SCALE LAW OF WELFARE WITH SIZE IN URBAN SYSTEMS: EVIDENCE AND THEORY

Francisco Martínez
Universidad de Chile
Instituto Sistemas Complejos de Ingeniería
Blanco Encalada 2002, Santiago, Chile
fmartine@ing.uchile.cl

July 29, 2011

ABSTRACT

The power laws relating nutrients with mass is a fundamental constant in living organisms and the basic relationship in the dynamics of ecological systems. That cities show an equivalent law for the dynamics social organizations is the surprising evidence emerging from empirical studies developed recently. In this paper the following question is explored: considering the known theories of discrete choice and random utility developed by McFadden, and the equilibrium conditions of auction land markets proposed by Alonso, do they provide a theoretical support to this evidence. This paper presents model of the urban system assuming rational households and firms- in all their actions in the social and economic system, competitive markets and a land auction market. All this processes are stylized assuming a Gumbel Type I distribution of utilities and profits. Under this assumption we conclude that the scale law of welfare against population that emerges after aggregating the welfare from the individuals' optimal choice processes is not a power law but a logarithm law, which increases faster than power laws within the range of current cities population size.

Key words: scaling laws, logit models, urban system dynamics, cities welfare

Total number of words: 5301

1. INTRODUCTION

The existence of a general law that governs the dynamic of all cities has been present in geographers and other urban analysis since the early evidence that cities complies with the Zipf's law (see Gabaix, 1999). More recently, however another regularity has been postulated: the power law. In several papers (Bettencourt et al., 2007; Bettencourt et al. 2008) the authors report studies of a number of cities across the developed world that support their argument that cities, like many other organisms in nature, consume and produce according to a scale or power law of the population: $y = N^{\beta}$. This result is striking because for the first time the evolution of cities resembles the basic evolutionary law of organisms, i.e. social structures replicating natural processes, or rational beings replicate the basic structure of "less intelligent" creatures. Thus, there is an important body of knowledge that can be imported from what is know about the evolution of species to help understanding the dynamic of cities. For example, that under some circumstances power laws in organisms led to catastrophic ends at critic level of development.

An interesting and relevant lesson from biology comes from the once enigmatic parameter β in organisms (see West 1999), empirically observed as a multiple of ¼ in almost all forms of life, defining metabolic rate (energy consumption), time scale (e.g. lifespan and heart rate) and sizes (aorta lengths or tree heights). Theoretical explanations followed using fractal geometry and the fact that the fractal structure at a micro scale ends in a basic dimension given by the size of the organism cell (West et al. 1999; 2001).

This result raises fundamental questions for urban research. Is there an underlying general theory of the dynamic of social organizations that explains the observed scale laws in cities? ii) If so, which are the expected evolution of cities as they develop in population, size or wealth?

In this paper I develop a theoretical explanation for a scale law in cities between welfare and population. It is based on the well-established discrete choice and random utility theories developed by McFadden and applied here to the specific conditions studied in urban economic theory. This theory needs only one basic assumption: humans are rational beings facing stochastic information of the environment, or equivalently, all agents in the city behave as to maximize a random utility. As we show, it is precisely the randomness of the optimizing behavior that reveals a power law and gives an explanation to the scale parameter beta.

The evidence

The evidence provided by Bettencourt et al. (2008) was obtained from a data of numerous urban indicators of American, European and Chinese cities and their variation was contrasted with city size using the following relation for a given indicator *y* and population (size) *N* along time *t*: $y(t) = y_0 N(t)^{\beta}$; here, y_0

represents a normalization constant that vary across cities. Their main conclusions are, first, that the scale factor β is statistically constant for the set of cities analyzed; second, that indicators of economic quantities that characterize the creation of wealth and innovation show increasing returns of scale, i.e. $\beta > 1$; third, that indicators of material infrastructure are characterized by economies of scale with size $\beta < 1$. Combined these results justify a strong tendency to concentrate population on large cities: per capita wealth increases and living costs (of infrastructure) decrease.

These results are replicated for Chilean cities, as shown in Table 1, were we have combined regional and urban data. These results replicate the existence of increasing returns on innovation indicators (wealthy population, tertiary education and research projects) and economies of scales on infrastructure indicators (road network size, total energy, city area and bus fleet). Particularly out of the range observed elsewhere is the scale parameter for research projects, with beta equal to 1,51 that reflects a potentially unsustainable condition compared with the more developed world; concentration of rich population is also notably high (1.25) and also potentially unsustainable.

