Available online at www.sciencedirect.com

SCIENCE DIRECT®

Regional Science and Urban Economics 36 (2006) 542–563

regional SCIENCE & urban ECONOMICS

www.elsevier.com/locate/econbase

Some foundations for Zipf's law: Product proliferation and local spillovers

Gilles Duranton

Department of Economics, University of Toronto, 150 Saint George Street, Toronto, Ontario, Canada M 5S 3G 7

Accepted 14 November 2005
Available online 12 June 2006

Abstract

This paper embeds the canonical model of endogenous growth with product proliferation developed by Romer [Romer, P.M., 1990. Endogenous technical change. Journal of Political Economy 98, S71–S102] into a simple urban framework. This yields a reduced form isomorphic to the popular statistical device developed by Simon [Simon, H., 1955. On a class of skew distribution functions. Biometrika 42, 425–440], which in turn can yield Zipf's law for cities. The stochastic outcomes of purposeful innovation and local spillovers can thus serve as foundations for random growth models.

JEL classification: O18; R11; R12
Keywords: City size distribution; Zipf's law; Endogenous growth; Simon's model

1. Introduction

Considerable differences in population size across cities have for a long time captured the attention of economists and geographers. Since Auerbach (1913), many have tried to approximate the distribution of city sizes with a Pareto distribution (a.k.a., a power law).[1] In this respect, the

E-mail address: gilles.duranton@utoronto.ca.
URL: http://individual.utoronto.ca/gilles/default.html.

[1] The standard approach is to rank cities in a country from the largest to the smallest and correlates this ranking with their population in the following manner:

$$\log \text{Rank} = \text{Constant} - \xi \log \text{Size}.$$

The estimated coefficient ξ then corresponds to the exponent of the Pareto distribution (a.k.a., the Zipf's exponent).

0166-0462/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2006.03.008

so-called Zipf's law has provided research with a 'useful benchmark' to think about the distribution of city sizes (Zipf, 1949). It basically states that the Pareto exponent of the distribution of city sizes is equal to unity.

The empirical validity of Zipf's law is debatable (see for instance Black and Henderson, 2003; Gabaix and Ioannides, 2004; Krugman, 1996; Soo, 2005). Even though it may be only a rough first approximation, Zipf's law nonetheless remains a useful benchmark in the following sense. Recent work by Gabaix (1999) shows how an urban system in which the growth of cities is independent of their size (i.e., urban growth satisfies Gibrat's law for the mean and variance of the growth rate) can generate Zipf's law in steady-state. Pushing the argument further, Córdoba (2003) proves that, in a growing economy, Zipf's law requires the mean of the growth rate of cities to be independent of their size (i.e., Gibrat's law for the mean must hold) and the variance to satisfy some particular condition (i.e., something less stringent than Gibrat's law for the variance). Crucially, Gabaix (1999) also shows how deviations of Zipf's law can be viewed as deviations from Gibrat's law. In a nutshell, despite its possible empirical shortcomings Zipf's law is a useful benchmark for city size distributions just like a constant natural rate of unemployment is a useful macroeconomic concept to think about the effects of macroeconomic policies on unemployment.[2]

Because of its simplicity, the model proposed by Simon (1955) remains to date a very popular way to generate Zipf's law.[3] In essence, Simon's model assumes that the urban population grows over time by discrete increments. With some probability, a new lump goes to form a new city. Otherwise it is added to an existing city. The probability that any particular city gets it is proportional to its population. This mechanism generates a Pareto distribution for city sizes. The Zipf's exponent falls to one at the limit as the probability of new cities being created goes to zero.

As an 'explanation' of Zipf's law, this model can be severely criticised on three grounds. The main criticism is that discrete population increments are postulated in a totally ad-hoc manner. There is no economic mechanism behind these shocks. This limitation is all the more serious since these shocks, and more specifically their proportionality to city size, are the main drivers of the results. With exogenous population shocks, Simon's model is simply telling us to look at changes in population to explain urban growth. This is an important step (Ioannides and Overman, 2003) but it is unlikely to be a very insightful exercise when trying to understand what ultimately drives the growth of cities. Note that this criticism applies not only to Simon's model but to random growth models in general.[4]

Second, there are some technical problems with Simon's model. As highlighted by Krugman (1996), Simon's model does not converge well and generates Zipf's law only for an infinite urban population with cities growing arbitrarily large and numerous. In a related criticism, Gabaix

[2] For an excellent review of the literature and an assessment of the empirical relevance of Zipf's law, see the recent survey of Gabaix and Ioannides (2004). For a review and some developments about the relevance of Zipf's law for firm size distributions, see Sutton (1998).

[3] However, a recent theoretical literature has instead built on Gabaix (1999). See below.

[4] By no means this implies that the recent statistical work on Zipf's law is irrelevant (Gabaix, 1999; Córdoba, 2003). Quite the contrary, this literature greatly clarifies the link between the distribution of city sizes and its proximate cause, i.e., urban population growth. In some sense, the present paper wishes to take this literature to the next step and look at the ultimate causes of urban growth, which would satisfy Zipf's law.

(1999) points out that the asymptotic variance of city growth is zero, running against empirical evidence (Ioannides and Overman, 2003).[5]

A third criticism of Simon's model is that cities play essentially no role and have no distinguishing feature. A 'city' in Simon's model is just a label tagged onto a unit of aggregation. It could well be replaced by another label such as 'region' or 'country'. But then Zipf's law is not a useful benchmark for neither regions nor countries. Hence, there is something specific about cities that is not captured by Simon's model.

The model introduced in this paper proposes some economic foundations for Zipf's law. By doing so, it also answers the first two criticisms of Simon's model and goes some way towards a response to the third. The main point of the paper is to argue that population shocks in cities can be the result of decisions made in cities by economic agents. More specifically, this paper views investments in research and development as the main driver of the growth of cities. To make this point in the least controversial manner, I use the streamlined version of Romer's (1990) endogenous growth model proposed by Grossman and Helpman's (1991, Chapter 3) and embed it into a very simple urban framework. This can generate Zipf's law under some particular condition. Under the same condition, Gibrat's law holds for the mean of the growth rate of cities but not for its variance.

The main intuition for this result is that (equalising) migration implies that the urban population is proportional to the number of differentiated varieties produced locally. Then, if investment in research and development at the city level is also proportional to the number of local firms, we get a model in which small but discrete new innovations occur in cities proportionately to their population. This model is then isomorphic to Simon's model with respect to new innovations. Over a given time period, some (lucky) cities will receive more than a proportionate share of new innovations whereas the (unlucky) others will receive less. The former will grow in relative population whereas the latter will decline. The steady-state of this model implies Zipf's law. Hence the model provides theoretically sound and empirically plausible foundations for Zipf's law.

Krugman's criticism does not apply to the model proposed here since the 'state variable' is the number of varieties produced locally rather than city population in Simon's model. That the number of varieties can grow arbitrarily large is a standard (and indeed desirable) feature of endogenous growth models. Gabaix' criticism does not apply either since my model does not in general generate a zero asymptotic variance for the growth of cities. This is because, although the development of a new variety becomes negligible with respect to the existing stock in any city, they occur more frequently because of exponential growth. Put differently, shocks to cities become smaller, as in Simon's model, but they also become more numerous.

Turning to the third criticism, it is important to note that the model relies on local spillovers, a standard feature of the modelling of cities. These local spillovers induce the location of production for new varieties to take place where they were developed. They also induce further research to co-locate with production. To derive the results in a simple fashion, the modelling of cities remains however fairly primitive and abstracts from other standard urban features such as other forms of agglomeration economies (such as urbanisation economies that take place between

[5] Arguing against the validity of Simon's model, Gabaix (1999) also highlights that it requires the number of cities to grow as fast as the population of the economy and faster than the size of a typical city. Although such a strong link between a country's demographics and its urban system is not a priori desirable, it may not be counterfactual. Controlling for education and a bunch of institutional characteristics, Henderson and Wang (2005) show that the equality between country population growth and the growth in number of cities cannot be rejected. They also show that the country population growth rates are well above the mean city growth rates.

sectors) and crowding costs. These issues are discussed in more details below. As an important caveat, note that a complete modelling of how and to which extent urban crowding and various forms of agglomeration economies not considered here may affect the distribution of city sizes is left for future work.

