i_{PR}^* satisfies $N^{1-\gamma} b'(N^{1-\gamma} i_{PR}^*) = K_I$. If

$$1 > x\frac{-b''\,(x)}{b'\,(x)} \text{ for all } x,$$

then $i_{PR}^* > i_{AN}^*$ and if

$$1 < x\frac{-b''\,(x)}{b'\,(x)} \text{ for all } x$$

then $i_{PR}^* < i_{AN}^*$.

When *N* gets small, i_{PR}^* approaches i_{AN}^* and welfare under anarchy must be higher than welfare under full prevention, since we have previously assumed that $K_P > 0$. As *N* goes to $1/\alpha v$ then any finite infrastructure expenditure effectively provides zero benefits, and welfare goes to *vB*, which is less than welfare under full protection.

The derivative of $b(N^{1-\gamma} i_{PR}^*) - K_P - K_I i_{PR}^* - b\left(N^{1-\gamma} i_{AN}^*(1 - \alpha v N)\right) - vB + K_I i_{AN}^*$ with respect to *N* is

$$(1 - \gamma)N^{-\gamma}\left(i_{PR}^* b'\left(N^{1-\gamma} i_{PR}^*\right) - i_{AN}^*(1 - \alpha v N) b'\left(N^{1-\gamma} i_{AN}^*(1 - \alpha v N)\right)\right) + \alpha v N^{1-\gamma} i_{AN}^* b'\left(N^{1-\gamma} i_{AN}^*(1 - \alpha v N)\right)$$

or

$$\frac{K_I}{N(1-\alpha v N)} \times (1-\gamma)\left(i_{PR}^* - i_{AN}^*\right)(1-\alpha v N) + \alpha v N i_{AN}^*,$$

which must be positive if either N is small or close to $1/\alpha v$ because the term $(1-\gamma)(i_{PR}^* - i_{AN}^*)(1-\alpha v N)$ will be sufficiently close to zero in either case. If γ is close to one, the derivative will be uniformly positive, so there will exist a unique level of population which determines whether full prevention dominates anarchy.

If

$$1 > x\frac{-b''(x)}{b'(x)} \text{ for all } x,$$

then $i_{PR}^* > i_{AN}^*$ and $1 > \gamma$, and there exists a unique crossing point. If

$$1 < x\frac{-b''(x)}{b'(x)} \text{ for all } x,$$

then $i_{PR}^* < i_{AN}^*$ and $1 < \gamma$ then there also exists a unique crossing point.

If a unique crossing point exists, then it is increasing with K_I if and only if $i_{PR}^* > i_{AN}^*$ which will always hold if

$$1 > x\frac{-b''(x)}{b'(x)}$$

and never hold if $1 < x(-b''(x))/b'(x)$. □

Proof of Proposition 9. If $d(P_L(N) - \underline{P}_L) < \bar{P}$, then private provision generates the first best, and total costs are $wL(A_{FB}) + P_L(N)A_{FB}$, where A_{FB} minimizes this amount. If $P_L(N) - \underline{P}_L > d\bar{P}$, then the total financial costs of the project equal $wL(A_{SUB}) + (\underline{P}_L + d\bar{P})A_{SUB}$, where A_{SUB} is again cost minimizing. Naturally, costs are lower if subversion occurs. In both cases, the financial costs of producing the project are less under private than under public provision. But the total social costs of the project are now $wL(A_{SUB}) + P_L(N)A_{SUB}$, and these must be compared with

$$\frac{wL\left(A_{PUB}\right)(1-zs^*)}{1-s^*} + P_L(N)A_{PUB},$$

where

$$-\left(\frac{wL'(A_{PUB})}{1-s^*}\right) = P_L(N).$$

If $\bar{P}$ goes to zero, the public entity costs go to infinity, while the private costs are bounded. As such, for very low levels of institutional quality, private provision—even with expropriation—dominates public provision. If N is close to the corruption threshold for subverting the adjudicator, then this will never hold, then private provision provides essentially the first best, while the public version entails losses as long as $z < 1$. The private entity will acquire more land than the public entity if and only if $\underline{P}_L + d\bar{P} < (1-s^*)P_L(N)$. As long as this condition holds, then if $\lambda = 0$, private provision dominates public provision. As N gets arbitrarily large, A_{PUB} goes to zero, and $wL(A_{SUB}) + P_L(N) A_{SUB}$ also becomes larger than $wL(0)(1 - zs^*)/(1 - s^*)$, which is the maximum social cost given public provision that is independent of city size.