Table 1: Scale parameters for Chilean cities and regions

Observation Unit	Nr of observations	Independent Variable	Parameter estimates	
			β	$\ln(y_0)$
County*	19	Roads network (km)	0.85 (9,72)	-3.80 (-3,57)
Region**	13	Total energy consumption (GWH)	0.87 (4,19)	-4.09 (-1,44)
City	148	Area (ha)	0.91 (74,41)	-3.20 (-26,35)
County	19	Urban bus fleet	0.93 (5,03)	-5.38 (-2,37)
County	19	Car fleet	1.00 (15,96)	-1.97 (-2,58)
Region	15	Residential energy consumption (GWh)	1.08 (7,17)	-8.96 (-4,43)
Region	15	Residential energy consumption (GWh)	1.08 (7,17)	-8.96 (-4,43)

Region	15	Terciary education vacancies	1.16 (8,80)	-5.61 (-3,17)
County	24	Rich Population	1.25 (12,48)	-5.79 (-4,72)
County	12	Research Projects	1.51 (5,35)	-17.42 (-4,71)

Test student t in parenthesis
* County: called municipality or "comuna".
** Region: Chile is divided in 15 regions
Source: Observatorio Urbano, Ministerio de la Vivienda y Urbanismo http://www.observatoriourbano.cl.

More evidence is needed to pursue a statistical analysis of the estimates for scales parameters, but the results already obtained clearly support the need for a theory able to explain the existence of scale laws in urban systems.

2. TOWARDS A THEORY OF SOCIAL ORGANIZATIONS

In the rest of the paper we present a theory of human organizations, which as rational beings behave maximizing a measure of satisfaction under constrained resources. In this theory we consider a large number of individual heterogeneous agents, or differentiated atoms, creating social structures out of their mutual dependency and their maximizing behavior in a spatial context.

A model of welfare scaling with city size emerges naturally from bottom-up integration of individuals' behavior, which resembles a fractal structure of the way nutrients flow in bodies down to the elemental cell. The analogy between cities and organisms is as follows: total welfare is the nutrients of cities, which in this case is endogenously produced by the city agents that represents the elemental societal unit called cells in the organism. Elemental or quantum units of welfare flow from individual agents across the city social and economic network (the body's blood system), which collects and aggregates welfare to feed the society that represents the organism's body in this analogy.

In what follows I first present the elemental behavioral model for rational agents, households and firms, based on the discrete choice and random utility theories proposed by McFadden (see Domencic and McFadden, 1975) and his followers. Then this behavior is integrated in the urban system and localized in the spatial context according to the urban economic theory applied to a set of differentiated agents under stochastic behavior.

We assume equilibrium is attained in all markets in the urban system, meaning that good and services, labor and land attain long-term equilibrium prices. Under standard urban equilibrium conditions, where all agents are allocated and floor-space densities are those that maximize utilities of households and firms, we measure the aggregated welfare of the system that, as it is demonstrated, scales with population.

Then, the specific aim is to test the hypothesis that the expected maximum welfare of the city (ω) scales with total population (N), and if the following power law of organism holds:

$$\omega = \beta \ln(N) \tag{1}$$

under the random utility discrete choice framework, were the population is an aggregation of households and firms.

2.1 The individuals' welfare function

Consider a region R partitioned in a set I of locations indexed by i, with a population of N inhabitants indexed by $n\varepsilon \mathrm{N}_R$ and partitioned into a set of households units H_R indexed by h.

