

A threshold model of urban development*

Alberto Vesperoni[§]

Paul Schweinzer^b

Abstract

We propose a simple model of distribution of economic activity across cities of endogenous size and number determined by individual incentives. The individuals populating our model are endowed with idiosyncratic entrepreneurial creativity, the realization of which requires urban agglomeration linked to a crowding cost. Focusing on the dynamics of urban development, our predictions include a U-shaped relation between well-known measures of urbanization and urban primacy, a hypothesis that we test empirically using World Bank data. Our findings complement a growing consensus on U-shaped relations between level and concentration of economic activity across a broad set of applications.

Keywords: *Agglomeration, Urbanization, Development.*

Journal of Economic Literature Classification: *C7, D71, O18, Q56.*

1 Introduction

Galor (2009) and Ray (2010), among others, illustrate mechanisms through which not only the level, but the distribution of economic prosperity greatly matters for institutional stability and economic growth of nations in the long-run. As argued in Glaeser and Henderson (2017), the increasingly uneven urban development of emerging economies such as China, India, and Nigeria appears as one of the crucial challenges of our times. The reasons are plenty and connected in a complex web of causalities, including the direct effects of urbanization and urban concentration on poverty alleviation, access to basic services, employment possibilities, socio-political tensions, pollution and the environment (see, e.g., Ravallion, 2015; Sun et al., 2016). Strikingly, the mounting rage felt by the impoverished provinces towards the so-called cosmopolitan elites of the capitals of the world is a phenomenon very much at the center of the rise of populism in recent times (see, e.g., Eatwell and Goodwin, 2018).

*We would like to thank İrem Bozbay, Carlo Fiorio, Frédéric Robert-Nicoud, Miguel Portela, and Yoichi Sugita for valuable comments. Schweinzer gratefully acknowledges the support of the Fulbright Scholar Program and the hospitality of the University of Minnesota, Twin Cities. [§]Department of Political Economy, King's College London, United Kingdom, alberto.vesperoni@kcl.ac.uk. ^bDepartment of Economics, Alpen-Adria-Universität Klagenfurt, Austria, paul.schweinzer@aau.at. (12-Jan-2023)

Our work is chiefly motivated by the need to understand dynamics of urban development which, to the best of our knowledge, are so far unexplored.¹ Taking a stylized approach in the tradition of coalition formation and threshold models of social interaction,² we model the agglomeration of a population into cities of endogenous number and size and study how their relative size changes when a larger fraction of the population moves from rural to urban areas. Our equilibrium predictions include a U-shaped relation between the level of urbanization (i.e., the fraction of the population living in urban areas) and urban primacy (i.e., the fraction of the urban population living in the largest city), a hypothesis that we test empirically using World Bank data. The economic, environmental, and socio-political implications of such an U-shaped trend are extensive. Assuming the bottom of the U has been reached sometimes in the late 20th century, we should now expect the urban population—and thus economic activity, wealth, and power—to increasingly concentrate in the capital city rather than the provinces. While this may feel true for many industrialized countries, the bottom of the U may not have been reached yet in many developing countries which may thus expect the opposite trend of empowerment of provinces.

To illustrate the relevance of our U-shaped hypothesis in a historical example, consider the long run effects of the industrial revolution on the level of urbanization and urban primacy in the United Kingdom through the last two-three centuries. Roughly speaking, before the industrial revolution, a large share of non-agricultural activity was concentrated in London and focused on services related to trade. When the industrial revolution took off, economic activity started diversifying across sectors and geographically spreading north towards growing industrial clusters such as Birmingham (automotive), Manchester (textile), and Newcastle (shipbuilding and steel).³ By the mid 20th century, these peripheral centers reached levels of economic prosperity never witnessed before, but their economic growth reached an apex sometime in the mid 1970s and never came back. With their economic decline becoming evident and urbanization still on the rise, London reacquired its uncontested centrality in the last decades, in line with the general pattern of the “renaissance of the metropolis” across the developed world (Glaeser, 2011). To summarize, as time progresses, we observe a steadily increasing trend in the level of urbanization and a U-shaped trend in the

¹ There is a vast literature that studies the spatial and agglomerative dimensions of urban distributions. See Fujita et al. (2001), Redding (2013), Duranton and Kerr (2018) for reviews of various subfields. This literature can be split into neoclassical general equilibrium models which include the urban economics approach of the system of cities (e.g., Picard and Tabuchi, 2013; Behrens et al., 2014; Davis and Dingel, 2020) and more stylized approaches in the tradition of regional science (e.g., de Palma et al., 2019; Albouy et al., 2019). Our contribution falls into the second category.

² Our theoretical framework can be seen as a coalition formation game with non-transferable utility (Peleg and Sudhölter, 2007). Specifically, it is a many-to-many matching game with hedonic preferences defined over an individual’s coalition size relative to an individual-specific threshold (see, e.g., Bogomolnaia and Jackson, 2002; Aziz et al., 2016). Seminal contributions to the threshold approach to coalition formation include Simon (1954), Schelling (1969) and Granovetter (1978); we refer to Watts and Dodds (2009) and Benhabib et al. (2010) for reviews of related threshold-based approaches to coalition formation and network formation across the social sciences.

³ For a thorough account of the geography of the industrial revolution in the United Kingdom, see Dodgshon and Butlin (2013). A historical dataset on urbanization trends in the United Kingdom is provided by Friedlander (1970).

level of urban primacy, with London remaining the largest city through the whole time span.

While our (anecdotal, and later statistical) evidence is only suggestive, it is remarkable that related fields in the literature found similarly U-shaped correspondences between the degree of concentration and the level of mobilization of resources akin to economic development. One group includes, among others, Imbs and Wacziarg (2003) for GDP per capita and concentration of economic activity across industrial sectors, and related papers on GDP per capita and sectoral concentration of exports.⁴ Another cluster revolves around inequality of income (or wealth) and GDP per capita as documented for instance by Piketty and Saez (2003) and Saez and Zucman (2016), which points directly against the well-known Kuznets hypothesis. Our paper provides a theory for this U-shaped relation in the context of urban development. While we do not claim one-to-one portability across fields, there are obvious interconnections between the distributions of people across space, those of people and resources across industrial sectors, and those of resources across people.

While other predictions of our model are broadly in line with stylized facts of urban economics,⁵ our U-shaped hypothesis directly contradicts the line of inquiry that, in reminiscence of the Kuznets curve, postulates an inverted U-shaped relation between urban primacy and the level of urbanization.⁶ Henderson (2003) provides an extensive review of this literature arguing that, although the inverted U-shape may still be present in the 1985-1995 decade, it is much noisier than in the 1965-1975 decade and it may be fading away in recent times (see, e.g., his Figure 4). We believe this fading effect is partly related to the aforementioned “renaissance of the metropolis” and the tumultuous urban transformation of certain emerging economies. Hence, in our view, there is scope for further debate on the empirical relation between urban primacy and urbanization, particularly in light of the opposite projections on future trends in urban concentration and the drastically different policy responses they may require.

Let us describe our framework in more detail. Our theoretical model considers a continuum of agents scattered on a territory, where the set of agents inhabiting a location is called a *city* if it has positive mass while it is called a *village* (or a solitary settlement) otherwise. For simplicity we assume that all locations are equally distant from each other, thus abstracting from the spatial dimension and exclusively focusing on the distribution of population across locations. With some narrative license, our agents can be interpreted as entrepreneurs with different business plans that are heterogeneous in

⁴ See Cadot et al. (2011) and references therein.

⁵ As we will see, our model complies with two established facts on the correlates of the size of a city: its positive relation with the mean and dispersion of the well-being of the city’s residents, and its negative relation with the number of cities of comparable size.

⁶ This research tradition originates with the influential cross-country analysis in Williamson (1965) and subsequent contributions that illustrate the empirical pattern with data spanning from the 1950s to the mid 1990s. In light of this evidence, Junius (1997) develops a general equilibrium framework where centripetal and centrifugal forces interact and provide a rationale for the inverted U-curve.

their degree of *ambition*.⁷ The core intuition is that while ambitious plans can lead to higher profits, they are more difficult to launch requiring more supportive stakeholders at early stages of implementation. Reflecting various frictions, these initial supporters are typically local and thus larger cities are more likely to provide the critical mass necessary to realize more ambitious plans.⁸ We thus assume that the ambition of an agent is a threshold (or *type*) such that her business plan is operative if and only if she inhabits a city of size larger than or equal to this type. Acknowledging that larger cities also typically lead to higher crowding costs (e.g., higher rents, congestion, pollution, etc), we then model the agent’s preferences such that her crowding cost is minimized conditional on her business plan being launched.

We define an urban distribution as a partition of the set of agents into cities and villages, and we call it an *equilibrium* if no agent prefers to leave her location. As the presence of agents in a city constitutes the very incentive for more agents to locate there, this naturally leads to the multiplicity of equilibria and potential coordination failures that are typical of the development discourse.⁹ We characterize the broad set of equilibria showing that the equilibrium distribution is always determined by an algorithm that lends itself to intuitive visualization in a simple diagram. Specifically, in every equilibrium, agents are sorted such that cities correspond to different intervals of types while villages are inhabited by the lowest and the highest (but non-utilized) types.¹⁰ Under mild restrictions, this implies that a larger city size positively affects the dispersion of the residents’ utility but not necessarily the mean, in line with the empirical observations in Eeckhout et al. (2014) and Gaubert (2018).¹¹

Further analysis shows that the distribution of agents that maximizes utilitarian welfare is necessarily an equilibrium, and this equilibrium must be *cost-efficient* in the sense that it minimizes the aggregate crowding cost for given profits of each agent. This insight delivers a one-to-one mapping between the levels of *urbanization* (i.e., the fraction of agents living in cities instead of villages) and the set of cost-efficient equilibria. A crucial feature of cost-efficient equilibria is the presence of an infinite number of arbitrarily small cities and a limited number of bigger cities of heterogeneous size,

⁷ See, e.g., Genicot and Ray (2017) for a formalization of the interaction of inherited wealth and aspirations in determining an individual’s ambition.

⁸ As pointed out in Carlino and Kerr (2015), among the three Schumpeterian business stages of invention, innovation, and commercialization, the second is geographically highly concentrated as it concerns the access to financial resources backed by specialized knowledge. For instance, while the software that is behind an internet platform can in principle be written and sold anywhere in the world, it is most likely to lead to an IT startup in Bengaluru. Similarly, Paul Krugman writes that since: “the 1980s America has experienced growing regional divergence. We have become a knowledge economy driven by industries that rely on a highly educated work force, and firms in those industries, it turns out, want to be located in places where there are a lot of highly educated workers already – places like the Bay Area.” (New York Times, 27-Aug-2021)

⁹ For seminal contributions, see e.g., Rosenstein-Rodan (1943) and Hirschman (1958).

¹⁰ Thus, villages and cities host “effective” or realized types. The presence of high but ineffective types in villages can be interpreted as these being “ahead of their time” in the sense that their ideas cannot be realized under the existing agglomeration structure.

¹¹ For a comparison with Eeckhout et al. (2014), the relevant variable in our model is utility (to match wages adjusted by housing prices), while for a comparison with Gaubert (2018) it is realized profits (to match firms’ revenues).

where the number of cities whose size falls within the interval $[s, s + k]$ naturally decreases in s for any k (ignoring any intervals entirely devoid of cities). We feel this delivers a fairly realistic and tractable framework roughly in line with the empirical evidence on the frequency of city sizes in relation to Gibrat's and Zipf's laws.¹²

Focusing on cost-efficient equilibria, we then engage in comparative statics that are relevant for urban development in the short to long run. We consider population replications that increase the mass of agents, and shifts in the distribution of ambition that lead to first-order stochastic dominance and mean-preserving spreads. In the short run,¹³ we determine that increases in the mass of agents systematically reduce *urban primacy* (i.e., the share of the urban population living in the largest city), upward shifts in the distribution of ambition have the opposite effect, while higher inequality in the distribution of ambition always leads to higher (lower) urban primacy if the level of urbanization is sufficiently high (low). By contrast, we find that the long run effects depend on specific assumptions and no general pattern can be discerned, with one crucial exception: we fully characterize how a change in the level of urbanization should affect urban primacy. Under fairly general conditions, this delivers a *U-shaped relation* between urban primacy and the level of urbanization across cost-efficient equilibria for any given distribution of ambition. We view this U-shaped relation to be the principal testable prediction of our paper, and using openly accessible World Bank data across all countries of the world from 1960 to 2016, we provide preliminary evidence in support of this hypothesis.

The paper develops as follows. Section 2 defines the basic model. The core equilibrium and welfare analyses are in Section 3, the comparative statics in Section 4, and the empirics in Section 5. Section 6 concludes. All proofs can be found in the Appendix.