If $\lambda = 0$, the derivative of social costs given public provision with respect to N equals

$$P_L'(N)\left(\frac{L'(A_{PUB})(1 - zs^*)}{L''(A_{PUB})} + A_{PUB}\right),$$

which is less than $P_L'(N) A_{PUB}$ because higher costs of land encourage more labor and labor is underconsumed relative to the social cost-minimizing level. The derivative of social costs given private provision with respect to N equals $P_L'(N) A_{SUB}$. As long as $A_{SUB} \geq A_{PUB}$ then the slope is higher for private provision, and there must exist a unique point with $A_{SUB} \geq A_{PUB}$ where private provision and public provision yield equivalent costs. If N is close to the corruption threshold for subverting the adjudicator, then this will never hold, but as N grows sufficiently large, the private entity will always acquire more land, and if $\lambda = 0$ and $z = 1$, then social costs of private provision will be higher than the social costs of public provision.

Since increases in d lower social costs under private provision, but not public provision, they will raise the threshold at which public provision is preferred to private provision. □

References

Ades, Alberto F. and Edward L. Glaeser (1995). "Trade and Circuses: Explaining Urban Giants." *Quarterly Journal of Economics*, 110, 195–227.

Alesina, Alberto and Romain Wacziarg (1998). "Openness, Country Size and Government." *Journal of Public Economics*, 69, 305–321.

Becker, Gary S. (1968). "Crime and Punishment: An Economic Approach." *Journal of Political Economy*, 76, 169–217.

Bolt, Jutta and Jan Luiten van Zanden (2013). "The First Update of the Maddison Project: Re-Estimating Growth Before 1820." Maddison Project Working Paper No. 4.

Ciccone, Antonio and Robert E. Hall (1996). "Productivity and the Density of Economic Activity." *American Economic Review*, 86(1), 54–70.

Cutler, David M. and Grant Miller (2006). "Water, Water Everywhere. Municipal Finance and Water Supply in American Cities." In *Corruption and Reform: Lessons from America's Economic History*, University of Chicago Press, pp. 153–184.

De Vries, Jan (1984). *European Urbanization 1500–1800*. Harvard University Press.

Djankov, Simeon, Edward Glaeser, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2003). "The New Comparative Economics." *Journal of Comparative Economics*, 31, 595–619.

Durand, John D. and Cesar A. Peláez (1965). "Patterns of Urbanization in Latin America." *Milbank Memorial Fund Quarterly*, 43, S166–196.

Engel, Eduardo, Ronald Fischer, and Alexander Galetovic (2013). "The Basic Public Finance of Public–Private Partnerships." *Journal of the European Economic Association*, 11, 83–111.

Ethier, Wilfred (1979). "Internationally Decreasing Costs and World Trade." *Journal of International Economics*, 9, 1–24.

Friedlander, Dov (1970). "The Spread of Urbanization in England and Wales, 1851–1951." *Population Studies: A Journal of Demography*, 24, 423–443.

Friedman, David (1999). "Why Not Hang Them All: The Virtues of Inefficient Punishment." *Journal of Political Economy*, 107, 259–269.

Glaeser, Edward L. (2013). "A World of Cities: The Causes and Consequences of Urbanization in Poorer Countries." NBER Working Paper No. 19745, revised March 2014.

Glaeser, Edward L. and Joshua D. Gottlieb (2008). "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity*, 39, 155–253.

Glaeser, Edward L., Giacomo A. M. Ponzetto, and Kristina Tobio (2010). "Cities, Skills, and Regional Change." Economics Working Paper No. 1255, Department of Economics and Business, Universitat Pompeu Fabra, revised March 2011.

Glaeser, Edward L. and Bruce Sacerdote (1999). "Why Is There More Crime in Cities?" *Journal of Political Economy*, 107, 225–258.

Glaeser, Edward L. and Andrei Shleifer (2003). "The Rise of the Regulatory State." *Journal of Economic Literature*, 41, 401–425.

Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath (2013). "Urbanization With and Without Industrialization." Working Paper No. 2014-01, The George Washington University, Institute for International Economic Policy.

Gomez-Ibanez, Jose and Sheikh Zahid Masood (2006). "Vision Mumbai." Harvard Kennedy School Case Study No. 1917.0.

Hart, Oliver, Andrei Shleifer, and Robert Vishny (1997). "The Proper Scope of Government: Theory and an Application to Prisons." *Quarterly Journal of Economics*, 112, 1127–1161.

Krugman, Paul (1991). *Geography and Trade*. MIT Press.

Krugman, Paul, and Raul Livas (1996). Trade Policy and the Third World Metropolis. *Journal of Development Economics*, 49, 137–150.

Matsuyama, Kiminori (1992). "Agricultural Productivity, Comparative Advantage, and Economic Growth." *Journal of Economic Theory*, 58, 317–334.

Sabine, E. L. (1937). "City Cleaning in Medieval London." *Speculum*, 12, 19–43.

United Nations (2012). *World Urbanization Prospects: The 2011 Revision*. Department of Economic and Social Affairs, Population Division.