Real Estate Papers   Wharton Faculty Research

7-2006

Some Foundations for Zipf's Law: Product Proliferation and Local Spillovers

Gilles Duranton
University of Pennsylvania

Some Foundations for Zipf's Law: Product Proliferation and Local Spillovers

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.

Keywords

Disciplines

Business | Economics | Real Estate | Social and Behavioral Sciences

Comments

The postprint version of this article, *Some economics for Zipf's law: Romer and Simon unified*, was published in its final version as *Some foundations for Zipf's law: Product proliferation and local spillovers*.

Some economics for Zipf's law:

Romer and Simon unified

Gilles Duranton[§][*]

London School of Economics

13 February 2004

ABSTRACT: This paper embeds the canonical model of endogenous growth with product proliferation developed by Romer (1990) into a simple urban framework. This yields a reduced form isomorphic to the popular statistical device developed by Simon (1955), which in turn can yield Zipf's law for cities. The uncertainties and unevenness associated with the outcomes of purposeful innovation can thus serve as microeconomic foundations for random growth models.

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 by a Pareto distribution (a.k.a., a power law).[1] In this respect, the 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 one.

The empirical validity of Zipf's law is debatable (see for instance Black and Henderson, 2003; Gabaix and Ioannides, 2004; Krugman, 1996; Soo, 2002). Even though it may only be a rough first approximation, Zipf's law still remains a useful benchmark in the following sense. Recent work by Gabaix (1999) shows that an urban system in which the growth of cities is independent of their size (i.e., urban growth satisfies Gibrat's law) generates Zipf's law in steady-state. Pushing the argument further, Córdoba (2003) proves that in a statistical sense, Zipf's law must be the outcome of Gibrat's law. 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 empirical shortcomings Zipf's law is a useful benchmark for city size distributions just like the Philips curve is a useful macroeconomic tool to think about unemployment and inflation.[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, with the probability

---

[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).

[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 some recent theoretical literature has instead built on Gabaix (1999). See below.

that any particular city gets it 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 their proportionality to city size are the main drivers of the results. With such 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 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.[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 on 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 proposed in this paper introduces some economics into Zipf's law. By doing so, it proposes a (complete) answer to the first two criticisms of Simon's model

---

[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.

[5]Arguing against the validity of Simon's model, Gabaix (1999) highlights that it also 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 (2003) show that the equality between population growth and the growth in number of cities cannot be rejected. They also show that population growth is well above mean city growth.

and at least a partial answer to the third one. The main point of the paper is to argue that population shocks in cities are the results of decisions made in cities by economic agents under uncertainty. 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. In other words, this paper shows that the canonical model of modern growth theory can generate Zipf's law after a few minor modifications.

The main intuition 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 cities will receive more than a proportionate share of new innovations whereas the other 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 microeconomic foundations for Zipf's law.

Krugman's criticisms 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.

Turning to the third criticism, it is important to note that the model requires a local externality in research. At the city level, the main trade-off takes place between this local externality, which favours agglomeration, and fixed natural resources that push towards dispersion. This type of trade-off is typical of economic geography and urban models. 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 agglomeration

economies in production 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 economies of agglomeration in production (as opposed to research) may affect the distribution of city sizes is left for future work.[6]

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.[7] Recent work by Rossi-Hansberg and Wright (2003) complements the present paper nicely. Their model is strong on the modelling of cities. However the productivity shocks they propose remain ad-hoc. Hence with respect to previous literature, the crucial advantage of the model proposed here is that the shocks are endogenous and rely on solid microeconomic foundations.

In a companion paper (Duranton, 2003) I use another growth model (the standard quality-ladder framework) to generate empirically relevant city size distributions. The main difference between the two papers is that this one focuses on how to generate Zipf's law whereas the companion paper has a more empirical focus and looks at the process of allocation and re-allocation of economic activity across cities. It also takes some steps towards the assessment of the effects of agglomeration economies and crowding on the distribution of city sizes.