Definitions: The individual n's welfare of performing an activity k, denoted ω_{nk}, with $k \in K$ and K the set of social or productive activities available in the region, is generically defined as the benefit minus the cost (including goods, services, land, etc.) necessary to perform the activity:

$$\omega_{nk} = b_{nk} - c_{nk}$$

If benefits and/or costs are differentiated by local environment conditions at location $j \in I_R$, then we define

$$\omega_{nkj} = b_{nkj} - c_{kj}$$

Consider that individual n resides at location $i \in I_R$. Then benefits are affected by transport costs (assuming travel do not induce extra benefits), such that the individuals' welfare yield by performing the activity is:

$$\omega_{nikj} = b_{nkj} - c_{kj} - c_{nij}$$

Hypothesis: All individuals are rational and make all their decisions in order to maximize welfare. Individuals' welfare are random variables subject to unpredictable shocks from idiosyncratic behavior and natural randomness on activities' conditions, then:

$$\bar{\omega}_{nikj} = b_{nkj} - c_{nikj} + \varepsilon_{nikj} = \omega_{nikj} + \varepsilon_{nikj} \tag{2}$$

with ω_{nikj} the systematic benefit and ε_{nikj} are stochastic shocks identically and independently distributed Gumbel Type I.

The justification for the assumption of independent and identical Gumbel distribution (IIG) of benefits and the basic properties of this model of behavior is resumed in the Annex. As mentioned below this assumption can be generalized to other extreme value pdf.

Individuals' welfare and their optimal choices.

Choice of spatial interactions: Conditional on the residential location of *n* at location *i* and on a given activity *k*, the expected maximum welfare of performing activities given the set of alternative locations C_{nik} where the activity can be performed is:

$$\bar{\omega}_{nik} = \frac{1}{\mu_{nik}} \ln\left(\sum_{j \in C_{nik} \subset I} \exp(\mu_{nik}\omega_{nikj}) \right) + \varepsilon_{nik} = \omega_{nik} + \varepsilon_{nik} \tag{3}$$

with ε_{nik} distributed IIG(0, μ_{nik}). Note that in equation (3) we dropped the Eulers' constant because it is innocuous in our model. The logit probability of choosing any location j_0 is:

$$P_{nikj_0} = \frac{d\omega_{nik}}{d\omega_{nikj}} = \frac{\exp(\mu_{nik}\omega_{nikj_0})}{\sum_{j \in C_{nik} \subset I} \exp(\mu_{nik}\omega_{nikj})} = \exp(\mu_{nik}(\omega_{nikj_0} - \omega_{nik})) \tag{4}$$

and the expected shares of benefits and costs by locations are*:*

$$b_{nikj} = b_{nik} \cdot \frac{d\omega_{nik}}{d\omega_{nikj}} = b_{nik} P_{nikj} \quad \text{and} \quad c_{nikj} = c_{nik} \cdot \frac{d\omega_{nik}}{d\omega_{nikj}} = c_{nik} P_{nikj}$$

Note that the subset $C_{nik} \subset I$ considers the lack of availability of activity *k* in some locations.

Choice of activities: Conditional on the residential location of *n* at location *i*, the maximum welfare yield by performing activities is:

$$\bar{\omega}_{ni} = \frac{1}{\mu_{ni}} \ln\left(\sum_{k \in C_{ni} \subset K} \exp(\mu_{ni}\omega_{nik}) \right) + \varepsilon_{ni} = \omega_{ni} + \varepsilon_{ni} \tag{5}$$

with ε_{ni} distributed IIG(0, μ_{ni}) and $C_{ni} \subset K$ the set of activities available for individual *n* with residence at *i.* The logit probability of performing an activity k_0 is:

$$P_{nik_0} = \frac{d\omega_{ni}}{d\omega_{nik_0}} = \frac{\exp(\mu_{ni}\omega_{nik_0})}{\sum_{k \in C_{ni} \subset K} \exp(\mu_{ni}\omega_{nik})} = \exp(\mu_{ni}(\omega_{nik_0} - \omega_{ni})) \tag{6}$$

and the shares of benefits and costs by activity are:

$$b_{nik} = b_{ni} \cdot \frac{d\omega_{ni}}{d\omega_{nik}} = b_{ni} P_{nik} \text{ and } c_{nik} = c_{ni} \cdot \frac{d\omega_{ni}}{d\omega_{nik}} = c_{ni} P_{nik}$$

Note that the subset $C_{ni} \subset K$ may be the result of external regulations on consumption /activities or caused by individuals constraints.