This paper is closely related to the recent theoretical literature on Zipf's law surveyed in Gabaix and Ioannides (2004). The work of Gabaix (1999) completed by Córdoba (2003) proposes a statistical explanation for Zipf's law. As argued above, their results are fundamental in making Zipf's law a useful benchmark. However, they propose very little economic explanation regarding the nature of the shocks and no real modelling of cities.[6] Recent work by Rossi-Hansberg and Wright (2005) and Eeckhout (2004) complement the present paper nicely. Their models are strong on the modelling of cities but the productivity shocks they propose remain ad-hoc. With respect to previous literature, the crucial advantage of the model proposed here is that the shocks are endogenous and rely on solid foundations.

This paper is also related to Duranton (2005), who builds on another growth model (the quality-ladder framework) to analyse systems of cities. The key differences between the two papers are the following. First, the focus here is to generate Zipf's law and more generally Pareto distributions of city sizes, a type of distribution that is not obtained in Duranton (2005). On the other hand, the focus of this companion paper is on the churning of industries across cities, a stylised fact that is not dealt with here. Given that the quality ladder and the product proliferation frameworks are dealing with different aspects of the growth process, these two papers must be viewed as complements rather than substitutes.

Finally, there is a large literature on urban growth building on the basic insights of urban economics. This literature is surveyed in Berliant and Wang (2005) and Henderson (2005). Like this paper, it also often views purposeful innovative activity as a key engine of urban growth. Unlike this paper and the ones mentioned above, this literature is not however concerned by Zipf's law.

The rest of the paper is as follows. The model is laid out in the next section. Section 3 derives the main results. Section 4 concludes.

2. Model

The model builds on the standard endogenous growth framework with expanding product variety developed by Romer (1990) and Grossman and Helpman (1991). This is arguably the canonical model of modern growth theory. With respect to preferences and technology, I follow Grossman and Helpman (1991, Chapter 3). The necessary adjustments to embed their model in an urban setting and generate Zipf's law are highlighted as the exposition unfolds.

2.1. Preferences

Consider a population of long-lived households whose mass is normalised to one. The instantaneous utility of the representative consumer is given by

$$u(t) \equiv \log \left[\sum_{z=1}^{n(t)} d(z,t)^{1-1/\sigma} \right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

[6] Gabaix (1999) relies on unspecified amenity shocks whereas Córdoba (2003) requires highly specific shocks on either some particular preference parameters or on the city-specific intensity of agglomeration economies.

where $d(z, t)$ is the consumption of product variety z at time t, $n(t)$ is the number of available varieties, and $\sigma(>1)$ is the elasticity of substitution between them. Total instantaneous consumption expenditure is given by

$$E(t) \equiv \sum_{z=1}^{n(t)} p(z,t)d(z,t), \tag{2}$$

where $p(z, t)$ is the price of variety z at time t.

The objective of consumers is to maximise the discounted sum of their future instantaneous utilities

$$U \equiv \int_0^\infty u(\tau)e^{-\rho\tau}d\tau, \tag{3}$$

subject to the intertemporal budget constraint

$$\int_0^\infty E(\tau)e^{-R(\tau)}d\tau \leq W(0), \tag{4}$$

where $R(\tau)$ is the cumulative interest factor between 0 and τ and $W(0)$ is the net present value of the stream of income plus the initial asset holdings.

The consumer's maximisation problem can be solved in two stages: first allocate instantaneous expenditure, $E(t)$, across varieties to maximise $u(t)$ and then choose the intertemporal allocation of expenditure. The maximisation of instantaneous utility (1) for any given level of expenditure implies the following instantaneous demand

$$d(z,t) = \frac{E(t)p(z,t)^{-\sigma}}{\sum_{z'=1}^{n} p(z',t)^{1-\sigma}}. \tag{5}$$

After defining the aggregate price index

$$P(t) \equiv \left(\sum_{z=1}^{n} p(z,t)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \tag{6}$$

and inserting (5) into (1), intertemporal utility becomes

$$U = \int_0^\infty [\log E(\tau) - \log P(\tau)]e^{\rho\tau}d\tau. \tag{7}$$

Eq. (7) can now be used to solve the optimal consumption path whose solution is characterised by

$$\dot{E}/E = \dot{R} - \rho, \tag{8}$$

together with the budget constraint and a transversality condition. After normalising total expenditure $E(t)$ to unity through the choice of numéraire, the nominal interest rate is always equal to the subjective discount rate, $\dot{R}=\rho$.

2.2. Technology

There are two activities: the manufacturing of existing varieties and the development of new blueprints, which make it possible to produce new varieties. Each blueprint is developed by a single successful innovator protected by an infinitely-lived patent.[7] Existing varieties are produced under constant returns to scale. By choice of units, manufacturing one unit of a variety requires one unit of labour. Facing the demand function (5), the objective of manufacturer of variety z is to maximise its instantaneous profits

$$\pi(z,t) = p(z,t)d(z,t) - w(t)d(z,t). \tag{9}$$

The number of varieties is large so that each monopoly is of negligible size and takes the aggregate price index $P(t)$ as given. Thus, profit maximisation implies $p(z,t) = \frac{\sigma}{\sigma-1}w(t)$, that is prices are a constant mark-up over marginal cost.[8] Since all manufacturers behave in the same way, the equilibrium aggregate price index is $P(t) = \frac{\sigma}{\sigma-1}w(t)$ and this pricing strategy implies instantaneous profits equal to

$$\pi(z,t) = \frac{1}{\sigma n(t)}. \tag{10}$$

This flow of profit is distributed continuously to shareholders as dividend. It allows them to recoup the cost of developing variety z.

Thus far the model is very similar to that of Grossman and Helpman (1991, Chapter 3); the only difference being the existence of a discrete number of varieties, instead of a continuum.[9] To remain consistent with such lumpiness, the standard framework of product proliferation needs to be amended slightly. Instead of being deterministic and smooth, the development of new blueprints is (realistically) taken to be discrete and stochastic as in the models of quality improvements.[10]

The micro-economic foundations of this research process also share some similarities with Weitzman (1998) who uses the metaphor of agricultural research where people work on existing plants and cross-pollinate them to obtain new plants. In this spirit, assume that the blueprint underlying each variety is a separate source of ideas. Re-stated, each existing blueprint constitutes a line of research that research firms can pursue in their attempts to develop new blueprints. Taken together, blueprints also constitute the general stock of knowledge, which, along with research labour, are the two factors for the production of new blueprints.

To summarise, each blueprint intervenes in three different instances in the production process. It serves first as the basis to manufacture a given variety for consumption. The rent associated with this can be fully appropriated by its innovator. Second, each blueprint constitutes a line of research which can be used by some research firms (but not all the degree of publicness of blueprints in this respect is described below). After the choice of this discrete number of varieties

[7] Rather than infinitely-lived patent, it would be equivalent to assume costly imitation and ex-post price competition.

[8] When each monopoly is not of negligible size, optimal pricing yields $p(z,t) = \frac{\sigma}{\sigma-1}w(t) + o(1/n(t))$. As will become clear below, this term in $o(1/n(t))$ only plays an asymptotically negligible role with respect to economic growth and no role whatsoever with respect to the size distribution of cities.

[9] This lumpiness plays a crucial role in the derivation of the main result because it prevents the law of large numbers from applying at the level of each city. Otherwise cities would experience parallel growth in the range of goods they produce and the distribution of city sizes would remain constant, as set by the initial conditions.

[10] See for instance Grossman and Helpman (1991, Chapter 4).

instead of a continuum, this constitutes the second difference with Grossman and Helpman's benchmark.[11] Third, each blueprint is part of the general stock of knowledge and, as such, it is a pure public good.[12]

Back to the metaphor proposed by Weitzman (1998), think of each variety as an existing species of plant used as a basis to create new hybrids. The assumptions are such that each species can be used to generate new varieties, which in turn offer potential for future development. In this light, Weitzman (1998) discusses the invention of the 'electric candle' by Edison. This key invention was generated by marrying the recently developed electric technology (a line of research in the language of this paper) with the concept of candle (arguably part of the general stock of knowledge). Such combinations of a given line of research, like for instance the polymerisation process, with existing concepts (part of the general stock of knowledge) like swimming or existing products like the cart, can lead to new products like swimming aids or tyres. Not all new combinations work however since working on polymerisation to generate food should not to go very far.