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 (2004) and Henderson (2004). 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.

---

[6]Hence the title '*Some* economics for the Zipf's law' instead of '*The* economics of Zipf's law'.

[7]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.

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 expending 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.

Preferences

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

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

where $d(z,t)$ is the consumption of product variety z at time t, $n(t)$ is the number of available varieties, and and σ (> 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

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

5

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

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

Equation (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, $\dot{R} = \rho$.

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.[8] Existing varieties are produced under constant returns to scale. By choice of units, manufacturing

---

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

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.[9] 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 profits 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.[10] 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 uncertain as in the models of quality improvements.[11] 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,

---

[9]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.

[10]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 city-size distribution would remain constant, as set by the initial conditions.

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

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 instead of a continuum, this constitutes the second difference with Grossman and Helpman's benchmark.[12] Third, each blueprint is part of the general stock of knowledge and, as such, it is a pure public good.[13]

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.

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. It is also assumed that a short-run increase in research labour on a variety leads to a less proportional increase in the probability of an innovation.

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 $b(\lambda(z),n)\lambda^k(z)dt$ where $\lambda(z)$ is the total research labour working on z.

---

[12]Such an assumption would play 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).

[13]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).

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: $\frac{\partial b(n,\lambda(z))}{\partial n} > 0$. To allow for 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. 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.

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 dimension of technologies, which is captured by the term $\lambda(z)n$ in expression (11), has a local dimension. This assumption of local knowledge spill-overs generates some agglomeration of research with 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 local knowledge spill-overs has received ample empirical support.[14] By contrast, the global stock of knowledge is made of codified knowledge, which can be accessed by everyone from everywhere.

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. Suppose temporarily and for simplicity that these locations are always different.[15] Thus each new first-nature blueprint implies the creation of a new city.[16] On the other hand, with a second-nature blueprint, the corresponding variety can only be produced where the blue-

---

[14]See Duranton and Puga (2004) and Rosenthal and Strange (2004) for reviews on the theoretical and empirical aspects of agglomeration economies.

[15]In the 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). This captures the idea that some industries are tied to particular natural resources in a better fashion.

[16]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.

print was developed. Each new blueprint is first-nature with probability α and second-nature with probability $1 - \alpha$. Note that this first-/second-nature dichotomy captures the difference between industries that rely strongly on some natural resources as opposed to more footloose industries. Note also that the case where all industries are second-nature is of particular interest below.

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 assumption are discussed at the end of the next Section.

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.*

From equation (10), producers all make the same profit in equilibrium so that the present value of the uncertain profit stream is the same across varieties: $v(z,t) = v(t)$ for all z. From equation (10), if an innovation takes place between t and $t + dt$, profits are such that $\pi(t + dt) = \frac{n(t)}{n(t)+1}\pi(t)$.[17] This implies that the value of any manufacturer is scaled down by the same factor: $v(t + dt) = \frac{n(t)}{n(t)+1}v(t)$. Hence, research firm k when investing $\lambda^k(z)$ units of research over dt at a cost $w\lambda^k(z)dt$ can expect to win $b(\lambda(z),n) \times \lambda^k(z) \times \frac{n}{n+1}vdt$. Then, profit maximisation by research firms implies that in equilibrium

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

---

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

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

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

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

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.

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 $\Lambda + \frac{1}{p} = 1$. Since $p = \frac{\sigma}{\sigma-1}w$, labour market clearing implies $w = \frac{\sigma-1}{\sigma}\frac{1}{1-\Lambda}$. Inserting this into equation (14) and using symmetry across varieties implies a first key equation relating the value of manufacturers to research employment:

$$v = \frac{\sigma-1}{\sigma}\frac{n+1}{\beta n^2}\frac{\Lambda^\phi}{1-\Lambda} .$$ (15)