2 Model

Urban distributions

We consider a continuum of agents of mass $a > 0$, denoted by the set $A \subset \mathbb{R}$. These agents are distributed on a territory constituted by an arbitrarily large set of locations. Assuming all locations are identical and abstracting from spatial distances, we define an *urban distribution* as a partition of A into a collection of sets of zero mass (*villages*) and positive mass (*cities*). We denote by \mathcal{D} the set of all urban distributions of agents (i.e., the set of all possible partitions of A). Note that, solely by our definition of city

¹² By Gibrat's law the growth rate of a city is independent of its size and by Zipf's law the size of a city is inversely proportional to its rank, which is consistent with the independence condition on growth (see, e.g., Gabaix, 1999; Eeckhout, 2004). Such inverse proportionality can be obtained within our model under suitable restrictions, and it is consistent with the general property that the frequency of city sizes within the interval $[s, s + k]$ decreases in s for any k .

¹³ To distinguish between short run and long run effects, we assume that the level of urbanization is fixed in the short run while, in the long run, it may or may not adjust to a different level depending on welfare-efficiency and coordination.

as a set of agents of positive mass, any urban distribution has countably many cities; these cities can be ranked in terms of the mass of agents they contain, and there can be multiple cities with equal mass of agents.

Let $D \in \mathcal{D}$ be any urban distribution. For each possible rank $k \in \mathbb{N}$ of a city in terms of mass of agents, we denote by $n_k^D \in \mathbb{N} \cup \{0\}$ the number of cities ranked k and by $m_k^D \in \mathbb{R}_+$ the mass of agents contained in each of them. If the number of cities in the urban distribution D is finite we write $m_k^D = n_k^D = 0$ for all ranks k larger than the rank of the city with the smallest mass of agents. Then, the *structure* of D is summarized by the sequence $\mathcal{S}(D) := (m_k^D, n_k^D)_{k=1}^\infty$.¹⁴

We define the *level of urbanization* of $D \in \mathcal{D}$ as the fraction of agents who are urban,

$$\mathcal{U}(D) := \frac{1}{a} \sum_{k=1}^{\infty} n_k^D m_k^D.$$

We think of the degree of urban concentration as a measure of the inequality of the distribution of the mass of the urban agents across cities. By the principle of transfers (i.e., the defining property of an inequality measure) urban concentration should not increase whenever a positive mass of agents is relocated from a larger city to a smaller city (or to a village that becomes a city), as long as this transfer is small enough so that the receiving city or village does not become larger than the providing city. It seems also desirable that a measure of urban concentration is scale invariant, in the sense that it remains constant whenever the mass of agents in each city is multiplied by the same positive factor (so that the proportions of mass of agents across cities are maintained). A measure of *urban concentration* that satisfies these properties is the generalized Herfindahl-Hirschman Index,

$$\mathcal{K}(D) := \sum_{k=1}^{\infty} n_k^D \phi \left(m_k^D / \sum_{k=1}^{\infty} n_k^D m_k^D \right),$$

where the function $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ satisfies $\phi(0) = 0$ and it is differentiable, increasing and strictly convex. Finally, we define the level of *urban primacy* as the fraction of urban population that inhabits one of the largest cities,

$$\mathcal{P}(D) := m_1^D / \sum_{k=1}^{\infty} n_k^D m_k^D.$$

Urban primacy is a crude but popular measure of urban concentration that is sensitive only to transfers of urban agents that involve the largest cities. As we will see, these three measures of urban development are intimately related to the workings and predictions of our model. Specifically, the measure of urban concentration $\mathcal{K}(D)$ will be crucial for the interpretation of the cost-efficient equilibria our analysis will focus on, while the level of urbanization $\mathcal{U}(D)$ and the level of urban primacy $\mathcal{P}(D)$ will be the core ingredients of our U-shaped prediction.

¹⁴ For example, if D has $n_1^D = 2$ cities with mass of agents $m_1^D = .3$ and $n_2^D = 1$ city with mass of agents $m_2^D = .2$ we write $\mathcal{S}(D) = (.3, 2; .2, 1; 0, 0; \dots)$.

Preferences

We think of the agents in our model as entrepreneurs, each endowed with a different idea or business plan. These business ideas are heterogeneous in their degree of ambition which affects both profits and implementability. More ambitious plans potentially lead to higher profits but require a higher critical mass of initial stakeholders (investors, customers, etc.) to become operative. We assume that, due to various frictions related to distance, these initial stakeholders are necessarily local and that larger cities can provide more (varied) resources. For each agent $i \in A$, we denote by the threshold $t_i \in \mathbb{R}$ the minimum city size that allows her business plan to realize, so that agent i makes profits if and only if she inhabits a city of mass larger than or equal to t_i . We refer to t_i as the *type* of agent $i \in A$, which is the critical mass required to implement her business plan and indicates her level of ambition.¹⁵

Our definition of agents' preferences is schematic but at the same time relatively general. We shall assume that each agent always prefers to make profits to not making profits, and because of increasing crowding costs she will prefer to live in the smallest available city that allows her to make profits. If she is unable to make profits in any available city, she will prefer to live in a village. These statements fully characterize the preferences that we will use in our general analysis, which are lexicographic with 'making profits' as the primary criterion and 'minimizing the crowding cost' as the secondary one.¹⁶ The basic idea is that, while an agent's profits may increase steeply in her degree of ambition, they should be relatively independent of the mass of the city she inhabits (once her business plan is operative) which seems to be a plausible simplification if a business operates on a national or global scale.

We now define the central element of our model, the *distribution of types*. For each possible city mass $m \in [0, a]$, we denote by $F(m)$ the total mass of agents whose types are lower than or equal to m , so that they all can make profits in any city of size m or larger. This cumulative mass function $F : [0, a] \rightarrow [0, a]$ is non-decreasing by construction and we shall assume it is increasing and twice differentiable on the pre-image of $[0, a)$, so that there is a density function $f(m) := dF(m)/dm$ that is positive and differentiable on such a domain. Denoting by \bar{m}_F the smallest $m \in [0, a]$ such that $F(m) = a$, we can then write $f(m) > 0$ if $m < \bar{m}_F$ and $f(m) = 0$ if $m \geq \bar{m}_F$.

Our examples of distributions of types will primarily focus on the case of $a = 1$, making use of well-known distributions from probability theory. A convenient example distribution is the Beta density

$$f(m) = \frac{m^{\alpha-1}(1-m)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx},$$

¹⁵ Here, we implicitly assume that each entrepreneur is associated with a single skill which may or may not realize into a business plan. The desirable generalization to the case of entrepreneurs endowed with multiple skills is considerably more complex and left to future research, requiring the present analysis as a prerequisite step.

¹⁶ Formally, agent $i \in A$ prefers a city (or village) of mass m to a city (or village) of mass m' if and only if one of the following conditions holds: (i) profits with m and no profits with m' ($m \geq t_i > m'$); (ii) profits with none of them and m smaller ($t_i > m' > m$); (iii) profits with both of them and m smaller ($m' > m \geq t_i$).

whose cumulative mass function satisfies $F(0) = 0$ and $F(1) = 1$ for all parameter configurations $\alpha, \beta > 0$. Another convenient distribution is based on the Gumbel density

$$f(m) = \frac{1}{\beta} e^{-(x-\alpha)/\beta} e^{-e^{-(x-\alpha)/\beta}},$$

which substantially differs from the Beta as $F(0) > 0$ and $F(1) < 1$ for all parameter configurations $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}_{++}$. Note that by $F(0) > 0$ there is a positive mass of types (non-positive) that can make profits even in villages, while by $F(1) < 1$ there is a positive mass of types (larger than $a = 1$) that cannot make profits in any contingency. With the aforementioned Beta distribution, instead, by $F(0) = 0$ and $F(1) = 1$ such cases have zero mass.

Welfare

We now present the various welfare criteria that we will employ in our analysis. Let $D, D' \in \mathcal{D}$ be any pair of urban distributions. We say that D Pareto dominates D' if a positive mass of agents prefers D to D' while no positive mass of agents prefers D' to D . While Pareto dominance leads to unquestionable welfare rankings, it typically leaves many pairs of urban distributions unranked. Hence, to sharpen our predictions, we impose some more structure. Let the function $\pi : \mathbb{R} \rightarrow \mathbb{R}_+$ define the potential profits of each agent depending on her type, and let the function $c : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ define the crowding cost of each agent depending on the mass of the city that she inhabits. We shall assume that these functions are twice differentiable and c satisfies $c(0) = 0$, is increasing and weakly convex, and that $\pi(x) > c(x)$ for all $x \in [0, a]$.¹⁷ We can now represent the preferences of each agent $i \in A$ by the utility function

$$u(t_i, m_{r(i)}^D) = \pi(t_i) I(t_i \leq m_{r(i)}^D) - c(m_{r(i)}^D),$$

in which $m_{r(i)}^D$ denotes the mass of the city inhabited by agent i in the urban distribution $D \in \mathcal{D}$ and $I(t_i \leq m_{r(i)}^D)$ is an indicator function that takes value 1 if $t_i \leq m_{r(i)}^D$ and 0 otherwise.¹⁸ Figure 1 is an illustration of these ideas.

We say that an urban distribution $D \in \mathcal{D}$ is *cost-efficient* if, for a given level of urbanization, it is not possible to decrease the aggregate crowding costs

$$C(D) := \sum_{k=1}^{\infty} n_k^D m_k^D c(m_k^D)$$

¹⁷ Let us remark on two points regarding π and c . First, while it seems reasonable that π is non-decreasing and we encourage the reader to follow this interpretation (as more ambitious plans are typically more profitable), we do not need this assumption for our core results to hold. Second, the assumption that profits strictly dominate costs is made only for convenience, in order to rule out situations in which some equilibria are infeasible for exogenous reasons. Our core results would carry over, for instance, to the case of π non-decreasing and weakly concave, as by the weak convexity of c there would be x^* such that the condition $\pi(x) > c(x)$ is satisfied for $x < x^*$ and violated for $x \geq x^*$.

¹⁸ The lexicographic preferences of each agent admit a utility representation because of the restrictions on the domain. This specific formulation of utility is chosen for tractability. In principle, the lexicographic preferences of each agent are compatible with a utility function where π depends on $m_{r(i)}^D$ as long as the derivative $\partial\pi/\partial m_{r(i)}^D$ is sufficiently small.

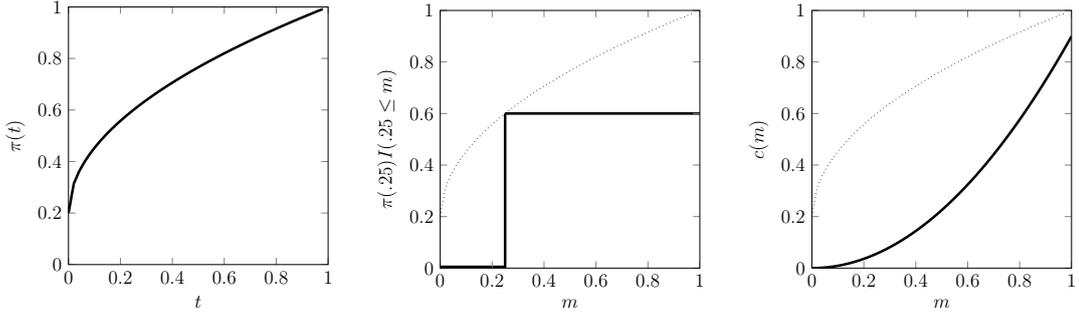


Figure 1: The solid lines in the left, central and right panels, respectively, represent the potential profits $\pi(t) = .2 + .8\sqrt{t}$ of an agent of type $t \in [0, 1]$, the actual profits $\pi(t)I(t \leq m)$ of an agent of type $t = .25$ in a city of size $m \in [0, 1]$, and the crowding cost $c(m) = .9m^2$ of an agent in a city of size $m \in [0, 1]$. Note that these specifications of potential profits, actual profits and crowding cost are consistent with our restrictions on preferences given $a = 1$.

without decreasing the profits of some agent. Note that the constrained minimization of $C(D)$ is equivalent to the minimization of urban concentration in the form of the generalized Herfindahl-Hirschman Index $\mathcal{K}(D)$, as urbanization is held constant in such minimization. Finally, we say that an urban distribution is welfare-efficient if it maximizes utilitarian welfare, which, for each $D \in \mathcal{D}$, is defined by the average utility

$$\begin{aligned} W(D) &:= \frac{1}{a} \int_{i \in A} u(t_i, m_{r(i)}^D) di \\ &= \frac{1}{a} \int_{i \in A} \pi(t_i) I(t_i \leq m_{r(i)}^D) di - \frac{1}{a} C(D). \end{aligned}$$

Note that cost-efficiency is a necessary condition for welfare-efficiency.

3 Equilibrium and welfare analysis

In this section we develop the core theoretical results, characterizing the subset of urban distributions to be used in the comparative statics analysis. Specifically, we start by characterizing the set of equilibria and then proceed by pinning down the subset of equilibria that are cost-efficient, arguing that the welfare-efficient urban distribution is one of them.

We say that an urban distribution $D \in \mathcal{D}$ is an *equilibrium* if no agent prefers to move from her city or village to another existing city or village. The basic idea is that individuals are free to move from one location to another but—being of sub-atomic size—take the existence and size of cities as given.

We say that an urban distribution $D \in \mathcal{D}$ is *assortative* if each of the following conditions holds: (i) for each rank $k \in \mathbb{N}$, the type of an agent inhabiting a city of mass m_k^D takes a value in $(m_{k+1}^D, m_k^D]$; (ii) the type of an agent inhabiting a village takes

a value in $(-\infty, 0]$ or $(m_1^D, +\infty)$. So, by assortativeness agents are segregated into cities according to their types, guaranteeing that each agent inhabits the smallest city where she can make profits while villages are inhabited by a mix of highly ambitious and highly unambitious agents.

We say that an urban distribution $D \in \mathcal{D}$ has *nested structure* if $F(m_{k+1}^D) = F(m_k^D) - n_k^D m_k^D$ for each rank $k \in \mathbb{N}$, which is a recurrence relation that determines the series of masses of cities $(m_k^D)_{k=1}^\infty$ given the largest city mass m_1^D and the series of numbers of cities $(n_k^D)_{k=1}^\infty$. Intuitively, this nestedness condition is intimately related to assortativeness.

Proposition 1 1. *An urban distribution is an equilibrium if and only if it is assortative.* 2. *Each equilibrium has nested structure.*

Note that, as all equilibria have nested structure, we can represent the structure of each equilibrium graphically using the recurrence relation of nestedness. In Figures 2 and 3, we consider two examples of distributions of types and the graphical representations of the corresponding equilibria. Each of them is useful to identify critical points to be addressed in the subsequent analysis.

Figure 2 illustrates the structures of six equilibria for the Beta distribution with parameters $(\alpha, \beta) = (2, 5)$. Together with the equilibrium with no cities, the figure fully characterizes the set of all seven equilibria in this example. All shown six equilibria Pareto dominate the equilibrium with no cities as they introduce new cities all else equal, and many other pairs of equilibria can be Pareto ranked (although not all of them).¹⁹

In the example of Figure 2, Pareto rankings are evident because the equilibria have a very limited number of cities (at most three). In reality, we typically observe a much higher number of cities on the territory of a country and, given that we have a continuum of agents in our model (a convenient approximation of a large finite population), it may seem natural to expect infinitely many cities in equilibrium. This can be achieved with suitable restrictions on the distribution of types that are introduced in the next example.

Figure 3 illustrates the structures of three equilibria for the Gumbel distribution with parameters $(\alpha, \beta) = (0, .05)$. As $F(0) = e^{-1} \approx .37$, there is a positive mass of agents that can make profits in villages, and the nested structure of each equilibrium must be identified using the shifted cumulative mass function $F(m) - F(0)$, represented by the dotted line. The maximum level of urbanization that can be achieved in equilibrium corresponds to the case of a single city of mass $m^* \approx .63$ in the left panel, where m^* is determined by the equation $F(m^*) - F(0) = m^*$. There are uncountably many other equilibria, at least one for each size of the largest city $m \in (0, m^*]$, each presenting

¹⁹ Specifically, each equilibrium in the bottom panels Pareto dominates the equilibrium in the top left panel, and the equilibrium in the top central panel Pareto dominates the equilibrium in the top left panel while it is Pareto dominated by the equilibria in the bottom central and bottom right panels. However, there is no Pareto dominance relation between the equilibrium in the top right panel and the equilibria in the other five panels.

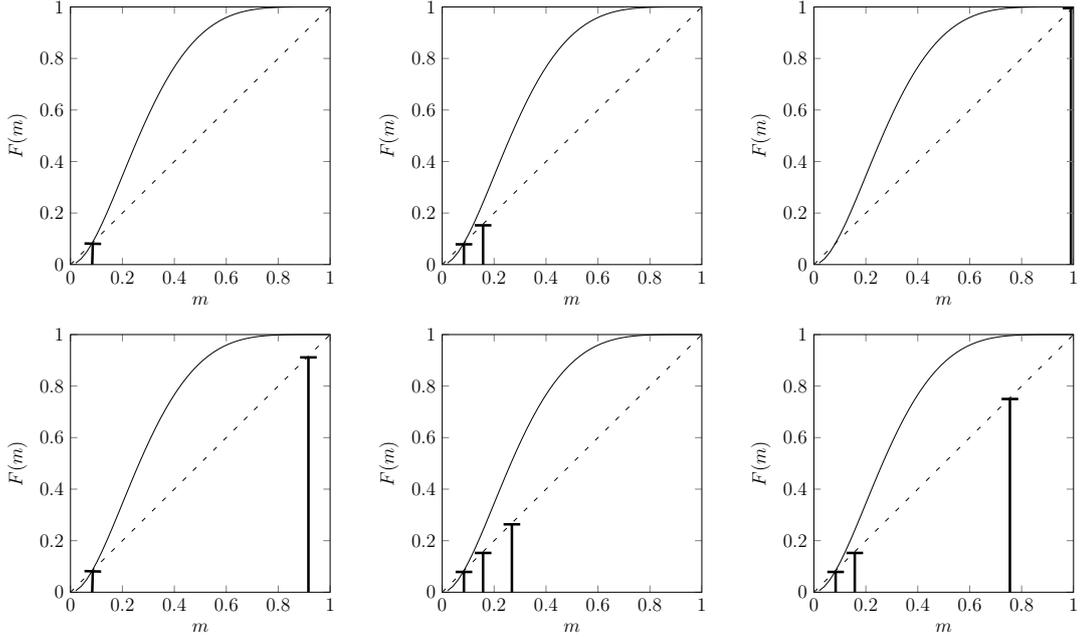


Figure 2: Given $a = 1$, $F(m)$ corresponds to the cumulative mass function of the Beta distribution with parameters $(\alpha, \beta) = (2, 5)$. Each panel depicts the nested structure of a different equilibrium, where the solid lines indicate the sizes of the various cities.

infinitely many cities and an urbanization level equal to $(F(m) - F(0))/a$. For instance, the central panel depicts an equilibrium with an infinite number of cities, each of different size, where the largest size is .2, while the right panel depicts another equilibrium with an infinite number of cities, each of different size except for the two largest ones, each of size .2. Note that there is no Pareto dominance across these three equilibria, although we may expect the equilibrium in the central panel to lead to higher welfare than the one in the right panel as it presents equal urbanization levels (which implies equal profits for all agents) while having much lower urban concentration (which implies lower aggregate crowding cost, by the weak convexity of c). These insights on efficiency and welfare will be formalized shortly, in Proposition 2. Before doing so, we briefly discuss desirable restrictions on the distribution of types.

As suggested by the example in Figure 3, one can show that, in our model, there exists an equilibrium with infinite number of cities if and only if $f(0) > 1$. Note that this implies the existence of $\epsilon > 0$ such that $m < F(m) - F(0)$ for each $m \in (0, \epsilon]$, that is, there is an excess of agents which can make profits in a city of size smaller than or equal to ϵ and cannot make profits in a village. In this spirit, we now consider a stronger condition on the distribution of types that allows to focus on equilibria with infinite number of cities for a broad set of urbanization levels.²⁰ We say that a distribution of types is *non-constraining* if $m < F(m) - F(0)$ for each $m \in (0, \bar{m}_F)$, which means that for each m in the pre-image of $(0, a)$ there is an excess of agents which can make profits in a city of size m and cannot make profits in a village. This greatly simplifies the

²⁰ The following restriction is purely for expositional convenience. It is straightforward to show that all our core results extend under the weaker assumption $f(0) > 1$.

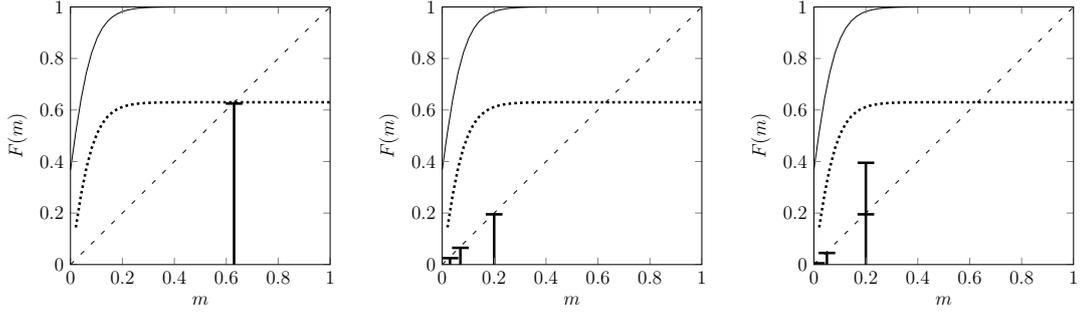


Figure 3: Given $a = 1$, $F(m)$ corresponds to the cumulative mass function of the Gumbel distribution with parameters $(\alpha, \beta) = (0, .05)$, represented by the solid curve, while the dotted curve represents $F(m) - F(0)$. Each panel depicts the nested structure of a different equilibrium, where the vertical lines indicate the sizes of the various cities.

analysis, leading to the general properties of equilibria discussed in and after Remark 1. Before we turn to this discussion, however, we state a brief observation on the stability of the equilibria our analysis concentrates on.

We motivate our focus on non-constraining distributions by the argument that, within our framework, they guarantee the existence of equilibria with the “realistic” feature of representing a high number of cities. Another way to motivate this focus is by the stability of the implied equilibria. The argument is that, if a distribution systematically leads to unstable outcomes, there will be forces—by evolution or design—pushing for a change towards stability. We briefly sketch the argument here, which is along the lines of Granovetter (1978) in our extended framework with crowding costs. For a given equilibrium, consider an exogenous marginal decrease in the size of a city. If, on the one hand, the distribution is non-constraining, such a marginal decrease does not affect the size of other cities of different size, as agents have no incentive to migrate to or from these cities. Conversely, it affects the size of other cities of equal size only minimally, as agents of these cities will marginally migrate to the perturbed city so that their sizes re-balance. In this sense, the equilibrium may thus be considered *stable*. If, on the other hand, the distribution is constraining, in a typical equilibrium there must be a city whose size is determined by F crossing the 45 degree line from below (for an example, see Figure 2 where this is the case for all equilibria except the one with no cities and the one with a single city containing the whole population). One can show that a marginal decrease in the size of such a city then leads to a chain reaction so that all residents leave the perturbed city for the villages. As this drastically alters the structure of the equilibrium, such a situation may thus be considered unstable.

Remark 1 *Given that the distribution of types is non-constraining:*

1. *For each $m \in (0, a - F(0))$, there exists an equilibrium with size of the largest city equal to m , infinite number of cities, and level of urbanization equal to $(F(m) - F(0))/a$ if $m \leq \bar{m}_F$ and equal to $(a - F(0))/a$ if $m > \bar{m}_F$.*
2. *There exist multiple equilibria exhibiting up to $n \in \mathbb{N}$ cities of same size $m \in (0, a - F(0))$ if and only if $nm \leq F(m) - F(0)$.*

Recall that, in the example of Figure 2, certain equilibria Pareto dominate others because they create new cities all else equal. Conversely, while there is no Pareto dominance across the equilibria of Figure 3, we may expect the equilibrium in the right panel to lead to higher welfare than the one in the central panel as it presents equal urbanization levels while having much lower urban concentration. These two intuitions are at the core of our welfare analysis.

We say that an urban distribution $D \in \mathcal{D}$ has *substantial structure* if $m_1^D \geq \underline{m}_F := F^{-1}(\max_{m \in [0, \bar{m}_F]} [F(m) - m])$, a condition which rules out particularly low levels of urbanization (e.g., no cities) because they are Pareto dominated.

We say that an urban distribution $D \in \mathcal{D}$ has *hierarchical structure* if $n_k^D = 1$ for each rank $k \in \mathbb{N}$ with $m_k^D > 0$, which means that there are no multiple cities of same size so that the aggregate crowding cost is minimized for a given urbanization level.

Proposition 2 *Given that the distribution of types is non-constraining:*

1. *An equilibrium is cost-efficient if and only if it has hierarchical structure and the size of the largest city is lower than or equal to \bar{m}_F .*
2. *An urban distribution is welfare-efficient only if it is an equilibrium (up to misallocation of zero mass of agents) that is cost-efficient and has substantial structure.*

Proposition 2 formalizes aforementioned intuitions on the optimality of substantial and hierarchical structures. Firstly, it states that cost-efficiency implies hierarchical structure, meaning that the urban distribution cannot present cities of equal size. The intuition is that, by the weak convexity of the cost function c and for a given mass of urbanized $U > 0$, the crowding cost

$$C(D) = \sum_{k=1}^{\infty} n_k^D m_k^D c(m_k^D)$$

is effectively a measure of urban concentration belonging the family of generalized Herfindahl-Hirschman indices,

$$\mathcal{K}(D) = \sum_{k=1}^{\infty} n_k^D \phi \left(m_k^D / \sum_{k=1}^{\infty} n_k^D m_k^D \right),$$

with $\phi(m/U) = mc(m)$. The crowding cost then naturally decreases when a larger city is dismantled (or reduced in size) by redistributing its population to smaller cities. This is exactly what happens in our model when transitioning from a non-hierarchical to a hierarchical equilibrium, implying a lower crowding cost. Secondly, Proposition 2 provides novel insights on the connection between the upper bound \bar{m}_F and the cost-efficient size of the largest city as well as the relation between welfare-efficiency and equilibrium (where the former implies the latter). The reason for the upper bound \bar{m}_F is best understood via the example in Figure 4, which shows that increasing the size of the largest city above \bar{m}_F leaves urbanization (and the profits of each agent) unchanged while it increases urban concentration (therefore increasing the aggregate crowding cost). Finally, regarding the stated relation between welfare-efficiency and

equilibrium in Proposition 2, the former implies the latter because there is an excess of agents in the population that can make profits in a city of any size, the distribution of types being non-constraining.²¹ This implies that agents can always be rearranged so that there is no need to keep anyone in a city unwillingly, that is, it is efficient to keep an individual in a city only if such individual actually wants to be there. Hence, in our model the only source of inefficiency is miscoordination on the wrong equilibrium, as the efficient structure is an equilibrium itself and thus self-sustaining.

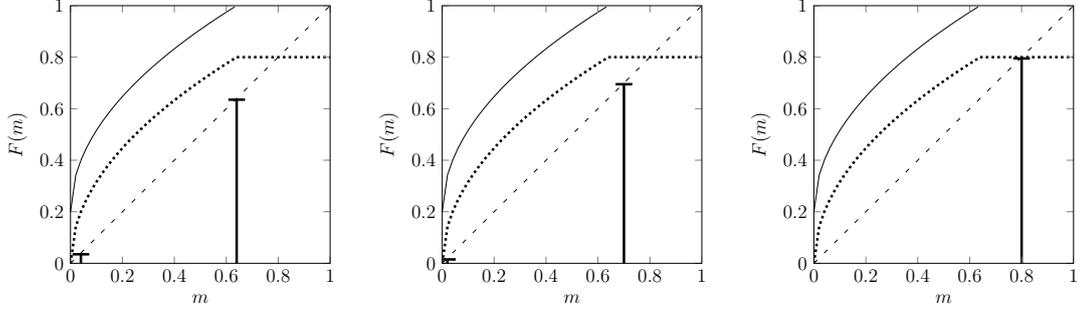


Figure 4: Given $a = 1$, $F(m) = .2 + \sqrt{m}$ corresponds to the cumulative mass function of the shifted Beta distribution with parameters $(\alpha, \beta) = (0, .5)$, represented by the solid curve, while the dotted curve represents $F(m) - F(0)$ where $\bar{m}_F = .64$. Each panel depicts the nested structure of a different equilibrium, where the vertical lines indicate the sizes of the various cities.

Proposition 2 greatly simplifies the maximization of utilitarian welfare. Suppose that the distribution of types is non-constraining. By Proposition 2, a cost-efficient equilibrium is fully characterized by the mass of the largest city, and a welfare-efficient urban distribution must be a cost-efficient equilibrium that is substantial. Then, denoting by $D^*(\mu_1) \in \mathcal{D}$ the cost-efficient equilibrium with mass of the largest city equal to $\mu_1 \in [\underline{m}_F, \bar{m}_F]$, the maximization of utilitarian welfare can be simply stated as

$$\begin{aligned} \max_{\mu_1 \in [\underline{m}_F, \bar{m}_F]} W(D^*(\mu_1)) &= \frac{1}{a} \int_0^{\mu_1} \pi(t) dF(t) - \frac{1}{a} \sum_{k=1}^{\infty} \mu_k c(\mu_k) \\ \text{s.t. } \mu_k &= F^{-1}(F(\mu_{k-1}) - \mu_{k-1}) \text{ for each } k \geq 2. \end{aligned}$$

It is noteworthy that, on the considered domain, choosing the size of the largest city μ_1 is equivalent to choosing the corresponding level of urbanization $\mathcal{U}(D^*(\mu_1)) = (F(\mu_1) - F(0))/a$, which by our previous considerations must take a value in

$$[(F(\underline{m}_F) - F(0))/a, (a - F(0))/a].$$

Going back to our examples, one can show that each of the equilibria with hierarchical and substantial structure depicted in the left and central panels of Figure 3 is welfare-efficient for some combination of cost and profit functions. This is because the

²¹ We wish to remark how these considerations rely on the distribution of types being non-constraining, as if it was not, there would not be such excess of agents and the whole argument would collapse, in the sense that it may be efficient to keep certain agents in a city against their will.

corresponding distribution of types is non-constraining. Conversely, if the distribution of types is constraining such as the one in Figure 2, it is possible that no equilibrium is welfare-efficient for a given combination of cost and profit functions.

4 Comparative statics of urban development

In this section we focus on welfare-efficient solutions and study how they should change with shocks to the fundamentals. Assuming F to be non-constraining, we exclusively consider cost-efficient equilibria, as the welfare-efficient urban distribution is one of them. Specifically, the two variables of interest are the level of urbanization and the level of urban primacy of cost-efficient equilibria, which can be written as

$$\mathcal{U}(D^*(\mu_1)) = (F(\mu_1) - F(0)) / a \text{ and } \mathcal{P}(D^*(\mu_1)) = \mu_1 / (F(\mu_1) - F(0))$$

for each size of the largest city $\mu_1 \in [0, \bar{m}_F]$. Note that $\mathcal{U}(D^*(\mu_1))$ and $\mathcal{P}(D^*(\mu_1))$ can be easily visualized graphically as the height of the function F evaluated at μ_1 (shifted by $F(0)$ and divided by a) and the fraction of this height that lies below the 45° line, respectively.

In what follows, we divide our comparative static analysis in short run and long run considerations. The *short run* is defined by a fixed level of urbanization, and we assume that any shock summarized by a change in the distribution of types from F' to F maps each cost-efficient equilibrium given the old distribution F' into the unique cost-efficient equilibrium with same urbanization level given the new distribution F . Within this framework, our short run analysis determines whether urban primacy should increase or decrease, depending on the specific shock. In the *long run*, we assume that urbanization can adjust to the welfare-efficient level (provided that coordination is achieved). While the analysis of the long run consequences of shocks to F does not lead to sharp predictions, we can fully determine the relation between the level of urbanization and the level of urban primacy across cost-efficient equilibria for any given distribution of types F . Intuitively, this relation is suggestive of the long run trends in the levels of urbanization and urban primacy of the welfare-efficient solution driven by shifts in the functions π and c , and more generally, of the relation between the levels of urbanization and urban primacy across different levels of development akin to the solution of coordination problems.

Short run considerations

In our short run analysis, we consider three shocks to the fundamentals that change the qualitative properties of the distribution of types.

We say that the distribution of types F is a *population replication* of the distribution of types F' corresponding to a mass of agents equal to a if there is $k > 1$ such that $F(t) = kF'(t)$ for all $t \in [0, a]$. Then, a population replication rescales the mass of agents by a factor of k while leaving the distribution of types unchanged (in relative terms).

We say that the distribution of types F is *more ambitious* than (first-order stochastically dominates) the distribution of types F' on $[0, a]$ if each of the following conditions holds: (i) $F(t) = F'(t)$ if $t \in \{0, a\}$; (ii) $F(t) < F'(t)$ if $t \in (0, a)$. This means that high types are relatively more abundant in F than in F' (while low types are relatively scarcer).

We finally consider a mean-preserving spread that transfers mass from the center of a distribution to the sides, leaving the mean unchanged. Formally, we say that the distribution of types F is an *expansion* of the distribution of types F' on $[0, a]$ if each of the following conditions holds: (i) $F(t) = F'(t)$ if and only if $t \in \{0, \int_0^a rdF(r), a\}$; (ii) $\int_0^t F'(r)dr > \int_0^t F(r)dr$ for all $t \in (0, a)$; (iii) $\int_0^a rdF(r) = \int_0^a rdF'(r)$.

Proposition 3 *Restricting attention to non-constraining distributions of types:*

1. *If the distribution of types F is a population replication of F' , urban primacy is lower in the cost-efficient equilibrium with F than in the cost-efficient equilibrium with F' for any given level of urbanization.*
2. *If the distribution of types F is more ambitious than F' on $[0, a]$, urban primacy is higher in the cost-efficient equilibrium with F than in the cost-efficient equilibrium with F' for any given level of urbanization.*
3. *If the distribution of types F is an expansion of F' on $[0, a]$, there is $\lambda^* \in (0, (a - F(0))/a)$ such that urban primacy is higher (lower) in the cost-efficient equilibrium with F than in the cost-efficient equilibrium with F' for any given level of urbanization that is higher (lower) than λ^* .*

Figure 5 is an illustration of the results summarized by Proposition 3. The left panel considers a population replication that doubles the population and compares the old cost-efficient equilibrium with the new cost-efficient equilibrium with equal level of urbanization. As shown by the dotted lines, the size of the largest city is left unchanged, which implies that the level of urban primacy decreases with the population replication (it becomes half). This illustrates Point 1 above.

The central panel of Figure 5 considers a shift in the distribution of types that leads the new distribution to first-order stochastically dominate the old. As shown by the dotted lines, for a fixed level of urbanization, the size of the biggest city is larger in the cost-efficient equilibrium of the new distribution, which implies that urban primacy is higher as predicted by Point 2 above.

Finally, the right panel of Figure 5 considers a shift in the distribution of types that leads the new distribution to be an expansion of the old. As shown by the dotted lines, for a fixed level of urbanization, the size of the largest city is smaller in the cost-efficient equilibrium of the new distribution than in the corresponding equilibrium of the old. Moreover, this remains true for any old size of the largest city below .5 (the old size is .4 in the example), while the opposite would be true if the old size of the largest city was above .5. As the level of urbanization is proportional to the size of the largest city (see Point 1 of Remark 1), this illustrates Point 3 above.

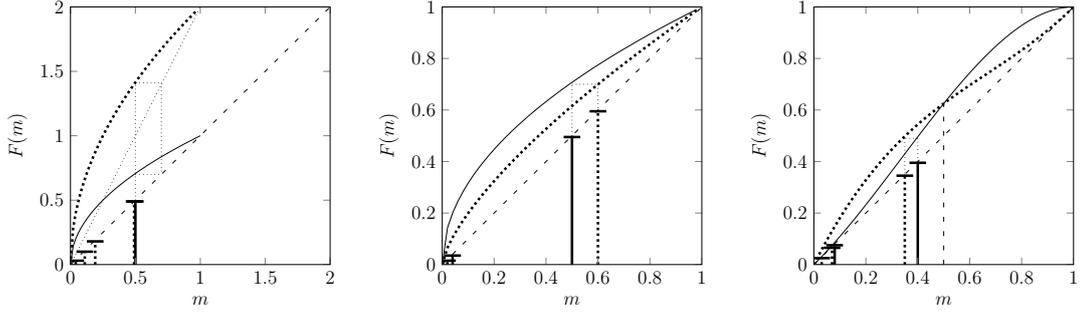


Figure 5: In the left panel the solid curve corresponds to the case $a = 1$, depicting the cumulative mass function of the Beta distribution with parameters $(\alpha, \beta) = (0, .5)$, while the dotted curve depicts a population replication that doubles the mass of agents. The central panel focuses on $a = 1$, depicting the cumulative mass functions of the Beta distributions with parameters $(\alpha, \beta) = (0, .5)$ (solid line) and $(\alpha, \beta) = (0, .7)$ (dotted line), where the second distribution first-order stochastically dominates the first. The right panel also focuses on $a = 1$, depicting the cumulative mass functions of the shifted Beta distributions $F'(m) = m + m^2(1 - m)$ (solid line) and $F(m) = m + m(1 - m)^2$ (dotted line), where F is an expansion of F' . Each panel depicts the nested structures of two different equilibria, where the vertical solid (dotted) lines indicate the sizes of the various cities that correspond to the equilibrium with the solid (dotted) cumulative mass function.

Long run considerations

We now consider long run trends in urban development, when the level of urbanization can adjust to the welfare-efficient level (provided that coordination is achieved). In principle, one can always identify the optimal level of urbanization by solving the constrained maximization problem stated at the end of Section 3. However, our attempts suggest that results crucially depend on specific assumptions on the functions F , π and c , and no general pattern emerges.²²

While we cannot generally predict whether urbanization increases or decreases in the long run as a consequence of shocks to F , we can determine how a change in the urbanization level should affect urban primacy across cost-efficient equilibria for a given F . Intuitively, by the constrained maximization problem at the end of Section 3, this analysis is suggestive of the long run trends in the levels of urbanization and urban primacy of the welfare-efficient solution due to rescaling of the functions π and c . More generally, it can indicate the relation between the levels of urbanization and urban primacy across different levels of development. In this context, we can think of developmental increments as the solution to coordination problems limiting the agglomeration of agents into cities. Recall that, in our model, the expectation of many agents inhabiting a city constitutes the very incentive for such agents to actually go and settle there.

Proposition 4 *Let F be non-constraining. For each $\mu_1 \in [0, \overline{m}_F)$, the relation between urban primacy and the level of urbanization of the cost-efficient equilibrium $D^*(\mu_1)$ is such that a marginal increase in the urbanization level leads to an increase (decrease)*

²² The only regularity is perhaps the positive (negative) effect on the optimal urbanization level of the scaling up by positive multiplication of function π (function c).

in urban primacy if

$$f(\mu_1) < (>) (F(\mu_1) - F(0)) / \mu_1. \quad (1)$$

To appreciate Proposition 4, it is fundamental to give meaning to the two variables $f(\mu_1)$ and $(F(\mu_1) - F(0)) / \mu_1$ which govern the long run relation between the level of urbanization and urban primacy across cost-efficient equilibria. On the one hand, $f(\mu_1)$ is the *marginal density* of the urbanized types in the cost-efficient equilibrium $D^*(\mu_1)$, which indicates the total mass of agents that would become urbanized if the level of urbanization was to be marginally increased. On the other hand, $(F(\mu_1) - F(0)) / \mu_1$ is the *average density* of the urbanized types in such an equilibrium, which indicates the relative abundance of agents that can make profits in the largest city. We are now ready to grasp the intuition of Proposition 4. Note that, by the nature of cost-efficient equilibria, an increase in urbanization must go hand in hand with a proportional increase in the size of the largest city. All newly urbanized agents must be residents of the largest city, but these may or may not be enough to match the new size of the largest city, and consequent migration in or out of the largest city may be triggered. Note that such migration must necessarily be from or to the smaller cities, not the villages, therefore involving the urban population only. Thus, by changing the fraction of urban population that resides in the largest city, these population movements directly affect urban primacy. Specifically, when $f(\mu_1) < (F(\mu_1) - F(0)) / \mu_1$, the mass of newly urbanized joining the largest city is relatively small, and a marginal increase in urbanization should lead to migration of agents from the smaller cities to the largest to fill in the vacant slots, thus increasing urban primacy. Conversely, when $f(\mu_1) > (F(\mu_1) - F(0)) / \mu_1$, the mass of newly urbanized joining the largest city is relatively large and the migration must go in the opposite direction, thus decreasing urban primacy.²³

We now argue that, under fairly general conditions, the mechanism identified in Proposition 4 predicts a *U-shaped relation* between urban primacy and the level of urbanization in a cost-efficient equilibrium. While Proposition 5 identifies a sufficient condition to state this formally, Figure 6 illustrates this in an example.

We say that a distribution of types F has a density f that is *single-peaked* on $(0, \bar{m}_F)$ if there is $m^* \in (0, \bar{m}_F)$ such that $df(m)/dm > (<)0$ if $m < (>)m^*$ for all $m \in (0, \bar{m}_F)$.

Proposition 5 *Let F be non-constraining and satisfying $f(\mu_1) = (F(\mu_1) - F(0)) / \mu_1$ for some $\mu_1 \in (0, \bar{m}_F)$.²⁴ If the density f is single-peaked on $(0, \bar{m}_F)$, the relation between urban primacy and the level of urbanization of cost-efficient equilibria is U-shaped.*

²³ This directly links to condition (1). This condition can be understood graphically as the marginal density $f(\mu_1)$ being the slope of the function $F(\mu_1) - F(0)$ evaluated at the point $(\mu_1, F(\mu_1) - F(0))$, while the average density $(F(\mu_1) - F(0)) / \mu_1$ being the slope of the line passing through the origin and the same point. Urban primacy $\mu_1 / (F(\mu_1) - F(0))$ can be understood as the reciprocal of this latter slope. In this setting, the relative magnitude of these slopes determines whether urban primacy should increase or decrease with the level of urbanization at a specific point.

²⁴ Alternatively, instead of this last condition, it is sufficient to assume $f(a) = 0$.

The crucial assumption behind Proposition 5 is to have a density f that is single-peaked on $(0, \bar{m}_F)$, which we now argue to be a plausible property of a distribution of types. Consider an extension of our model where F is endogenously determined in a pregame interaction in which individuals choose their types by maximizing expected utility under strategic uncertainty on the formation of the urban distribution. Although this extension is far from obvious,²⁵ we can immediately see that certain predictions should hold generally and serve to justify the single-peakedness of f . Intuitively, if a distribution of types emerges from the maximization of expected utility, business plans of intermediate ambition should be the most common as they are close to the optimal compromise in the trade-off between higher profits and lower crowding costs. Conversely, highly or minimally ambitious plans should be relatively scarce due to excessive crowding costs and the insufficient profits, respectively. So, in this setup, we should expect f to be single-peaked in the interior, and the peak of f should coincide with the ex-ante optimal type.

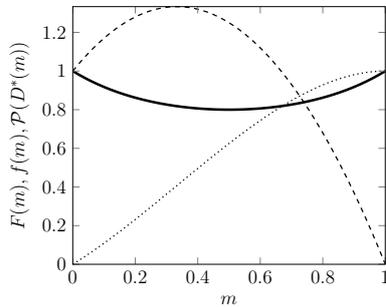


Figure 6: Given $a = 1$, the dotted, dashed and solid lines respectively depict the (non-constraining) shifted Beta distribution $F(m) = m + m^2(1 - m)$, its density function $f(m) = 1 + 2m(1 - m) - m^2$, and the level of urban primacy corresponding to the cost-efficient equilibrium with the largest city of size $\mu_1 = m$.

As a final note, we wish to point out that the converse of Proposition 5 can also hold under different assumptions. Roughly speaking, if we consider a single-dipped density f (i.e., if there is $m^* \in (0, \bar{m}_F)$ such that $df(m)/dm < (>)0$ if $m < (>)m^*$ for all $m \in (0, \bar{m}_F)$), a Kuznets-type inverted U-shaped relation between urban primacy and level of urbanization is generated by the same arguments of Proposition 4. While in the following section we concentrate on the U-shaped relation using 20th and 21st century observations, the opposite could follow from a bi-modal distribution of ambition ascribed to the lack of access to education of large parts of pre-20th century populations. More generally, Kuznets-type cycles of inverted U-shaped and then U-shaped relations between urban primacy and level of urbanization can be generated as a consequence of the introduction of new technologies and the subsequent growth of access to education for the use of such technologies (see Chapter 2 in Milanovic, 2016 for a related approach).

²⁵ A challenge is the formalization of the expectation with respect to the formation of the urban distribution under strategic uncertainty, which can be conceptualized within the framework of global games (see, e.g., Carlsson and Van Damme, 1993; Frankel et al., 2003).

5 An empirical pattern

To test the predicted U-shaped relation empirically, we base the analysis of this section on the World Bank’s dataset, topic “Urban Development,” which includes a panel reporting the levels of urbanization and urban primacy for each country in the world, annually from 1960 to 2016.²⁶ As predicted by Proposition 5, the scatter plot in Figure 7 suggests a U-shaped empirical relation between the level of urbanization and urban primacy. While this scatter plot is based on cross-country average data, the rest of this section tests this hypothesis further using econometric analysis of a panel consisting of all 218 covered countries of the world through the last 60 years.

Our analysis is similar in spirit to the highly influential Imbs and Wacziarg (2003) on stages of economic development. They document a remarkably robust U-shaped relation between sectoral concentration and GDP per capita. Since industrial sectors typically cluster in specialized cities according to increasing returns from spatial proximity, and since higher levels of GDP per capita typically coincide with higher levels of urbanization as joint manifestations of higher levels of economic development, we would like to pose our model as a common theoretical foundation for the empirical observations in Imbs and Wacziarg (2003) and ours. With some caution, one may also link our prediction to the empirical U-shaped relation between the inequality of income (or wealth) and GDP per capita as documented for instance by Piketty and Saez (2003) and Saez and Zucman (2016). Intuitively, when economic resources concentrate in fewer cities and industries, it may also be that income concentrates in the hands of the fewer individuals who dominate these cities and industries.

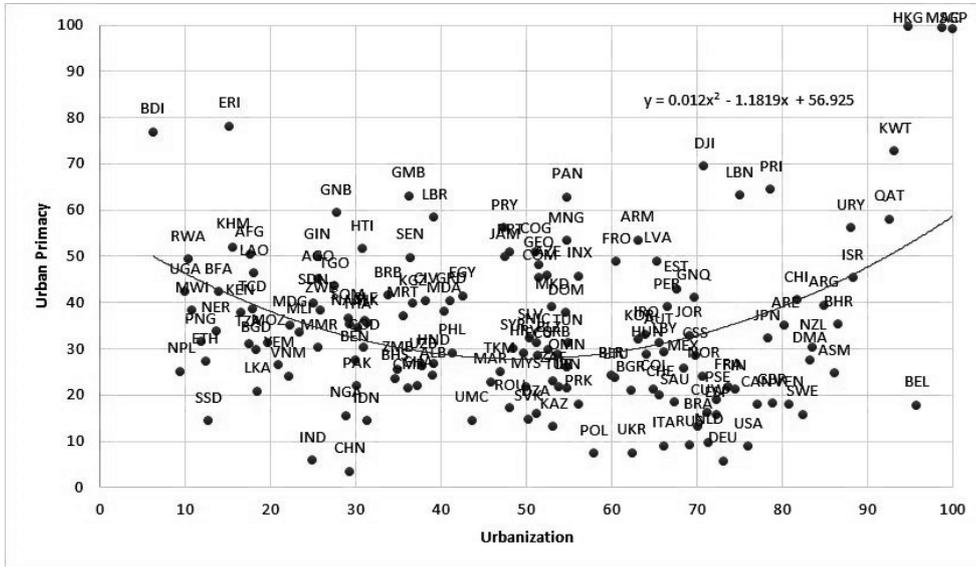


Figure 7: U-shaped cross-country relation between the average level of urbanization and the average urban primacy, where these averages are computed within each country across the years 1960-2016. Source: Own calculations based on World Bank data.

²⁶ This data is publicly available from <https://data.worldbank.org/topic/urban-development>.

In the following, our empirical strategy consists of a linear regression with the level of urban primacy of each country and year as dependent variable and the level of urbanization and the level of urbanization squared in the same country and year as the two main independent variables. We start by considering basic econometric specifications with robust standard errors with fixed effects for year and continent/country.²⁷ The resulting estimations are in Table 1. As shown in columns (1) and (2), the specifications

Table 1: Relation between urban primacy and urbanization in the world sample.

Urban primacy	(1)	(2)	(3)
Urbanization	-.9504*** (.0408)	-.6149*** (.0406)	-.0816** (.0384)
Urbanization squared	.0096*** (.0004)	.0073*** (.0003)	.0011*** (.0003)
Observations	8689	8689	7562
R^2	0.1039	0.2127	0.9289

Notes: Columns (1) to (3), respectively, correspond to the specifications (1) without fixed effects, (2) with year fixed effects and continent fixed effects, (3) with year fixed effects and country fixed effects excluding the 21 countries belonging to the “continent” Middle East and North Africa. Standard errors are heteroscedastically robust; ***, **, and * indicate statistical significance at the levels of 1%, 5%, and 10%, respectively.

which do not include country fixed effects yield statistically significant estimations of the two coefficients of interest which are negative for urbanization and positive for urbanization squared, and are thus in line with our predictions. Most notably, the specification in column (2) with year and continent fixed effects confirms the U-shaped relation. These estimations are robust to marginal changes of the empirical specification such as excluding certain countries from the sample, like e.g., the ones in the top-right corner of Figure 7. However, when we introduce country fixed effects the evidence is somewhat weakened as the significance of the estimations depends on the exact empirical specification. For instance, the empirical pattern continues to hold as long as we exclude from the sample the countries that belong to the continent-label ‘Middle East and North Africa’, as shown in column (3), while the empirical pattern is blurred when these countries are included. Intuitively, other dynamics than those captured by our analysis may be at play in these countries as many of them have been systematically plagued by political turmoil, civil war, and international conflict.

One weakness of the above estimations is that, when we consider the relation between urban primacy and urbanization within a country and across time, the distribution of types is generally not constant as assumed in Proposition 5. This motivates our second empirical exercise where we introduce into a standard regression with country and year fixed effects control variables roughly corresponding to the shocks to the distribution considered in Proposition 3. As within the World Bank’s dataset these controls are reported only for a relatively small subset of rich countries and recent years,

²⁷ The World Bank’s dataset on Urban Development codes countries according to the continent they belong to.

we exclusively focus on the corresponding subsamples within Europe and Central Asia and the world.²⁸ The resulting estimations are shown in Table 2 which considers two alternative sets of three control variables as empirical proxies for the three shocks. In these alternative specifications, ‘population replication’ is either population density

Table 2: Relation between urban primacy and urbanization on restricted samples with additional control variables.

Urban primacy	(1)	(2)	(3)	(4)
Urbanization	-1.4579*** (.4806)	-.5322*** (.1581)	-.6450** (.2771)	-.6862*** (.1154)
Urbanization squared	.0113*** (.0034)	.0019 (.0012)	.0039* (.0020)	.0028*** (.0010)
Population density	-.0562** (.0262)	.0184*** (.0057)	- -	- -
Tertiary education exp.	.0545 (.0438)	.0711* (.0381)	- -	- -
Income ineq. (Gini coeff.)	.0136 (.0352)	.0072 (.0466)	- -	- -
Total population	- -	- -	1.71e-07*** (5.30e-08)	1.71e-08 (2.06e-08)
Tertiary education enroll.	- -	- -	-.0351** (.0164)	-.0061 (.0176)
Income ineq. (top 10%)	- -	- -	-.0164 (.0628)	-.0462 (.0479)
Observations	219	465	344	709
R^2	.9933	.9768	.9877	.9752

Notes: Columns (1) and (3) correspond to the subsamples of available observations for Europe and Central Asia, while columns (2) and (4) for the world. Regressions include year and country fixed effects; standard errors are heteroscedastically robust; ***, **, and * indicate statistical significance at the levels of 1%, 5%, and 10%, respectively.

or total population, ‘more ambition’ is either tertiary education expenditure (as % of total government expenditure on education) or tertiary education enrollment (as % of the age group that is entitled to enrollment), and ‘expansion’ is income inequality measured either as Gini coefficient or as income share held by the top 10%.²⁹ As shown in Table 2, no matter which set of controls we choose or whether we focus on ‘Europe and Central Asia’ or the world, our empirical estimations are systematically consistent with the U-shaped hypothesis.

To conclude, the econometric exercises in Tables 1 and 2 together with the scatter plot in Figure 7 are suggestive of an empirical pattern that is consistent with the U-shaped hypothesis predicted by Proposition 5. We provide additional evidence in the Appendix, in which Figure 8 demonstrates the robustness of the pattern across time

²⁸ Roughly speaking, by introducing these control variables we lose about 90–95% of the observations almost exclusively focusing on years after 1990 and on a subset of countries within the continents ‘Europe and Central Asia,’ ‘North America,’ and ‘South Asia.’

²⁹ All these control variables are from World Bank datasets corresponding to the topics Health, Education and Poverty, respectively, which are publicly available from <https://data.worldbank.org/indicator>.

(with scatter plots for the time periods 1960-1979, 1980-1999, 2000-2016) while Table 3 and Figure 9 show that the U-shape persists with polynomial specifications of higher order. Arguably, our handful of plots and regressions are far from a comprehensive analysis as many alternative empirical specifications can be chosen in terms of, e.g., subsamples and control variables. However, in combination with the much more robust evidence in Imbs and Wacziarg (2003) on the U-shaped relation between sectoral concentration and the level of economic development and the related findings in Piketty and Saez (2003) and Saez and Zucman (2016), we believe this is sufficient to motivate our model as empirically relevant.

6 Conclusions

We take a novel approach to urban development in the tradition of threshold models of social interaction. In our model, the number and the sizes of cities are endogenously determined by the incentives of agents to freely move across municipalities, where settlers in larger cities face a trade-off between higher productivity and higher crowding costs. In this setup, we characterize the set of equilibria, study their welfare properties, and analyze the equilibrium relation between three key measures of urban development: urbanization, urban concentration, and urban primacy. One appealing feature of our model is that all equilibria are defined by a simple recursive algorithm that can be represented graphically with an intuitive diagram, and welfare-efficiency corresponds to an urban distribution with an infinite number of cities of heterogeneous size.

Focusing on welfare-efficient solutions (and the weaker concept of cost-efficiency, which does not require the level of urbanization to be welfare-efficient) we find that in the short run population replications tend to decrease urban primacy, while the short run effects on urban primacy of changes in population characteristics are positive if they come in the form of first-order stochastic dominance, and positive/negative depending on the high/low level of urbanization if they come in the form of mean-preserving spreads. Although we cannot generally pin down the long run effects of these shocks, we can fully determine how changing the level of urbanization should affect other variables. Assuming that the distribution of types is single-peaked in the interior, our findings suggest a U-shaped relation between the level of urbanization and urban primacy. We find preliminary confirmation of this prediction considering a panel of all countries of the world through the last 60 years.

Due to its simplicity and versatility, our model of urban development has potential for various applications and extensions. One possibility is to explore the conflict of interest across cities. While here we have focused on welfare-efficient solutions, in practice these may be difficult to implement because of the necessary compensation of the ‘losers’ using part of the gains of the ‘winners’ of a welfare improvement. As these compensatory transfers should occur across cities in our model, they may be often infeasible and motivate an analysis of second-best solutions. From an empirical viewpoint, an interesting application would be to estimate the distribution of types of a country from the distribution of city sizes assuming that the nestedness condition holds.

This would allow for more extensive testing of our predictions as one could monitor how the estimated distribution of types changes across time and countries, and whether these patterns are broadly in line with what we know from other sources.

Appendix

Proof of Proposition 1

1. Assortativeness. Recall that an urban distribution $D \in \mathcal{D}$ is assortative if and only if each of the following conditions holds: (i) for each rank $k \in \mathbb{N}$, the type of an agent inhabiting a city of mass m_k^D takes a value in $(m_{k+1}^D, m_k^D]$; (ii) the type of an agent inhabiting a village takes a value in $(-\infty, 0]$ or $(m_1^D, +\infty)$. Consider any assortative urban distribution. Note that each urban agent is located in a city of the smallest available size that is sufficiently high for her to make profits (so that her type is lower than or equal to such size but higher than the size of any smaller city). So, no urban agent prefers to move to another existing city (or village) as either it is too small for her to make profits or it is unnecessarily large, leading to the same profits but a higher crowding cost. On the other hand, no villager prefers to move to an existing city as either she cannot make profits in there (as her type is higher than the size of such city) or she already makes profits in the village (therefore moving to the city only increases the crowding cost). So, any assortative urban distribution is an equilibrium. We now prove the converse: that any urban distribution that is not assortative is not an equilibrium. It is easy to verify that for any urban distribution that is not assortative one of the following statements must be true: there is an agent in some city that does not make profits or that makes profits but can make profits in some other existing city that is smaller (i.e., condition (i) is violated); there is an agent in some village that does not make profits but can make profits in some existing city (i.e., condition (ii) is violated). As each of these statements is in contradiction with the definition of equilibrium (as there is an agent that prefers to move), this proves that an urban distribution is an equilibrium if and only if it is assortative.

2. Nestedness. Finally, we need to show that all equilibria have nested structure. Let $D \in \mathcal{D}$ be any equilibrium. As D is necessarily assortative, by condition (ii) of assortativeness a mass $a - (F(m_1^D) - F(0))$ on agents is in villages. Of the remaining mass $F(m_1^D) - F(0)$ of urban agents, a mass $n_k^D m_k^D$ is in cities of rank $k \in \mathbb{N}$. By condition (i) of assortativeness, $n_k^D m_k^D = F(m_{k+1}^D) - F(m_k^D)$ for each $k \in \mathbb{N}$, which can be rearranged into the recurrence relation of nestedness $F(m_{k+1}^D) = F(m_k^D) - n_k^D m_k^D$, thus concluding our proof. \square

Proof of Proposition 2

1. Cost-efficiency. We start by showing that an equilibrium $D' \in \mathcal{D}$ is cost-efficient if and only if its structure is hierarchical. Let the structure of D' be hierarchical. Our strategy is to prove that any other equilibrium with same the same level of urbanization and same profits for each agent presents higher aggregate crowding cost than D' . For a

contradiction, suppose that there is another equilibrium $D \in \mathcal{D}$ with same urbanization level and same profits as D' such that

$$C(D) = \sum_{h=1}^{\infty} n_h^D m_h^D c(m_h^D) \leq C(D') = \sum_{h=1}^{\infty} n_h^{D'} m_h^{D'} c(m_h^{D'}). \quad (2)$$

As F is non-constraining, the levels of urbanization take value $\mathcal{U}(D') = (F(m_1^{D'}) - F(0))/a$ and $\mathcal{U}(D) = (F(m_1^D) - F(0))/a$. By assumption, $f(m) > 0$ if $m \leq \bar{m}_F$ and $f(m) = 0$ if $m > \bar{m}_F$. We then divide our analysis in two cases: $m_1^{D'} < \bar{m}_F$ and $m_1^{D'} \geq \bar{m}_F$.

Consider $m_1^{D'} < \bar{m}_F$. Since $f(m_1^{D'}) > 0$, $\mathcal{U}(D') = \mathcal{U}(D)$ implies $m_1^D = m_1^{D'}$. As D' is hierarchical $n_k^{D'} = 1$ for all $k \in \mathbb{N}$. Letting $k \in \mathbb{N}$ being the smallest number such that $n_k^D > 1$, condition (2) can then be rewritten as

$$(n_k^D - 1)m_k^D c(m_k^D) + \sum_{h=k+1}^{\infty} n_h^D m_h^D c(m_h^D) \leq \sum_{h=k+1}^{\infty} m_h^{D'} c(m_h^{D'}). \quad (3)$$

Now, let $k' \in \mathbb{N}$ be the highest number such that $F(m_{k'}^{D'}) > F(m_{k+1}^D)$. Note that such k' exists by the assumption that F is non-constraining. By $F(m_{k+1}^D) = F(m_k^D) - n_k^D m_k^D$, $F(m_k^D) = F(m_{k'}^{D'})$ and $F(m_{k'}^{D'}) > F(m_{k+1}^D)$, we must have

$$n_k^D m_k^D = F(m_k^D) - F(m_{k+1}^D) > F(m_{k'}^{D'}) - F(m_{k'}^{D'}).$$

As assortativeness implies $F(m_k^{D'}) - F(m_{k'}^{D'}) = \sum_{h=k}^{k'-1} m_h^{D'}$, and $m_k^D = m_k^{D'}$, we obtain $(n_k^D - 1)m_k^D > \sum_{h=k+1}^{k'-1} m_h^{D'}$. Therefore, there is $\rho \in [0, 1)$ such that

$$(n_k^D - 1)m_k^D = \sum_{h=k+1}^{k'-1} m_h^{D'} + (1 - \rho)m_{k'}^{D'}, \quad (4)$$

$$\sum_{h=k+1}^{\infty} n_h^D m_h^D = \rho m_{k'}^{D'} + \sum_{h=k'+1}^{\infty} m_h^{D'}. \quad (5)$$

Note that $m_k^D > m_h^{D'}$ for all $h \in \{k+1, \dots, k'\}$ and $m_{k+x}^D > m_{k'+x}^{D'}$ for all $x \geq 1$, which by (4) and (5) respectively imply

$$(n_k^D - 1)m_k^D c(m_k^D) > \sum_{h=k+1}^{k'-1} m_h^{D'} c(m_h^{D'}) + (1 - \rho)m_{k'}^{D'} c(m_{k'}^{D'}),$$

$$\sum_{h=k+1}^{\infty} n_h^D m_h^D c(m_h^D) > \rho m_{k'}^{D'} c(m_{k'+1}^{D'}) + \sum_{h=k'+1}^{\infty} m_h^{D'} c(m_h^{D'}).$$

Then, a necessary condition for (3) to hold is

$$\rho m_{k'}^{D'} \left(c(m_{k'}^{D'}) - c(m_{k+1}^D) \right) \geq m_{k'-1}^{D'} \left(c(m_k^D) - c(m_{k'-1}^{D'}) \right).$$

As $\rho m_k^{D'} < m_{k'-1}^{D'}$, this is possible only if $c(m_{k'}^{D'}) - c(m_{k+1}^D) \geq c(m_k^D) - c(m_{k'-1}^{D'})$. However, since $m_k^D + m_{k+1}^D \geq m_{k'-1}^{D'} + m_{k'}^{D'}$, $m_k^D > m_{k'-1}^{D'} > m_{k'}^{D'} > m_{k+1}^D$ and the function c is weakly convex, this condition is never fulfilled. So, we can conclude that given $m_1^{D'} < \bar{m}_F$ the equilibrium D' is cost-efficient if and only if its structure is hierarchical.

Consider $m_1^{D'} \geq \bar{m}_F$. As D' is an equilibrium we must have $m_1^{D'} \leq a - F(0)$, and any equilibrium $D \in \mathcal{D}$ that has the same level of urbanization as D' must satisfy $m_1^D \in [\bar{m}_F, a - F(0)]$. We are going to show that D' is cost-efficient if and only if it is hierarchical and $m_1^{D'} = \bar{m}_F$. Suppose D satisfies such properties. Firstly, it follows from arguments similar to the above that any other equilibrium $D \in \mathcal{D}$ with $m_1^D = \bar{m}_F = m_1^{D'}$ and which is non-hierarchical has higher aggregate crowding cost than D' . Secondly, suppose $m_1^{D'} = \bar{m}_F$ and let $D \in \mathcal{D}$ be any other equilibrium with $m_1^D > \bar{m}_F$. If D is non-hierarchical, by arguments analogous to our previous analysis it must lead to an aggregate crowding cost $C(D)$ that is higher than the one of the equilibrium $D'' \in \mathcal{D}$ that is hierarchical and has largest city of same size as m_1^D . On the other hand, D'' can be derived from D' via a series of mass transfers from larger cities to smaller ones, which implies $C(D') < C(D'')$ by the convexity of c . Then, $C(D') < C(D'') < C(D)$ and condition (2) never holds. So, combining these results with our previous analysis we can conclude that D' is cost-efficient if and only if its structure is hierarchical and $m_1^{D'} \leq \bar{m}_F$.

2. Welfare-efficiency. We now show that an urban distribution that is welfare-efficient must be a cost-efficient equilibrium (up to misallocation of zero mass of agents). Since welfare-efficiency implies cost-efficiency, to do so it is sufficient to show that a welfare-efficient urban distribution is necessarily an equilibrium. Let $D \in \mathcal{D}$ be a welfare-efficient urban distribution. If D has nested structure it must be an equilibrium, otherwise aggregate profits can be increased by reshuffling individuals across cities and villages without changing the structure and therefore without affecting the aggregate crowding cost. Suppose D has non-nested structure, which implies that $F(m_{k+1}^D) \neq F(m_k^D) - n_k^D m_k^D$ for some $k \in \mathbb{N}$. We divide our analysis in two cases: $F(m_{k+1}^D) < F(m_k^D) - n_k^D m_k^D$ and $F(m_{k+1}^D) > F(m_k^D) - n_k^D m_k^D$. If $F(m_{k+1}^D) < F(m_k^D) - n_k^D m_k^D$, welfare can be augmented by decreasing by some arbitrarily small $\epsilon > 0$ the mass of a city of size m_k^D and increasing by the same amount ϵ the mass of a city of size m_{k+1}^D , while reshuffling agents across cities and villages so that aggregate profits are unchanged while the aggregate crowding cost decreases. Note that this reshuffling is always possible as the distribution of types is non-constraining, while the aggregate crowding cost decreases as

$$(m_{k+1}^D + \epsilon)c(m_{k+1}^D + \epsilon) + (m_k^D - \epsilon)c(m_k^D - \epsilon) < m_{k+1}^D c(m_{k+1}^D) + m_k^D c(m_k^D),$$

since by assumption the function c is weakly convex. On the other hand, if $F(m_{k+1}^D) > F(m_k^D) - n_k^D m_k^D$ there must be a positive mass of urban agents that do not make profits in some city. Then, welfare can be augmented by moving an arbitrarily small fraction of these agents to a village, which reduces the aggregate crowding cost, while reshuffling agents across cities and villages so that aggregate profits are unchanged. Again, this reshuffling is always possible as the distribution of types is non-constraining. This

proves our desired result.

Finally, we are going to show that, given that an urban distribution D is welfare-efficient, the structure of D must be substantial, that is,

$$F(m_1^D) \geq \max_{m \in [0, \bar{m}_F]} F(m) - m.$$

We already know that D is an equilibrium (up to misallocation of zero mass of agents) whose structure is hierarchical. Suppose for a contradiction that

$$F(m_1^D) < \max_{m \in [0, \bar{m}_F]} F(m) - m.$$

Since F is non-constraining, there is $m' \in (m_1^D, \bar{m}_F)$ such that $F(m_1^D) = F(m') - m'$, which implies that there is another equilibrium which is identical to D except that there is a new city of size m exclusively composed of agents who are villagers in D and that can make profits in this new city. Note that this would constitute a Pareto improvement on D , and that welfare-efficiency implies Pareto efficiency. Then, if D is welfare-efficient, it must have substantial structure. \square

Proof of Proposition 3

For any non-constraining distribution of types F , let $D_{\lambda, F} \in \mathcal{D}$ denote the cost-efficient equilibrium that corresponds to the level of urbanization $\lambda \in (0, (a - F(0))/a)$. Note that, as F is non-constraining, the size of the largest city is $m_1^{D_{\lambda, F}} = F^{-1}(\lambda a + F(0))$ and urban primacy takes value $\mathcal{P}(D_{\lambda, F}) = m_1^{D_{\lambda, F}} / (a\lambda)$.

1. Population replication. Consider a distribution F that is a population replication of another distribution F' which rescales the mass of agents by a factor of $k > 1$, so that the new mass of agents is $a = ka'$ and the new distribution of types is $F(t) = kF'(t)$ for all $t \in [0, a]$. Given that urbanization is constant,

$$(F(m_1^{D_{\lambda, F}}) - F(0))/a = \lambda = (F'(m_1^{D_{\lambda, F'}}) - F'(0))/a',$$

so that we obtain $m_1^{D_{\lambda, F}} = m_1^{D_{\lambda, F'}}$ which implies the desired result

$$\mathcal{P}(D_{\lambda, F}) = \frac{m_1^{D_{\lambda, F}}}{a\lambda} < \mathcal{P}(D_{\lambda, F'}) = \frac{m_1^{D_{\lambda, F'}}}{a'\lambda}.$$

2. First-order stochastic dominance (more ambitious). Consider a distribution F that first-order stochastically dominates another distribution F' on $[0, a]$, so that $F(t) = F'(t)$ if $t \in \{0, a\}$ and $F(t) < F'(t)$ if $t \in (0, a)$. Given that urbanization is constant,

$$F(m_1^{D_{\lambda, F}}) = a\lambda + F(0) = F'(m_1^{D_{\lambda, F'}}),$$

so that we obtain $m_1^{D_{\lambda, F}} > m_1^{D_{\lambda, F'}}$ which implies the desired result $\mathcal{P}(D_{\lambda, F}) > \mathcal{P}(D_{\lambda, F'})$.

3. Expansion (mean-preserving spread). Consider a distribution F that is an expansion of another distribution F' on $[0, a]$, so that (i) $F(t) = F'(t)$ if and

only if $t \in \{0, \int_0^a rdF(r), a\}$; (ii) $\int_0^t F'(r)dr > \int_0^t F(r)dr$ for all $t \in (0, a)$; (iii) $\int_0^a rdF(r) = \int_0^a rdF'(r)$. Being a mean-preserving spread, one can show the cumulative mass functions F, F' must satisfy

$$F(x) > (<)F'(x) \text{ if } x < (>)\tilde{m} := \int_0^a rdF(r).$$

Given that urbanization is constant, $F(m_1^{D_{\lambda,F}}) = a\lambda + F(0) = F'(m_1^{D_{\lambda,F'}})$. These considerations jointly imply

$$m_1^{D_{\lambda,F}} < (>)m_1^{D_{\lambda,F'}} \text{ if } m_1^{D_{\lambda,F}}, m_1^{D_{\lambda,F'}} < (>)\tilde{m},$$

or equivalently if $\lambda < (>)\lambda^* := (F(\tilde{m}) - F(0))/a$. It follows that $\mathcal{P}(D_{\lambda,F}) < (>)\mathcal{P}(D_{\lambda,F'})$ if $\lambda < (>)\lambda^*$, which proves the desired result. \square

Proof of Proposition 4

Given that F is non-constraining, the set of cost-efficient equilibria is characterized by the unique equilibrium $D^*(\mu_1)$ with hierarchical structure for each size of the largest city $\mu_1 \in [0, \bar{m}_F]$. Take any such μ_1 and consider the corresponding cost-efficient equilibrium. As in a cost-efficient equilibrium there is a single largest city, the level of urbanization is

$$\mathcal{U}(D^*(\mu_1)) = (F(\mu_1) - F(0)) / a.$$

Differentiating with respect to the size of such city,

$$d\mathcal{U}(D^*(\mu_1))/d\mu_1 = f(\mu_1)/a > 0,$$

we directly see that a marginal increase in urbanization goes hand in hand with a marginal increase in the size of the largest city. Then, the level of urban primacy

$$\mathcal{P}(D^*(\mu_1)) = \mu_1 / (F(\mu_1) - F(0))$$

increases (decreases) with a marginal increase in the level of urbanization if

$$d\mathcal{P}(D^*(\mu_1))/d\mu_1 = 1 / (F(\mu_1) - F(0)) - \mu_1 f(\mu_1) / (F(\mu_1) - F(0))^2 > (<)0,$$

which is equivalent to $f(\mu_1) < (>) (F(\mu_1) - F(0)) / \mu_1$ and directly leads to condition (1). \square

Proof of Proposition 5

The argument directly follows from the application of Proposition 4 to the cost-efficient equilibria corresponding to the considered distribution of types. From the discussion right after Proposition 4, recall that the average density

$$\Phi(\mu_1) := (1/\mu_1) \int_0^{\mu_1} f(m)dm = (F(\mu_1) - F(0)) / \mu_1$$

is the average of the marginal densities $f(m)$ in the interval $m \in [0, \mu_1]$. This implies $\lim_{\mu_1 \rightarrow 0} \Phi(\mu_1) = \lim_{\mu_1 \rightarrow 0} f(\mu_1)$. By the stated assumptions on the distribution of types,

the density function f is single-peaked on $(0, \bar{m}_F)$ so f must be first increasing and then decreasing, and the cumulative mass function F is non-constraining and satisfying

$$f(\mu_1^*) = \Phi(\mu_1^*) \text{ for some } \mu_1^* \in (0, \bar{m}_F).$$

These considerations jointly imply that

$$f(\mu_1) > \Phi(\mu_1) \text{ for } \mu_1 \in (0, \mu_1^*) \text{ and } f(\mu_1) < \Phi(\mu_1) \text{ for } \mu_1 \in (\mu_1^*, \bar{m}_F),$$

where the critical point μ_1^* lies in the subdomain of f where the density is decreasing.³⁰ Now, by Proposition 4, urban primacy increases (decreases) with a marginal increment in the urbanization level if $f(\mu_1) < (>)\Phi(\mu_1)$. Combined with the analysis above, we then have that urban primacy decreases with urbanization for $\mu_1 \in (0, \mu_1^*)$, while it increases for $\mu_1 \in (\mu_1^*, \bar{m}_F)$. As in any cost efficient equilibrium there is a single largest city, the urbanization level can be written as $\mathcal{U}(D^*(\mu_1)) = (F(\mu_1) - F(0)) / a$ and thus urbanization increases with μ_1 . We can conclude that the relation between urban primacy and urbanization is U-shaped. \square

³⁰ For a graphical visualization of the argument, as already stated in Footnote 23, note that $f(\mu_1)$ is the slope of the function $F(\mu_1) - F(0)$ at the point $(\mu_1, F(\mu_1) - F(0))$, while $\Phi(\mu_1)$ is the slope of the line passing through the origin and the same point.

Robustness of U-shaped results

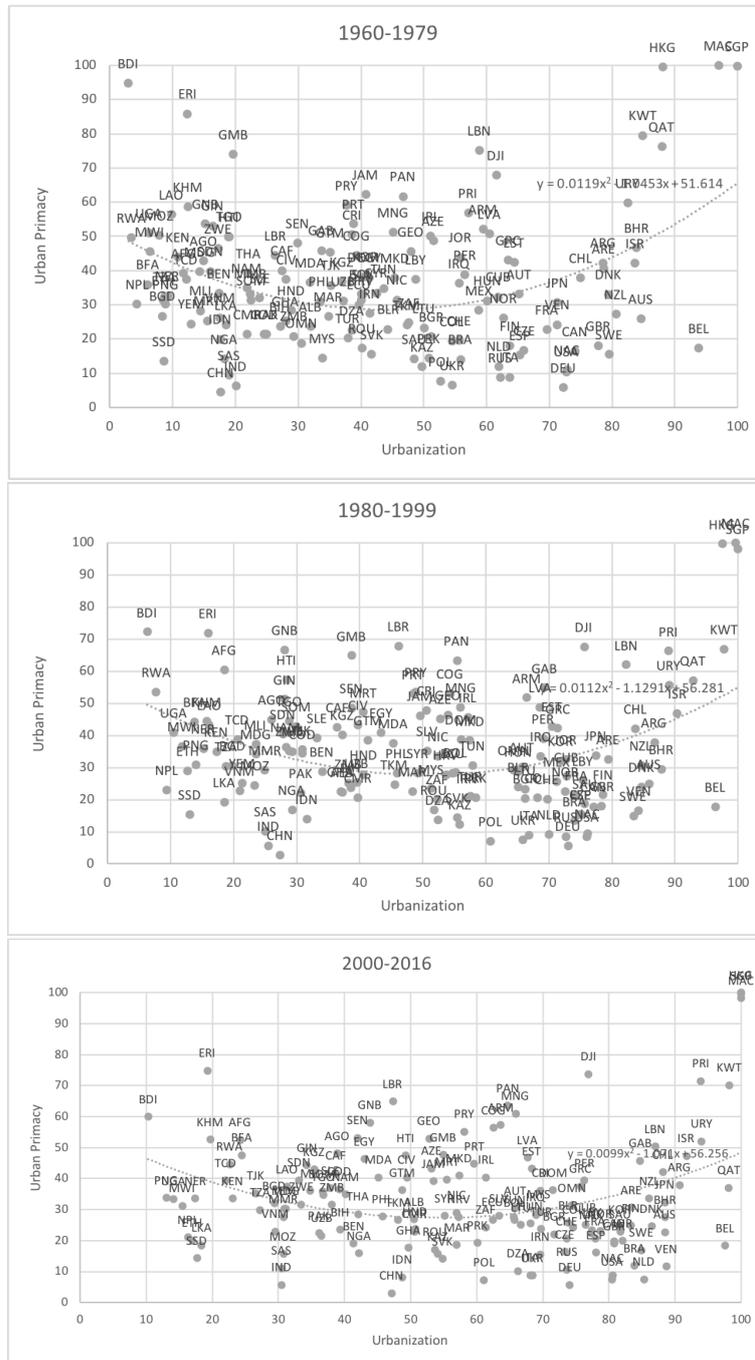


Figure 8: U-shaped cross-country relation between the average level of urbanization and the average urban primacy, where averages are computed within each country across the years 1960-1979, 1980-1999, and 2000-2016. Source: Own calculations based on World Bank data.

Table 3: Relation between urban primacy and urbanization in polynomial form of different degree (one to five).

Urban primacy	(1)	(2)	(3)	(4)	(5)
Constant	34.2584*** (1.6350)	47.3656*** (1.7815)	30.7888*** (2.0235)	61.7247*** (2.2639)	52.3944*** (3.0844)
Urbanization	.1917*** (.0143)	-.6149*** (.0406)	.8846*** (.1021)	-3.7317*** (.19405)	-1.9250*** (.41468)
Urbanization ²	-	.0073*** (.0003)	-.0284*** (.0022)	.1532*** (.0069)	.04702** (.0215)
Urbanization ³	-	-	.0002*** (.0001)	-.0024*** (.0001)	.0001 (.00049)
Urbanization ⁴	-	-	-	.0001*** (4.70e-07)	-1.49e-05*** (5.06e-06)
Urbanization ⁵	-	-	-	-	1.07e-07*** (1.92e-08)
Observations	8689	8689	8689	8689	8689
R ²	0.1643	0.2127	0.2506	0.3243	0.3275
AIC	73900.81	73384.86	72957.36	72060.67	72019.06
BIC	74360.35	73851.47	73431.04	72541.42	72499.8

Notes: All columns correspond to specifications in the world sample with year fixed effects and continent fixed effects. Standard errors are heteroscedastically robust; ***, **, and * indicate statistical significance at the levels of 1%, 5%, and 10%, respectively.

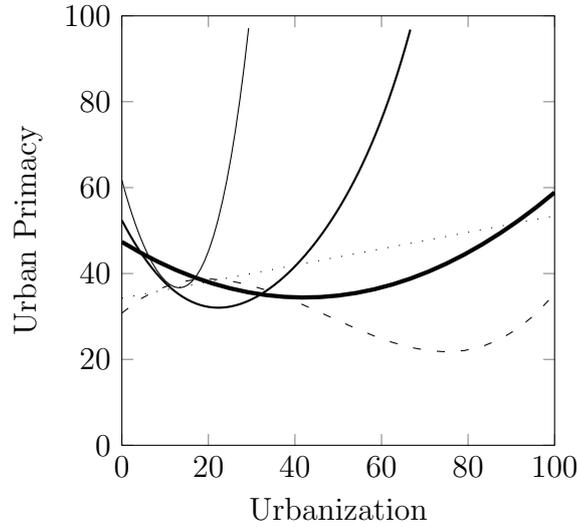


Figure 9: The five lines depict the relation between urban primacy and urbanization in polynomial form of different degrees. The second degree polynomial is represented in ultra thick as it is our chosen specification, while the dotted, dashed, solid, and thick lines respectively correspond to the polynomial in first, third, fourth, and fifth degrees. The coefficients of these five polynomial specifications are determined by the five regressions in Table 3.

References

- Albouy, D., K. Behrens, F. Robert-Nicoud, and N. Seegert, “The optimal distribution of population across cities,” *Journal of Urban Economics*, 2019, 110, 102–13.
- Aziz, H., R. Savani, and H. Moulin, “Hedonic games,” in F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. Procaccia, eds., *Handbook of Computational Social Choice*, Cambridge University Press, 2016, p. 356–376.
- Behrens, K., G. Duranton, and F. Robert-Nicoud, “Productive cities: sorting, selection, and agglomeration,” *Journal of Political Economy*, 2014, 122 (3), 507–553.
- Benhabib, J., A. Bisin, and M. Jackson, *Handbook of Social Economics*, Elsevier, 2010.
- Bogomolnaia, A. and M. Jackson, “The stability of hedonic coalition structures,” *Games and Economic Behavior*, 2002, 38 (2), 201–230.
- Cadot, O., C. Carrère, and V. Strauss-Kahn, “Export diversification: what’s behind the hump?,” *Review of Economics and Statistics*, 2011, 93 (2), 590–605.
- Carlino, G. and W. Kerr, “Agglomeration and innovation,” in “Handbook of Regional and Urban Economics,” Vol. 5, Elsevier, 2015, pp. 349–404.
- Carlsson, H. and E. Van Damme, “Global games and equilibrium selection,” *Econometrica*, 1993, pp. 989–1018.
- Davis, D. and J. Dingel, “The comparative advantage of cities,” *Journal of International Economics*, 2020, 123, 103291.
- de Palma, A., Y. Papageorgiou, J-F. Thisse, and P. Ushchev, “About the origin of cities,” *Journal of Urban Economics*, 2019, 111, forthcoming.
- Dodgshon, R. and R. Butlin, *Historical Geography of England and Wales*, Elsevier, 2013.
- Duranton, G. and W. Kerr, “The logic of agglomeration,” in “The New Oxford Handbook of Economic Geography,” Oxford University Press, 2018, chapter 18, pp. 347–65.
- Eatwell, R. and M. Goodwin, *National Populism: The Revolt Against Liberal Democracy*, Penguin UK, 2018.
- Eeckhout, J., “Gibrat’s law for (all) cities,” *American Economic Review*, 2004, 94 (5), 1429–1451.
- , R. Pinheiro, and K. Schmidheiny, “Spatial sorting,” *Journal of Political Economy*, 2014, 122 (3), 554–620.
- Frankel, D., S. Morris, and A. Pautner, “Equilibrium selection in global games with strategic complementarities,” *Journal of Economic Theory*, 2003, 108 (1), 1–44.
- Friedlander, D., “The spread of urbanization in England and Wales, 1851-1951,” *Population Studies*, 1970, 24 (3), 423–43.
- Fujita, M., P. Krugman, and A. Venables, *The Spatial Economy*, MIT Press, 2001.
- Gabaix, X., “Zipf’s law for cities: an explanation,” *Quarterly Journal of Economics*, 1999, 114 (3), 739–767.
- Galor, O., *Inequality and Economic Development: The Modern Perspective*, Edward Elgar Publishing, 2009.
- Gaubert, C., “Firm sorting and agglomeration,” *American Economic Review*, 2018, 108 (11), 3117–53.
- Genicot, G. and D. Ray, “Aspirations and inequality,” *Econometrica*, 2017, 85 (2), 489–519.
- Glaeser, E., *Triumph of the city*, Penguin, 2011.
- and J. Henderson, “Urban economics for the developing world: an introduction,” *Journal of Urban Economics*, 2017, 98, 1–5.
- Granovetter, M., “Threshold models of collective behavior,” *American Journal of Sociology*, 1978, 83 (6), 1420–1443.
- Henderson, J., “Urbanization and economic development,” *Annals of Economics and Finance*, 2003, 4, 275–342.
- Hirschman, A., *The Strategy of Economic Development*, Yale University Press, 1958.

- Imbs, J. and R. Wacziarg**, “Stages of diversification,” *American Economic Review*, 2003, *93* (1), 63–86.
- Junius, K.**, “Economic development and industrial concentration: an inverted U-curve,” *SSRN Working Paper 1695*, 1997.
- Milanovic, B.**, *Global inequality: a new approach for the age of globalization*, Harvard University Press, 2016.
- Peleg, B. and P. Sudhölter**, *Introduction to the Theory of Cooperative Games*, Vol. 34, Springer Science & Business Media, 2007.
- Picard, P. and T. Tabuchi**, “On microfoundations of the city,” *Journal of Economic Theory*, 2013, *148* (6), 2561–82.
- Piketty, T. and E. Saez**, “Income inequality in the United States: 1913–1998,” *Quarterly Journal of Economics*, 2003, *118* (1), 1–41.
- Ravallion, M.**, *The Economics of Poverty: History, Measurement, and Policy*, Oxford University Press, 2015.
- Ray, D.**, “Uneven growth: a framework for research in development economics,” *Journal of Economic Perspectives*, 2010, *24* (3), 45–60.
- Redding, S.**, “Economic geography: a review of the theoretical and empirical literature,” in D. Bernhofen, R. Falvey, D. Greenaway, and U. Kreickemeier, eds., *Handbook of International Trade*, London, UK: Palgrave Macmillan, 2013, chapter 16, pp. 497–531.
- Rosenstein-Rodan, P.**, “Problems of industrialization of Eastern and Southeastern Europe,” *Economic Journal*, 1943, *53* (2010/211), 202–211.
- Saez, E. and G. Zucman**, “Wealth inequality in the United States since 1913: evidence from capitalized income tax data,” *Quarterly Journal of Economics*, 2016, *131* (2), 519–578.
- Schelling, T.**, “Models of segregation,” *American Economic Review*, 1969, *59* (2), 488–493.
- Simon, H.**, “Bandwagon and underdog effects and the possibility of election predictions,” *Public Opinion Quarterly*, 1954, *18* (3), 245–253.
- Sun, Y., X. Zhang, G. Ren, F. Zwiers, and T. Hu**, “Contribution of urbanization to warming in China,” *Nature Climate Change*, 2016, *6* (7), 706.
- Watts, D. and P. Dodds**, “Threshold models of social influence,” in P. Bearman and P. Hedström, eds., *Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press, 2009, chapter 20, pp. 475–497.
- Williamson, J.**, “Regional inequality and the process of national development: a description of the patterns,” *Economic Development and Cultural Change*, 1965, *13* (4), 1–84.