Household units: The residence choice is a collective choice among the households' members. Consider a set of individuals C_h belonging to the household unit h with $C_h \in H_R$. The maximum utility that the household unit can obtain from their members' activities, conditional on the residential location, is:

$$\bar{\omega}_{hi} = \frac{1}{\mu_h} \ln\left(\sum_{n \in h} \exp(\mu_h \omega_{ni})\right) + \varepsilon_{hi} = \omega_{hi} + \varepsilon_{hi} \tag{7}$$

with ε_{hi} distributed IIG $(0,\mu_{hi})$. The assumption in this equation is that the household valuation of individuals' welfare is identical across individual members. The distribution of welfare among households' members is given by:

$$P_{n_0 i} = \frac{\partial\omega_{hi}}{\partial\omega_{n_0 i}} = \frac{\exp(\mu_{hi}\,\omega_{n_0 i})}{\sum_{n \in h} \exp(\mu_{hi}\,\omega_{ni})} = \exp(\mu_{hi}(\omega_{n_0 i} - \omega_{hi})) \tag{8}$$

and the shares of benefits and costs by household member are:

$$b_{ni} = b_{hi} \cdot \frac{\partial\omega_{hi}}{\partial\omega_{ni}} = b_{hi} P_{ni} \quad \text{and} \quad c_{ni} = c_{hi} \cdot \frac{\partial\omega_{hi}}{\partial\omega_{ni}} = c_{hi} P_{ni}$$

Households' location: Conditional on h, the maximum welfare of living in the region R at a land costs r_i is:

$$\bar{\omega}_h = \frac{1}{\mu_h} ln\left(\sum_{i \epsilon C_i \subset I} \exp(\mu_h(\omega_{hi} - r_i))\right) + \varepsilon_h = \omega_h + \varepsilon_h \tag{9}$$

with ε_h distributed IIG(0, μ_h) and $C_h \subset I$ the set of locations available for household h after zoning regulations. The logit probability of choosing the residential location i_0 is:

$$P_{hi_0} = \frac{d\omega_h}{d\omega_{hi_0}} = \frac{\exp(\mu_h(\omega_{hi_0} - r_{i_0}))}{\sum_{i\epsilon C_i \subset I} \exp(\mu_h(\omega_{hi} - r_i))} = \exp(\mu_h(\omega_{hi} - r_i - \omega_h)) \quad (10)$$

Note that the subset $C_{hi} \subset I$ may be the result of regulations on consumption /activities or caused by individuals constraints.

Population's welfare: For any one of the *H* households in the region *R* the maximum welfare is:

$$\bar{\omega}_H = \frac{1}{\mu_H} ln\left(\sum_{i\epsilon C_i \subset I} \exp(\mu_H(\omega_{hi} - r_i))\right) + \varepsilon_H = \omega_H + \varepsilon_H \quad (11)$$

with ε_R distributed IIG(0, μ_R). The logit probability of welfare among households is:

$$P_{h_0} = \frac{d\omega_R}{d\omega_{h_0}} = \frac{\exp(\mu_R \omega_{h_0})}{\sum_{h \in H} \exp(\mu_R \omega_h)} = \exp(\mu_R(\omega_{h_0} - \omega_R)) \quad (12)$$

and the distribution of benefits and costs in the population of households is:

$$b_h = b_R \cdot \frac{d\omega_R}{d\omega_h} = b_R P_h \quad \text{and} \quad c_h = c_R \cdot \frac{d\omega_R}{d\omega_h} = c_R P_h$$

2.2 The firms' behavior

Consider again the region *R* partitioned in the set *I* of locations indexed *i*, with an economy of *N* firms, indexed by $n \in \mathbf{N}$, partitioned into a set of industries *M* indexed by *m*.

Definitions: The firm's production in industry *k*, with $k \in K$, and *K* the set of economic sectors in the region's economy is generically defined as the profit: the benefit minus the cost of the production, which is differentiated by local environment conditions at location $i \in I$, such as economies of agglomeration, and by transport cost to reach input and outputs. Then

$$\pi_{nik} = b_{nik} - c_{nik}$$

Transport costs of inputs and outputs to reach demand points (assuming travel do not induce extra benefits) are $\tilde{c}_{nik} = Min_{j \in I}\ c_{nij}(y)$, the minimum transport cost to produce and deliver the firm's optimal output *y*.

Hypothesis: All firms are rational and make all their decisions in order to maximize profit. Firms' profits are random variables distributed IIDG (0, μ_{nik}).