This optimal investment equation relates the value of firm to key parameters and variables of the model. A higher elasticity of substitution between varieties, σ, is positively related to the value of firms, v. 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 v 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 easier to develop new blueprints. Since the expected returns to investing in blueprint development must be

equal to 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, v, 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 πdt over dt. The value of a manufacturer appreciates by $\dot{v}dt$ when no research firm succeeds in developing a new variety, whereas it decreases by a factor $\frac{n}{n+1}$ in the opposite case.[18] This loss occurs with probability ι, the aggregate probability of any research firm being successful as defined in equation (12). Thus, with any manufacturer, the expected rate of return for a shareholder is $\pi + \dot{v} - \iota\frac{\pi}{n+1}$. This return is risky but can be perfectly diversified since by equation (10) aggregate profit is constant and equal to $\frac{1}{\sigma}$. Consequently, manufacturers are valued so that their stock-market return is equal to the safe interest rate, $\dot{R}$, 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:

$$\pi + \dot{v} - \iota\frac{\pi}{n+1} = \rho v .\tag{16}$$

Then note that in steady-state the absence of arbitrage opportunity for investors also implies that the value of firms must remain constant between t and $t + dt$ if no new variety is developed. Inserting this together with (10), (12), and symmetry into equation (16) implies

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

This no-arbitrage condition is the second key equation relating the value of manu-facturers 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

---

[18]Again the case of two or more innovations occurring between t and $t + dt$ can be neglected by ignoring the terms in $(dt)^2$.

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 makes further increases in the number of varieties more likely. This second effect also implies a depreciation of v. 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.

The steady-state values of Λ and v solve equations (15) and (17). Λ is given by:

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

By inspection of equation (18), Λ is unique and interior. Figure 1 depicts the evolution of the economy. The (NN) locus is the no-arbitrage condition (17) and the (II) locus is the optimal investment condition (15). First with profit maximising research firms, equation (15) must always be satisfied. Hence, the economy must always lie along (II). If the economy is below (NN), this implies $\dot{v} < 0$ in equation (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 $\dot{v} > 0$. The value of firms would then tend to infinity, which is not sustainable. Hence the economy must always be in steady-state and jump to point S. Note finally from equations (11) and (12) that the expected growth rate g is equal to $E(\iota/n) = \beta\Lambda^{1-\phi}$ where Λ is determined by equation (18).

The comparative statics of the implicit equation (18) is straightforward. It indicates that aggregate research labour (and hence the rate of innovation) decreases with ρ because of discounting. It also decreases with σ since a higher elasticity of substitution across varieties reduces profits for manufacturers. The effect of the efficiency of innovation, β, is ambiguous. 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. Equation (18) shows that the first effect dominates when β is small whereas the second effect dominates when it is large. 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.
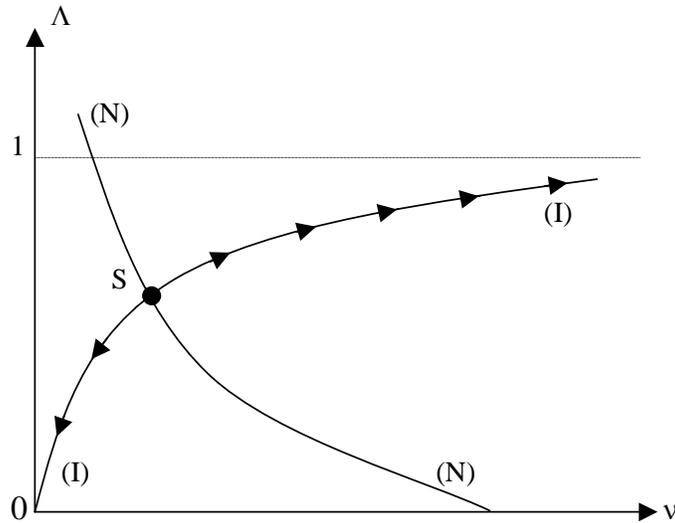
14

Figure 1. Determination of the equilibrium

At the same time, more congestion also decreases the rate of innovation, which raises the value of existing blueprints.

