CESifo WORKING PAPERS

7326
2018

November 2018

A Threshold Model of Urban Development

Alberto Vesperoni, Paul Schweinzer

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editors: Clemens Fuest, Oliver Falck, Jasmin Gröschl
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7326*
Category 9: Resource and Environment Economics

A Threshold Model of Urban Development

Abstract

We propose a simple model of distribution of economic activity across cities of endogenous size and number determined by individual incentives in the tradition of threshold models of social interaction. The individuals populating our model are endowed with idiosyncratic entrepreneurial creativity the realization of which requires urban agglomeration linked to a crowding cost. As the latter is higher in cities of larger size, this leads to a trade-off between productivity and congestion. While our focus on distributive aspects comes at the cost of highly stylized behavior, we aim to provide a tractable framework to think about the interlinkages between various measures of urban development which became increasingly available through accessible data sets. Our predictions include an U-shaped relationship between the well-known measures of urbanization and urban primacy, a hypothesis that we test empirically using World Bank data.

JEL-Codes: C700, D710, O180, Q560.

Keywords: agglomeration, urbanization, development.

Alberto Vesperoni
Department of Economics
Alpen-Adria-University
Klagenfurt / Austria
alberto.vesperoni@aau.at

Paul Schweinzer
Department of Economics
Alpen-Adria-University
Klagenfurt / Austria
paul.schweinzer@aau.at

October 22, 2018
This draft is preliminary, incomplete, and subject to change; please do not circulate. We would like to thank Írem Bozbay, Carlo Fiorio, Frédéric Robert-Nicoud, Miguel Portela, and Yoichi Sugita for the valuable comments.

1 Introduction

The Mesopotamian city of Uruk is often portrayed as the prototype of the modern city.[1] At the peak of its influence around 2900 BCE, it featured 50,000–80,000 inhabitants in an enclosed, protected area of 6km^2. Located at the intersection of important trading routes, Uruk is seen as the first agglomeration which possessed all hallmarks of the modern city: mass production with standardized work patterns, division of labor, effective administration and bureaucracy, archival of and access to written knowledge. These characteristics—equally stressed in their importance by Marshall (1920), Fujita et al. (2001), and Glaeser (2011)—allow for the specialization of tasks, services, and products within the metropolis while towns and rural villages can still coexist in the neighborhood. While the academic discourse on the underlying agglomeration process ranges back to at least von Thünen (1826), the theoretical literature is still struggling with the identification of precisely which elements are needed in order to explain this process. Of even greater importance than the intellectual puzzle presented by the historical reality of urban agglomeration is the pressing need to explain some spatial aspects of recent economic development. Ray (2010) illustrates various mechanisms through which not only the level, but the distribution of economic prosperity greatly matters for institutional stability and the long-run growth of nations. Among the many imbalances of the development process, Glaeser & Henderson (2017) argue that the tumultuous and uneven urban development of China, India, Nigeria and other emerging economies is one of the crucial challenges of our times.

In this paper we propose a theory of agglomeration which requires the potential of individual human creativity to be realized through interaction. In our model, this leads to the clustering of economic activity in what may be termed cities, villages, or any other conceivable social structure which leads to specialization. Hence, the proposed environment is rich enough to enable the study of urban agglomeration, that is, a countable set of cities not fully crowding out the rural village from both a historical and a developmental perspective. The proposed mechanism is motivated by a simple observation: in order to be productive, individuals often need the presence of other people. While a medieval farmer may have been able to gainfully reap the fruits of the lands without much interaction outside her family, the same is hardly true for an automotive worker today or, indeed, a university teacher or social media specialist. In an influential book on the future of urban development and the environment, Glaeser (2011) celebrates the city as the archetype of civilization and economic prosperity, but also acknowledges that it is the place where the worst living conditions can be experienced due to various inefficiencies related to congestion. Our analysis develops around this trade-off between enhanced productivity and diminished living conditions.

The core idea of our model is that, on the one hand, an individual (or firm) locates

[1] For details and references, see Crawford (2004).

in a city only if a critical mass of other individuals has already done so and there are no other cities that provide the same critical mass for better living conditions. Among the many well-known mechanisms that incentivize to locate in close proximity to other individuals by enhancing productivity (e.g., demand-supply linkages, specialization, sharing), we choose to focus on the economics of innovation because of their timeless role in the generation of economic prosperity and the increasing relevance in today's economy for developed as well as developing countries. On the other hand, we think of living conditions in terms of crowding costs worsening in city size, which can materialize in the form of higher rents, congestion of public services, pollution, etc. Note that, as we assume that the presence of individuals in a city constitutes the very incentive for more individuals to locate there, our framework naturally leads to the multiplicity of equilibria and the coordination failures that are typical of the development discourse.[2] In our analysis of the trade-off between innovation economies and crowding costs, we particularly focus on efficient solutions and how they should change with fundamentals such as population growth, technological improvement, and rising inequality.

Let us describe our framework in more detail. There is a continuum of agents scattered on a territory constituted by a continuum of locations, where the set of agents inhabiting a location is considered a city if it has positive mass and a village (or a solitary settlement) otherwise. We interpret agents as entrepreneurs with different business plans that are heterogeneous in their degree of ambition, where we think of ambition as jointly determined by inherited wealth and aspirations.[3] We assume that the ambition of a business plan affects the potential profits positively and its implementability negatively. Roughly speaking, ambitious plans can lead to higher profits once established, but are more difficult to launch and may require more supportive stakeholders at early stages of implementation. We crucially assume that, due to various frictions related to distance, these initial supporters are necessarily local, and that larger cities are more likely to provide the critical mass to launch an ambitious plan. As pointed out in Carlino & 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.[4] So, in our model, larger cities also lead to higher crowding costs, and each agent prefers to locate in the smallest city which presents a sufficiently large mass of residents for her business plan to become operative.

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 city or village

[2] For seminal contributions, see e.g., Rosenstein-Rodan (1943) and Hirschman (1958). As can be inferred from our brief literature overview in the next section, our proposed analysis focuses more on the study of potential mobilization of dormant resources rather than a standard neoclassical analysis of the efficient use of mobilized resources.

[3] See, e.g., Genicot & Ray (2017) for a formalization of this interaction.