Competitive research firms face constant returns to scale. However, in each line of research, there are aggregate decreasing returns. The justification is that, although every line of research has the same potential to generate new ideas regardless of how fruitful it has been in the past, an increase in research labour on a given line of research leads to some duplication. This duplication of the research effort is viewed as a negative congestion externality that is not internalised by research firms.

In practice, each research firm starts by choosing a line of research and then performs some cross-fertilisation experiments using the general stock of knowledge. A small fraction of the experiments will turn out to be successful. Each successful experiment then leads to a new variety, generates a new line of research, and adds to the general stock of knowledge. As noted by Weitzman (1998), this type of combinatorial process can easily generate explosive economic growth since the number of potentially successful experiments increases with the square of the size of the stock of knowledge. There are two forces that prevent explosive growth. First, research firms do not co-ordinate their research effort so that some experiments are duplicated. It is also the case that experimenting requires research labour. Consequently, despite the multiplication of opportunities, the research effort spreads ever more thinly in the absence of a sustained growth in the number of researchers (which does not occur in steady-state). In turn, this limits the development of new varieties.[13]

More formally, any research firm k working on variety z and investing $\lambda^k(z)$ units of research labour for a time interval of length dt succeeds in inventing a new variety with probability $b(\lambda(z), n)\lambda^k(z)\mathrm{d}t$ where $\lambda(z)$ is the total research labour working on z. Because of duplication, the individual hazard function $b(n, \lambda(z))$ decreases with $\lambda(z) : \frac{\partial b(n,\lambda(z))}{\partial \lambda(z)} < 0$. It also increases with the total stock of knowledge, which is measured by the number of varieties $n : \frac{\partial b(n,\lambda(z))}{\partial n}$. To allow for

[11] Such an assumption would be neutral in the standard product proliferation framework. It is important here because, together with local spill-overs (introduced below), it pins down the location of research, which would otherwise be indeterminate. Note also that this idea of 'line of research', which research uses to develop new products, is a standard feature in quality-ladders models of growth (Grossman and Helpman, 1991, Chapter 4).

[12] Recall that self-sustaining growth requires the expected number of new blueprints to be proportional to the number of existing varieties. Hence, such general stock of knowledge is necessary to make growth self-sustainable with an ever expending number of varieties (Grossman and Helpman, 1991, Chapter 3).

[13] Weitzman's (1998) model predicts explosive growth. This then leads him to argue that growth is eventually limited by 'our ability to process an abundance of new ideas into usable form'—an argument not far from the one developed here.

self-sustaining and non-explosive growth, the aggregate hazard function for any line of research z is assumed to take the following functional form:

$$B(\lambda(z), n) \equiv b(\lambda(z), n)\lambda(z) \equiv \beta[\lambda(z)n]^{1-\phi}, \tag{11}$$

where β is an efficiency shifter for the innovation process and $\phi \in (0, 1)$ is the intensity of congestion in research.[14] Finally aggregating across lines of research yields the instantaneous probability of an innovation taking place in the economy

$$\iota = \sum_{z=1}^{n} B(\lambda(z), n). \tag{12}$$

This idea of decreasing returns for each line of research is a third departure from the standard product proliferation framework. Constant returns to innovation in all lines of research would make the distribution of research across varieties irrelevant. Instead, decreasing returns pin down the location of research.

For the sake of clarity, the assumptions presented here stick as closely as possible to the canonical model of product proliferation developed by Romer (1990) and simplified by Grossman and Helpman (1991): a multi-product model where firms compete and invest in research in order to reap the monopoly profits associated with new varieties. Self-sustaining and non-explosive growth is possible since new innovations are neither more difficult nor easier than past ones. This well-known model can now be embedded in the simplest possible urban setting. With firms located in different cities, uneven product development will provide the basis for population changes in cities.

2.3. Cities

The spatial organisation of population, innovation, and production must now be spelled out. There is a continuum of sites that can potentially be used as cities. Let $m(t)$ denote the (discrete) number of cities. As new cities can be created, this number can increase over time. Workers are freely mobile and final goods freely tradable across cities. Initially there are more varieties than cities and each city hosts the production of at least one variety.

Regarding research, recall that the development of a new variety requires each research firm to work on a line of research. In turn, to work on a given line of research, a research firm must locate in the same city as the firm that developed the corresponding blueprint. In other words, the public good aspect of technologies, which is captured by the term $\lambda(z)n$ in expression (11), is local. This

[14] What really matters for the main result is the existence of decreasing returns to research for each line. For simplicity, I use a simple congestion externality. A first alternative would be to use the microfoundations of search theory (Pissarides, 1979). At each point in time, the n varieties that form the stock of knowledge each constitute an opportunity within a line of research. If the $\lambda(z)$ researchers allocate themselves randomly across opportunities, some duplication is bound to take place. The expected number of explored opportunities within a line of research z is then $n[1 - (1-1/n)^{\lambda(z)}]$. Put differently, with random search there would be decreasing returns to scale in $\lambda(z)$. The expected number of innovations would then be proportional to this quantity. The expression implied by this alternative differs from (11). This would affect the computation of the growth rate but not the steady-state distribution of city sizes. As suggested by a referee, another alternative modelling strategy is to assume that only producers can conduct research on their on line of research under decreasing returns. This other alternative would complicate slightly the computation of the steady-state growth rate since profit maximisation by producers implies that ϕ would appear both multiplicatively and as a power to determine the growth rate. It would also imply a fixed number of firms.

assumption of local knowledge spill-overs implies the colocation of research and production. Turning to the global stock of knowledge (n), it is assumed to benefit research labour regardless of its location.

There is no contradiction between these two assumptions. The idea is that one may learn about the details of a technology only by observing directly how it is implemented or through small-talk with workers involved in production. Access to such tacit knowledge requires physical proximity. Note that this assumption of localised knowledge spill-overs has received ample empirical support.[15] By contrast, the global stock of knowledge is made of codified knowledge, which can be accessed by everyone from everywhere. Put differently, cities act as research laboratories in which each variety produced locally offers a line of research for local research firms that then draw on the general stock of knowledge to develop new varieties.

Any new blueprint is of one of two types: first-nature or second-nature. With a first-nature blueprint, its innovator must go to a specific location drawn from existing sites. Only at this location can the blueprint be implemented. This assumption generates some dispersion of production without having to call on more complex mechanisms such as land developers or diseconomies to city size. This assumption is justified by the fact that a new variety may sometimes require proximity to a specific natural resource. For instance, a new technology to extract oil from oil sands can only be implemented in the rare places that have an abundance of oil sands.[16] Given that there is a discrete number of (active) cities and a continuum of sites, each new first-nature blueprint implies almost surely the creation of a new city.[17,18] On the other hand, with a second-nature blueprint, the corresponding variety can only be produced where the blueprint was developed.

Each new blueprint is first-nature with probability α and second-nature with probability $1-\alpha$. Note that this first- vs. second-nature dichotomy captures the difference between industries that rely strongly on some natural resources as opposed to more footloose industries. The case where all industries are second-nature will turn out to be of particular interest.

Assume finally that any city can accommodate any number of workers at zero cost and that there is no advantage to city size with respect to production. The effects of these last two assumptions are discussed at the end of the next section. Fig. 1 summarises the assumptions of the model.

3. Steady-state

Local knowledge spill-overs directly imply that research on a blueprint and the manufacturing of the associated variety co-locate in the same city:

Lemma 1. *In equilibrium, the workers engaged in the manufacturing of a variety and those working on the corresponding line of research are located in the same city.*

[15] See Rosenthal and Strange (2004) for a review.

[16] The development of technologies to extract oil from oil sands in Canada led to the transformation of Fort McMurray (Alberta) where the presence of oil sands had been known for a long time from a village of 1000 or so inhabitants into a boomtown of around 60,000 people today. To repeat, the real discovery is the extraction process not the resource itself.

[17] The urban landscape in $t=0$ was taken as exogenous above. Nonetheless it is possible to argue that initially each existing city is defined by a sole first-nature industry.