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 first 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.

Steady-state city size distribution

From equation (10) and Lemma 1-2, it is immediate that the population of a city where i blueprints are implemented is $\frac{i}{n}$. 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. 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.[19] 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. Finally any new variety increases the denominator of m_i/n. This implies the following steady-state condition:

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

Equation (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:

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

---

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

This implies that the number of cities of size greater than i/n is proportional to $i^{-1/(1-\alpha)}$. 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 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 and over a short time period, all cities get a new variety with a probability proportional to their 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.

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 the description.

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 risky research, which yields new product varieties. Empirically, this type of purposeful innovation process is potentially a key driver of urban growth. Although to my knowledge there is no work documenting directly and rigourously the link between local innovative activity and urban growth, there is a highly significant positive correlation between population growth in cities and the proportion of university graduates, arguably a reasonable proxy for innovative activity (see Glaeser, Scheinkman, and Schleifer, 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.[20] 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 growth process, 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

---

[20]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.

of patents should grow without bound was instead the chief goal of endogenous growth modelling!

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, 2003), 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.[21]

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 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 $i-1$ varieties can also now gain one first-nature blueprint just like a city producing i varieties. Following this we obtain a new steady-state condition

---

[21]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.

replacing equation (19):

$$E \left(\frac{\Delta(m_i/n)}{\Delta n} \right) = \frac{(1-\alpha)(i-1)m_{i-1} + \alpha m_{i-1}\frac{n}{m} - (1-\alpha)im_i - \alpha m_i\frac{n}{m} - m_i}{n^2}. \tag{22}$$

This can be re-written as

$$\frac{m_i - m_{i-1}}{m_{i-1}/i} = -1 - \frac{1 - \frac{1}{i} - \frac{\alpha n}{im}}{1 - \alpha + \frac{1}{i} + \frac{\alpha n}{im}}. \tag{23}$$

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, equation (23) 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 over a long period a city not receiving any new variety (and thus remaining very small) is more unlikely than a city receiving a few new varieties. Interestingly this variant of the model predicts some mean reversal in urban growth (the expected growth rate of cities decline with their size) and a concave Zipf's curve for the distribution of city sizes in a log size - log rank plot. These features are relatively common empirically (Gabaix and Ioannides, 2004).

The intuition behind this result is easy to understand. Recall that when the expected growth rate of cities (for both mean and variance) is independent of their size, the steady-state distribution of city sizes follows Zipf's law (Gabaix, 1999; Córdoba, 2003). 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. It is thus not surprising that compared to a Zipf distribution there should be relatively less dispersion in the upper tail of the distribution and more dispersion in the lower tail.

To look at the sort of magnitude involved, consider the following parameter values. Take 800 cities and $\alpha = 10\%$. To fit loosely 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 an 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%.[22] 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 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 further features (discussed below) are expected to lead to further changes in 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.

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 contains some urban micro-foundations. 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

---

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

leading production to locate in empty sites). This type of tension, which determines the equilibrium size(s) for cities, is typical from 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 crowding costs. The modelling of the economies of agglomeration in research is also quite primitive. A full-fledged exploration of the effects of crowding costs and a richer modelling of agglomeration economies is beyond the scope of this paper.[23] However, 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 occurs in steady-state. Consider economies of agglomeration in production whereby the workforce is more productive in larger cities. This would make production more profitable in such cities. In turn, more research would be performed there. The same result would be achieved if the productivity of research labour increased with local employment in research. In both cases, innovation becomes more than proportional to city size. When expected growth and its variance increase with city size, this generates distributions of city sizes more skewed than Zipf's law (Gabaix, 1999; Córdoba, 2003).