[4] 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.

to move to another existing city or village. We characterize the set of equilibria and show that for each equilibrium the exact distribution of city sizes is determined by a recursive algorithm which can be visualized with an intuitive diagram. It turns out 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. Under fairly general conditions, this implies that the number of cities is infinite and there are no cities of equal size. So, urban concentration (i.e., the inequality of the distribution of urban population across cities) is minimized and the welfare-efficient urban distribution is fully characterized by our recursive algorithm plus the optimal level of urbanization (i.e., the fraction of agents living in cities instead of villages). Focusing on cost-efficient equilibria (as the welfare-efficient urban distribution is one of them), we engage in comparative statics that are relevant for urban development in the short to long run. 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 adjust to higher or lower levels. In our comparative statics exercises, we consider population replications that increase the mass of agents all else equal, and shifts in the distribution of ambition that lead to first-order stochastic dominance and mean-preserving spreads. In the short run, for any fixed level of urbanization, we determine that increases in the mass of agents (caused by, e.g., population growth or institutional and technological developments that increase mobility of people across regions) systematically reduce urban primacy (i.e., the share of urban population living in the largest city), upward shifts in the distribution of ambition (caused by, e.g., analogous shifts in the distribution of inherited wealth and aspirations, education, or technological improvement) 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 identified. However, although we cannot generally say whether urbanization increases or decreases in the long run, we can fully pin down how a change in the level of urbanization should affect urban primacy. Roughly speaking, under fairly general conditions our model delivers an U-shaped relation between urban primacy and the level of urbanization across cost-efficient equilibria for a given distribution of ambition-types, which is the principal testable prediction of our paper.

We provide preliminary empirical evidence in support of this U-shaped hypothesis using openly accessible World Bank data across all countries of the world through the last 60 years. Our findings roughly confirm the U-shaped relation across the world's sample using year and continent/country fixed effects, and within a restricted sample of rich countries where control variables are included to control for shifts in a country's distribution of ambition across time. There are related fields in the economics literature which have found similarly U-shaped correspondences between concentration and the degree of mobilization of resources (akin to the level of development). One group includes, among others, Imbs & Wacziarg (2003) for GDP per capita and sectoral concentration and related papers on concentration of exports. Another cluster revolves around inequality of income (or wealth) and GDP per capita as documented for instance by Piketty & Saez (2003) and Saez & Zucman (2016). Our paper provides a theory for this non-linear relationship for the context of urban development. While we do not claim one-to-one portability of our insights across fields, there is an obvious correlation between the distributions of people in space, those of industrial sectors, and of income.

The rest of the paper develops as follows. Section 1 reviews the literature and Section 2 defines the basic model. The core equilibrium and welfare analyses are in Section 3 while Section 4 focuses on the comparative statics. The empirical analysis is in Section 5. Section 6 concludes. All proofs are in the Appendix.

Related literature

An overview of the classical literature on spatial economics is contained in the first part ("Urban economics" and "Regional science") of Fujita et al. (2001). The rest of the same book—the standard textbook reference in this field—provides an excellent introduction to what has been called the "New Economic Geography," that is, the utilization of economic theories of trade and growth for the explanation of geographic realities through the three-way interaction between increasing returns, transportation costs, and the movement of productive factors. A recent survey of theoretical and empirical work of research in economic geography is Redding (2013), accompanied by the more specialized Duranton & Puga (2004), Behrens & Robert-Nicoud (2015), and Duranton & Kerr (2018). Within this literature the workhorse model of determination of the urban distribution is Henderson (1974), which consists of a neoclassical general equilibrium setting where the optimal size of a city is determined by fundamentals, and cities form in heterogeneous sizes because they host industries that differ in these fundamentals. More recent versions of this line of modeling are Behrens & Robert-Nicoud (2014) and Behrens et al. (2014), where (among other things) the framework is extended to allow for within-city heterogeneity. General equilibrium models have on their side elegance and consistency. However, the heavy machinery of general equilibrium can severely constrain the tractability of a model, obstructing the analysis of the distributive aspects of urban development which are the main focus of this paper. The study of urbanization in emerging economies calls for a model representing the multiple equilibria and coordination failures that are typical of the development discourse, which is one of our core motivations. Moreover, the analysis of interlinkages between the three core measures of urban development (i.e., urbanization, urban concentration, and urban primacy) requires a framework in which city sizes are highly interdependent.

Our theoretical framework is rooted in the tradition of so called "threshold" models of social interaction. The first appearance of an identifiable threshold model we found in the economics literature is Simon (1954), which puts forward an election framework where voters' preferences depend on the share of the population that supports a certain candidate. Other seminal applications of threshold models of social interaction include Schelling (1969), Akerlof (1970), Granovetter (1978), and Arthur (1989), respectively on racial segregation, quality of traded goods, rioting and technological standards. Roughly speaking, the common features of these models are that: (*i*) there is a large population of individuals and each individual must choose from the same discrete set of alternatives; (*ii*) each individual's preference over the alternatives depends on the population share choosing each of the alternatives; (*iii*) the thresholds of population shares that determine preferences differ across individuals and are summarized by a threshold distribution which is known. Due to the versatility and tractability of the basic framework, threshold models have been applied to a variety of topics in economics and other social sciences leading to a vast literature that is still vibrant today, sometimes under the designations of agent-based modeling or discrete choice with social interactions.[5] However, despite the wide popularity of threshold models, as far as we know, we are the first to apply these ideas to agglomeration and introduce crowding costs in the basic multinomial setup. In a nutshell, this modeling technique allows to achieve general solutions to the related issues of equilibrium and welfare efficiency, and to make broad distributional statements on macro-structures at the cost of more stylized micro-behavior compared to general equilibrium models.

A recent example of the vast set of empirical investigations of the determinants of the urban landscape is Henderson et al. (2018), which also provides an excellent summary of the recent empirical literature. The number of contributions investigating the causes and consequences of urbanization in OECD countries is high but the corresponding set of publications focusing on the developing world is comparatively small, as remarked in Glaeser & Henderson (2017). Among the various empirical contributions that focus on the determinants of urbanization and urban concentration, the seminal contribution Davis & Henderson (2003) is particularly relevant. Firstly, it shows that increasing urbanization goes hand in hand with higher income per capita and the bulk of economic activity shifting away from the agricultural sector towards industry and services, thus supporting the conceptual link between our prediction of an U-shaped relation between urban primacy and the level of urbanization and Imbs & Wacziarg (2003)'s evidence regarding the U-shaped relationship between sectoral concentration and GDP per capita. Secondly, our empirical exercise is similar in spirit to part of their empirical analysis, although the results diverge at times since there are important differences such as the different datasets and their focus on the logarithm of absolute urban population instead of urbanization as we define it.[6] Another relevant empirical contribution is Ades & Glaeser (1995), which demonstrates a solid positive relationship between urban primacy (measured as the logarithm of total population in the largest city) and the autocratic nature of government. For the time being our model abstracts from such political variables although they are clearly relevant.

[5] For seminal contributions to a wide range of economic problems see, e.g., Glaeser et al. (1996); Lindbeck et al. (1999); Brock & Durlauf (2001). For a recent survey, see Watts & Dodds (2009).