[18] In an extension of this analysis below, I also use an alternative assumption whereby the location of production for a first-nature blueprint is randomly drawn across existing cities (and independently of their sizes) instead of leading to the creation of a new city. This may capture the idea that some industries are tied to particular natural resources in a better fashion since new first-nature varieties need not be associated with new resources.

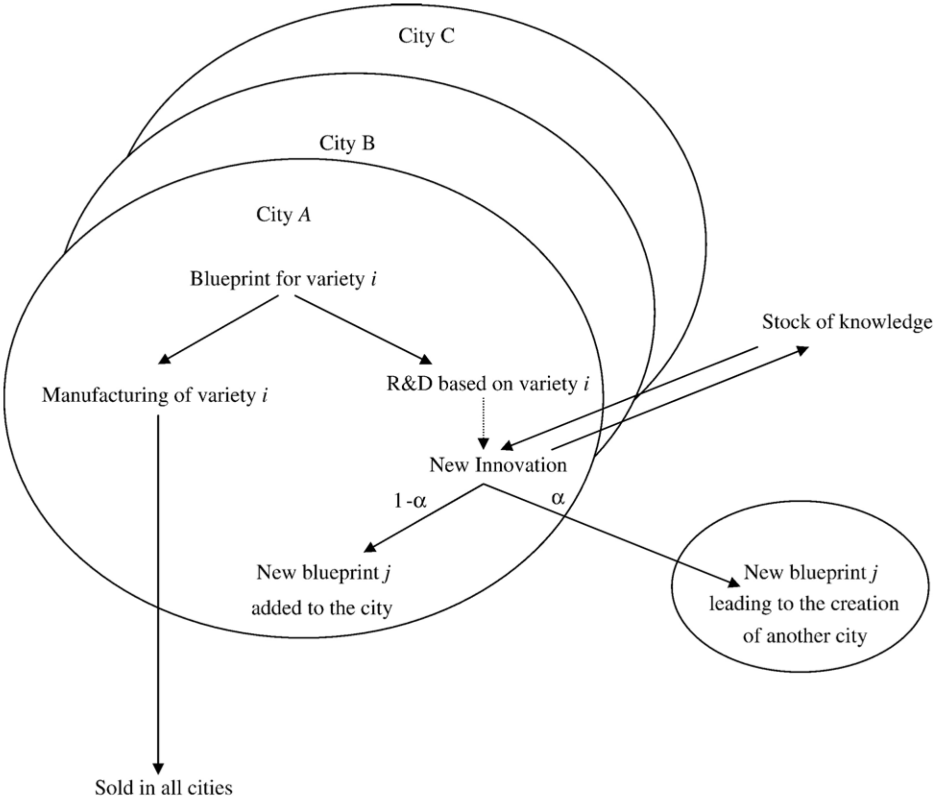

Fig. 1. Structure of the model.

Then, from Eq. (10), producers all make the same profit in equilibrium so that the present value of the uncertain profit stream is the same across varieties: $\upsilon(z, t)=\upsilon(t)$ for all z. From Eq. (10), if an innovation takes place between t and $t+\mathrm{d}t$, profits are such that $\pi(t+td)=\frac{n(t)}{n(t)+1}\pi(t)$.[19] This implies that the value of any manufacturer is scaled down by the same factor: $\upsilon(t+\mathrm{d}t)=\frac{n(t)}{n(t)+1}\upsilon(t)$. Hence, research firm k when investing $\lambda^k(z)$ units of research over $\mathrm{d}t$ at a cost $w\lambda^k(z)\mathrm{d}t$ can expect to win $b(\lambda(z),n)\times\lambda^k(z)\times\frac{n}{n+1}\upsilon\mathrm{d}t$. Then, profit maximisation by research firms implies that in equilibrium

$$w=b(\lambda(z),n)\frac{n}{n+1}\upsilon. \tag{13}$$

Inserting Eq. (11) into (13) and re-arranging yields:

$$\upsilon=\frac{n+1}{\beta n^{2-\phi}}\left[\lambda(z)\right]^{\phi}w. \tag{14}$$

After denoting Λ, total research labour, this equation implies the following lemma:

[19] As it is commonplace in this literature, the case of two or more innovations taking place between t and $t+\mathrm{d}t$ can be neglected. Formally, the probability of exactly k innovation happening over $\mathrm{d}t$ is given by $(\iota\mathrm{d}t)^k e^{-\iota\mathrm{d}t}/k!$ where ι is the aggregate probability of any research firm being successful defined by Eq. (12). The time interval $\mathrm{d}t$ can be made arbitrarily small so that the probability of two or more innovations taking place between t and $t+\mathrm{d}t$ can be neglected since it is a function of $(\mathrm{d}t)^2$ and terms of higher order.

Lemma 2. *In equilibrium, the same quantity of research labour is employed in all lines of research:*
$$\forall z, \lambda(z) = \lambda = \frac{\Lambda}{n}.$$

This lemma will prove crucial to derive the main result with respect to city-size distribution. Before turning to this, the rest of the model can be solved first.

3.1. Steady-state growth

With aggregate expenditure normalised to unity, each of the n producers sells of $\frac{1}{np}$ unit of its variety and thus employs $\frac{1}{np}$ unit of labour. With Λ unit of labour employed in research, labour market clearing implies $\Lambda + 1/p = 1$. Together with $p = \frac{\sigma}{\sigma - 1} w$, this implies $w = \frac{\sigma - 1}{\sigma}\frac{1}{1 - \Lambda}$. Inserting this last equation into (14) and using symmetry across varieties implies a first key equation relating the value of manufacturers to research employment:
$$\upsilon = \frac{(\sigma - 1)(n + 1)\Lambda^{\phi}}{\sigma\beta n^2(1 - \Lambda)}. \tag{15}$$

This optimal investment equation relates the value of producers to key parameters and variables of the model. A higher elasticity of substitution between varieties, σ, is positively related to the value of firms, υ. This is because in equilibrium, the cost of developing a new blueprint, which depends on the wage w, must be proportional to the value of firms. In turn, because of product market competition, wages are positively related to the elasticity of substitution between varieties since profits (which are negatively affected by σ) and wages must sum to unity. Hence optimal investment implies a positive relation between υ and σ. The value of firms is also negatively related to the number of varieties n. This is because a larger n implies more lines of research. In turn, this makes it easier to develop new blueprints. Since the expected returns to investing in blueprint development must equal the costs (which are independent from n), an increase in the number of investment opportunities, n, must be compensated in equilibrium by a lower value for new varieties. By the same argument, optimal investment implies that a higher efficiency parameter for innovations, β, is also negatively correlated with the value of firms, ceteris paribus. Finally, the value of firms, υ, must also be positively related to research labour since a higher research investment can only be justified by higher returns.

Turning to the stock-market valuation of firms, manufacturers pay a dividend $\pi \mathrm{d}t$ over $\mathrm{d}t$. The value of a manufacturer appreciates by $\dot{\upsilon}\mathrm{d}t$ when no research firm succeeds in developing a new variety, whereas it decreases by a factor $\frac{n}{n+1}$ in the opposite case.[20] This loss occurs with probability ι, the aggregate probability of any research firm being successful as defined in Eq. (12). Thus, with any manufacturer, the expected rate of return for a shareholder is $\frac{\pi}{\upsilon} + \frac{\dot{\upsilon}}{\upsilon} - \frac{\iota}{n+1}$. This return is stochastic but can be perfectly diversified since by Eq. (10) aggregate profit is constant and equal to $1/\sigma$. Consequently, manufacturers are valued so that their stock-market return is equal to the safe interest rate, R, which is itself equal to the subjective discount rate, ρ. Hence the absence of arbitrage implies a second key equation relating the value of manufacturers to research employment:
$$\frac{\pi}{\upsilon} + \frac{\dot{\upsilon}}{\upsilon} - \frac{\iota}{n + 1} = \rho. \tag{16}$$

[20] Again the case of two or more innovations occurring between t and $t + \mathrm{d}t$ can be neglected by ignoring the terms in $(\mathrm{d}t)^2$.