On the other hand, crowding costs reduce the profitability of new innovations in larger cities. Hence with crowding costs, less research would occur in larger cities leading to fewer innovations relative to city size. Crowding costs would thus achieve the opposite result and lead to distributions of city sizes less skewed than Zipf's law. Ideally agglomeration economies and crowding costs should be part of 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

---

[23]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.

less skewed than Zipf's law.[24] 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 micro-founded model whose reduced form boils down to Simon's 1955 model. The micro-economic foundations are based on a canonical model of purposeful innovation generating endogenous growth (Romer, 1990; Grossman and Helpman, 1991). The model also relies on knowledge spillovers and the opposition between footloose and resource-driven industries.

In the benchmark, the main result is that 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.

As made clear above, 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 this paper 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.

---

[24]For some steps illustrating how the trade-off between crowding costs and agglomeration economies may explain the shape of the distribution of city sizes, see Duranton (2003).

References

Auerbach, F. 1913. Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen* 59:73–76.

Barro, Robert J. and Xavier Sala-i-Martin. 2003. *Economic Growth*. Second edition. Cambridge, MA: MIT Press.

Berliant, Marcus and Ping Wang. 2004. Dynamic urban models: Agglomeration and growth. In Roberta Capello and Peter Nijkamp (eds.) *Advances in Urban Economics*. Amsterdam: Elsevier (forthcoming).

Black, Duncan and J. Vernon Henderson. 2003. Urban evolution in the US. *Journal of Economic Geography* 3(4):343–372.

Córdoba, Juan-Carlos. 2003. On the distribution of city sizes. Processed, Rice University.

Dobkins, Linda Harris and Yannis M. Ioannides. 2000. Dynamic evolution of the size distribution of US cities. In Jean-Marie Huriot and Jacques-François Thisse (eds.) *Economics of Cities: Theoretical Perspectives*. Cambridge: Cambridge University Press, 217–260.

Duranton, Gilles. 2003. City size distribution as a consequence of the growth process. Processed, London School of Economics.

Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland (forthcoming).

Gabaix, Xavier. 1999. Zipf's law for cities: an explanation. *Quarterly Journal of Economics* 114(3):739–767.

Gabaix, Xavier and Yannis M. Ioannides. 2004. The evolution of city size distributions. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland (forthcoming).

Glaeser, Edward L., José A. Scheinkman, and Andrei Schleifer. 1995. Economic growth in a cross-section of cities. *Journal of Monetary Economics* 36(1):117–143.

Grossman, Gene and Elhanan Helpman. 1991. *Innovation and Growth in the World Economy*. Cambridge, MA: MIT Press.

Henderson, J. Vernon. 2004. Urbanization, economic geography, and growth. In Philippe Aghion and Steven N. Durlauf (eds.) *Handbook of Economic Growth*, volume 4. Amsterdam: North-Holland (forthcoming).

Henderson, J. Vernon and Hyoung Gun Wang. 2003. Urbanization and city growth. Processed, Brown Univserity.

Ioannides, Yannis M. and Henry G. Overman. 2003. Zipf's law for cities: an empirical examination. *Regional Science and Urban Economics* 33(2):127–137.

Krugman, Paul. 1996. Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economies* 10(4):1120–1171.

Romer, Paul M. 1990. Endogenous technical change. *Journal of Political Economy* 98(5(2)):S71–S102.

Rosenthal, Stuart S. and William C. Strange. 2004. Evidence on the nature and sources of agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland (forthcoming).

Rossi-Hansberg, Esteban and Mark L. J. Wright. 2003. Urban structure and growth. Processed, Stanford University.

Simon, Herbert. 1955. On a class of skew distribution functions. *Biometrika* 42(2):425–440.

Soo, Kwok Tong. 2002. Zipf's law for cities: A cross country investigation. Processed, London School of Economics.

Sutton, John. 1998. *Technology and market structure: theory and history*. Cambridge, Mass.: MIT Press.

Weitzman, Martin L. 1998. Recombinant growth. *Quarterly Journal of Economics* 113(2):331–360.

Zipf, George K. 1949. *Human behavior and the principle of least effort: an introduction to human ecology*. Cambridge, Mass.: Addison Wesley.