Therefore,

$$\tilde{\pi}_{nik} = b_{nik} - c_{nik} + \varepsilon_{nik} = \pi_{nik} + \varepsilon_{nik}$$

where $\pi_{nik} = \underset{y_k \in T_k}{Max}\, \pi(y_k) = \pi_{nik}(y^*)$, with y^* the optimum production under the technology T_k available for the production of goods k.

Notice that technology innovations expand the production possibilities and the optimum production benefit from these innovations.

Firms' profit and optimal choices:

Choice of spatial interactions or input-outputs: Conditional on the residence of firm n at location i and on a given good production k, the logit probability of demanding inputs from -and supplying outputs to- industry k_0 at location j_0, are respectively given by:

$$P^{D}_{nik-k_0 j_0} = \frac{\partial \pi_{nik}}{\partial c_{nik-k_0 j_0}} = \frac{\exp(\mu_{nik} c_{nik-k_0 j_0})}{\sum_{\substack{k' \in C_k \subseteq K \\ j \in I}} \exp(\mu_{nik} c_{nik-k' j'})} = \exp(\mu_{nik}(c_{nik-k_0 j_0} - c_{nik})) \qquad (13)$$

$$P^{S}_{nik-k_0 j_0} = \frac{\partial \pi_{nik}}{\partial b_{nik-k_0 j_0}} = \frac{\exp(\mu_{nik} b_{nik-k_0 j_0})}{\sum_{\substack{k' \in C_k \subseteq K \\ j \in I}} \exp(\mu_{nik} b_{nik-k' j'})} = \exp(\mu_{nik}(b_{nik-k_0 j_0} - b_{nik})) \qquad (14)$$

where the costs and benefits include transport delivery costs to every location j. The following shares of benefits are obtained from, and expenditure is allocated to location j:

$$b_{nikj} = b_{nik} \cdot \sum_{k'} \frac{d\pi_{nik}}{db_{nik\_k'j}} = b_{nik} \sum_{k'} P_{nik\_k'j}$$

and $$c_{nikj} = c_{nik} \cdot \sum_{k'} \frac{d\pi_{nik}}{dc_{nik-k'j}} = c_{nik} \sum_{k'} P_{nik-k'j}$$

Note that the subset $C_{nik} \subseteq I$ may be the result of the lack of availability of activity k inputs/outputs in some locations.

Distribution of production into firms: Conditional on the production of k at location i, the maximum profit from differentiated firms in set C_{nik} is:

$$\tilde{\pi}_{ki} = \frac{1}{\mu_{ki}} \ln\left(\sum_{n} \exp(\mu_{ki} \pi_{nik}) \right) + \varepsilon_{ki} = \pi_{ki} + \varepsilon_{ki} \qquad (15)$$

with ε_{ni} distributed IIG(0, μ_{ni}) and $C_{nik} \subseteq K$ the set of production options for individual n with residence at i. The logit probability of producing k_0 is:

$$P_{nik} = \frac{\square\pi_{ki}}{\square\pi_{nik}} = \frac{\exp(\mu_{ki}\pi_{nik})}{\underset{n'}{\square}\exp(\mu_{ki}\pi_{n'ik})} = \exp(\mu_{nik}(\pi_{nik} - \pi_{ki})) \qquad (16)$$

Location Choice: Conditional on product *k*, the maximum profit from producing in the region *R* after paying land rents r_i is:

$$\tilde{\pi}_k = \frac{1}{\mu_k}\ln\left(\sum_{i\epsilon C_i \subset I}\exp(\mu_k(\pi_{ki} - r_i)\right) + \varepsilon_k = \pi_k + \varepsilon_K \qquad (17)$$

with ε_k distributed G(0,μ_k) and $C_{ki} \in I$ the set of locations to produce *k.* This magnitude represents the maximum profit attainable by the production of a given product given resources (material and human) and technology available.

The logit probability of choosing the location i_0 is:

$$P_{ki_0} = \frac{d\pi_h}{d\pi_{hi_0}} = \frac{\exp(\mu_h(\pi_{ki_0} - r_{i_0}))}{\sum_{i\epsilon C_i \subset I}\exp(\mu_h(\pi_{ki} - r_i))} = \exp(\mu_h(\pi_{ki_0} - r_{i_0} - \pi_k)) \qquad (18)$$

Note that the subset $C_{ki} \in I$ may be the result of zoning regulations on production activities.