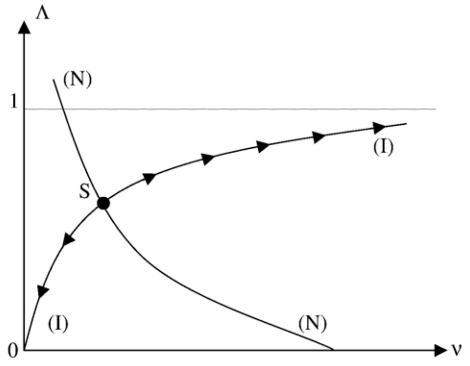

Fig. 2. Determination of the equilibrium.

Inserting Eqs. (10)–(12) into (16) implies

$$\upsilon = \left(\frac{1}{\sigma n} + \dot{\upsilon}\right)\left(\rho + \frac{n}{n+1}\beta\Lambda^{1-\phi}\right)^{-1}. \tag{17}$$

This no-arbitrage condition is the second key equation relating the value of manufacturers and research employment. A higher rate of time preference, ρ, has a negative effect on the value of manufacturers through the discounting of future profits. A higher elasticity of substitution, which affects profits negatively, also has a negative impact on the value of a manufacturer. A higher number of varieties reduces profits on the product market and thus the value of firms. A higher n also makes further increases in the number of varieties more likely. This second effect also implies a depreciation of υ. Finally, an increase in the efficiency of research also has a negative effect on the value of manufacturers since an increase in future innovations reduces future profits.

Fig. 2 depicts the evolution of the economy. The (*NN*) locus is the no-arbitrage condition (17) under $\dot{\upsilon}=0$ and the (*II*) locus is the optimal investment condition (15). First, with profit maximising research firms, Eq. (15) must always be satisfied. Hence, the economy must always lie along (*II*). If the economy is below (*NN*), this implies $\dot{\upsilon}<0$ in Eq. (16). The economy would then converge to a situation where there is no research labour and firms are worthless. This is a contradiction since in absence of innovation the value of a firm is $\frac{1}{\rho\sigma n}$. Similarly, if the economy is above (*NN*), (16) implies $\dot{\upsilon}>0$. The value of firms would then tend to infinity, which is not sustainable. Hence the economy must always be in steady-state with $\dot{\upsilon}=0$ and jump to point *S*.

The steady-state values of Λ and υ solve Eqs. (15) and (17) under $\dot{\upsilon}=0$. Λ is given by:

$$\sigma\Lambda + \frac{n}{n+1}\frac{\sigma-1}{\beta}\rho\Lambda^{\phi} - 1 = 0. \tag{18}$$

Straightforward inspection of this equation shows that Λ is unique and interior. Note finally from Eqs. (11) and (12) that the expected growth rate g is equal to $E(\imath/n)=\beta\Lambda^{1-\phi}$ where Λ is determined by Eq. (18).[21]

The comparative statics of the implicit Eq. (18) is straightforward. It indicates that aggregate research labour (and hence the rate of innovation) decreases with ρ because of discounting. It also

[21] Note that without duplication ($\phi=0$) and with a continuum of goods (i.e., $n/(n+1)\rightarrow 1$), the growth rate is given by the standard formula $g=\beta\Lambda=\frac{\beta-(\sigma-1)\rho}{\sigma}$.

decreases with σ since a higher elasticity of substitution across varieties reduces profits for manufacturers. The effect of the efficiency of innovation, β, is less obvious. On the one hand, a higher β makes research more efficient and thus reinforces the incentive to invest. On the other hand, a higher rate of innovation depreciates the value of innovations. Eq. (18) shows however that the first effect dominates. The effect of ϕ, a measure of the decreasing returns in research, is also ambiguous for similar reasons. An increase in ϕ implies more strongly decreasing returns in research and thus a lower incentive to invest. At the same time, more congestion also decreases the rate of innovation, which raises the value of existing blueprints. However, the first effect dominates so that more strongly decreasing returns imply less innovative activity.

Regarding welfare, there are four sources of inefficiency in this model. First, research firms do not take into account the surplus accruing to consumers when a greater number of varieties is available. Nor do they take into account the negative effect of new blueprints on existing profits. With CES preferences these two distortions exactly offset each other. The third distortion stems from research firms not internalising the effect of their innovations upon future innovations. Such intertemporal spill-overs imply that too little research labour is employed in equilibrium. The magnitude of this inefficiency can be shown to rise with σ (Grossman and Helpman, 1991, Chapter 3). These three inefficiencies are standard. The last inefficiency is specific to this paper. Research firms can expect to get their average and not their marginal expected returns since they do not internalise the congestion externality in research. This leads to over-investment whose magnitude increases with ϕ, the intensity of congestion. Overall the outcome is ambiguous. There may be too much or too little research in equilibrium depending on the relative values of ϕ and σ.

3.2. Steady-state city size distribution

From Eq. (10) and Lemmas 1 and 2, it is immediate that the population of a city where i blueprints are implemented is i/n. Consequently, the population of a city is exactly proportional to the number of varieties it manufactures. This holds thanks to symmetry across varieties and the co-location of the manufacturing of a variety and the associated research. As a consequence, a new innovation in a city leads to an increase in its population size while an innovation elsewhere implies a size reduction. This is because the existing workforce is divided over a larger number of varieties. Thus, a city that keeps the same number of blueprints sees its population decline as new blueprints are developed elsewhere.

The second crucial property of the model stems directly from Lemma 2. Research labour in any city is proportional to its number of blueprints and thus to its population. Hence, conditional on an innovation taking place, the probability that any particular city gets it is proportional to its population. With respect to the number of blueprints, this model is thus isomorphic to Simon (1955):

Proposition 1. (Upper tail Pareto distribution)

In steady-state, the upper tail of the distribution of city sizes follows a Pareto distribution with exponent $\frac{1}{1-\alpha}$.

A short proof of this result is as follows.[22] In steady-state, ratio of the number of cities producing i varieties, m_i, to the total number of varieties n must be constant. This ratio can change for three reasons. A city producing $i-1$ varieties may gain one blueprint leading m_i to increase by one unit. A city producing i varieties may gain one blueprint leading m_i to decrease by one unit.

[22] See also Simon (1955), Krugman (1996), and Gabaix (1999) for a derivation of the same result.

Finally any new variety increases the denominator of m_i/n. This implies the following steady-state condition:

$$E\left(\frac{\Delta(m_i/n)}{\Delta n}\right) = \frac{(1-\alpha)(i-1)m_{i-1} - (1-\alpha)im_i - m_i}{n^2} = 0. \tag{19}$$

Eq. (19) immediately yields:

$$\frac{m_i}{m_{i-1}} = \frac{(1-\alpha)(i-1)}{(1-\alpha)i+1}. \tag{20}$$

This can be re-written as $\frac{m_i m_{i-1}}{m_{i-1}} = \frac{\alpha - 2}{(1-\alpha)i+1}$. Then for i large enough, the following approximation holds:

$$\frac{m_i - m_{i-1}}{m_{i-1}/i} \cong -1 - \frac{1}{1-\alpha}. \tag{21}$$

This implies that the number of cities of size greater than i/n is proportional to $i^{-1}/(1-\alpha)$, i.e., the counter-cumulative has an exponent equal to $-1/(1-\alpha)$.[23] Put differently the number of cities producing i varieties, m_i, in the upper tail is distributed approximately like a Pareto distribution with exponent $\frac{1}{1-\alpha}$. Since, as shown above, the population of a city is proportional to its number of blueprints, the distribution of city sizes also follows a Pareto distribution with the same exponent. The case $\alpha=0$ implies immediately Zipf's law.

In equilibrium, each city gets a new variety with a probability proportional to its size. A city twice as big as another one is twice as likely to receive a new innovation. But since new innovations are of fixed-size and city population is proportional to the number of innovations implemented in that city, this implies that the expected population growth rate of the two cities is the same (actually zero since aggregate population is assumed to be fixed). Consequently the growth process faced by cities is scale-invariant. This implies that the size distribution of cities in steady-state must also be scale invariant. In other words, it must follow a Pareto distribution (or power law).

When the proportion of first-nature varieties (α) is high, a large fraction of new innovations leads to the creation of new cities. Put differently, when α is high, the rate of growth in the number of cities is large compared to the growth in the number of innovations within existing cities. This implies a low degree of skewness for the city size distribution (and thus a high Pareto exponent). As α converges to zero the steady-state distribution becomes more skewed. At the limit when α goes to zero, the steady-state converges to Zipf's law. Empirically, there are good reasons to think that α is small: only a few industries such as mining and other extractive activities are strongly tied to some particular locations.[24]

This main result shows that the canonical mechanism used to model economic growth can also be used to generate Zipf's law. It only requires the following minor modifications: innovations are discrete rather than continuous; each existing blueprint constitutes a different line of research; and there is some crowding in each line of research. These three extra features, although not part of the canonical model, have been considered in the rest of the literature (Barro and Sala-i-Martin, 2003). None of them changes anything to the qualitative properties of the basic growth process either. If anything, they add greater realism to its description.

[23] The difference between the exact value of $\frac{m_i m_{i-1}}{m_{i-1}/i}$ given by Eq. (20) and its approximation given in Eq. (21) is $\frac{2-\alpha}{(1-\alpha)[(1-\alpha)i+1]}$. For α close to zero, this implies a difference of about 0.02 for a city producing 100 varieties. Empirically, this implies a 2% difference locally for ξ, the estimated coefficient in the standard Zipf regressions (see footnote 1).

[24] Interestingly, the declining importance of industries strongly tied to specific locations may explain the decline of the Zipf's exponent, a well documented fact in many countries (Gabaix and Ioannides, 2004).

The key feature of the model is that shocks are endogenous and result from optimised behaviour by profit-maximising firms and utility-maximising consumers. More specifically, the savings of consumers are invested in research, which stochastically yields new product varieties. Taking a narrow interpretation, the growth model that is used here is mostly relevant to think about modern growth in developed economies (and thus about cities in modern developed economies). However, the heart of the model–purposeful innovation–has a much broader spatial and temporal applicability. In this respect, purposeful innovation and some key proportionalities are what matter to generate Zipf's law rather than the detailed modelling of the research process. It would be easy to show that a model of accidental growth based on, say, learning-by-doing (e.g., Romer, 1986) could also be used to generate the same results.

It is well-known that the growth model used here has the following undesirable property. An increase in overall population size leads to a higher rate of blueprint development (although here scale effects are attenuated by the duplication externality). This paper does not add anything to this debate. Two things are nonetheless worth noting. First, standard growth models, like the one used here, can be amended to get rid of scale effects at the cost of greater analytical complexity (see Temple, 2003, for an introduction to this literature and references). Second, the role of the stock of knowledge could be 'weakened' in the innovation function (11) (by elevating n to an exponent inferior to $1-\phi$) to make the model much closer to the 'semi-endogenous' growth model proposed by Jones (1995). In this case, the development of new blueprints could only be sustained with a growing population (and thus a rising workforce in research).

Empirically, the type of purposeful innovation process used here is potentially a key driver of urban growth. Anecdotal evidence suggests that the growth of San Jose (CA) in the 1990s has much to do with developments in the software industry, while the 'resurgence' of New York may be the consequence of the successful development of new financial products and business services. To my knowledge, there is unfortunately no work documenting directly and rigorously the link between local innovative activity and urban growth. Nonetheless, there is a highly significant positive correlation between population growth in cities and the proportion of university graduates, arguably a proxy for innovative activity (see Glaeser et al., 1995, and the subsequent literature).

Interestingly, it is the unevenness in the growth process that drives urban change not the level of aggregate growth. Regardless of its pace, parallel growth in new blueprints across cities would leave the distribution of population unchanged and determined only by initial conditions. Hence, the rate of growth in itself has no effect on the steady-state of the urban system.[25] Among the parameters of the model, it is only the proportion of first-nature industries (α) that plays a role to determine the steady-state distribution of city sizes. The intuition for this complete separation between the steady-state growth and the steady-state city sizes distribution is the following. What matters for the steady-state city sizes distribution is the relative level of research investment in cities, not its absolute value. Hence provided investment remains symmetric across lines of research, cities face the same expected growth, which is independent of their size. This type of growth process, regardless of the aggregate growth rate, generates the same size distribution of cities.

The other desirable property of the model is that the 'degeneracy' feature highlighted by (Krugman, 1996) does not occur with respect to population but instead with respect to the number of blueprints. That population should be allowed to grow without bound for the desired outcome to emerge is certainly a drawback for a model. That the number of patents should grow without bound was instead the chief goal of endogenous growth modelling!

[25] However, the aggregate growth rate determines first how fast the urban system converges to the steady-state and second, within steady-state, how fast cities move up and down the urban hierarchy.

This feature also answers Gabaix's (1999) critique about a related shortcoming of Simon's model. Namely, Simon's model predicts that the rate of new city creation should be as high as the population growth rate. Although this prediction is not rejected empirically (Henderson and Wang, 2005), having a theory of urban systems that relies only on demographic changes and excludes everything else certainly runs against a lot of our knowledge about cities. Because here the state variable is not population but blueprints, the present model (unlike Simon's) does not have this un-desirable property: city creation and aggregate population growth are independent.[26]

3.3. Empirical implications for the steady-state distribution of cities

An important task for future empirical research will be to assess the importance of purposeful innovation in urban growth, which is the ultimate driver of urban growth in this model. Before that, the model also offers some predictions about the proximate causes of the steady-state distribution (i.e., predictions about the growth of cities as a function of their size).

As alluded to above, Gibrat's law for the mean of the growth of cities holds.[27] To see this when $\alpha=0$, note that the expected growth rate of a city of size i can be computed as follows. Over a short time interval $\mathrm{d}t$, the probability of an innovation occurring is $gn\,\mathrm{d}t$ where g is the expected rate of growth computed from (18). The innovation can go to the city of size i under consideration with (conditional) probability i/n. In this case, this city will grow by $\frac{(i+1)/(n+1)}{i/n}-1$. Alternatively, it can go to another city with (conditional) probability $(n-i)/n$. Then, the city under consideration will grow by $\frac{i/(n+1)}{i/n}-1$. This yields:

$$E[g(i)] = gn\mathrm{d}t \times \left[\frac{i}{n}\left(\frac{(i+1)/(n+1)}{i/n}-1\right)+\frac{n-i}{n}\left(\frac{i/(n+1)}{i/n}\right)\right]. \tag{22}$$

Simple algebra then shows that this simplifies to zero:

$$E[g(i)] = 0. \tag{23}$$

Put differently: Since the probability of innovating is proportional to city size and total population is constant, the expected relative population growth rate of each and every city is zero. This implies that the growth of cities should show no correlation with their initial size.

This relation has been much examined empirically (see Gabaix and Ioannides, 2004, for a review). Although the correlation between initial size and subsequent growth is usually small (and negative), it is often found to be significant.[28] Just like with Zipf's law for the distribution, a zero correlation may be a reasonable first-order description of the growth process of cities. Nonetheless, the extensions below show that it is easy to amend the model to generate the sort of mean-reversion of city sizes usually found in empirical studies.

[26] Instead, the model requires the rate of apparition of new cities to be as large as the growth rate for new varieties (see Gabaix, 1999, for a proof). This growth rate for new varieties is one component of the growth rate in total factor productivity (the other being process innovation/quality improvement). Hence according to the model we expect total factor productivity growth to be above the growth rate of cities. According to Dobkins and Ioannides (2000), the number of US 'cities' in 1900 was 112 against 344 in 1990, which implies a growth rate of about 1% a year. This is below most estimate for US TFP growth for the period. Hence the model does not contradict the empirical reality on this aspect.

[27] This applies to established cities only. The treatment of new cities is more problematic since they enjoy an infinite growth rate upon creation.

[28] A highly significant, negative coefficient (indicating mean-reversion) is conspicuous in US data for metro areas (Black and Henderson, 2003, Duranton, 2005).

Turning to the variance of the growth of a city producing i varieties, it can be computed as:

$$\mathrm{var}(i) = \left(\frac{1}{i} + \frac{1}{n-i}\right)\left(\frac{n}{n+1}\right)^2\left(\frac{m}{m-1}\right)^2 g\mathrm{d}t. \tag{24}$$

Appendix A gives a complete derivation for this equation. Empirically, (24) implies:

$$\mathrm{var}(i) = \text{Constant} \times \left(\frac{1}{\text{Population}_i} + \frac{1}{\text{Total Population} - \text{Population}_i}\right). \tag{25}$$

There is both a cross-section and a time-series dimension to this expression. In cross-section, it is immediate to see from (24) that the variance decreases with i for $i<n/2$.[29] That the variance of city growth should decline with size is not empirically counterfactual. Duranton (2005) provides some preliminary evidence for a large cross-section of countries. Ioannides and Overman (2003) find a constant variance for small and mid-sized US cities over the 20th century but with very large confidence intervals. In time-series, the number of varieties, n, increases over time without bound. Hence the variance decreases with time since it decreases with n. However, this decline is very mild since the second part of the first term $(\frac{1}{n-i})$ is likely to be small to start with. The important result is that, unlike with Simon's model, the asymptotic variance of city growth is not zero. Again, this is because the number of new innovation increases exponentially with time.

3.4. A first deviation from Zipf's law

The assumption that each new first-nature blueprint leads to a new city is ad-hoc if not counterfactual. More realistically, it can be assumed instead that first-nature blueprints must be implemented in an existing city rather than lead to the creation of a new city. Specifically, each first-nature blueprint must be implemented in a city randomly drawn from the existing distribution of cities with equal probability.

With this alternative assumption, the ratio of the number of cities producing i varieties m_i to the total number of varieties n can now change for five reasons. As previously, a city producing $i-1$ varieties may gain one second-nature blueprint (leading m_i to increase by one unit); a city producing i varieties may gain one second-nature blueprint (leading m_i to decrease by one unit); and any new variety increases the denominator of m_i/n. Furthermore, a city producing i varieties can also now gain one first-nature blueprint (leading mi to decrease by one unit) just like a city producing $i-1$ varieties (leading instead mi to increase by one unit). Following this, we obtain a new steady-state condition replacing Eq. (19):

$$E\left(\frac{\Delta(m_i/n)}{\Delta n}\right) = \frac{(1-\alpha)(i-1)m_{i-1} - (1-\alpha)im_i - m_i - \alpha m_i\frac{n}{m} + \alpha m_{i-1}\frac{n}{m}}{n^2} = 0. \tag{26}$$

This can be re-written as

$$\frac{m_i - m_{i-1}}{m_{i-1}/i} = -1\frac{1-\frac{1}{i}\frac{\alpha n}{im}}{1-\alpha+\frac{1}{i}+\frac{\alpha n}{im}}. \tag{27}$$

[29] The fact that the variance increases with size for cities with more than half the total number of varieties is empirically irrelevant since no country with a well developed urban system has a city hosting more than half its population. It is more important to note that this declining variance property would disappear if varieties only had a finite life-time. This is because a finite life-time would limit their accumulation, and thus limit the growth of n.

This expression shows that the Zipf's exponent is no longer constant. To see this, it is useful to note first that when i tends to infinity, Eq. (27) becomes equivalent to (21). Hence for arbitrarily large cities, the Zipf's exponent of the city size distribution is locally equal to $\frac{1}{1-\alpha}$. As i decreases, the local Zipf's exponent along the distribution declines to zero. It even takes negative values below a certain threshold. This is because first-nature blueprints are allocated to cities independently of their size. Consequently, the probability over a long period of observing a city not receiving any new variety (and thus remaining very small) is lower than observing a city receiving one or more new varieties. Put differently, Eq. (27) implies a concave Zipf's curve for the distribution of city sizes in a log size–log rank plot.

This extension of the model also predicts that the growth rate of cities should be decreasing with their size since the expected growth rate of cities is a function of α/i. Fortunately, mild mean-reversal (which would imply a small α here) is also a common empirical feature as argued above. In fact, these two properties are tightly related: a concave Zipf's curve is the steady-state counterpart of mean-reversal (Gabaix, 1999).[30]

The intuition behind the two results is easy to understand. Recall that when the expected growth rate of cities is independent of their size and the expected variance is as given by Eq. (24), the steady-state distribution of city sizes follows Zipf's law as showed by Proposition 1. In the variant of the model proposed here, large cities receive proportionately fewer positive shocks. This is because the rate of innovation is proportional to city size but first-nature blueprints are uniformly distributed across all cities. Hence, the expected mean growth rate for large cities is lower than for small cities. More shocks in smaller cities also imply a greater variance for them. So with respect to the benchmark, both the growth rate and variance are lower for larger cities are larger for smaller cities. This should imply relatively less dispersion in the upper tail of the distribution and more dispersion in the lower tail compared to a Zipf distribution.

To look at the sort of magnitude involved, consider the following parameter values. Take 800 cities and $\alpha=10\%$. To loosely fit this exercise on real life data, consider an urban population of 250 million—not far from that of the US in 2000. In this case, for a hypothetical New-York (with a population around 20 million and hence a relative population of $i/n=0.08$), the predicted local Zipf's exponent is 1.24. For an hypothetical Cincinnati (of population around 2 millions), the predicted local Zipf's exponent is 1.16. For cities with population around 0.5 million, the predicted Zipf's exponent is equal to 0.94. Finally, for small cities of population 0.2 million, the predicted local Zipf's exponent is 0.62.

For the US in 2000, in a log size–log rank plot, a quadratic form fits the distribution of the population of metropolitan areas very well since it has a R^2 of 99.8%.[31] The local slope of this quadratic function can be used to approximate the real local Zipf's exponents along the distribution. For New-York, the local Zipf exponent is 1.50. For Cincinnati, it is 1.06. For cities of population 0.5 million, it is 0.81. Finally, for cities of population 0.2 million, the local Zipf's

[30] Another possible extension may imply that each first-nature industry is not tied to a specific location but to a specific industry, which for some reason (e.g., face-to-face interactions) has to be in the same city. In this case, first-nature industries would be randomly allocated to cities in proportion to their size. When deriving the steady-state as in Eq. (26), the last two terms become $-\alpha i m_i$ and $\alpha(i-1)m_{i-1}$. After simplification, this steady-state equation boils down to $m_i/m_{i-1}=(i-1)/(i+1)$. This extension would thus implies Zipf's law just like the baseline model when assuming $\alpha=0$ for the latter. I am grateful to a referee for pointing this out to me.

[31] Using data for US consolidated metropolitan areas from the US Census Bureau (2000 decennial census), the equation is $\text{logRank}=-0.2118\ \text{logPopulation}^2+1.6021\ \text{logPopulation}-0.3101$.

exponent is 0.64. Although the predicted steady-state values do not fit perfectly with the real values, the magnitudes are rather close. Note further that this simple exercise is indeed not expected to match real life data perfectly. The first reason is that the confidence intervals associated with this type of process are fairly large (Gabaix and Ioannides, 2004). Hence, the steady-state defined above is only ‘steady’ in a statistical sense. The second reason is that other features (discussed below) are expected to lead to further changes to the steady-state. Put differently, this exercise has no pretension of being a full-fledged calibration. It only aims to show that the proposed change to the model makes it empirically more relevant and not less so.

3.5. Further deviations from Zipf's law

One of the three criticisms levelled at Simon’s model in the Introduction is its lack of urban content. By contrast, the model proposed above considers some key ‘urban’ features. In particular there is a tension between an agglomeration force (knowledge spill-overs leading research to cluster with production) and a dispersion force (some natural features leading production to locate in empty sites). This type of tension, which determines the equilibrium size(s) of cities, is typical in the urban economics literature. From the perspective of this literature, the model enriches it by proposing a well-recognised source of shocks leading to a non-degenerate distribution of city sizes in steady-state.

However, to keep the modelling simple and transparent some very standard urban features were left aside. More specifically, the model lacks economies of agglomeration in production and urban crowding costs. These two features are empirically important. In their review of the literature on agglomeration economies, Rosenthal and Strange (2004) assert that a doubling of city size leads to a 3–8% increase in productivity. On the other hand, larger cities are more costly for consumers since they typically imply higher rents and longer commutes.

A full exploration of the effects of crowding costs and a richer modelling of agglomeration economies is beyond the scope of this paper.[32] Nonetheless, it is possible to propose a few conjectures on the basis of existing results. In the benchmark proposed here, innovation is proportional to city size. When all new varieties remain in the city where their blueprints were developed, Zipf’s law arises in steady-state.

Consider now economies of agglomeration in production whereby the work-force is more productive in larger cities. Then final producers face a lower wage cost per efficiency unit of labour supplied (since each worker can supply more efficiency units of labour at the same wage). Using the profit maximisation (9), it is easy to show that with agglomeration economies final producers employ more workers and are more profitable in larger cities. The first effect directly implies a more skewed distribution of city sizes. Higher profits in larger cities then imply more research labour through Eq. (13). Again, this directly increases the workforce per variety in larger cities. This also increases the probability of getting a new innovation. Put differently, with positive agglomeration economies, new innovations occur more than proportionately to city size. With the expected growth of cities increasing with city size, the steady-state distribution will be more skewed than Zipf’s law as shown by previous literature (Gabaix, 1999; Córdoba, 2003).

[32] Unfortunately a complete modelling of how and to which extent urban crowding and economies of agglomeration may affect the distribution of city sizes would involve solving an endogenous growth model in which the research sector faces costs, efficiency, and returns that vary across places and over time.

On the other hand, crowding costs increase the cost of labour per efficiency unit. In turn, this reduces employment per variety in larger cities and reduces the profitability of new innovations in larger cities. Hence with crowding costs, less research occurs in larger cities leading to fewer innovations relative to city size. Crowding costs thus achieve the opposite result of agglomeration economies in production and lead to distributions of city sizes less skewed than Zipf's law.

Ideally agglomeration economies and crowding costs should be considered in the same framework. In size regions where agglomeration economies dominate, the distribution is expected more skewed than Zipf's law, whereas in regions where crowding costs dominate the distribution is expected less skewed than Zipf's law. Considering cities of increasing sizes, it natural to expect agglomeration economies to be first more important before crowding costs start dominating. Then, this would imply a concave Zipf's curve in a log size–log rank plot. This type of shape is observed in many countries, including the US as shown above.

4. Concluding comments

This paper offers some economics for Zipf's law by proposing a model whose reduced form boils down to Simon's (1955) model. The foundations for this model are based on a canonical model of purposeful innovation generating endogenous growth (Romer, 1990; Grossman and Helpman, 1991). The model also relies on knowledge spill-overs and the opposition between footloose and resource-driven industries.

The main result is that the steady-state size distribution of cities is Pareto and Zipf's law occurs when the production of any new variety is always located where the blueprint for this variety was developed. Some extensions of the model are also considered. These extensions drive the model away from Zipf's law but arguably towards greater empirical realism.

This paper only takes some small steps towards the integration of standard features from the urban and growth literatures in models that generate realistic distribution for city sizes. In particular, the modelling of the endogenous shocks that lead cities to grow or decline is quite detailed, whereas the modelling of cities is more primitive. A richer modelling of the urban side of the model is certainly an important priority for future research in this area.

Acknowledgements

I am grateful to Antonio Ciccone, Vernon Henderson, Tomoya Mori, Pierre M. Picard, four referees, and conference and seminar participants in Manchester and Lisbon for their comments and suggestions. Financial support from the Leverhulme Trust is also gratefully acknowledged.

Appendix A. A Computation of the variance of city growth

The sample variance over dt of the growth of a city with i varieties when $\alpha = 0$ is

$$\mathrm{var}(i) = gn\frac{i}{n}\left(\frac{\frac{i+1}{n+1}}{\frac{i}{n}} - \mu_i\right)^2 \mathrm{d}t + gn\frac{n-i}{n}\left(\frac{\frac{i}{n+1}}{\frac{i}{n}} - \mu_{-i}\right)^2 \mathrm{d}t, \tag{A.1}$$

where g is the expected rate of growth computed from (18), μ_i is the sample mean when this city receives an innovation, and μ_{-i} is the sample mean when the innovation occurs in another city. The sample mean μ_i can be computed as follows:

$$\mu_i = \frac{1}{m}\left[\sum_{j\neq i}\frac{\frac{I(j)}{n+1}}{\frac{I(j)}{n}} + \frac{\frac{i+1}{n+1}}{\frac{i}{n}}\right], \tag{A.2}$$

where $I(j)$ is the number of blueprints in city j. After simplification:

$$\mu_i = \frac{n}{n+1}\left(1+\frac{1}{im}\right). \tag{A.3}$$

(Note that the sample mean when the city receives an innovation is above unity when the city is below mean size and below unity when $i>n/m$.) In turn, μ_{-i} is such that $\mu = \frac{i}{n}\mu_i + \frac{n-i}{n}\mu_{-i}$ with μ being the mean growth of cities, which is equal to $\mu = \sum_j m_j \frac{j}{n}\mu_j = 1$. Then μ_{-i} can be easily computed from the last two expressions and inserted together with (A.3) into (A.1). After simplification, the variance is equal to expression (24) given in the text.

References

Auerbach, F., 1913. Das Gesetz der Bevölkerungskonzentration. Petermanns Geographische Mitteilungen 59, 73–76.
Barro, R.J., Sala-i-Martin, X., 2003. Economic Growth, Second edition. MIT Press, Cambridge.
Berliant, M., Wang, P., 2005. Dynamic urban models: Agglomeration and growth. In: Capello, R., Nijkamp, P. (Eds.), Advances in Urban Economics. Elsevier, Amsterdam, pp. 533–581.
Black, D., Henderson, J.V., 2003. Urban evolution in the US. Journal of Economic Geography 3, 343–372.
Córdoba, J.-C., 2003. On the Distribution of City Sizes. Processed, Rice University.
Dobkins, L.H., Ioannides, Y.M., 2000. Dynamic evolution of the size distribution of US cities. In: Huriot, J.-M., Thisse, J.-F. (Eds.), Economics of Cities: Theoretical Perspectives. Cambridge University Press, Cambridge, pp. 217–260.
Duranton, G., 2005. Urban Evolutions: The Fast, the Slow, and the Still. Processed, University of Toronto.
Eeckhout, J., 2004. Zipf's law for (all) cities. American Economic Review 94, 1429–1451.
Gabaix, X., 1999. Zipf's law for cities: an explanation. Quarterly Journal of Economics 114, 739–767.
Gabaix, X., Ioannides, Y.M., 2004. The evolution of city size distributions. In: Henderson, V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. North-Holland, Amsterdam, pp. 2341–2378.
Glaeser, E.L., Scheinkman, J.A., Schleifer, A., 1995. Economic growth in a cross-section of cities. Journal of Monetary Economics 36, 117–143.
Grossman, G., Helpman, E., 1991. Innovation and Growth in the World Economy. MIT Press, Cambridge.
Henderson, J.V., 2005. Urbanization and growth. In: Aghion, P., Durlauf, S.N. (Eds.), Handbook of Economic Growth, vol. 1B. North-Holland, Amsterdam, pp. 1543–1591.
Henderson, J.V., Wang, H.G., 2005. Urbanization and City Growth: The Role of Institutions. Processed, Brown University.
Ioannides, Y.M., Overman, H.G., 2003. Zipf's law for cities: an empirical examination. Regional Science and Urban Economics 33, 127–137.
Jones, C.I., 1995. R&D based models of economic growth. Journal of Political Economy 103, 759–784.
Krugman, P., 1996. Confronting the mystery of urban hierarchy. Journal of the Japanese and International Economies 10, 1120–1171.
Pissarides, C.A., 1979. Job matchings with state employment agencies and random search. Economic Journal 89, 818–833.
Romer, P.M., 1986. Increasing returns and long-run growth. Journal of Political Economy 94, 1002–1037.
Romer, P.M., 1990. Endogenous technical change. Journal of Political Economy 98, S71–S102.
Rosenthal, S.S., Strange, W.C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics, vol. 4. North-Holland, Amsterdam, pp. 2119–2171.
Rossi-Hansberg, E., Wright, M.L.J., 2005. Urban Structure and Growth. Processed, Stanford University.

Simon, H., 1955. On a class of skew distribution functions. Biometrika 42, 425–440.
Soo, K.T., 2005. Zipf's law for cities: a cross country investigation. Regional Science and Urban Economics 35, 239–263.
Sutton, John, 1998. Technology and Market Structure: Theory and History. MIT Press, Cambridge.
Temple, J., 2003. The long-run implications of growth theories. Journal of Economic Surveys 17, 497–510.
Weitzman, M.L., 1998. Recombinant growth. Quarterly Journal of Economics 113, 331–360.
Zipf, George K., 1949. Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology. Addison-Wesley, Cambridge.