[6] See, e.g., Table 1 in Davis & Henderson (2003).

From a historical perspective, the literature on urban development emphasizes that the early stages of urban growth have often coincided with agricultural reform (see, e.g., Childe, 1950; Diamond, 1998). One popular interpretation is that improvements in agricultural productivity created the necessary surplus to sustain larger urban populations that engaged in activities other than subsistence. This led to specialization and trade, which in turn fostered innovation in a spiral of technological improvement that reinforced the city. On the one hand, our model can be seen as a very stylized version of this story, but it is clearly too simple to capture all complexities of the symbiotic relation between the rural and the urban, which include context-dependent variables such as natural resources, technology, political institutions, opportunity to trade within and across national borders, etc. On the other hand, our context-free approach can be seen as a plus, providing a unified theory of urban development which abstracts from such historical contingencies.

2 Model

Urban distributions

We consider a continuum of agents of mass $a > 0$ denoted by the set A. These agents are distributed on a territory constituted by a continuum of locations. We define an urban distribution of agents as a partition of A into a collection of sets of zero mass (villages, each containing rural agents who share a location with countably many dwellers) and a collection of sets of positive mass (cities, each containing urban agents who share a location with uncountably many fellows), where each of these sets (village or city) is assigned to a different location.

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 any urban distribution in $\mathcal{D}$ 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, denote by n_k^D the number of cities ranked k and by m_k^D the mass of agents contained in each of them. If the number of cities in 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 an urban distribution $D \in \mathcal{D}$ is summarized by the sequence $\mathcal{S}(D) := \left(m_k^D, n_k^D\right)_{k=1}^{\infty}$.[7]

Let $D \in \mathcal{D}$ be any urban distribution. We define the level of urbanization of D as the fraction of agents who are urban,

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

[7] 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 $\mathcal{S}(D) = (.3, 2; .2, 1; 0, 0; \ldots)$.

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 $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ 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.

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 ambition-type of agent $i \in A$, which is the critical mass required to implement her business plan.[8]

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

[8] The interpretation of t_i as the level of ambition of i's business plan may break off in the upper tail of the distribution where we may find agents whose type is so high not because of bold ambition but because of lack of entrepreneurial capacity (so that they can't be entrepreneurs in any possible city). While in our general discourse we typically interpret profits as non-decreasing in t_i, our results are independent of such a restriction, and we can always accommodate non-feasible 'dreamer' types by letting profits decline for very high t_i.

making profits. While these profits may increase steeply with the degree of ambition of the project (i.e., with the threshold t_i), they should be relatively independent of the mass of the city inhabited by the agent once the business plan is operative (i.e., given that the city mass is larger than or equal to t_i).[9] Because of increasing crowding costs, while an agent always prefers to make profits to not making profits, she will also prefer to live in the smallest available city that allows her to make profits. If an agent is unable to make profits in any city she will prefer to be in a village to minimize the crowding cost. These last two statements fully characterize the preferences that we will use in our general analysis. 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 ($m' > m \geq t_i$).

We now define the central element of our model, the distribution of ambition-types. For each possible city mass $m \in [0, a]$, we denote by $F(m) := |\{i \in A : t_i \leq m\}|$ the total mass of agents whose ambition-types are lower than or equal to m, so that they all can make profits in any city of size m or larger. The cumulative mass function (or distribution of ambition-types) $F : [0, a] \to [0, a]$ is non-decreasing by construction. We shall assume that F 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. Then, we can write $f(m) > 0$ if $m \leq \overline{m}_F$ and $f(m) = 0$ if $m > \overline{m}_F$, where $\overline{m}_F$ denotes the smallest $m \in [0, a]$ such that $F(m) = a$. Our examples of distributions of ambition-types will primarily focus on the case of $a = 1$, making use of well-known distributions from probability theory. A convenient distribution is the Beta density

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

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

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

which substantially differs from the Beta as $F(0) > 0$ and $F(1) < 1$ for all parameter configurations $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}_{++}$.

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

[9] We believe that this is a plausible simplification in a world of increasingly integrated markets where the profits of an established business can be inelastic to local demand given that the business operates on a national or global scale (as local demand becomes negligible).

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. To sharpen our welfare criteria we must impose some more structure. Let the function $\pi : \mathbb{R} \to \mathbb{R}_+$ define the potential profits of each agent depending on her ambition-type, and let the function $c : \mathbb{R}_+ \to \mathbb{R}_+$ 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 $\pi(t) > c(t)$ for all $t \in [0, a]$.[10] While it seems reasonable that π is non-decreasing as more ambitious plans are typically more profitable, we do not need this assumption for our results. Then, we can represent the preferences of each agent $i \in A$ by the utility function

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

in which $m^D_{r(i)}$ denotes the mass of the city inhabited by agent i in the urban distribution $D \in \mathcal{D}$ and $I(t_i \leq m^D_{r(i)})$ is an indicator function that takes value 1 if $t_i \leq m^D_{r(i)}$ and 0 otherwise.[11] Figure 1 is an illustration of these ideas.

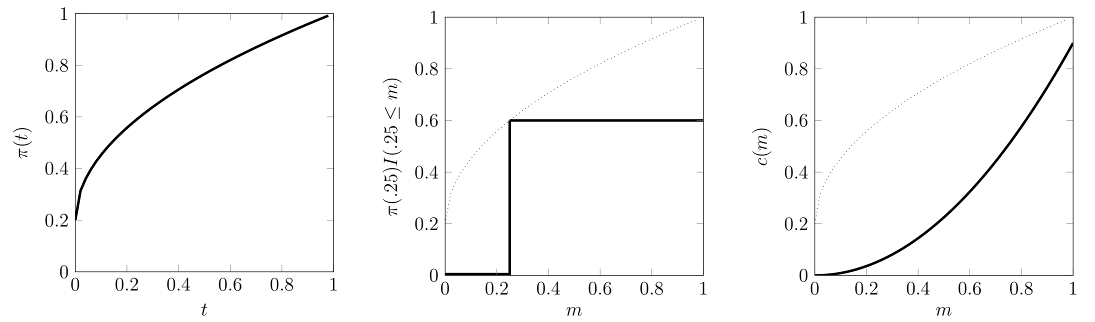

Figure 1: The solid lines in the left, central and right panels respectively represent the potential profits $\pi(t) = .2 + .8\sqrt{t}$ of an agent of ambition-type $t \in [0, 1]$, the actual profits $\pi(t)I(t \leq m)$ of an agent of ambition-type $t = .25$ in a city of size $m \in [0, 1]$, and the 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 $a = 1$.

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

$$C(D) := \sum_{k=1}^{\infty} n^D_k m^D_k c(m^D_k)$$

[10] This last 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.

[11] This parsimonious formulation of utility is chosen for tractability. A micro-foundation of the distribution of ambition-types could be achieved by adding a pregame interaction where F is the outcome of an education process of individuals choosing their ambition-types maximizing expected utility under strategic uncertainty on city size.

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 $\mathcal{K}(D)$.[12] 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

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

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

3 Equilibrium and welfare analysis

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 take the existence and size of the 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 ambition-type of an agent inhabiting a city of mass m^D_k takes a value in $\left(m^D_{k+1}, m^D_k\right]$; (ii) the ambition-type of an agent inhabiting a village takes a value in $(-\infty, 0]$ or $\left(m^D_1, +\infty\right)$. So, by assortativeness agents are segregated into cities according to their ambition 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 $F(m^D_{k+1}) = F(m^D_k) - n^D_k m^D_k$ for each rank $k \in \mathbb{N}$, which is a recurrence relation that determines the series of masses of cities $\left(m^D_k\right)^\infty_{k=1}$ given the largest city mass m^D_1 and the series of numbers of cities $\left(n^D_k\right)^\infty_{k=1}$.

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

As all equilibria have nested structure, we can represent the structure of each equilibrium graphically. Figure 2 illustrates the structures of six equilibria for the Beta distribution with parameters $(\alpha, \beta) = (2, 5)$. Each of these equilibria exhibits at most three cities and, together with the equilibrium with no cities, they fully characterize the set of equilibria in this example. All these equilibria Pareto dominate the equilibrium with no cities, each equilibrium in the bottom panels Pareto dominates the equilibrium in the top left panel, and the equilibrium in the top central

[12] This is because constant profits of each agent imply constant urbanization.

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.

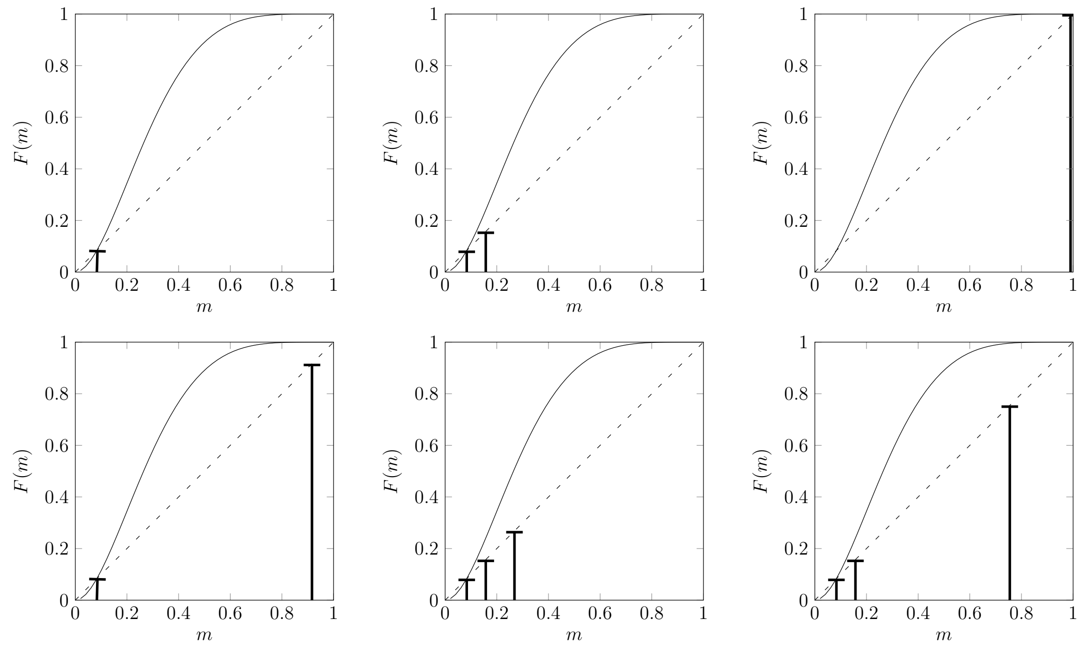

Figure 2: Given $a = 1$, $F(m)$ corresponds to the cumulative mass function of the Beta distribution with parameters $(\alpha, \beta) = (2, 5)$. Each panel depicts the nested structure of a different equilibrium, where the solid lines indicate the sizes of the various cities.

In the aforementioned example in Figure 2 all equilibria have a very limited number of cities (at most three). In reality, we typically observe a very high number of cities on the territory of a country or a region 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 opportune restrictions on the distribution of ambition-types which we consider shortly.

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 infinitely many cities and an urbanization level equal to $(F(m) - F(0))/a$. For instance, the central panel depicts an equilibrium with infinite number of cities, each of different size, where the largest size is .2, while the right panel depicts another equilibrium with 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, and that the equilibria in the central and right panels present levels of urbanization equal to each other.

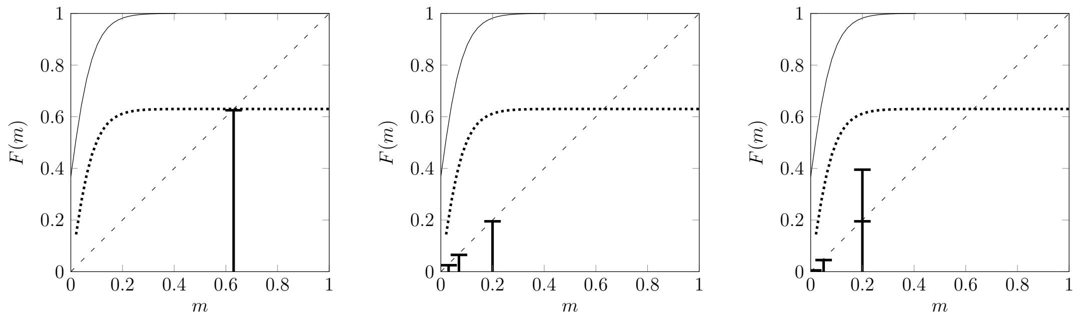

Figure 3: Given $a = 1$, $F(m)$ corresponds to the cumulative mass function of the Gumbel distribution with parameters $(\alpha, \beta) = (0, .05)$, represented by the solid curve, while the dotted curve represents $F(m) - F(0)$. Each panel depicts the nested structure of a different equilibrium, where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

One can show that there exists an equilibrium with infinite number of cities if and only if $f(0) > 1$, as this implies that there is $\epsilon > 0$ such that $m < F(m) - F(0)$ for each $m \in (0, \epsilon]$. In this spirit, we now consider a stronger condition on the distribution of ambition-types that allows to focus on equilibria with infinite number of cities for a broad set of urbanization levels.

We say that a distribution of ambition-types is non-constraining if $m < F(m) - F(0)$ for each $m \in (0, \overline{m}_F)$, which means that for each m in the pre-image of $(0, a)$ there is an excess of agents which can make profits in each city of size m and which cannot make profits in a village.

This greatly simplifies the analysis, leading to the following.

Remark 1 *Given that the distribution of ambition-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 $m \leq \overline{m}_F$ and equal to $(a - F(0))/a$ if $m > \overline{m}_F$.*

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 in Figure 2 certain equilibria Pareto dominate other equilibria because they create new cities where ex-villagers start making profits all else equal. On the other hand, while there is no Pareto dominance relation across the equilibria in the example in 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 (which implies equal profits for all agents) while having much lower urban concentration (which implies lower aggregate crowding cost). 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 $m_1^D \geq \underline{m}_F := F^{-1}\left(\max_{m\in[0,\overline{m}_F]}[F(m) - m]\right)$, 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 ambition-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 $\overline{m}_F$.*
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.*

Besides formalizing the aforementioned intuitions on the optimality of substantial and hierarchical structures, Proposition 2 provides novel insights on the connection between the upper bound $\overline{m}_F$ 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). As the distribution of ambition-types is assumed non-constraining, welfare-efficiency implies equilibrium because there is an excess of agents in the population that can make profits in a city of any size, therefore agents can always be rearranged so that there is no need to keep anyone in a city unwillingly. The reason for the upper bound $\overline{m}_F$ is best understood via the example in Figure 4, which shows that increasing the size of the largest city above $\overline{m}_F$ leaves urbanization (and the profits of each agent) unchanged while it increases urban concentration (therefore increasing the aggregate crowding cost).

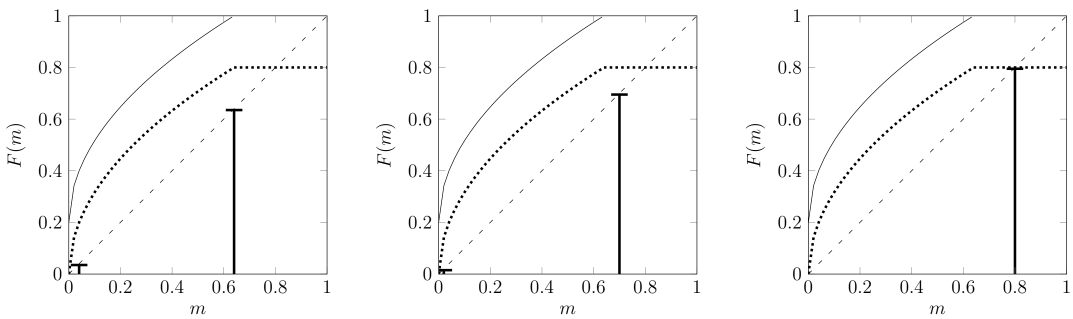

Figure 4: Given $a = 1$, $F(m) = .2 + \sqrt{m}$ corresponds to the cumulative mass function of the shifted Beta distribution with parameters $(\alpha, \beta) = (0, .5)$, represented by the solid curve, while the dotted curve represents $F(m) - F(0)$ where $\overline{m}_F = .64$. Each panel depicts the 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 ambition-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 $\mu_1 \in [\underline{m}_F, \overline{m}_F]$, the maximization of utilitarian welfare can be simply stated as

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

It is noteworthy that, on the considered domain, choosing the size of the largest city μ_1 is equivalent to choosing the level of urbanization $\mathcal{U}(D^*(\mu_1)) = (F(\mu_1) - F(0))/a$, therefore the maximization of welfare can be rewritten with respect to the level of urbanization λ, that is,

$$\max_{\lambda \in [\frac{F(\underline{m}_F)-F(0)}{a}, \frac{a-F(0)}{a}]} W(D^*(\mu_1)) = \frac{1}{a}\int_0^{\mu_1} \pi(t)dF(t) - \frac{1}{a}\sum_{k=1}^{\infty} \mu_k c(\mu_k)$$
$$\text{s.t. } \mu_1 = F^{-1}(a\lambda + F(0)),$$
$$\mu_k = F^{-1}\left(F(\mu_{k-1}) - \mu_{k-1}\right) \text{ for each } k \geq 2.$$

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 corresponding distribution of ambition-types is non-constraining. On the other hand, if the distribution of ambition-types is constraining, it is possible that no equilibrium is welfare-efficient for a given combination of cost and profit functions. For instance one can show that, for many cost and profit functions, none of the equilibria of the example in Figure 2 is welfare-efficient.

4 Comparative statics of urban development

We now consider three shocks to the fundamentals that change the qualitative properties of the distribution of ambition-types. Focusing on non-constraining distributions and cost-efficient equilibria (as the welfare-efficient urban distribution is one of them) we divide our analysis in short run and long run considerations, where the short run is defined by a fixed level of urbanization while we assume that urbanization can adjust to the welfare-efficient level in the long run.

We say that the distribution of ambition-types F is a population replication of the distribution of ambition-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 cardinality of the set of agents by a factor of k while leaving the distribution of ambition-types unchanged (in relative terms).

We say that the distribution of ambition-types F is more ambitious than (first-order stochastically dominates) the distribution of ambition-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 ambition-types are relatively more abundant in F than in F' (while low ambition-types are relatively more scarce).

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 ambition-types F is an expansion of the distribution of ambition-types F' on $[0, a]$ if each of the following conditions holds: (i) $F(t) = F'(t)$ if and only if $t \in \left\{0, \int_0^a r dF(r), a\right\}$; (ii) $\int_0^t F'(r)dr > \int_0^t F(r)dr$ for all $t \in (0, a)$; (iii) $\int_0^a r dF(r) = \int_0^a r dF'(r)$.

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

1. *If the distribution of ambition-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 ambition-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 ambition-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 in 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. The central panel considers a shift in the distribution of ambition-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 largest city is higher in the cost-efficient equilibrium of the new distribution, which implies that urban primacy is higher as predicted by Point 2. Finally, the right panel considers a shift in the distribution of ambition-types that leads the new distribution to be an expansion of the old. As shown by the dotted lines, for a fixed the level of urbanization, the size of the largest city is lower in the cost-efficient equilibrium of the new distribution than in the one of the old. Moreover, it is straightforward that this holds 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.

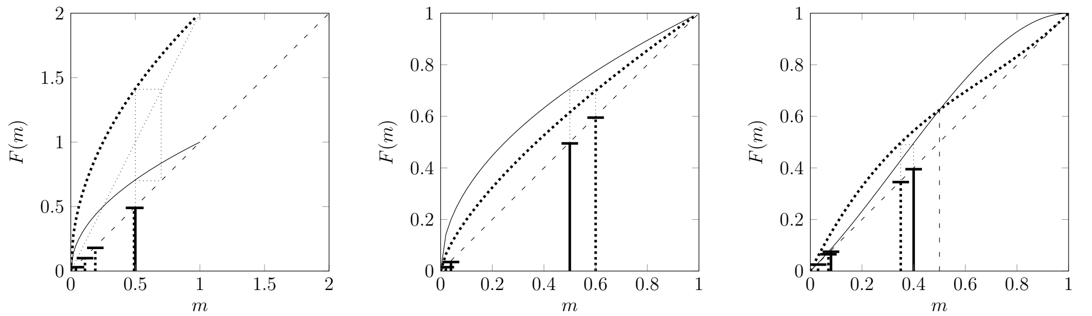

Figure 5: In the left panel the solid curve corresponds to the case $a = 1$, depicting the cumulative mass function of the Beta distribution with parameters $(\alpha, \beta) = (0, .5)$, while the dotted curve depicts a population replication that doubles the mass of agents. The central panel focuses 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 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 the second is an expansion of the first. 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.

We now consider long run trends in urban development, when the level of urbanization can adjust to the welfare-efficient level. In principle, one can always identify the optimal level of urbanization by solving the constrained maximization problem stated at the end of the previous section. However, it turns out that results crucially depend on specific assumptions on the functions F, π and c, and no clear pattern emerges. While we cannot generally predict whether urbanization increases or decreases in the long run, we can at least determine how a change in the urbanization level should affect other variables. More specifically we now show that, under fairly general conditions, our model predicts an U-shaped relation between urban primacy and the level of urbanization of a cost-efficient equilibrium. The crucial assumption behind this result is to have a density f that is single-peaked on $(0, a)$, which seems reasonable. Intuitively, business plans of intermediate ambition may be the most common, while highly or minimally ambitious plans may be relatively scarce due to the higher risks associated with their implementation and the lower returns, respectively. For instance, in an extension of our model where F is endogenously determined in a pregame interaction where individuals choose their ambition-types by maximizing expected utility under strategic uncertainty on city size, the peak of f may coincide with the ex-ante optimal type. The following remark formalizes our U-shaped prediction and Figure 6 illustrates it in an example.

Remark 2 *Let F be non-constraining and satisfying $F(0) = 0$, $F(a) \leq a$ and*

$F(m) = mf(m)$ for some $m \in (0, a)$.[13] *If the density f is single-peaked on $(0, a)$, the relationship between urban primacy and the level of urbanization of cost-efficient equilibria is U-shaped.*

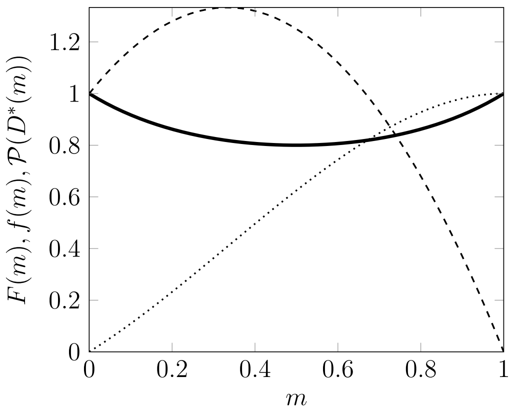

Figure 6: Given $a = 1$, the dotted, dashed and solid lines respectively depict the (non-constraining) shifted Beta distribution $F(m) = m + m^2(1 - m)$, its density function $f(m) = 1 + 2m(1 - m) - m^2$, and the level of urban primacy corresponding to the cost-efficient equilibrium with the largest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

5 An empirical pattern

As predicted by Remark 2, the scatter plot in Figure 7 suggests an U-shaped empirical relationship 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 of all countries of the world through the last 60 years.

Our analysis is similar in spirit to the seminal Imbs & 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 & Wacziarg (2003) and ours. With some caution, one may 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 & Saez (2003) and Saez & 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.

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

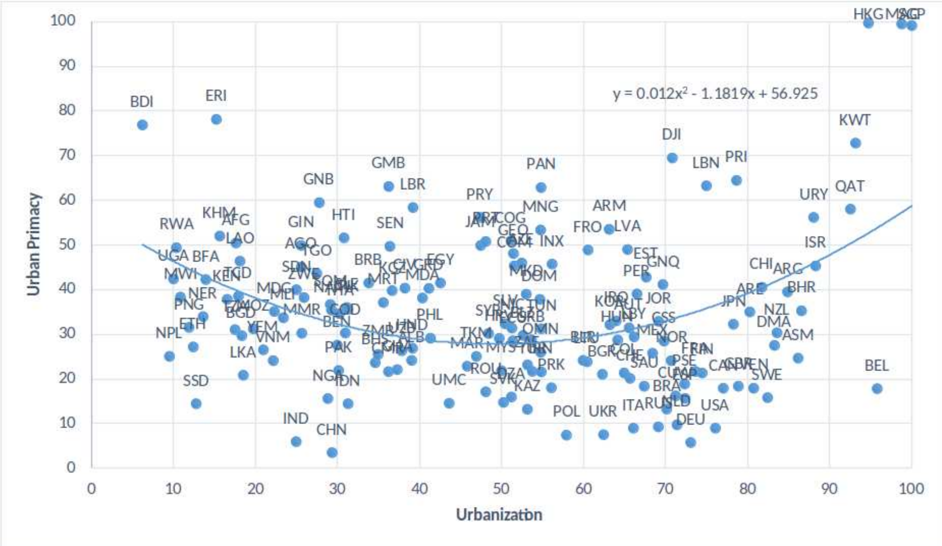

Figure 7: U-shaped cross-country relationship between the average level of urbanization and the average urban primacy, where these averages are computed within each country across the years 1950-2017. Source: Own calculations based on World Bank data.

To test the U-shaped relation empirically, we base our econometric analysis 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 1950 to 2017.[14] 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.[15] The resulting estimations are shown in Table 1. As shown in columns (1) and (2), the specifications 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, thus in line with our predictions. Most notably, the specification in column (2) with year and continent fixed effects confirms the U-shaped relation between urbanization and urban primacy predicted by our theory and suggested by the scatter plot of Figure 7. These estimations are robust to marginal changes to 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

[14] This data is publicly available from `https://data.worldbank.org/topic/urban-development`.
[15] The World Bank's dataset on Urban Development codes countries according to the continent they belong to.

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

Urban primacy	(1)	(2)	(3)
Urbanization	-.9504***	-.6149***	-.0816**
	(.0408)	(.0406)	(.0384)
Urbanization squared	.0096***	.0073***	.0011***
	(.0004)	(.0003)	(.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 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.