The profit yield is distributed among shareholders. The profit of shareholders residing in the city increases their income, which feeds back into the individuals' behavior; for absentee shareholders the profit, is a capital flow out of the system. In this paper we assume a static equilibrium with no inter-temporal savings and investments.

Industry's profit: For the set of production activities in the region *R* the maximum profit is:

$$\tilde{\pi}_R = \frac{1}{\mu_R} ln\left(\sum_k \exp(\mu_R \pi_k)\right) + \varepsilon_R = \pi_R + \varepsilon_R \qquad (19)$$

with ε_R distributed IIG (0,μ_R).

2.3. Section Remarks

The above model of decision making leads to the conclusion that households and firms' choice making processes can be represented by a generic agent that makes optimal choices in a large set of options of social and economic activities. Important is to notice that all these choices are interdependent: interaction occur between consumers, between suppliers, and between consumers and suppliers; hence the allocation of resources and market price signals are the result of a complex non-linear system equilibrium that is implicit in our approach. Notice particularly, that implicit in this model is the labor market, where residents produce labor supply consumed by firms at a wage rate cost.

For simplicity we adopted the Gumbel Type I assumption to model random behavior, which opens the field of modeling using other extreme value distributions, such as Type II and III or a generalized extreme value pdf. (see Mattson, et al. 2011).

3. LOCATION EQUILIBRIUM

In this section the McFadden's random utility approach is combined with Alonso's urban economic principles to allocate the urban land resource to different residential and non-residential agents. Each of these agents' behavior is assumed modeled as in the previous section conditional on the location choice; such last choice is modeled here to complete the urban system model.

3.1 The auction of land

Consider a region *R* partitioned in a set *I* of locations indexed *i*, with population of N_h households and an economy with N_f firms. Households and firms are generically called agents in the land market, and are indexed by *n*, with $N_a=N_h+N_f$ the number of agents in the region. Agents' welfare is a vector $\omega = (\omega_{ni})_{ni} = ((\omega_{hi})_{hi}, (\pi_{fi})_f;\ h = 1, \dots, N_h, f = 1, \dots, N_f, i\epsilon I)$.

From urban economics the land rent is assigned by the auctioneer of land to the maximum bidder (or willingness to pay) at each location. An important point in this process is that the set of bidders is given, or that the allocation solution is conditional on the bidders' set. Assuming a static equilibrium with **no savings over time**, the agent's maximum bid for a given location is her maximum monetary benefit accumulated from all activities conditional on residing at that location, ω_{ni}, which is distributed Gumbel Type I (0, μ_R). Then, the expected maximum bid, i.e. the land rent, is:

$$\bar{r}_i = \frac{1}{\mu_R} ln\left(\sum_{n=1,\dots,N_a} \exp(\mu_R \omega_{ni})\right) + \varepsilon_R = r_i + \varepsilon_R \qquad (21)$$

The probability of landlord at i_0 to allocate the land to an agent indexed by n_0 is the logit probability of this agent to be the best bidder, given by:

$$Q_{ni_0} = \frac{dr_i}{d\omega_{n_0 i}} = \frac{\exp\left(\mu_R(\omega_{n_0 i})\right)}{\sum_{n=1,\dots,N_a} \exp\left(\mu_R(\omega_{ni})\right)} = \exp\left(\mu_R\left(\omega_{n_0 i} - r_i\right)\right) \qquad (22)$$

After Alonso's (1964) seminal paper, the consensus on the standard urban economic theory is that the urban equilibrium condition is that "all residents are allocated somewhere". This is represented by:

$$\sum_{i\epsilon I} Q_{ni} = \sum_{i\varepsilon I} \exp\big(\mu_R(\omega_{ni} - r_i)\big) = 1 \qquad \forall n \tag{23}$$

It has been demonstrated by Martínez (1992) that at this equilibrium condition, the bid-auction location probability of equation (21) is equivalent to the household and firms choice probabilities (equations 10 and 18 respectively). This means that at equilibrium the maximum bid allocation of agents replicates their maximum utility spatial distribution if under the land auction, i.e. equation (23) holds.

Replacing the agent's welfare defined in equation (9) and (17) above we obtain:

$$\sum_{i\epsilon I} