NBER WORKING PAPER SERIES

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

Edward L. Glaeser

Working Paper 19745
http://www.nber.org/papers/w19745

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2013

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. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

A World of Cities: The Causes and Consequences of Urbanization in Poorer Countries
Edward L. Glaeser
NBER Working Paper No. 19745
December 2013
JEL No. R0

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 thirty 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 tradeoff 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.

Edward L. Glaeser
Department of Economics
315A Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
eglaeser@harvard.edu

I. Introduction

Between 1950 and 2010, the world's urbanization rate increased from under thirty percent to over fifty percent (United Nations, 2012). In countries such as China and Korea, urbanization accompanied income growth, following a familiar historic pattern. But the more surprising fact is that there are many countries, including Pakistan, Haiti and the Democratic Republic of the Congo, in which significant urbanization occurred despite persistent poverty and problematic politics. Why did poor mega-cities, like Karachi and Kinshasa emerge, and how do their policy challenges differ from those faced by the world's wealthier cities?

In Section II, I document the rise of poor country urbanization. In 1960 the urbanization rate was under ten percent in the majority of nations where per capita annual incomes were below $1,000 in current dollars, but no similarly poor country is that rural today. The U.S. only became one-third urbanized in 1890, when per capita Gross Domestic Product (GDP) was over $5000, but the poorest urbanizing countries today have hit that threshold with per capita GDPs under $1200. There have been million person cities at such low levels of income historically, such as classical Rome, Baghdad and Kaifeng, but typically these were capitals of capably governed empires. Despite the increasing urbanization of poor countries, the cross-country link between urbanization and income remains as strong, because richer countries have also urbanized dramatically.

After documenting these facts, Section III proposes a simple explanation of these phenomena that draws heavily on Matsuyama (1992) and Gollin, Jedwab and Vollrath (2013). Like those previous papers, I argue that globalization radically changes the process of urbanization. In an age of autarky, nations needed to develop agricultural surpluses and strong domestic transport networks in order to feed their cities. Today, globalization means that Port-Au-Prince can be fed with imported American rice. The model shows that in closed economies, urbanization typically increases with agricultural productivity or transport improvements, but in open economies, these comparative statics reverse themselves. In an open economy, rural deprivation can mean increased urban growth, as in Kinshasa today.

To test the globalization hypothesis, I examine the link between agricultural productivity, country size and urbanization in 1961 and 2010. I take small country size to be a proxy for openness, and do indeed find that in both years, agricultural productivity is far more strongly correlated with urbanization in large countries than in small countries. I also find a sharp decline in the connection between local agricultural productivity and urbanization between 1961 and 2010, which is compatible with the hypothesis that global food supply has reduced the need to develop a domestic agricultural surplus before building cities.

Despite their low levels of economic development, the new, poor mega-cities still grapple with the same adverse urban externalities that have troubled western cities for centuries, such as contagious disease, traffic congestion and crime. They also exhibit the high costs of housing that come from density. Yet they must face these problems with neither economic wealth nor capable government. Section IV and V of this paper address the challenges of governing the world's poor mega-cities.

Section IV focuses on controlling the externalities that emerge when people live close to one another, through use of ex ante prevention or ex post punishment. In the model, individuals take externality-creating actions, such as careless disposing of human waste or driving on crowded roads, which have a larger social cost when the size of the city population is larger. A critical assumption is that the probability of catching a perpetrator declines with city size, which is an empirical regularity noted by Glaeser and Sacerdote (1999). I also assume that institutional quality acts as a limit on the ability to punish public officials who misbehave.

Together these assumptions create an institutional possibilities frontier, as in Djankov, Glaeser, LaPorta, Lopez-de-Silanes and Shleifer (2003), but unlike that previous paper, city size is an important determinant of the shape of that frontier. When city sizes are smaller, it is easier to maintain the rule of law with small penalties, because the probability of detection is larger. As a result, the large cities that have the most to gain by controlling their externalities also have the least ability to enforce good behavior through effective punishment.

The first sub-section allows the government to respond to harmful behavior either by ignoring it, or assessing large penalties or assessing light penalties. Strong penalties are always preferable to light penalties in the model, because they eliminate the bad behavior entirely, but they may not be feasible either if institutions are too weak or if city sizes are too large. When only ex ante punishment is available, the model predicts that smaller cities will always prefer non-regulation, no matter how good their institutions may be, because the costs of establishing a system are lower than the benefits. In the largest cities, as institutional quality improves, cities march through a progression of no penalties, weak penalties and then strong penalties. Medium sized cities leap immediately from no penalties to strong penalties.

Yet historically, there have been some examples of cities that respond to chaos with draconian means, even at relatively low levels of institutional development. The early stages of communist China and Russia may provide one example, and Julius Caesar's banning of wheeled vehicles from Rome's streets may be another. As in Glaeser and Shleifer (2003), I assume that the state can respond to an externality either by imposing an ex post punishment or taking some form of ex ante prevention, preventing people from engaging in a harmless action in order to eliminate any possibility of a harmful action.

At low levels of institutional development, draconian prevention or anarchy become the only two options and prevention carries lower costs if cities are sufficiently big. At higher levels of institutional development, ex post punishment becomes possible, first in smaller cities and then in bigger cities. When ex post punishment is possible, it creates fewer social costs than ex ante prevention.

These results change somewhat when I assume that bribes can be large without limit. In this case, any policy can be enforced at any level of institutional development, as long as penalties are set to be sufficiently high. The downside of such high penalties is that they create the potential for police extortion. These results echo those in Djankov, Glaeser, LaPorta, Lopez-de-Silanes and Shleifer (2003) who argue that countries have an institutional possibilities frontier that trades the costs of disorder, which are represented by the externality in the model, with the costs of dictatorship, which is reflected by the costs of extortion. Ex ante prevention somewhat limits the costs of extortion, because people end up avoiding any occasion in which they can be help up by the police, but the costs of repressing harmless, beneficial activity can be considerable.

Section V turns to the provision of infrastructure and its relationship with city size, wealth and institutional strength. Infrastructure provision, such as sewers, aqueducts, and extra highways reduce the downsides of density, but the marginal returns to these investments may either rise or fall when harmful behavior is controlled. Harmful behavior reduces the effective level of infrastructure, which makes the provision of infrastructure more valuable because of its scarcity. Harmful behavior also reduces the effective infrastructure that is produced with a unit of investment. Institutional quality will make infrastructure more attractive if the second effect dominates the first, which depends on the degree of concavity in the infrastructure benefit function.

This section also follows Engel, Fisher and Galetovic (2013) and examines the private provision of public infrastructure through devices such as public-private partnerships. Public-private partnership can offer a means of avoiding the losses due to waste and corruption that occur with public provision at low levels of institutional quality. However, the downside of private provision is the possibility that the private infrastructure builder will itself corrupt the state, especially if the service requires implicit or explicit subsidies. I specifically focus on the subversion of the land acquisition process, which has been a particularly important issue in the development of Chinese cities. The model suggests that public provision will be more attractive when urban populations are large, because large populations increase the social costs of expropriating too much land, while private provision will be more attractive when city sizes are smaller. Private provision also carries lower costs when institutions are strong.

The finally part of Section V turns to the issue of housing price and land use regulation. Housing costs are minimized when state action protects land from private expropriation and lightly regulates land use. In the west, private protection of property generally preceded significant land use controls, but many developing countries manage to combine weak protection from private incursions (e.g. squatting) along with significant restrictions on new building.

This combination is understood as reflecting the objective function of a grasping public sector. Protecting private property from squatters is costly to enforce ex ante; restricting new dense development is far easier. As a result, the public sector regulates new building to extract concessions from developers, but does a poor job of protecting private property.

Section VI 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.

II. The Rise of Poor Mega-Cities

In 1960, poor nations were overwhelmingly rural nations and South Korea was one of only a few poor states with an urbanization rate over 25 percent. 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 went from 18 percent in 1950 to 47 percent in 2011 (United Nations, 2012). In fifty years, Botswana has grown three to sixty percent urban. China is now over fifty percent urban, but it was only 16 percent urban in 1960.

These two examples of rapid urbanization are unsurprising given the equally rapid economic growth in these countries. Real per capita incomes are 19 times higher in Botswana today than they were in 1960 and China's real income has increased nearly eight-fold. The historic link between urbanization and industrialization seen in the U.S. and Europe has led us to expect city growth in countries with rapidly growing incomes.

Yet urbanization is also occurring in countries like Bangladesh and Kenya, where per capita incomes have increased by less than $250 over the past 50 years. Despite stagnant incomes, urbanization has increased from five to 28 percent in Bangladesh and from seven percent to 24 percent in Kenya. Also remarkably, the urbanization of these poorer places has been particularly centered on a dominant primate city. Three million people inhabit the Nairobi agglomeration and Dhaka is home to 15 million inhabitants.

Figures 1A and 1B provide a visual confirmation of the phenomenon. Both figures show the relationship between per capita incomes in 2012 dollars and urbanization across countries with per capita incomes below $5000. Figure 1A shows the strong positive relationship for 1960, where there are no really poor places with high levels of urbanization. Figure 1B shows the same relationship for 2010. Not only has the overall level of urbanization increased, but the growth has been particularly dramatic among particularly poor places.

These graphs do not mean that the estimated slope linking urbanization and the logarithm of per capita income has declined over time. To the contrary, the estimated impact of urbanization on the logarithm of per capita GDP was 3.6 (standard error of .24) in 1960 and 5.3 (standard error of .35) in 2010, across all countries. Even among poor countries (defined as having per capita GDP levels below $5000 in 2012 dollars), the estimated coefficient when the logarithm of per capita GDP is regressed on urbanization has been roughly constant, around 2.6, over the 50 year period. However, the r-squared of that regression has dropped significantly from .54 to .33, reflecting the increasing number of extremely poor, urbanization nations.

Moreover, there is a strong link between GDP growth and initial urbanization among the poorest places, shown in Figure 2. A ten percent higher level of urbanization among these poor countries in 1960 is associated with .23 log points faster growth between 1960 and 2010. This coefficient declines by one-third, but remains statistically significant, when I also control for the years of schooling in this sample of countries in 1960. I am not suggesting that the link between urbanization and GDP growth among poorer nations is causal, but the robust correlation should give us pause before embracing policies aimed at reducing the level of urbanization in a country.

To give a more concrete sense of places that are poor but urbanized, Table 1 lists the seven most extreme nations in my sample, where per capita incomes are below $1250, populations are over ten million and urbanization is over one-third: the Democratic Republic of the Congo, Zimbabwe, Mali, Haiti, Pakistan, Senegal and the Cote D'Ivoire. Five of these countries are in Africa. Haiti and Pakistan are the two non-African countries. Every one of these countries has an agglomeration with more than one million inhabitants, and there are three cities of four million or more. There are another eight smaller countries that also have incomes below $1250 and urbanization rates below one-third, including Liberia, a country almost as poor as the Democratic Republic of the Congo, and the Kyrgyz Republic, another non-African example.

In 1960, there were only two countries where per capita incomes were under $1250 in current dollars and urbanization rates were over one third: Egypt and Nicaragua. The number of countries rises to ten if we raised the income cutoff to $2000, and it would include seven nations from Latin America (including Brazil and Peru) and three from the near East (Syria and Iraq as well as Egypt). Indeed, in 1960, urbanization in the typical South American nation was about 11 percent higher than its income level would predict, and the region's high level of urbanization was considered to be a puzzle at the time (Durand and Paláez, 1965), as was the extreme size of the region's biggest cities. The unusually high degree of urbanization in Latin America remains, although since the region has gotten much wealthier, it is no longer the epicenter of impoverished urbanization.

To put the urbanization of today's poorer nations in perspective, it is helpful to turn to two different types of history. Figures 3 and 4 show the time path of urbanization for the United States and then for England and Wales (based on Friedlander, 1970). The implied time series coefficient when urbanization is regressed on income is .24 in the U.S. and .25 for England and Wales, which are both relatively close to the cross-sectional coefficients discussed earlier. The U.S. was one-third urbanized in 1890 when its income was nearly $6,000 and one-half urbanized in the 1920s when its income was close to $10,000.

The U.K. became one-third urbanized in 1861 when its income was somewhat lower, around $5,000, and one-half urbanized in 1881, when its income levels were closer to $6,000. Both of these figures are substantially higher than the incomes reached by the poorer urbanized places today, although it is notable that the U.S.—a more closed economy— was more prosperous than the U.K. before it urbanized. France, Germany, and the Netherlands are closer to the U.S. where according to De Vries (1984) urbanization levels didn't reach 50 percent until income levels were well over $5000.

There were, of course, mega-cities in the distant past which had substantially lower income levels. Table 2 lists the cities that are thought to have reached a population of around one million before 1875. Rome reached 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 and incomes in Italy were somewhat higher. The capitals of Beijing and Baghdad reached that population level between 800 and 1200, at substantially lower national income levels (Maddison via 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 didn't reach that threshold until their national incomes reached $3,000 and $4,565 respectively.

The striking fact about this list is that every one of these cities, except for New York, was the capital of a large empire. These empires existed precisely because their public sector was capable of conquering and administering vast land areas. They may not have had per capita incomes that are comparable to those in the United States in 1880, but these countries were able to use the powers of the state 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. Rome was also famous for its governmentally produced waterworks, as was Baghdad and Kaifeng. These places didn't have wealth, but they did have a competent public sector, precisely because they never would have grown so great with a capable government.

Today's poor but urbanized nations cannot rely on such public competence. The overall correlation between government effectiveness and urbanization is significant, but that reflects only the well-known correlation between public sector competence and income. Among countries, with income levels less than $5,000 there is no correlation between governmental effectiveness and urbanization. Figure 5 shows the weak negative correlation between the 2010 estimate of government effectiveness and urbanization across the 36 countries with per capita incomes under $1500 and populations over two million. Indeed, Haiti and the Democratic Republic of the Congo are among the nations with the lowest ratings of governmental competence in the world.