'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 ambition-types is generally not constant as assumed in Remark 2. 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, we exclusively focus on the corresponding subsamples within Europe and Central Asia and the world.[16] 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 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%.[17] 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 our U-shaped hypothesis.

[16] 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.'

[17] 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`.

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

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

Notes: Columns (1)-(3) and (2)-(4) respectively correspond to the subsamples of available observations for (1)-(3) Europe and Central Asia, (2)-(4) 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.

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 our U-shaped hypothesis. Arguably our handful of 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 & Wacziarg (2003) on the U-shaped relation between sectoral concentration and the level of economic development and the related findings in Piketty & Saez (2003) and Saez & Zucman (2016) on the relation between income (or wealth) inequality and GDP per capita, we believe this is sufficient to motivate our theoretical model as empirically relevant. It is well-known that the empirical measurement of urbanization and urban concentration is a difficult task due to the many degrees of freedom in defining the actual borders between cities and country side, and the World Bank's data we utilize is undeniably a very noisy signal of what is actually happening in terms of urban development. In line with these considerations, we believe that a proper assessment of the empirical validity of our U-shaped hypothesis requires a much more involved empirical effort which we prefer to leave to future research.

6 Conclusions

In this paper 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 them, where settlers in larger cities face a trade-off between higher productivity and 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 infinite number of cities of heterogeneous size. Focusing on welfare-efficient solutions 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 at fully determine how a marginal change in urbanization should affect other variables. Assuming that the distribution of ambition-types is single-peaked in the interior (so that business plans of some intermediate level of ambition are the most common), our findings suggest an U-shaped relationship between the level of urbanization and urban primacy. We find preliminary confirmation of this prediction in the data 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 view point, an interesting application would be to estimate the distribution of ambition-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 ambition-types changes across time and countries and whether these changes are broadly in line with what we know about these economies from other sources.

Appendix

Proof of Proposition 1

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 ambition-type of an agent inhabiting a city of mass m_k^D takes a value in $\left(m_{k+1}^D, m_k^D\right]$; (ii) the ambition-type of an agent inhabiting a village takes a value in $(-\infty, 0]$ or $\left(m_1^D, +\infty\right)$. 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 ambition-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 ambition-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. 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. Then, by combining these conditions we obtain $F(m_{k+1}^D) = F(m_k^D) - n_k^D m_k^D$ for each $k \in \mathbb{N}$, which concludes our proof. $\square$

Proof of Proposition 2

Assume that the distribution of ambition-types is non-constraining and let the equilibrium $D' \in \mathcal{D}$ be cost-efficient. For a contradiction, suppose that there is an urban distribution $D \in \mathcal{D}$ among the ones with the same level of urbanization that presents lower aggregate crowding cost than D' and non-lower profits for each agent. As D' is an equilibrium, all urban agents make profits and all villagers with ambition-types smaller than or equal to 0 make profits, while the remaining villagers are the only agents that do not make profits. Then, the profits of each agent are non-lower in D if and only if D is another equilibrium with equal level of urbanization. This means that, to prove that an equilibrium is cost-efficient, it is sufficient to compare it with other equilibria with same urbanization level.

We now show 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

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

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 $m \leq \overline{m}_F$ and $f(m) = 0$ if $m > \overline{m}_F$. We then divide our analysis in two cases: $m_1^{D'} < \overline{m}_F$ and $m_1^{D'} \geq \overline{m}_F$.

Consider $m_1^{D'} < \overline{m}_F$. Since $f(m_1^{D'}) > 0$, $\mathcal{U}(D') = \mathcal{U}(D)$ implies $m_1^D = m_1^{D'}$. It follows that condition (1) can be rewritten as

$$(n_k^D - 1) m_k^D c(m_k^D) + \sum_{h=k+1}^{\infty} n_h^D m_h^D c(m_h^D) \leq \sum_{h=k+1}^{\infty} m_h^{D'} c(m_h^{D'}). \tag{2}$$

Let $k' \in \mathbb{N}$ be the highest number such that $F(m_{k'}^{D'}) > F(m_{k+1}^D)$, where the existence of k' is guaranteed by our assumption that F is non-constraining. As $F(m_{k+1}^D) = F(m_k^D) - n_k^D m_k^D$ and $F(m_k^{D'}) = F(m_k^D)$, we must have $n_k^P m_k^D > F(m_k^{D'}) - F(m_{k'}^{D'})$. 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 $(n_k^D - 1) m_k^D > \sum_{h=k+1}^{k'-1} m_h^{D'}$. Then, there is $\rho \in [0, 1)$ such that

$$(n_k^D - 1) m_k^D = \sum_{h=k+1}^{k'-1} m_h^{D'} + (1 - \rho) m_{k'}^{D'}, \tag{3}$$

$$\sum_{h=k+1}^{\infty} n_h^D m_h^D = \rho m_{k'}^{D'} + \sum_{h=k'+1}^{\infty} m_h^{D'}. \tag{4}$$

Note that $m_k^D > m_h^{D'}$ for all $h \in \{k+1, \ldots, k'\}$ and $m_{k+x}^D > m_{k'+x}^{D'}$ for all $x \geq 1$, which by (3) and (4) 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 (2) 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 $m_1^{D'} < \overline{m}_F$ the equilibrium D' is cost-efficient if and only if its structure is hierarchical.

Consider $m_1^{D'} \geq \overline{m}_F$. 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 $m_1^D \in [\overline{m}_F, a - F(0)]$. We are going to show that D' is cost-efficient if and only if it is hierarchical and $m_1^{D'} = \overline{m}_F$. Suppose D satisfies such properties. Firstly, it is straightforward by arguments similar to the above that any other equilibrium $D \in \mathcal{D}$ with $m_1^D = \overline{m}_F = m_1^{D'}$ and which is non-hierarchical has higher aggregate crowding cost than D'. Secondly, suppose $m_1^{D'} = \overline{m}_F$ and let $D \in \mathcal{D}$ be any other equilibrium with $m_1^D > \overline{m}_F$. 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 (1) 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 $m_1^{D'} \leq \overline{m}_F$.

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 ambition-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 ambition-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,

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

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 $F(m_1^D) < \max_{m \in [0, \overline{m}_F]} F(m) - m$. Since F is non-constraining, there is $m' \in (m_1^D, \overline{m}_F)$ 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. □

Proof of Proposition 3

For any non-constraining distribution of ambition-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 $\mathcal{P}(D_{\lambda,F}) = m_1^{D_{\lambda,F}}/(a\lambda)$.

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 ambition 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

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

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 $\mathcal{P}(D_{\lambda,F}) > \mathcal{P}(D_{\lambda,F'})$.

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 $t \in \left\{0, \int_0^a r dF(r), a\right\}$; *(ii)* $\int_0^t F'(r)dr >$

$\int_0^t F(r)dr$ for all $t \in (0, a)$; (iii) $\int_0^a r dF(r) = \int_0^a r dF'(r)$. Given that urbanization is constant,

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

which implies $m_1^{D_{\lambda,F}} < m_1^{D_{\lambda,F'}}$ if and only if $\lambda < (F(\overline{m}) - F(0))/a$, where $\tilde{m} := \int_0^a r dF(r)$. Then, it is straightforward that

$$\mathcal{P}(D_{\lambda,F}) < \mathcal{P}(D_{\lambda,F'}) \text{ if } \lambda < (F(\tilde{m}) - F(0))/a, \text{ while}$$
$$\mathcal{P}(D_{\lambda,F}) > \mathcal{P}(D_{\lambda,F'}) \text{ if } \lambda > (F(\tilde{m}) - F(0))/a,$$

which proves the desired result, where $\lambda^* = (F(\tilde{m}) - F(0))/a$. □

References

Ades, A., & Glaeser, E. (1995). Trade and circuses: explaining urban giants. *Quarterly Journal of Economics*, *110*(1), 195–227.

Akerlof, G. (1970). The market for lemons: quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.1007/978-1-349-24002-9_9

Arthur, W. (1989). Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, *99*(394), 116–131. https://doi.org/10.2307/2234208

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: sorting, selection, and agglomeration. *Journal of Political Economy*, *122*(3), 507–553.

Behrens, K., & Robert-Nicoud, F. (2014). Survival of the fittest in cities: urbanisation and inequality. *Economic Journal*, *124*(581), 1371–1400.

Behrens, K., & Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In *Handbook of Regional and Urban Economics* (Vol. 5, pp. 171–245). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00004-0

Brock, W., & Durlauf, S. (2001). Discrete choice with social interactions. *Review of Economic Studies*, *68*(2), 235–260. https://doi.org/10.1111/1467-937X.00168

Carlino, G., & Kerr, W. (2015). Agglomeration and innovation. In *Handbook of Regional and Urban Economics* (Vol. 5, pp. 349–404). Elsevier. https://doi.org/10.1016/B978-0-444-59517-1.00006-4

Childe, V. (1950). The urban revolution. *Town Planning Review*, *21*(1), 3.

Crawford, H. (2004). *Sumer and the Sumerians*. Cambridge University Press.

Davis, J., & Henderson, J. (2003). Evidence on the political economy of the urbanization process. *Journal of Urban Economics*, *53*(1), 98–125.

Diamond, J. (1998). *Guns, germs and steel: a short history of everybody for the last 13,000 years*. Random House.

Duranton, G., & Kerr, W. R. (2018). The logic of agglomeration. In *The New Oxford Handbook of Economic Geography* (pp. 347–65). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198755609.013.14

Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of Regional and Urban Economics* (Vol. 4, pp. 2063–2117). Elsevier. https://doi.org/10.1016/S1574-0080(04)80005-1

Fujita, M., Krugman, P., & Venables, A. (2001). *The spatial economy*. MIT Press.

Genicot, G., & Ray, D. (2017). Aspirations and inequality. *Econometrica*, *85*(2), 489–519. https://doi.org/10.3982/ECTA13865

Glaeser, E. (2011). *Triumph of the city*. Penguin.

Glaeser, E., & Henderson, J. (2017). Urban economics for the developing world: an introduction. *Journal of Urban Economics*, *98*, 1–5. https://doi.org/10.1016/j.jue.2017.01.003

Glaeser, E., Sacerdote, B., & Scheinkman, J. (1996). Crime and social interactions. *Quarterly Journal of Economics*, *111*(2), 507–548. https://doi.org/10.2307/2946686

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*(6), 1420–1443. https://doi.org/10.1086/226707

Henderson, J. (1974). The sizes and types of cities. *American Economic Review*, *64*(4), 640–656.

Henderson, J., Squires, T., Storeygard, A., & Weil, D. (2018). The global distribution of economic activity: nature, history, and the role of trade. *Quarterly Journal of Economics*, *133*(1), 357–406. https://doi.org/10.1093/qje/qjx030

Hirschman, A. (1958). *The strategy of economic development*. Yale University Press.

Imbs, J., & Wacziarg, R. (2003). Linking conflict to inequality and polarization. *American Economic Review*, *93*(1), 63–86. https://doi.org/10.1257/000282803321455160

Lindbeck, A., Nyberg, S., & Weibull, J. (1999). Social norms and economic incentives in the welfare state. *Quarterly Journal of Economics*, *114*(1), 1–35. https://doi.org/10.1162/003355399555936

Marshall, A. (1920). *Principles of economics*. MacMillan.

Piketty, T., & Saez, E. (2003). Income inequality in the United States 1913–1998. *Quarterly Journal of Economics*, *118*(1), 1–41. https://doi.org/10.1162/00335530360535135

Ray, D. (2010). Uneven growth: a framework for research in development economics. *Journal of Economic Perspectives*, *24*(3), 45–60. https://doi.org/10.1257/jep.24.3.45

Redding, S. J. (2013). Economic geography: A review of the theoretical and empirical literature. In D. Bernhofen, R. Falvey, D. Greenaway, & U. Kreickemeier (Eds.), *Handbook of International Trade* (pp. 497–531). London, UK: Palgrave Macmillan. https://doi.org/10.1007/978-0-230-30531-1_16

Rosenstein-Rodan, P. (1943). Problems of industrialization of Eastern and Southeastern Europe. *Economic Journal*, *53*(2010/211), 202–211.

Saez, E., & Zucman, G. (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, *forthcoming*. https://doi.org/10.3386/w20625

Schelling, T. (1969). Models of segregation. *American Economic Review*, *59*(2), 488–493. Retrieved from `http://www.jstor.org/stable/1823701`

Simon, H. (1954). Bandwagon and underdog effects and the possibility of election predictions. *Public Opinion Quarterly*, *18*(3), 245–253. https://doi.org/10.1086/266513

United Nations. (2014). World urbanization prospects: The 2014 revision. *Department of Economic and Social Affairs, Population Division*, *(ST/ESA/SER.A/352)*. Retrieved from `http://esa.un.org/unpd/wup/index.htm`

von Thünen, J. H. (1826). *Der isolierte Staat in Beziehung auf Landschaft und Nationalökonomie, oder Untersuchungen über den Einfluss, den die Getreidepreise, der Reichthum des Bodens und die Abgaben auf den Ackerbau ausüben*. Hamburg, Deutschland: Friedrich Perthes.

Watts, D., & Dodds, P. (2009). Threshold models of social influence. In P. Bearman & P. Hedström (Eds.), *Oxford Handbook of Analytical Sociology* (pp. 475–497). Oxford, UK: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199215362.013.20