III. Why Have Poor Mega-Cities Spread?

I now turn to a positive model of urbanization that is meant to shed light on the rise of urbanization in countries that are both poor and poorly managed. The model draws heavily on Krugman (1991) and the vast literature that followed his start. The ideas in the model are highly indebted to the work of Matsuyama (1992) and Gollin, Jedwab and Vollrath (2013).[2] The fundamental result is that openness can reverse the link between prosperity and urbanization, helping to explain why we have seen this unprecedented rise in poor mega-cities.

[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.

I consider first the possibility of a single city, situated in the middle of a line of $2\overline{d}$ units of one agricultural land. The city occupies no land and makes manufactured goods that can be shipped at no cost. Farmers on the agricultural land make farm goods that have iceberg transportation costs, so that if one unit of the goods are shipped, then $e^{-\tau d}$ goods arrive at a place "d" distance away. The total population equals N.

The agricultural sector is characterized by a distribution of land density (which is uniform), a distribution of population density n(d), which will be a function of distance from the city (d), and a distribution of production. Output per worker equals $\underline{C} + A_A L^{\gamma}$, where $\underline{C}$ is a constant, L represents land per worker, A_A represents agricultural productivity and $0 < \gamma < 1$. Land per worker equals the inverse of density, so output per worker will equal $\underline{C} + A_A n(d)^{-\gamma}$, and total output at distance d will equal $\underline{C}n(d) + A_A n(d)^{1-\gamma}$.

Individual welfare is determined by Stone-Geary preferences, so that utility is $\frac{\theta_i}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}\left(C - \right.$ $\underline{C}\alpha M 1-\alpha$, where θi represents local amenity levels which equal 1 in the rural sector, C represents consumption of the primary, agricultural good and M represents consumption of the composite manufacturing good. For algebraic convenience, I have assumed that the minimal level of farm production 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 P_M.

A farmer at distance d from the city will have a welfare level of $A_A n(d)^{-\gamma} P_M^{\alpha-1} e^{-(1-\alpha)\tau d}$. In a spatial equilibrium, where land is not rationed by price, the welfare levels for the farmers will have to be equal across space, which implies that $n(d) = e^{-\psi d} n(0)$, where ψ denotes $\frac{(1-\alpha)\tau}{\gamma}$.[3] If there are N_A farmers in this zone, then welfare levels in rural sector equal $A_A N_A^{-\gamma} P_M^{\alpha-1}\left(2 - 2e^{-\psi\overline{d}}\right)^{\gamma}\psi^{-\gamma}$, which is declining in the size of the agricultural sector.

If the urban sector has a population level of N_U, I assume ϵN_U workers operate independent firms producing differentiated products. Each manufacturing worker produces A_M units of

[3] I have implicitly assumed that property rights are weak, and all farmers are squatters who freely occupy land, but must share it with other farmers. This will typically lead to overcrowding in the rural sector. The alternative assumption, which is more realistic in economies with well defined property rights, is that land is rented.

manufacturing goods. Following Ethier (1979), these products are costlessly aggregated into the non-differentiated manufactured good, so that $M = (\int (x(i)^{\sigma} di)^{\frac{1}{\sigma}}$. Manufactured goods cannot be shipped until they are aggregated 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 $A_M N_U^{\frac{1}{\sigma}}(1-\epsilon)\epsilon^{\frac{1-\sigma}{\sigma}}$. Workers wages, denominated in the agricultural good, will equal $\sigma(\epsilon N_U)^{\frac{1-\sigma}{\sigma}} A_M P_M$ and entrepreneurial profits will equal $(1-\sigma)(1-\epsilon)\epsilon^{\frac{1-2\sigma}{\sigma}} N_U^{\frac{1-\sigma}{\sigma}} A_M P_M$.

In the spirit of Krugman (1991), there could be free entry into entrepreneurship at the cost of one's time, and in that case wages and entrepreneurial profits are equal and $1-\sigma=\epsilon$. Alternatively, ϵ could be a fixed proportion of the population, so that each worker has a random chance of becoming an entrepreneur.[4] I will assume that quality of life in the city equals $\theta_0 e^{-\delta N_U}$, which can capture a range of negative effects of crowding including disease, congestion and even housing costs.[5] 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 and 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. Proposition1 reminds us 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.

[4] 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-founding of human capital externalities (Glaeser, Tobio and Ponzetto, 2010).
[5] 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.

The transportation system is also a critical element in the urbanization process in a closed economy, because if food doesn't make it to market, then the cities cannot be fed. Those transportation networks in turn often depended on the power of the state, such as imperial Rome, which enabled the transfer of food over vast distances.

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 fed enough. A natural parallel to this in European urban life is that it is alleged that the Black Death actually made urbanization easier by increasing the available per capita food supply.

Is there too much or too little urbanization in equilibrium? The derivative of average social welfare with respect to the urbanization rates equals:

$$(1)\quad (v_u - v_a) - \delta N v_u + \gamma v_a + \frac{1-\sigma}{\sigma}\vartheta v_u + \Big((1-\alpha)(1-u)v_a - \alpha u v_u - (\vartheta - 1)u v_u\Big)\left(-\frac{1}{P_M}\frac{\partial P_M}{\partial u}\right),$$

where v_u and v_a reflect welfare in the urban and rural sectors respectively, u is the urbanization rate, and ϑ equals $\frac{(1-\epsilon)(\epsilon N_U)^{\frac{1-\sigma}{\sigma}} A_M P_M}{(1-\epsilon)(\epsilon N_U)^{\frac{1-\sigma}{\sigma}} A_M P_M - \underline{C}}$.

The first term $(v_u - v_a)$ vanishes in a spatial equilibrium, as it reflects the net improvement in utility for the marginal migrant. The second term reflects the social losses from decreasing amenities levels in the urban area from in-migration. The third term reflects the reduction in crowding in the countryside, which offsets slightly, the congestion in the city. This effect ultimately represents the weak definition of property right in the rural sector, and would diminish or disappear entirely if property right there were better defined. The fourth term reflects the agglomeration economies in the city. The term $\frac{1-\sigma}{\sigma}$ is the elasticity of productivity with respect to city size. The terms ϑv_u multiply to give us marginal utility times expected earnings in the city.

The fifth term reflects the impact of a pure change in the price of the manufactured good caused by urbanization ($\frac{\partial P_M}{\partial u} < 0$). This term is essentially a pure transfer between the two sectors and it would vanish if $\vartheta = 1$ and $\alpha = 1 - u$. If $\alpha < 1 - u$, then the agricultural share of the population is greater than its share of the economy, and that means that lower manufactured prices essentially redistribute from the few to the many, which increases social welfare given this utilitarian social welfare function with no diminishing returns to income at the individual level. If $\vartheta > 1$, then urbanites have a greater marginal utility of income, and that makes the price decrease associated with urbanization less desirable.

Ignoring these purely distributional effects, there are three distinct market failures that co-exist in the system: adverse urban diseconomies, benefits from more land per worker in the rural sector and agglomeration economies. At this point, we are far from being confident about the relative magnitude of these effects, especially in the developing world.[6] As a result, it is difficult to know whether the growing cities of the world's poorest countries are too big or too small.

In the spirit of Matsuyama (1992), we 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 P_M, even if no one lives in the city. I assume that while $\frac{1}{\sigma}$ may be greater than $1 + \delta$ (providing increasing returns in the urban sector), it cannot be greater than $2 + \delta$.

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 $(1 - \epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} A_M P_M > \underline{C}$, 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 $\overline{d}$, δ and $\underline{C}$. The value of A_A^* is also rising with ϵ as long as $1 - \sigma > \epsilon$ and N as long as δ is sufficiently small. If there is an equilibrium with a positive level of urbanization, then generically, there will exist two levels of urbanization that satisfy the spatial equilibrium condition. At the equilibrium with the lower urbanization level, increases in urbanization will cause welfare in the urban sector to be higher than welfare in the agricultural sector. At the equilibrium with the higher urbanization level, increases in urbanization will cause the welfare in the urban sector to be lower than the welfare in the agricultural sector.

[6] Glaeser and Gottlieb (2008) discuss the estimation of these magnitudes in the U.S. 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.

If there exists no level of urbanization at which welfare in the urban sector is higher than welfare in the rural sector, then everyone farms. The possibility of a city exists if $(1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} A_M P_M > \underline{C}$ and if A_A is not too high. The first condition requires that the urban sector—even if completely dominant—can earn enough to feed itself by selling its goods on the world market. The second condition requires that the rural sector is not too productive. Factors that make the urban sector more productive or pleasant increase the maximum level of agricultural productivity that permits urbanization.

Proposition 2 highlights the fact that as long as there is any urbanization at all, there are multiple equilibria that come from agglomeration. In the closed economy, a small city means an expensive manufactured good and that ensures that some people want to work in the manufacturing sector. In the open economy, that price effect doesn't exist, and only agglomeration economies matter. As such, in the open economy setting, zero urbanization is a decided possibility. Complete urbanization is not, because agricultural productivity becomes infinite in that case. The nature of this multiple equilibrium setting also suggests that small changes in underlying parameters, such as small increases in urban productivity or decreases in rural productivity, can create far more massive urban change than in the open economy model, where urbanization can occur with astonishing rapidity. These aspects of the model seem to fit the rapid switch of the developed world from being overwhelming rural to being significantly urban.

I now consider the comparative statics in the second intuitively "stable" equilibrium, where 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, $\overline{d}$, and $\underline{C}$. Urbanization will be rising with ϵ as long as $1-\sigma > \epsilon$ and N as long as δ is sufficiently small.

As such, many of the previous comparative statics are reversed. Urbanization now depends quite strongly on the variables that determine productivity in the urban sector, such as A_M, P_M and θ_0. Variables that decreases the productivity in the rural sector will now cause the urbanization level to increase, such as A_A or τ. Unsurprisingly, the urban quality of life also determines the level of urbanization in the city.

Together Propositions 2 and 3 highlight the fact that in open economies, urbanization can be the result of misery, rather than productivity, especially in the rural sector. Because trade has alleviated the need for agricultural productivity, cities can develop despite enormously poor hinterlands. This creates both an opportunity—the ability to escape terribly poor rural land—and a challenge. Massive cities can develop at far lower levels of income as we see across the world today.

If we repeat the welfare exercise that we did before, we see that with a fixed manufacturing price, the social benefits of increasing urbanization equals $v_u - v_a - \delta v_u + \gamma v_u + \frac{1-\sigma}{\sigma}\vartheta v_u$. The open economy eliminates the redistributive effects that operate through the price of the manufactured good.

Mega-Cities versus Dispersed Urbanization

Will urbanization lead to a dispersed set of urban areas, as in Europe or the U.S., and a dominant mega-city as in much of Latin America and sub-Saharan Africa? This question is discussed by Ades and Glaeser (1995), and indeed many developing world mega-cities are capitals that have historically benefitted from governmental largesse. Krugman and Livas (1996) argue that the third world's mega-cities reflect may reflect import substitution policies that reduce international trade. The discussion that follows illustrates that the reverse can be also be true, and that globalization can lead 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 $\overline{d}$ units of agricultural land. In the case of the open economy, both cities will be able to trade with the outside world. In the case of the closed economy, we focus on the symmetric equilibrium where farmers will trade only with the nearest city. In both cases, the segments essentially amount to economies identical to those discussed above, but each with half as much land and half as many people. To simplify matters, I also assume that $\underline{C} = 0$.

It follows immediately, that a symmetric two city equilibrium will always exist in the closed economy, for this equilibrium is equivalent to two one city equilibria with half as many people and have as much land. I prove in the appendix that a symmetric two city equilibrium is actually 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 relatively to rural life when there are multiple cities.

In Proposition 4, I ask whether urban pioneers face strong incentives to start a second city, and whether a two city equilibrium is superior to a one city equilibrium. An equilibrium will generally exist in which only one city exists. The agglomeration structure implies that if there is no one in the city, and the price of the manufactured good is finite, then there will be no incentive to move to the city, as long as that city is sufficiently small. Thus, to answer the first question, I will focus on the minimum scale at which a second city can exist and be competitive with the first city. In the open economy model, in which urban welfare doesn't depend on the size of the agricultural sector, this amounts to just asking what is the minimum city size needed to generate the same welfare as the existing city, assuming that the existing city represents the larger of the two equilibrium outcomes discussed in Proposition 2. In the open economy case, I again ask what is the minimum scale needed for a second city to deliver equal utility to the first city, assuming that the level of urbanization is constant.

In both cases, the formation of a second city will generate advantages for farmers, by enabling some of those farmers to save transport costs. In the closed economy model, this will also yield benefits for the urbanites who will face a better price for their manufactured goods. In the open economy model, the urban pioneers will receive none of these advantages since the manufactured good is determined by world prices. In the closed economy model, the urban pioneers face a strong incentive to be as far from the old city as possible, and to locate on the opposite side of the circle. In the open economy model, urban pioneers face no such incentives. Typically, they would just choose a good port, but for symmetry's sake, I assume that they too locate at the opposite side of the circle.

In the closed economy, urban residents themselves benefit by having cities interspersed throughout the countryside, since this improves the terms of trade by lowering the agricultural produce lost in transport. In an open economy, there is benefit to the rural residents by having dispersed cities, but the only benefit to urban residents is the reduction in congestion.

Proposition 4: (a) In the open economy, if the equilibrium city size (at the higher urbanization level) with one city is less than city size that maximizes urban welfare, then a new city must be at least as large as the old city to attract urban pioneers, and if the equilibrium city size (at the higher urbanization level) with one city is greater than the city size that maximizes urban welfare, then a second city must be sufficiently large so that if it were the only city, urban welfare would be greater than rural welfare.

(b) In a closed economy, then a second city of any size can attract urbanites, as long as $\tau\overline{d}$ is sufficiently large and $1 < (1+\alpha)\sigma$.

(c) In the open economy case, then welfare is always higher in the stable equilibrium two cities than with one city if the city size with two cities is higher than the city size that maximizes urban welfare, and welfare is always higher with one city than with two cities if the city size with one city is less than the city size that maximizes urban welfare, which will always be the case if $\delta = 0$. In an open economy, welfare is always higher with two cities, even if $\delta = 0$, as long as $\tau\overline{d}$ is sufficiently large and $\sigma(1+\alpha\gamma) + \alpha(1-\sigma) > 1$.

In an open economy, the appeal of multiple cities depends essentially on the battle between agglomeration economies and urban congestion. If agglomeration economies are stronger than congestion, so that the city sizes are smaller than the city size that maximizes urban welfare, then a second city would have to be as big as the initial city to attract new residents. If congestion effects are strong, which seems quite relevant in the developing world, then a new city would still need substantial scale. At a minimum, it 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, the price of the manufactured good is rising with the number of cities, because decreased distances between the city and the hinterland is increasing the amount of available agricultural surplus. The benefits of increasing the number of cities—to the urbanites themselves—are always higher in the closed economy. Moreover, an increase in the number of cities will always increase the urbanization rate, even if there are no urban disamenities. A sufficiently dispersed number of cities can justify a wide range of urbanization levels.

The difference between open and closed economies helps, to us at least, explain the nature of urbanization in Europe, where cities are dispersed and 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 developed during a period of high transportation costs and generally limited global trade, at least in core agricultural products. This led to a dispersed set of smaller cities.

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. My summary measure of agricultural productivity is the product of cereal yield per hectare in the country and arable land per capita in the country. As these measures are influenced by the level of technological development which is in turn potentially influenced by the level of urbanization, these results must be seen as being merely suggestive.

Table 3 regresses:

(2) $Urbanization = a + b_1 Log(Prod_{Ag}) + b_2 Log(Pop) + b_3 Log(Prod_{Ag}) * Log(Pop)$

where $Prod_{Ag}$ represents the agricultural productivity measure and population 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 are highly likely to be dependent on the level of development and urbanization, country population is at least somewhat more independent. My interest is not in population itself, but the interaction between population and agricultural productivity, testing the hypothesis that the connection between urbanization and productivity will be weaker in smaller 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. For example, in a country that is one log point larger than the international mean, the estimated impact of one extra log point of agricultural productivity is urbanization rises to 13.3 percentage points. Holding these other variables constant, more populous countries were less urbanized.

Regression 2, like regression 4, includes continent dummies and this causes the estimate coefficient on agricultural productivity to drop by almost fifty percent. About half of the cross-national relationship between agricultural productivity and urbanization in 1961 is explained by differences across continents. That effect, however, remains significant and the interaction with population also remains significant.

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

These results hardly prove the hypothesis illustrated by the model, but they also fail to reject the model's key implications. The link between agricultural productivity and urbanization has declined over time, and it is also weaker in small countries than in big countries. By 2010, there is even a weak 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 is poor, poorly governed nations.

IV. Externalities and Density

Cities can create great positive externalities, not just the benefits of low shipping costs, but the spread of knowledge that creates everyday human capital and collaborative chains of creativity. Yet cities can also generate negative externalities as well. The same urban proximity that speeds the flow of goods and knowledge also speeds the flow of infectious disease and facilitates crime. The negative externality of congestion, for example, is practically synonymous with density.

We now turn to the causes of low quality of life in urban areas—the externalities that are associated with density. These include traffic congestion, water-borne or other contagious diseases, and crime. High housing costs are also a downside of urban living, not an externality but rather an inevitable consequence of high land costs. Still, urban policies, such as restrictions on land use, can artificially increase the costs of living.

Historically, these downsides of density have been fought with infrastructure investment or behavioral modification (i.e. punishment and fines) or both, and almost all of these problems can be solved by competent governments with enough money. Even the densest agglomerations can provide a virtually unlimited supply of safe clean water by desalination and spending enough on waste water treatment, like Singapore. Congestion can be eliminated by sufficient road building, public transport and electronic road pricing. American cities have become safe through a combination of large scale imprisonment and effective, if expensive, policing. An essentially unlimited supply of usable space can be provided even in tiny land masses by building up.

Yet all of these solutions require wealth or good government or both, and developing world mega-cities have neither. To address their issues, it is critical to recognize the institutional limitations that challenge the ability to deal effectively with negative urban externalities. I first discuss crime and punishment in urban areas, and then turn to infrastructure in Section V.

This model builds on the models of Djankov, Glaeser, LaPorta, Lopez-de-Silanes and Shleifer (2003) and Glaeser and Shleifer (2003) in which the ability to punish bad behavior is limited by the institutional strength of a particularly society. The added contribution here is to embed that insight in an urban setting in which the costs of bad behavior are a function of city size. The model will produce implications about the appropriate response to negative social problems, although sometimes the appropriate response will be to do nothing. This may sadly suggest that the awful conditions in many developing world mega-cities are the best that can be expected given the current level of institutional quality. While all of the models that follow are connected, they are totally distinct from the model described in Section III, and as such, I reuse for different purposes some of the notation used above.

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." Individuals do not know whether they will have the opportunity to undertake the "harmful action" at the time that they undertake the "harmless action." The harmful action might be introducing waste into a potential water source, or stealing goods from a neighbor.

If the harmful action is undertaken, then the action either brings private benefits of B+A or B, and the individual knows the benefits before undertaking the action. A share πv of the population will receive the higher benefit level if they undertake the action, and the remainder will receive the lower benefit level. This heterogeneity is meant to reflect the possibility that some people will benefit more undertaking the action than others, and may consequently be more difficult to deter.

The social cost of the harmful action is denoted C(N), where N reflects the city size and C'(N)>0. I will assume through that C(0)>A+B, so that it is always socially preferable that crime does not occur. I assume, throughout, that the action does more damage if the city is larger or denser. If the action is waste disposal or driving on a crowded street, then the connection with city population is obvious. If the waste spreads disease, then there are more people who can potentially be infected by the disease.

If the action is crime, then the connection with city population is less clear. After all, the direct loser from the crime is just the person who is robbed of his property. However, if we broaden the costs to include the self-protection that urbanites take in response to crime, then the downsides of crime would indeed be higher in cities, although formally modeling those costs would require a somewhat different and more complicated model. An alternative assumption is that the potential for crime is larger in cities because of proximity to abundant victims (Glaeser and Sacerdote, 1999), which would produce similar results. In the next section, I will allow infrastructure to alter the costs of this damage.

The most natural response to this action is to impose a punishment of size P on people who are caught doing the harmful action. I assume that if a punishment of size P is imposed, then the social cost of that punishment equals λP, with λ may be less than one (if the punishment is a fine) or potentially greater than one (if the punishment is prison). Everyone is assumed to be risk neutral, so that if the probability of arrest is denoted d(N), then, as in Becker (1968), the punishment deters those individuals who have low benefits from perpetrating the action if d(N)P>B and the punishment deters everyone if d(N)P>A+B.

Significantly, I assume that the probability of arrest is also a function of city size, and that is grounded in both fact and the long-standing understanding of urban anonymity. Glaeser and Sacerdote (1999) report that arrests per reported crime decline steadily with city size across American cities, and this correlation probably understates the true negative correlation because people in large cities report crime less often, perhaps because a report is less likely to lead to an arrest. One explanation for this link is that police often solve crimes by considering a range of possible suspects, and the number of possible suspects is much larger in a big city than in a small town. Anonymity is a fact of urban life—it is hard to know everyone—and that makes it more difficult to track down malefactors.

The cost of creating a criminal justice system that engages in ex post punishment equals NK_P. If there were no institutional limitations, so the system worked perfectly, then at a cost NK_P, a system could be established that would deter all crime, by setting a penalty equal to (A+B)/d(N). This cost must be compared with the social costs of anarchy, which equals $N(v(C(N) - B - \pi A)$. If $v(C(0) - B - \pi A) < K_P$, but $v(C(N) - B - \pi A) > K_P$, for N that is sufficiently large, which I assume, then it will be optimal to develop a criminal justice system that completely deters crime if and only if the city size gets sufficiently large. Since it costs no more to deter all crime than to deter the criminals with weak incentives, if there are no enforcement challenges, then it is preferable to set penalties high and deter everyone.

This result would only be strengthened if I assumed, as is probably realistic, that there are fixed costs in establishing the criminal justice system that could be shared across a wider city population. The increasing costs of negative externalities in big cities may help explain why criminal justice and policing emerges first in cities, including both the cities of the classical world, and during the post-medieval period, but it does little to help us understand the current problem of developing world mega-cities.

To introduce institutional limitations, I assume that the maximum expected punishment that can be imposed on a law enforcement official who takes a bribe is $\overline{P}$. This punishment is meant to summarize the level of institutional development, not because countries with weak institutions can't have draconian punishments on their books, but because those countries will have great difficulty enforcing such penalties. In countries with weak institutions, or in corrupt cities within generally well-governed countries, public officials collude in ways that enable them to extract resources without penalty. Police refuse to rat on other police and judges ignore official malfeasance.

I go further and initially assume that law-breakers, who are caught, make take-it-or-leave-it offers to policemen, who then accept those offers if and only if they are greater than this maximum punishment level $\overline{P}$. The ability to punish police therefore essentially creates an upper bound on the ability to punish criminals as well. No matter how high the penalties for crime are officially, the maximum punishment will in practice be no more than the minimum bribe level that will be accepted be a law enforcer: $\overline{P}$. This limitation can also be interpreted as the maximum ability of individuals to pay, based on their level of wealth. I will drop the assumption of limited bribe levels later.

These assumptions suggest that city size and institutional quality come together to produce an "institutional possibility frontier," that determines the possible behavior that can be effectively deterred. If $\overline{P}d(N) < B$, then the combination of low probability of arrest, caused by large city size, and weak institutions mean that no misbehavior can be deterred. If $B < \overline{P}d(N) < A + B$, then misbehavior by those with weak incentives to misbehavior can be deterred but not the misbehavior of those with strong incentives. In that case, it is sensible to set the penalty at no more than $B/d(N)$, since higher penalties lead to higher social costs with no offsetting benefits. If $\overline{P}d(N) > A + B$, then it is possible to deter all misbehavior, and the optimal penalty is no more than $(A + B)/d(N)$.

Figure 6 shows the two institutional possibility frontiers implied by those two equations. The top curve is an example of $d^{-1}\left(B/\overline{P}\right)$ and the bottom curve illustrates $d^{-1}\left((A+B)/\overline{P}\right)$. These curves illustrate the highest level of city size that permits penalties to be higher enough to deter crime. Higher $\overline{P}$ values reduce the probability of arrest, and therefore require a higher ex post penalty, as such there is a cutoff value of city size, that rises with the level of institutional development, that determines whether law and order can conceivably be enforced. When city sizes rise beyond the level implied by $d^{-1}\left((A+B)/\overline{P}\right)$ then criminals with strong incentives cannot be deterred. When city size rises beyond the level implied by $d^{-1}\left(B/\overline{P}\right)$ then no crime can be deterred.

Even if crime can potentially be deterred, it is not obvious that prevention is better than anarchy, as illustrated in Proposition 5. To avoid the possibility that governments would prefer bribing policemen to standard punishment, I assume that the social cost of a bribe equals $\lambda\overline{P} + \delta(Bribe - \overline{P})$, which is always strictly greater than the social cost of a punishment equal to the bribe. The logic of this assumption is that the expected punishment to the policemen carries the same social cost as other forms of punishment (i.e. $\lambda\overline{P}$) and that there may be some social costs to bribes that go higher than $\overline{P}$, perhaps because policemen's wages don't adjust downward fully in response to the expected bribes. The $\delta(Bribe - \overline{P})$ will be irrelevant here, when the accused make take-it-or-leave-it offers, but will matter when I later allow bribes to be determined more flexibly. I will assume that $\lambda \geq \delta$, based on the idea that if punishment is a fine that it is presumably equally costly socially to the losses from the net surplus created by a bribe to a policeman, but if punishment represents prison time, then the social costs will be higher.

Proposition 5: There exists two threshold levels of city population, denoted N_{LP}^{A} and N_{FP}^{A}, with $N_{LP}^{A} > N_{FP}^{A}$, and for cities that have less people than N_{FP}^{A}, anarchy is always cost-minimizing. For cities with population levels between N_{LP}^{A} and N_{FP}^{A}, then full punishment is possible and cost minimizing is $\overline{P} > \frac{A+B}{d(N)}$ and anarchy is the cost-minimizing option that is available. For cities with populations above than N_{LP}^{A}, then if $\overline{P} < \frac{B}{d(N)}$, anarchy is the only option. If $\frac{A+B}{d(N)} > \overline{P} > \frac{B}{d(N)}$, then low penalty punishment is the lowest cost available option and if $\frac{A+B}{d(N)} < \overline{P}$, full penalty punishment is available and cost-minimizing. 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. For levels of population below the level horizontal line, which illustrates N_{FP}^{A}, then 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. As a result, these cities descend into anarchy despite the significant benefits that law could bring. For cities of intermediate size, 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.

Finally, for the largest cities there are three regions depending on city size. At low levels of institutional quality, anarchy is the only option. For somewhat higher levels of institutional quality, low level punishment is possible and this reduces costs relative to anarchy. At high levels of institutional quality, full punishment is possible and cost-reducing.

The graph suggests that larger cities may see some form of externality-preventing law before smaller cities, but it may be imperfect. However, it also is possible that as city sizes grow, holding institutional quality constant, cities cross the institutional possibility frontier and law no longer becomes impossible. This is one interpretation of the breakdown in law and order that many cities experience as they grow.

The proposition and figure suggests that higher levels of institutional quality are uniformly related to more law and order, 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.

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

To allow more draconian policies that carry larger social costs, I now extend the model and allow the government to act against the harmless action (walking the streets for example) that carries private benefits and no social costs. I refer to this approach as ex ante prevention and 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 (prevention). Traffic congestion in classical Rome was reduced by Julius Caesar's ban on driving in 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.

People who take the harmless action do not know if they will also have the opportunity to commit a crime and as such, their net expected benefits of engaging in the harmless action equal b plus the net private benefits to committing the crime. In a first-best world, the state would leave the harmless activity alone, and only penalize crime. But it is possible to also ban the harmless action and if that ban is effective then it will also deter all other crime. We assume that the probability of catching someone engaged in the harmless action is $d_0(N) > d(N)$ and the maximum penalty is again $\overline{P}$, whether or not the person is also committing a crime. The probability $d_0(N)$ captures the probability of arrest, whether or not the individual engages in the harmful act and I assume that $lim_{N\to\infty} d_0(N) = 0$ and $lim_{N\to\infty} \frac{d_0(N)}{d(N)} = 1$. The cost of creating these policies again equals K_P.

Individuals will engage in the harmless act, and the harmful act if given the opportunity, if $h + vB + v\pi A > d_0(N)P$, as such the harmless act can be deterred if and only if $\frac{b+vB+v\pi A}{d_0(N)} < \overline{P}$. I assume that $\frac{b+vB+v\pi A}{d_0(N)} < \frac{B}{d(N)}$, 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 Proposition 6:

Proposition 6: If $h < \pi\left(\frac{K_P+v\lambda B}{1-\pi} - vA\right)$, then if city size if greater than a threshold, N_{PR}^A, then it is threshold that is increasing with K_P.A, B. h and π and decreasing with v. If city population is greater than this threshold, then it is cost-minimizing to bar ex ante prevention if $\frac{b+vB+v\pi A}{d_0(N)} < \overline{P} < \frac{A+B}{d(N)}$, and to engage in full ex post punishment if $\overline{P} > \frac{A+B}{d(N)}$. If city population is less than

N_{PR}^{A}, then it is optimal to engage in full ex post punishment if $\overline{P} > \frac{A+B}{d(N)}$ and to accept anarchy otherwise.

If $h > \pi\left(\frac{K_P+v\lambda B}{1-\pi} - vA\right)$, then if population is greater than N_{PR}^{A} it is cost-minimizing to engage in ex ante prevention if $\frac{b+vB+v\pi A}{d_0(N)} < \overline{P} < \frac{B}{d(N)}$, to engage in ex post punishment with low penalties if $\frac{B}{d(N)} < \overline{P} < \frac{A+B}{d(N)}$, and to engage in full ex post punishment if $\overline{P} > \frac{A+B}{d(N)}$. If the population is less than N_{PR}^{A} but greater than N_{LP}^{A}, then it is never optimal to engage in ex ante prevention, but it is optimal to engage in low penalty ex post punishment if $\frac{B}{d(N)} < \overline{P} < \frac{A+B}{d(N)}$ and full ex post punishment if $\overline{P} > \frac{A+B}{d(N)}$.

This proposal suggests that draconian rather than light policies can be the cost-minimizing strategy when institutional quality is 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 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."

The police can be seen as targeting the "harmless action," merely walking around the city. Anyone who appeared a threat 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 a necessary means of addressing 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.

Flexible Bribery

We now drop the assumption that people make take-it-or-leave it offers to policemen but rather that there is Nash bargaining between the criminal and the policemen. If the official penalty for a crime is P, and the policemen suffers $\overline{P}$ in expected value from taken a bribe, then the bargained bribe will equal $\sigma\overline{P}+(1-\sigma)P$, if $P>\overline{P}$. If $P<\overline{P}$, then there will be no bribery and the arrested individual will pay the official penalty.

In the absence of any other issues, it is then possible to always achieve the first best. If $\sigma\overline{P}+(1-\sigma)P=A+B$, then all crime will be deterred and neither punishment nor bribery will occur in equilibrium. Meeting this threshold only requires that the official penalty is sufficiently high. However, 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 $\sigma\overline{P}+(1-\sigma)\psi P$ as long as $\psi P>\overline{P}$. As before, I assume that bribery leads to a social loss of λ times the bribe. For simplicity, I also assume here that $\pi=1$, and $\nu C'(N)-d'_0(N)\left(\frac{\delta\psi B}{d(N)}+(\lambda-\delta(1-\sigma(1-\psi)))\overline{P}\right)+\frac{\delta\psi B d_0(N)d\prime(N)}{(d(N))^2}$ which essentially limits the slope of d(N) relative to $d_0(N)$ and C(N).

Proposition 7: Prevention is always possible and this reduces 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. An ex post punishment strategy that deters the harmful action, but does not lead to extortion, is possible if and only if $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)}<\overline{P}$, and this carries lower costs than ex ante prevention, and carries lower cost than anarchy as long as N is greater than a threshold that is rising with K_P and B and falling with v. If $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)} > \overline{P}$, and $\overline{P} > \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right)$, then ex post punishment with the minimal effective punishment will also deter the harmless action because of extortion, but if $\overline{P} < \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right)$ then extortion will not deter the harmless action, will generate costs lower than prevention if and only if $\overline{P} < \frac{1}{\lambda-\delta(1-\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\delta\psi B}{d(N)}\right)$ and will generate costs lower than anarchy if N is greater than a threshold level that is increasing K_P, B, λ, ψ, σ and $\overline{P}$.

The presence of more flexible bribery means that it is possible to deter anything. If the government sets a high enough punishment, then even that punishment serves to increase the size of the bribe that can be extracted, and that larger bribe can essentially deter any action. As such, it is now possible for even the weakest government to deter anything, even though the punishment will never be used.

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

As before, there exists an institutional possibilities frontier defined by $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)} = \overline{P}$, which determines whether it is possible to have ex post punishment with no extortion of the innocent. If $\overline{P}$ is sufficiently high, then ex post punishment without extortion is possible and always preferable to ex ante prevention and is preferable to anarchy if city size is sufficiently large.

If $\overline{P}$ is lower than this threshold, then ex post punishment will always be accompanied by extortion. If the extortion is sufficiently high, which occurs when $\overline{P} > \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right)$, then it will deter the harmless action as well as the harmful action. In this case, ex post punishment essentially becomes ex ante prevention. The threat of being held up by the police will deter individuals from venturing forth. Notably, this threshold is met when $\overline{P}$ is higher, as long as $\overline{P}$ is less than $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)}$, because higher values of $\overline{P}$ increase the bribes that have to be paid to avoid extortion. If $\overline{P} < \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right)$, then extortion is not large enough to deter the harmless action and it is always true that ex post punishment dominates ex ante prevention.

If ex post punishment is compatible with individuals undertaking the harmless action, then the per capita costs of extortion equal are increasing with λ, ψ, B, σ and $\overline{P}$, 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 N is uncertain, because higher values of N make it less likely that any individual will be caught and extorted by the police, but 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.

As long as $\frac{h}{d_0(N)} > \frac{\delta\psi B}{d(N)}$, which is necessary but not sufficient for extortion not to deter the harmless action, then with low enough levels of $\overline{P}$, ex post punishment with extortion, is preferable to ex ante prevention because the social costs of bribery are sufficiently low. So at very low levels of institutional quality, ex post punishment is preferable if it is possible. It is, however, also possible, as long as $\overline{P} < \frac{1}{\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\psi B}{d(N)}\right)$ then extortion will not deter the harmless action, will generate costs lower than prevention if and only if $\overline{P} < \frac{1}{\lambda-\delta(1-\sigma(1-\psi)}\left(\frac{h}{d_0(N)} - \frac{\delta\psi B}{d(N)}\right)$. If $\left(1 - \frac{(1-\delta)\sigma(1-\psi)}{\lambda-\delta}\right)\frac{h}{d_0(N)} > \frac{\psi B}{d(N)}$, then there will exist a region where ex post punishment with extortion is possible, but more costly than ex ante prevention, but if that condition does not hold, then ex ante prevention will always be more costly than prevention.

Figure 7 illustrates the linear case where $d_0(N) = \frac{d_0^1}{N}$ and $d(N) = \frac{d^1}{N}$, 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 vertical 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 $\overline{P}$ 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 $\overline{P}$, 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 $\overline{P}$, prevention dominates, and for still higher levels of $\overline{P}$, 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 punishment, but it must be expected that this will lead to extortion and bribery. At intermediate levels of development, ex ante prevention is either preferable or 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.

V. 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 at which urbanites, by importing water from other less dense areas.

In this section, I consider three aspects of infrastructure provision in the developing world. In the first section, 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. In this section, I simply assume that infrastructure is provided at a cost and do not model the challenges of actually delivering infrastructure. In the second section, I turn to the problems that poor governance creates for the provision of infrastructure. In the third section, I discuss issues of land use and housing, but do not present a formal model.

Infrastructure and Externalities

I now turn to urban infrastructure, and its relationship with externalities. If the total quantity of infrastructure provided is denoted "I", then if the proportion of the population that take the harmful action is denoted $\hat{v}$, then the per capita created by infrastructure equals $b\left(\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}\right)$, where $\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}$ represents the effective infrastructure and b(I) is increasing, and weakly concave and equal to zero of $I(1-\alpha\hat{v}N)$ =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, by polluting the water or congesting the street. 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 γ >1, and rising with city size if γ <1. This connects with the previous model, if b(.) is a linear function $b_0\frac{I(1-\alpha\hat{v}N)}{N^{\gamma}}$ and then $b_0\alpha N^{1-\gamma}$ =C(N).

If the cost of providing I units of infrastructure is just $K_I I$, 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 optimal 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, it is necessary to 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 that there is no extortion and no ex ante prevention, and those who are accused of a harmful action make take-it-or-leave-it offers to policemen, which ensures that the maximum penalty is $\overline{P}$. This implies that $\hat{v}$ either equals v, if there is anarchy, or 0, if there is full prevention, and full prevention is only possible if $\frac{\overline{P}}{d(N)} > B$. 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 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 $1 > x\frac{-b"(x)}{b\prime(x)}$ for all x or if $1 < \gamma$ and $1 < x\frac{-b"(x)}{b\prime(x)}$ for all x, then there will exist a unique population threshold, which determines whether anarchy carries lower costs than punishment. If $1 > x\frac{-b"(x)}{b\prime(x)}$, then infrastructure is higher under punishment than under anarchy and increases in K_I cause the punishment threshold to rise, but if $1 < x\frac{-b"(x)}{b\prime(x)}$, then infrastructure is lower under punishment than under anarchy and increases in K_I cause the punishment threshold to rise. As before, it is only possible to have full prevention when $\frac{B}{d(N)} > \overline{P}$.

The proposition suggests that the connection between infrastructure and rule-of-law depends on the shape of the function b(.). If b(.) is extremely concave, then anarchy leads to more, not less, infrastructure. This means in turn that 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(.) is less concave, for example $b(x) = x^{\sigma}$, with $1 > \sigma > 0$, then rule of law is a complement with infrastructure, meaning that infrastructure is higher if the society has a sufficient punishment to prevent harmful behavior and higher costs of infrastructure increase the minimal threshold needed for rule-of-law to reduce costs.

The comparative statics connection per capita infrastructure and population depend also on the shape of the function b and the value of γ. For example if $1 > x\frac{-b''(x)}{b'(x)}$, then per capita infrastructure is rising with N almost everywhere if $\frac{1-\gamma}{2-\gamma} > \alpha v N > 0$ and decreasing with N almost everywhere if $\frac{1-\gamma}{2-\gamma} < 0 < \alpha v N$. If $\frac{1-\gamma}{2-\gamma} > 0$ but for some value of N, less than the threshold at which rule of law becomes desirable $\alpha v N > \frac{1-\gamma}{2-\gamma}$, then infrastructure is first rising and then falling with city population. If $1 > x\frac{-b''(x)}{b'(x)}$, then there is always a discontinuous increase in infrastructure when rule of law is adopted. Similarly, if $\overline{P}$ rises, making rule of law possible, there will be a similar increase in the level of optimal infrastructure.

If $1 < x\frac{-b''(x)}{b'(x)}$ 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 $x\frac{-b''(x)}{b'(x)}$ 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 more 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, $x\frac{-b''(x)}{b'(x)}$ must be less than one.

Public vs. 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. To switch focus, I assume 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 just 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 kickback 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 will 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 there is a tradeoff between the two. Providing the infrastructure on less land requires more labor, specifically if A units of land are acquired than L(A) units of effective labor are needed. The value of L(0) is finite.

The opportunity cost of labor is denoted 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 $\overline{P}$. If the benefit of shirking is equal to the opportunity cost of time w times a multiplier$z < 1$, then they will maximize $zws - \phi(s)\overline{P}$, which leads to an optimal level of shirking for the public workers determined by $\delta w = \phi'(s^*)\overline{P}$. The labor cost of the project is therefore $\frac{wL(A)}{1-s^*}$, because to get one effective unit of labor, the public must pay for $\frac{1}{1-s^*}$ units of labor. The total cost of the project, therefore, holding land use constant is strictly decreasing in the value of $\overline{P}$.

The total social costs of the project however is $\frac{wL(A)(1-zs^*)}{1-s^*} + \frac{\phi(s^*)\lambda\overline{P}}{1-s^*}$, reflecting both the loss of value of time from shirking and the costs of punishment.

One natural 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 project then chooses the amount of land to minimize financial costs $\frac{wL(A)}{1-s^*} + P_L(N)A$, which leads to a first order condition of $-\frac{wL'(A)}{1-s^*} = P_L(N)$. I am assuming that the public manager is choosing land to minimize total costs to the public sector not 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 $\underline{P_L}$, so there is a potential gain of $P_L(N) - \underline{P_L}$ 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 $\overline{P}$ and has a probability of being caught of d per unit of land. As such, bribery will occur when $d\left(P_L(N) - \underline{P_L}\right) > \overline{P}$. 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 $\underline{P_L} + d\overline{P}$, 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 $d\left(P_L(N) - \underline{P_L}\right) < \overline{P}$, then private provision generates the first best, and if $d\left(P_L(N) - \underline{P_L}\right) > \overline{P}$, 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.

Land Ownership and Housing

In this last section, I will briefly discuss the issues of land tenure and housing provision in developing world cities. Governments impact land markets in two major ways. They establish property rights, which are preconditions for development, and they limit development through new regulation, presumably to address externalities. One public action protects private owners from private expropriation, which is called disorder in the framework of Djankow et al. (2003), and the other exposes owners to public takings, which is called dictatorship in the same framework.

In the west, the history progressed from property protection to regulation. The early history of English Common Law is replete with institutional interventions meant to settle the matter of who owned what and to facilitate private control over private property. For example, Henry II's extensive use of juries in the 12^{th} century served in part to establish long-standing property rights over land that may have been taken during the chaotic reign of King Stephen. While there were regular attempts by the sovereign to expropriate private property, particularly during the reign of Richard II, the general trend was towards more firmly established private ownership. The privatization of the commons provides a particularly infamous example of this process.

The rise of state property regulation was, by contrast, a more modern affair. Development was reasonably un-regulated at the beginning of the 19th century in the U.S. and U.K., although there had been limited earlier regulations related to war-making (crenellated walls were occasionally forbidden in medieval England) and fire (Boston forbid thatched roofs in 1631). Localities became more assiduous in applying rules to property in the 19th century (Novak, 1996) and then land use regulation became far more heavy handed during the 20th century. New York City's far-reaching zoning ordinance of 1916 is a particularly salient example.

The logic of this progression is relatively easy to understand. During the earlier period, the public sector understood the widespread social losses associated with combat over land rights. Establishing clear property ownership was a means of keeping the peace, and owners were typically willing to pay for their privileges. As city sizes increased, the potential downsides from negative externalities increased. Fire was the first major externality, but later building restrictions worried about public health and even the amount of light that would be blocked by a skyscraper.

All of this is clearly understandable within the scope of the model. As institutional quality increases, it is sensible to take actions that reduce the externalities from conflict. As city sizes increased and externalities became more costly, it made sense to increase the number of building structures that would be regulated.

But the pattern in developing world mega-cities reverses this historic trend. In a large number of developing world cities, private property rights are at best imperfectly protected. By contrast, public regulations tend to be stringent. For example, the World Bank has catalogued the number of official procedures needed to build a warehouse in different parts of the world. Some of the most extreme cases occur in the world's poorest metropolitan areas. Yet in these areas, squatters can often occupy private land with impunity, at least from public retribution. This combination makes land development particularly difficult, because of the double threat of public regulation and private expropriation. This may help explain why purchasing power adjusted Class A office rents in Mumbai are among the highest in the world (Gomez-Ibanez and Masood, 2006).

Yet this phenomenon becomes more understandable in light of enforcement costs in weak institutional environments. Regulations that limit large structures are cheap to enforce. The structures are large and obvious, and even if there is only a small probability that a building will get torn down due to a zoning violation, this may be enough to deter construction.

By contrast, pushing large numbers of poor people off undeveloped land is far more difficult. It can be difficult to enforce and lead to embarrassing journalism. In an environment where institutions are weak, the costs of limiting construction seem likely to be far less than the costs of enforcing rights to private property.

Yet while this claim helps us to understand why property rights aren't protected against squatters, it does not explain why developing world mega-cities restrict new development so intensively. There are three plausible explanations: externality-mitigation, historical accident and rent extraction. The externality-mitigation explanation is that these regulations are fundamentally benevolent, because of the externalities that high buildings create. This view is plausible, but seems unlikely given that newer, taller, more densely situated buildings have the capacity to mitigate disaster risk and reduce traffic congestion.

The historic accident view postulates that these rules were borrowed from the west, where they were relatively benevolent. Indian land use planners, for example, were heavily influenced by the British Garden City model. Little attention was paid to the different conditions in Indian cities, and as a result, inappropriate institutions may have been imported. Those institutions remain because they are relatively easy to enforce.

The third hypothesis is that land use regulations persist because they provide political power and rents. A land use regulation is an invitation to bribery and there have been many scandals in which developers have paid their way to more lenient rules. Future research is needed to assess which of these theories is more likely to be correct.

VI. Conclusion

In this paper, I have argued that the rise of poor world mega-cities is relatively surprising based on the history of the world. Typically, wealth, or at least political competence, preceded the development of large urban agglomeration. Yet over fifty years, the world has seen 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 was needed to feed urbanites. The cities of the developing world today can feed themselves based on mineral exports or aid from abroad. 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 examine the consequences when negative urban externalities collide with weak institutions. An institutional possibilities frontier came about based on the combination of city population and institutional quality. Larger cities are more likely to need interventions to alleviate externalities, but urban anonymity also makes it less likely that a given level of institutions will produce good outcomes. When I allowed for both ex ante prevention and ex post punishment, I found the more dictatorial approach of ex ante restrictions might be more effective than ex post punishment when institutions are weak and city sizes are large.

I also discussed infrastructure, which alleviates 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 do more to increase the stock of effective infrastructure, thereby alleviating the need to spend more, or do more to increase the marginal gains from building more infrastructure. I also discussed private and public provision, and suggested that if the costs of public provision are shirking by public employees and the cost of private provision is subversion of the land acquisition process, then public provision would dominate when city sizes are sufficiently large and private provision of infrastructure would dominate with smaller cities.

Given all of 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 might suggest that such limitations would be beneficial, the dynamics effects of cities 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 seems to suggest that urbanization is one tool for institutional development. The connections and social movements that form readily in the dense confines of urban areas can ultimately be strong enough to change and discipline government. The rural past in the poorer world has shown little evidence of supporting institutional development. The cities of the developed world have been wellsprings for the growth of politically-minded groups. Indeed, it is possible that an urban contribution to the growth of institutional quality may be the most important urban benefit.

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

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

Welfare in the city equals $\theta_0 e^{-\delta u N}\left((1-\epsilon)(\epsilon N_U)^{\frac{1-\sigma}{\sigma}} A_M P_M - \underline{C}\right) P_M^{\alpha-1}$ or $\frac{\theta_0 e^{-\delta u N}}{\alpha}\left((1-\alpha)A_A g(\tau,\overline{d})\frac{(1-u)^{1-\gamma}}{uN^{\gamma}} - \underline{C}\right) P_M^{\alpha-1}$ while welfare in the countryside equals $A_A\left((1-u)N\psi\right)^{-\gamma}\left(2-2e^{-\psi\overline{d}}\right)^{\gamma} P_M^{\alpha-1}$. The value of $(1-\alpha)A_A g(\tau,\overline{d})\frac{(1-u)^{1-\gamma}}{uN^{\gamma}} - \underline{C}$ goes to infinity as u goes to zero, and is negative at sufficiently high levels of u. Moreover as long as $(1-\alpha)A_A g(\tau,\overline{d})\frac{(1-u)^{1-\gamma}}{uN^{\gamma}} > \underline{C}$, then $\frac{\theta_0 e^{-\delta u N}}{\alpha}\left((1-\alpha)A_A g(\tau,\overline{d})\frac{(1-u)^{1-\gamma}}{uN^{\gamma}} - \underline{C}\right)$ is strictly decreasing in u, while $A_A\left((1-u)N\psi\right)^{-\gamma}\left(2-2e^{-\psi\overline{d}}\right)^{\gamma}$ 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:

(A1) $e^{-\delta u N}\left(\frac{1-u}{u}(1-\alpha)g(\tau,\overline{d}) - (1-u)^{\gamma}\frac{N^{\gamma}\underline{C}}{A_A}\right) = \alpha\left(2-2e^{-\psi\overline{d}}\right)^{\gamma}\frac{\psi^{-\gamma}}{\theta_0}$

The derivative of the left hand side with respect to u is negative whenever the left hand side is positive. Equation (A1) can be rewritten as

(A1) $$h(u(z),z) = \alpha\left(2-2e^{-\psi\overline{d}}\right)^{\gamma}\frac{\psi^{-\gamma}}{\theta_0}$$

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 $\underline{C}$. Urbanization is independent of A_M and ϵ.

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

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

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 $\frac{1-u}{u}\frac{2^{\gamma}\gamma^{\gamma-1}(1-\alpha)^3}{\alpha\overline{d}(1-\alpha+\alpha\gamma)}\frac{e^{-\frac{(1-\alpha+\alpha\gamma)\tau}{\gamma}\overline{d}}}{\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right)^2}\left(\frac{\alpha\gamma}{1-\alpha}\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right) -\right.$

$e^{\alpha\tau\bar{d}} + 1\Big)$ minus $\left(1 - \left(1 + \frac{(1-\alpha)\bar{d}\tau}{\gamma}\right) e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}\right) \frac{\gamma\tau^{\gamma-1}}{\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}\right)^{\gamma+1}} (1-\alpha)^{\gamma} \frac{\left((1-u)N/q\right)^{\gamma}\underline{C}}{A_A\alpha}$. 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

$$\text{(A2)} \qquad \left((1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} u^{\frac{1-\sigma}{\sigma}} A_M P_M - \underline{C}\right)(1-u)^{\gamma} e^{-\delta u N} = \frac{N^{-\gamma}A_A}{\theta_0}\left(\frac{2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}$$

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 $(1-u)^{\gamma} e^{-\delta u N}\left[\left((1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} u^{\frac{1-\sigma}{\sigma}} A_M P_M\right)\left(\frac{1-\sigma}{\sigma u} - \frac{\gamma}{1-u} - \delta N\right) - \left(\frac{\gamma}{1-u} + \delta N\right)\underline{C}\right]$

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

$(1-u)^{\gamma} e^{-\delta u N}\left((1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} u^{\frac{1-\sigma}{\sigma}} A_M P_M - \underline{C}\right) > \frac{N^{-\gamma}A_A}{\theta_0}\left(\frac{2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}$ 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 $(1-u)^{\gamma} e^{-\delta u N} N^{\gamma}\left((1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} u^{\frac{1-\sigma}{\sigma}} A_M P_M - \underline{C}\right)$, is greater than $\frac{A_A}{\theta_0}\left(\frac{2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}$. If $A_A = 0$, this will occur, as long as $(1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}} A_A P_M > \underline{C}$, but as A_A goes to infinity, the condition will never hold. As the right 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

$$(1-\hat{u})^{\gamma}e^{-\delta\hat{u}N}N^{\gamma}\left((1-\epsilon)(\epsilon N)^{\frac{1-\sigma}{\sigma}}\hat{u}^{\frac{1-\sigma}{\sigma}}A_M P_M-\underline{C}\right)=\frac{A_A^*}{\theta_0}\left(\frac{2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right)}{(1-\alpha)\tau}\right)^{\gamma}$$, 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 right hand side to rise. As such, A_A^* is rising with A_M, P_M θ_0, and τ and falling with A_A, δ , $\overline{d}$, and $\underline{C}$. A_A^* is rising will be rising with ϵ as long as $1-\sigma>\epsilon$, and N is and only if $\gamma>\delta\hat{u}N$.

Proof of Proposition 3:

At any spatial equilibrium (A2) 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, $\overline{d}$, δ and $\underline{C}$. Urbanization will be rising with ϵ as long as $1-\sigma>\epsilon$ and N if and only if $\gamma>\delta\hat{u}N$.

Proof of Proposition 4:

Welfare in the city equals $\theta_0 e^{-\delta N_U^*}(1-\epsilon)(\epsilon N_U^*)^{\frac{1-\sigma}{\sigma}}A_M P_M^{\alpha}$, where N_U^* reflects the equilibrium level of population discussed in Proposition 2, which must satisfy

(A2') $$(N_U^*)^{\frac{1-\sigma}{\sigma}}(N-N_U^*)^{\gamma}e^{-\delta N_U^*}=\frac{A_A\left(2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right)\right)^{\gamma}}{\theta_0(1-\epsilon)(\epsilon)^{\frac{1-\sigma}{\sigma}}A_M P_M((1-\alpha)\tau)^{\gamma}}$$

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

The derivative of urban welfare with respect to city size is $\left(\frac{1-\sigma}{\sigma N_U}-\delta\right)(1-\epsilon)(\epsilon N_U)^{\frac{1-\sigma}{\sigma}}A_M\theta_0 e^{-\delta N_U}P_M^{\alpha}$, 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 $\frac{1-\sigma}{\sigma\delta}$ which maximizes well-being in the city.

If $N_U^*<\frac{1-\sigma}{\sigma\delta}$, which requires that $\left(\frac{1-\sigma}{\sigma\delta}\right)^{\frac{1-\sigma}{\sigma}}(N-\frac{1-\sigma}{\sigma\delta})^{\gamma}e^{-\frac{1-\sigma}{\sigma}}<\frac{A_A\left(2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right)\right)^{\gamma}}{\theta_0(1-\epsilon)(\epsilon)^{\frac{1-\sigma}{\sigma}}A_M P_M((1-\alpha)\tau)^{\gamma}}$, then the second city must be as large as the first city to attract urban pioneers.

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

$\frac{1-\sigma}{\sigma\delta})^{\gamma} e^{-\frac{1-\sigma}{\sigma}} > \frac{A_A\left(2\gamma\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}\right)\right)^{\gamma}}{\theta_0(1-\epsilon)(\epsilon)^{\frac{1-\sigma}{\sigma}} A_M P_M((1-\alpha)\tau)^{\gamma}}$then the minimum size for the new city must be a fraction of the size of the old city, φ,that satisfies $-Ln\varphi/(1-\varphi) = \frac{\sigma\delta}{1-\sigma}$, so this fraction is decreasing in σ and δ .

Part (b): In the closed economy, when $\underline{C} = 0$, then the welfare of urban residents in any second location equals $e^{-\delta(N_{U2}-N_{U1})}\left(\frac{N_{U2}}{N_{U1}}\right)^{\frac{1-\sigma}{\sigma}}\left(\frac{P_{M2}}{P_{M1}}\right)^{\alpha} V_1$, 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 ration of the price of the manufactured goods in city 1 to city 2 or $\frac{P_{M1}}{P_{M2}}$ will equal $\left(\frac{N_{U2}}{N_{U1}}\right)^{\frac{1}{\sigma}}\left(\frac{1-e^{-\frac{(1-\alpha+\alpha\gamma)}{\gamma}(1-\Delta)\tau\bar{d}}}{1-e^{-\frac{(1-\alpha+\alpha\gamma)}{\gamma}\Delta\tau\bar{d}}}\right)$, 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 $\left(\frac{N_{U2}}{N_{U1}}\right)^{\frac{1}{\sigma}}\left(\frac{1-e^{-\frac{(1-\alpha+\alpha\gamma)}{2\gamma}(\tau\bar{d}-k)}}{1-e^{-\frac{(1-\alpha+\alpha\gamma)}{2\gamma}(\tau\bar{d}+k)}}\right) = e^{-k}$, because the left hand size 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 $e^{-\delta(N_{U2}-N_{U1})}\left(\frac{N_{U2}}{N_{U1}}\right)^{\frac{1-(1+\alpha)\sigma}{\sigma}}\theta V_1$. 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 $1 < (1+\alpha)\sigma$

Part (c) In either open or closed economies, the condition for spatial equilibrium in the two city case is

$$\text{(A2')} \quad (N_U^*)^{\frac{1-\sigma}{\sigma}} e^{-\delta N_U^*}\theta_0(1-\epsilon)(\epsilon)^{\frac{1-\sigma}{\sigma}} A_M P_M = A_A\left(\frac{2\gamma}{(1-\alpha)\tau(.5N-N_U^*)}\left(1-e^{-\frac{(1-\alpha)\tau}{2\gamma}\bar{d}}\right)\right)^{\gamma}$$

The difference between the open and closed economy is that the value of P_M is endogenous in the closed city case.

The curve $\frac{1-e^{-\frac{(1-\alpha)\tau}{2\gamma}\bar{d}}}{.5N-N_U^*}$ is strictly greater than $\frac{1-e^{-\frac{(1-\alpha)\tau}{2\gamma}\bar{d}}}{.5N-.5N_U^*}$which is weakly greater than $\frac{1-e^{-\frac{(1-\alpha)\tau}{\gamma}\bar{d}}}{N-N_U^*}$, and rural welfare is convex in urban population. As such, in the open economy, it is strictly less likely that an urban equilibrium exists where there are two cities, and if an equilibrium exists, the city size, at the larger equilibrium level, city sizes will be smaller and at the lower equilibrium level, city sizes will be larger. Welfare must be strictly higher at the lower equilibrium level with two cities than with one city, because a larger city sizes, guarantees a higher rural welfare level which must also be equal to the urban welfare level. A sufficient condition for the new welfare level to be higher than the old welfare level is that the new city size is greater than N_U^{Max}, while a sufficient condition for the new welfare level to be less than the old welfare level is that the old city size is less than N_U^{Max}. If $\delta = 0$, then all cities sizes are less than N_U^{Max} and welfare must be higher than with one city than with two.

In the closed economy, welfare is higher for two cities than for one city if $2^{\alpha\gamma - \frac{(1-\alpha)(1-\sigma)}{\sigma}}$ is greater than

$$\frac{\frac{\theta_0(1-\alpha)}{1-\alpha+\alpha\gamma}\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right)^{-\alpha(1-\gamma)}\left(1-e^{-\frac{(1-\alpha+\alpha\gamma)\tau}{\gamma}\overline{d}}\right)^{\alpha}+\left(1-e^{-\frac{(1-\alpha)\tau}{\gamma}\overline{d}}\right)^{1-\alpha(1-\gamma)}\left(1-e^{-\frac{(1-\alpha+\alpha\gamma)\tau}{\gamma}\overline{d}}\right)^{\alpha-1}}{\frac{\theta_0(1-\alpha)}{1-\alpha+\alpha\gamma}\left(1-e^{-\frac{(1-\alpha)\tau}{2\gamma}\overline{d}}\right)^{-\alpha(1-\gamma)}\left(1-e^{-\frac{(1-\alpha+\alpha\gamma)\tau}{2\gamma}\overline{d}}\right)^{\alpha}+\left(1-e^{-\frac{(1-\alpha)\tau}{2\gamma}\overline{d}}\right)^{1-\alpha(1-\gamma)}\left(1-e^{-\frac{(1-\alpha+\alpha\gamma)\tau}{2\gamma}\overline{d}}\right)^{\alpha-1}}$$

For sufficiently large values of $\tau\overline{d}$, this must always be true if $\sigma(1+\alpha\gamma)+\alpha(1-\sigma)>1$

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 $K_P + vB\big(1-\pi(1-\lambda)\big) < v(1-\pi)C(N)$. 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 $K_P + vB\big(1-\pi(1-\lambda)\big) = v(1-\pi)C(N_{LP}^A)$. 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 as 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 $v(1-\pi)C(N) - vB\big(1-\pi(1-\lambda)\big) < K_P$, 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 $\overline{P} = \frac{B}{d(N)}$ (which is increasing with B and N) at which limited punishment is possible and desirable and a second higher value of $\overline{P} = \frac{A+B}{d(N)}$ (which is increasing with A, B and N), at which full punishment is possible add desirable. For all values of N less than N_{LP}^A, but greater than N_{FP}^A there exists a value of $\overline{P} = \frac{A+B}{d(N)}$, 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 punishment are always less 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 $K_P + b + v(B + \pi A) = vC(N_{PR}^A)$ as N_{PR}^A. The value of N_{PR}^A is increasing with K_P.A, B. b and π and decreasing with v. The value of N_{PR}^A is greater than N_{LP}^A if and only if $b > \pi\left(\frac{K_P+v\lambda B}{1-\pi} - vA\right)$. 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_{PR}^A, then it is cost –minimizing to impose ex ante prevention when $\frac{h+vB+v\pi A}{d_0(N)} < \overline{P}$ and then engage in ex post punishment with full penalties when $\frac{A+B}{d(N)} < \overline{P}$. If b is high, then for cities with population levels above N_{LP}^A,it is cost-minimizing to engage in ex ante prevention when $\frac{b+vB+v\pi A}{d_0(N)} < \overline{P} < \frac{B}{d(N)}$, and then ex post punishment with low penalties when $\frac{B}{d(N)} < \overline{P} < \frac{A+B}{d(N)}$ and then ex post punishment with high penalties when $\overline{P} > \frac{A+B}{d(N)}$.

Proof of Proposition 7: If the policy is to prevent ex ante prevention, then there are no false accusations, since individuals need to be engaging in the 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 $+vB < d_0(N)(\sigma\overline{P} + (1-\sigma)P\ )$ and there are no costs from false arrests. As such a value of P equal to $\frac{h+vB}{(1-\sigma)d_0(N)} - \frac{\sigma\overline{P}}{1-\sigma}$ 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.

Alternatively, full ex post punishment can be accomplished if $\leq d(N)(\sigma\overline{P} + (1-\sigma)P\ )$, which is always possible. If $\overline{P} > B/d(N)$, then a second best is achievable with a fixed punishment that is less than $\overline{P}$, leading to no bribery and no extortion at a flat cost of NK_P.

If $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)} < \overline{P}$, then a second best is similarly achievable, by setting $P = \frac{B}{(1-\sigma)d(N)} - \frac{\sigma\overline{P}}{1-\sigma}$, which implies that $\left(\sigma\overline{P} + (1-\sigma)\psi P\right) < \overline{P}$, as long as $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)} < \overline{P}$. Ex post punishment with full enforcement dominates anarchy if and only if $K_P < v(C(N) - B)$. As such, there exists a value of N, denoted N_{FP}^{A2} at which anarchy and full punishment yield equivalent costs, and for higher levels of N full punishment yields lower costs, and for lower levels of N, anarchy yields lower costs. The value of N_{FP}^{A2} is rising with K_P and B and falling with v. There also exists a higher value of population N_{PR}^{A2} at which anarchy and prevention yield equal costs and prevention generations lower costs if and only if N is greater than N_{PR}^{A2}. The value of N_{PR}^{A2} is rising with K_P, b and B and falling with v.

The curve $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)} = \overline{P}$ is a similar institutional possibility frontier, that determines a minimal level of $\overline{P}$ needed to enforce the law without corruption. This defines a minimum value of $\overline{P}$, that is rising in N, B and ψ and falling with σ at which full prevention is possible without extortion.

If $\frac{\psi B}{(1-\sigma+\sigma\psi)d(N)} > \overline{P}$, then it is impossible to eliminate crime without generating extortion. A fraction $d_0(N)$ of individuals who are engaged in the harmless action will also be extorted and will have to pay $\frac{\psi B}{d(N)} + \sigma(1-\psi)\overline{P}$. If $d_0(N)\left(\frac{\psi B}{d(N)} + \sigma(1-\psi)\overline{P}\right) > b$, then the extortion will also eliminate the harmless action, and as such, ex post penalties are equivalent to ex ante prevention. As such, ex ante prevention and ex post punishment become equivalent over some values and both carry costs of $N(K_P + h)$.

If $d_0(N)\left(\frac{\psi B}{d(N)} + \sigma(1-\psi)\overline{P}\right) < b$, then ex post punishment carries social costs $N\left(K_P + \lambda d_0(N)\overline{P} + \delta d_0(N)\left(\frac{\psi B}{d(N)} - (1-\sigma(1-\psi))\overline{P}\right)\right)$, which is less than ex ante prevention as $(\lambda - \delta(1-\sigma(1-\psi)))\overline{P} + \frac{\delta\psi B}{d(N)} < \frac{b}{d_0(N)}$. Ex post punishment with extortion will reduce costs relative to anarchy if and only if $v(C(N) - B) - K_P - d_0(N)\left(\frac{\delta\psi B}{d(N)} + (\lambda - \delta(1-\sigma(1-\psi)))\overline{P}\right) > 0$. As long as $vC'(N) - d'_0(N)\left(\frac{\delta\psi B}{d(N)} + (\lambda - \delta(1-\sigma(1-\psi)))\overline{P}\right) + \frac{\delta\psi B d_0(N) d'(N)}{(d(N))^2} > 0$, 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, greater than N_{FP}^{A2} but less than N_{PR}^{A2} at which ex post punishment with extortion provides equal costs to anarchy. This value is increasing with K_P, B, λ, ψ, σ and $\overline{P}$.

Proof of Proposition 8:

I ignore the welfare related to the harmless action as it is unchanged under the scenarios. 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 $\frac{(1-\alpha v N)}{N^{\gamma-1}} b'(N^{1-\gamma} i_{AN}^*(1-\alpha v N)) = K_I$. 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 $1 > x\frac{-b''(x)}{b'(x)}$ for all x, then $i_{PR}^* > i_{AN}^*$ and if $1 < x\frac{-b''(x)}{b'(x)}$ 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 previous 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 $b(N^{1-\gamma} i_{PR}^*) - K_P - K_I i_{PR}^* - b(N^{1-\gamma} i_{AN}^*(1-\alpha v N)) - vB + K_I i_{AN}^*$ with respect to N is $(1-\gamma)N^{-\gamma}\left(i_{PR}^* b'(N^{1-\gamma} i_{PR}^*) - i_{AN}^*(1-\alpha v N) b'(N^{1-\gamma} i_{AN}^*(1-\alpha v N))\right) + \alpha v N^{1-\gamma} i_{AN}^* b'(N^{1-\gamma} i_{AN}^*(1-\alpha v N))$ or $\frac{K_I}{N(1-\alpha v N)}$ times $(1-\gamma)(i_{PR}^* - i_{AN}^*)(1-\alpha v N) + \alpha v N i_{AN}^*$, 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 $1 > x\frac{-b"(x)}{b\prime(x)}$ for all x, then $i_{PR}^* > i_{AN}^*$ and $1 > \gamma$, and there exists a unique crossing point. If $1 < x\frac{-b"(x)}{b\prime(x)}$ 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 $1 > x\frac{-b"(x)}{b\prime(x)}$ and never hold if $1 < x\frac{-b"(x)}{b\prime(x)}$.

Proof of Proposition 9:

If $d\left(P_L(N) - \underline{P_L}\right) < \overline{P}$, 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 $P_L(N) - \underline{P_L} > d\overline{P}$, then the total financial costs of the project equal $wL(A_{SUB}) + \left(\underline{P_L} + d\overline{P}\right)A_{SUB}$, 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 $\frac{wL(A_{PUB})(1-zs^*)}{1-s^*} + P_L(N)A_{PUB}$, where $-\frac{wL\prime(A_{PUB})}{1-s^*} = P_L(N)$. If $\overline{P}$ 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 $\underline{P_L} + d\overline{P} < (1 - s^*)P_L(N)$. As long as his condition holds, then if $\lambda = 0$, then private provision dominate 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 $\frac{wL(0)(1-\delta s^*)}{1-s^*}$, 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 $P'_L(N)\left(\frac{L\prime(A_{PUB})(1-\delta)s^*}{L"(A_{PUB})} + A_{PUB}\right)$, 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." *The Quarterly Journal of Economics* 110(1): 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)(2): 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 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*, Chicago: University of Chicago Press: 153-184.

De Vries, Jan. 1984. *European Urbanization 1500-1800*. Cambridge, MA: 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(4): 595-619.

Durand, John D., and Cesar A. Peláez. 1965. "Patterns of Urbanization in Latin America." *Milbank Memorial Fund Quarterly* 43(4): Suppl: 166-196.

Engel, Eduardo, Ronald Fischer, and Alexander Galetovic. 2013. "The Basic Public Finance of Public–Private Partnerships." *Journal of the European Economic Association* 11(1): 83-111.

Ethier, Wilfred. 1979. "Internationally decreasing costs and world trade." *Journal of International Economics* (9)(1): 1-24.

Friedlander, Dov. 1970. "The Spread of Urbanization in England and Wales, 1851-1951." *Population Studies* November: 423-443.

Friedman, David. 1999. "Why Not Hang Them All: The Virtues of Inefficient Punishment." *Journal of Political Economy* (107)(6): S259-S269.

Glaeser, Edward L. and Joshua D. Gottlieb. 2008. "The Economics of Place-Making Policies." *Brookings Papers on Economic Activity* 39(1): 155-253.

Glaeser, Edward L., Giacomo A. M. Ponzetto, and Kristina Tobio. 2010. "Cities, Skills, and Regional Change." Economics Working Papers 1255, Department of Economics and Business, Universitat Pompeu Fabra, revised Mar 2011.

Glaeser, Edward L. and Bruce Sacerdote. 1999. "Why Is There More Crime in Cities?" *Journal of Political Economy* 107(S6): S225-S258.

Glaeser, Edward L. and Andrei Shleifer. 2003. "The Rise of the Regulatory State." *Journal of Economic Literature* 41(2): 401-425.

Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath. 2013. "Urbanization with and without Industrialization." Submitted.

Gomez-Ibanez, Jose and Sheikh Zahid Masood. 2006. "Vision Mumbai." Harvard Kennedy School Case Study 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(4): 1127-1161.

Krugman, Paul. 1991. *Geography and Trade*. Cambridge, MA: MIT Press.

Krugman, Paul and Raul Livas. 1996. "Trade Policy and the Third World Metropolis." *Journal of Development Economics* 49(1): 137-150.

Matsuyama, Kiminori. 1992. "Agricultural productivity, comparative advantage, and economic growth." *Journal of economic theory* (58)(2): 317-334.

Novak, William. 1996. *The People's Welfare: Law and Regulation in Nineteenth Century America.* Charlotte: University of North Carolina Press.

Sabine, E. L. 1937. "City cleaning in medieval London." *Speculum* 12: 19–43.

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

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%	$236	$410
Zimbabwe	Harare 1.5 million	39%	12%	$775	missing
Mali	Bamako 1.9 million	35%	12%	$700	$1,283
Haiti	Port-au-Prince 2.1 million	53%	21%	$743	$1,227
Pakistan	Karachi 13.5 million	36%	8%	$1,217	$2,859
Senegal	Dakar 2.9 million	43%	23%	$1,146	$1,956
Cote D'Ivoire	Abidjan 4.15 million	51%	21%	$1,223	$1,931

Sources:
World Bank data and United Nations (*World Urbanization Prospects: The 2011 Revision*)

Table 2:
World's Largest Cities Through History

World's Largest Cities Throughout History	Year Reached 1 million (World Maximum until 1800)	Approximate per Capita Income (2012 $)	Notes
Roman Empire	1	$1,400 - Italy $1,000 - Roman Empire	Fed with grain from Spain and Egypt, provided because of imperial power.
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 non-political mega-city

Sources: 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

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)
Observations	119	119	139	139
R-squared	0.189	0.531	0.085	0.304

Notes:
Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

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. Data comes from the World Bank. Standard errors are in parentheses. A constant is included in the regression but not reported.

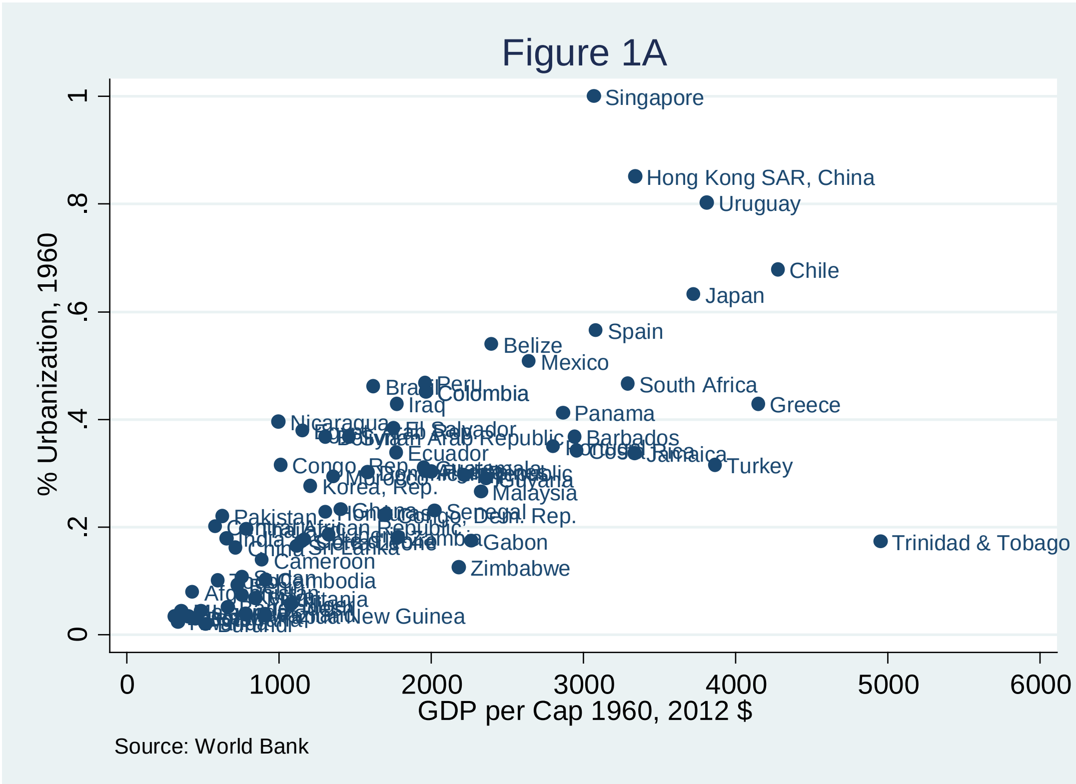
Figure 1A
% Urbanization, 1960
GDP per Cap 1960, 2012 $
0
.2
.4
.6
.8
1
0
1000
2000
3000
4000
5000
6000
Singapore
Hong Kong SAR, China
Uruguay
Chile
Japan
Spain
Belize
Mexico
South Africa
Greece
Colombia
Iraq
Panama
Nicaragua
Barbados
Turkey
Ecuador
Korea, Rep.
Malaysia
Pakistan
Senegal
Gabon
Trinidad & Tobago
Cameroon
Zimbabwe
Papua New Guinea
Burundi
Source: World Bank

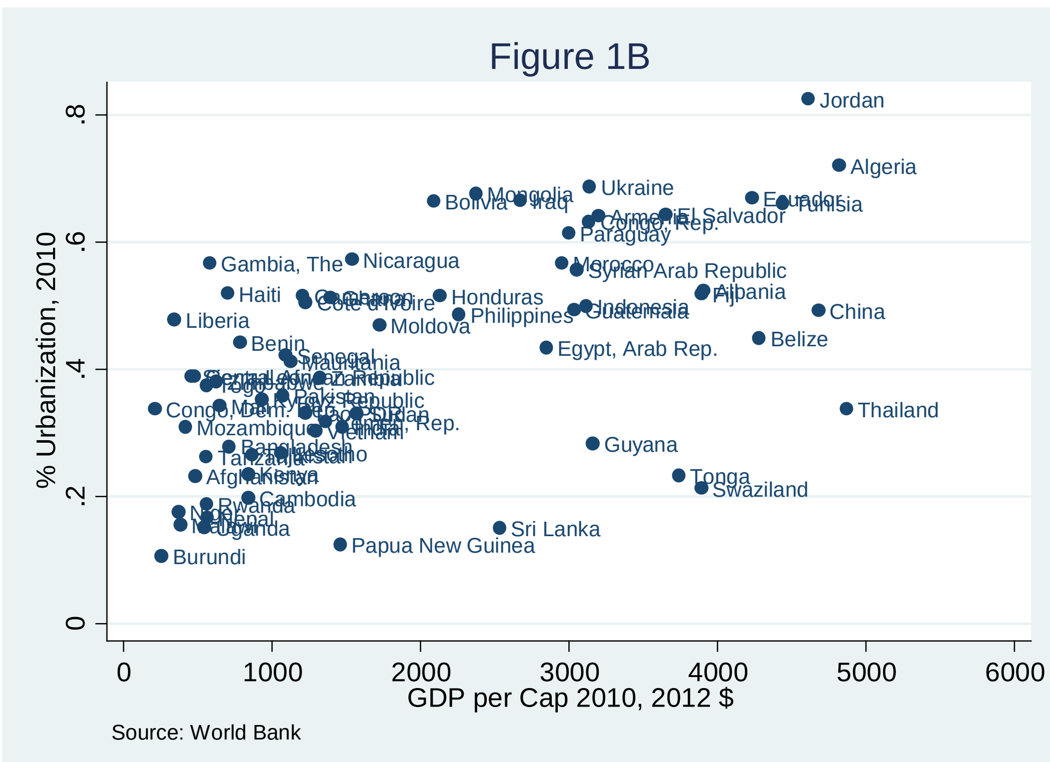
Figure 1B
% Urbanization, 2010
GDP per Cap 2010, 2012 $
0
.2
.4
.6
.8
0
1000
2000
3000
4000
5000
6000
Jordan
Algeria
Ukraine
Mongolia
Bolivia
Iraq
Tunisia
El Salvador
Congo, Rep.
Paraguay
Gambia, The
Nicaragua
Morocco
Syrian Arab Republic
Albania
Fiji
Haiti
Honduras
Cote d'Ivoire
Philippines
Indonesia
Guatemala
China
Liberia
Moldova
Belize
Benin
Egypt, Arab Rep.
Mauritania
Thailand
Pakistan
Mozambique
Vietnam
Guyana
Tanzania
Afghanistan
Kenya
Tonga
Swaziland
Cambodia
Rwanda
Nepal
Uganda
Sri Lanka
Papua New Guinea
Burundi
Source: World Bank

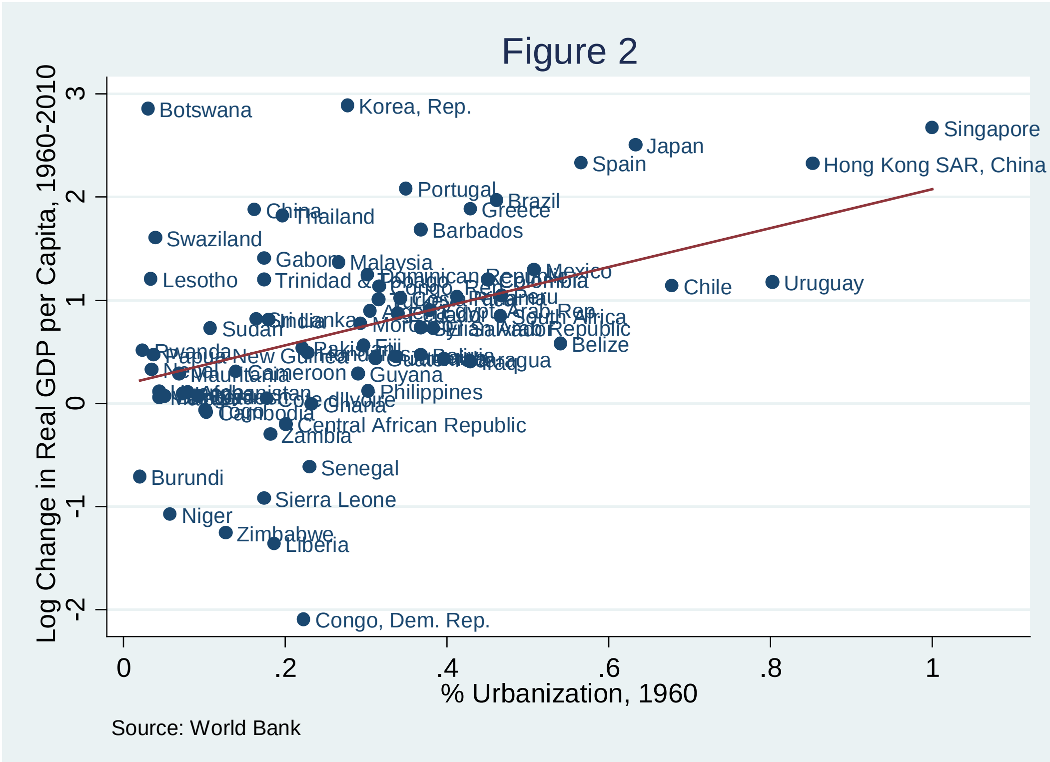
Figure 2
Log Change in Real GDP per Capita, 1960-2010
% Urbanization, 1960
-2
-1
0
1
2
3
0
.2
.4
.6
.8
1
Botswana
Korea, Rep.
Singapore
Japan
Spain
Hong Kong SAR, China
Portugal
Brazil
Greece
Thailand
Barbados
Swaziland
Malaysia
Mexico
Lesotho
Chile
Uruguay
Sudan
Belize
Cameroon
Guyana
Philippines
Central African Republic
Zambia
Senegal
Burundi
Sierra Leone
Niger
Liberia
Congo, Dem. Rep.
Source: World Bank

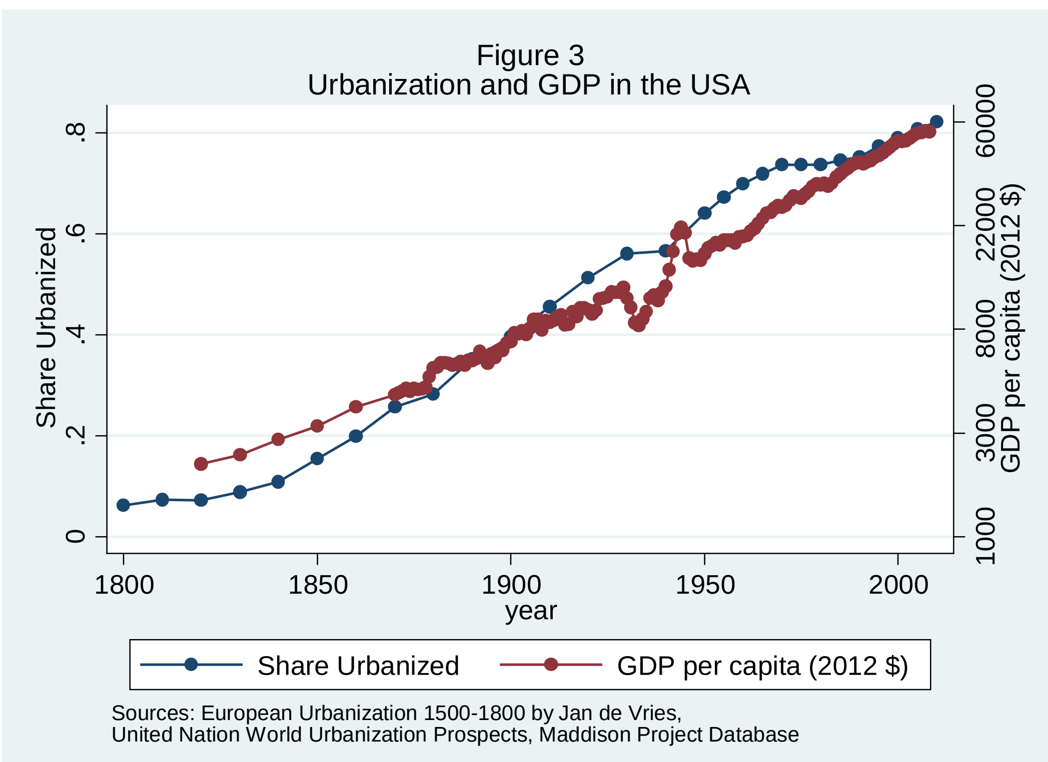
Figure 3
Urbanization and GDP in the USA
Share Urbanized
GDP per capita (2012 $)
year
0
.2
.4
.6
.8
1000
3000
8000
22000
60000
1800
1850
1900
1950
2000
Share Urbanized
GDP per capita (2012 $)
Sources: European Urbanization 1500-1800 by Jan de Vries,
United Nation World Urbanization Prospects, Maddison Project Database

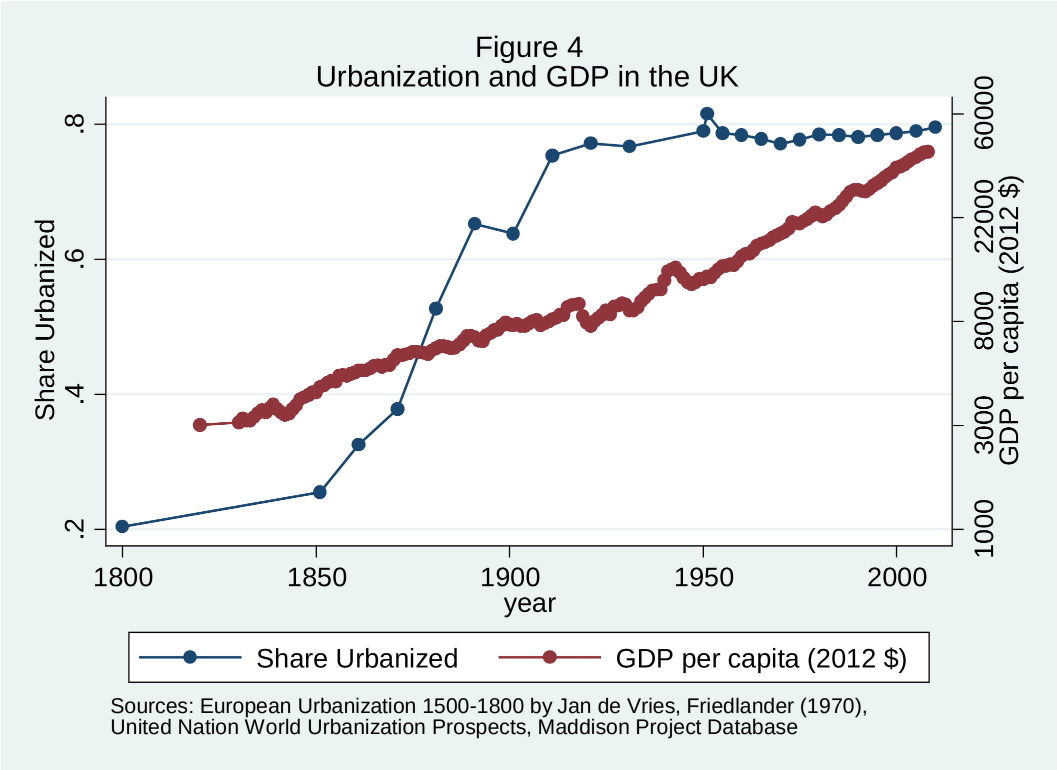
Figure 4
Urbanization and GDP in the UK
Share Urbanized
.2
.4
.6
.8
GDP per capita (2012 $)
1000
3000
8000
22000
60000
1800
1850
1900
1950
2000
year
Share Urbanized
GDP per capita (2012 $)
Sources: European Urbanization 1500-1800 by Jan de Vries, Friedlander (1970),
United Nation World Urbanization Prospects, Maddison Project Database

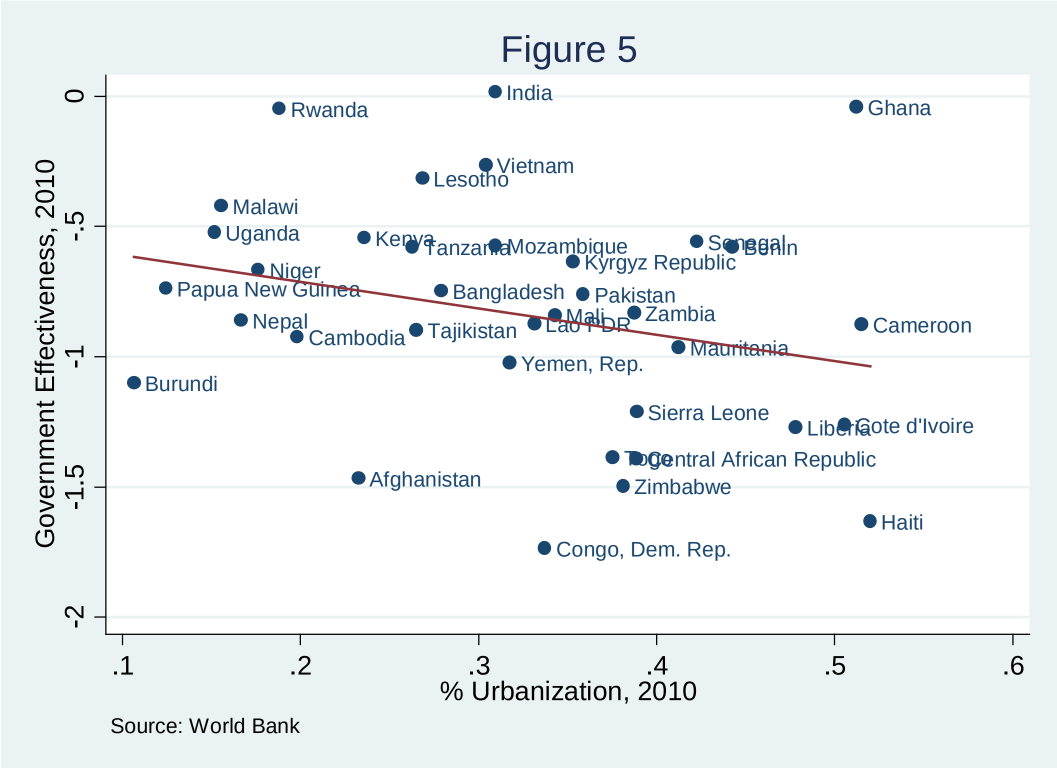
Figure 5
Government Effectiveness, 2010
0
-.5
-1
-1.5
-2
% Urbanization, 2010
.1
.2
.3
.4
.5
.6
India
Rwanda
Ghana
Vietnam
Lesotho
Malawi
Uganda
Kenya
Tanzania
Mozambique
Senegal
Benin
Kyrgyz Republic
Niger
Papua New Guinea
Bangladesh
Pakistan
Nepal
Cambodia
Tajikistan
Lao PDR
Mali
Zambia
Cameroon
Mauritania
Yemen, Rep.
Burundi
Sierra Leone
Liberia
Cote d'Ivoire
Togo
Central African Republic
Afghanistan
Zimbabwe
Haiti
Congo, Dem. Rep.
Source: World Bank

Figure 6: Institutional Quality and City Size

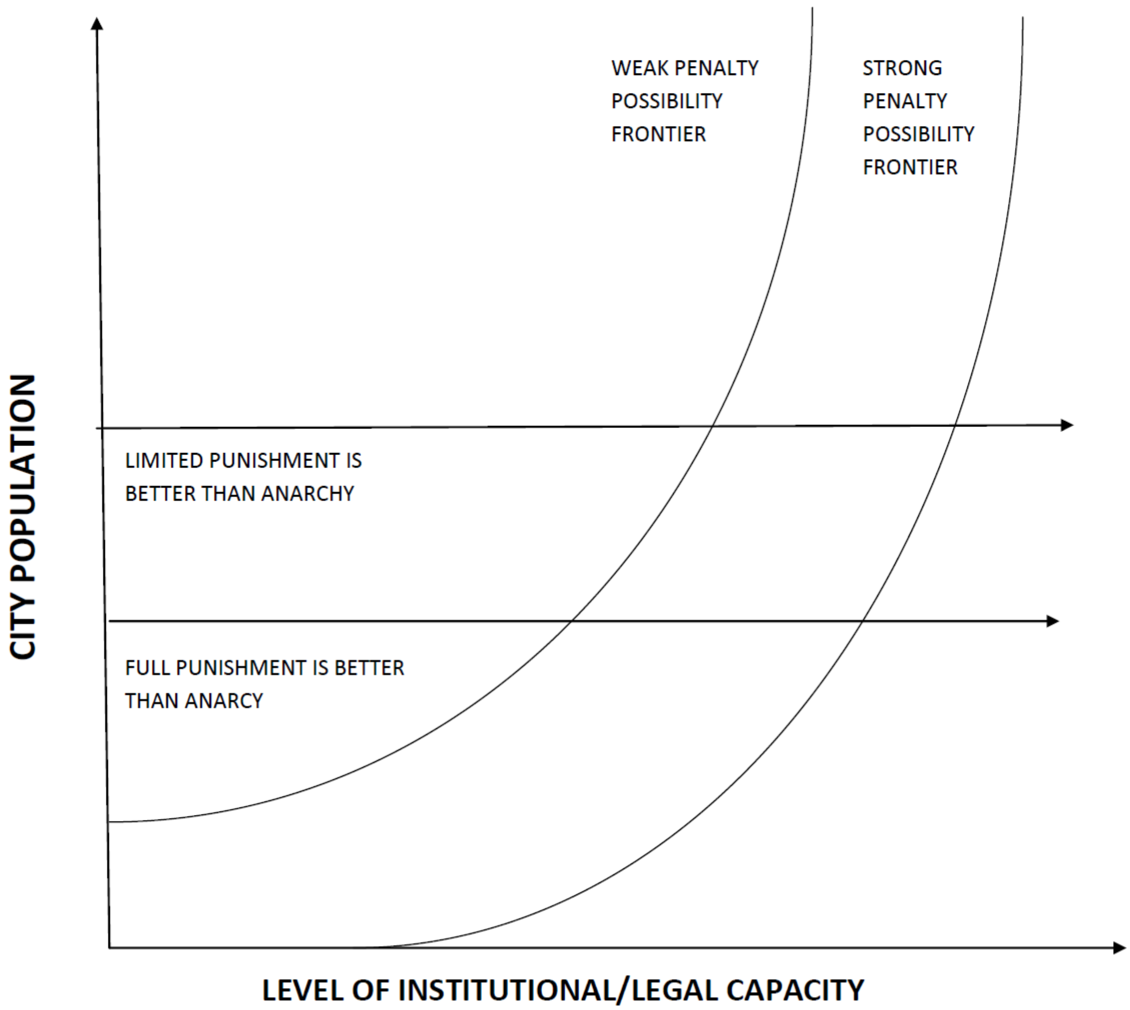

Figure 7: Prevention, Punishment and City Size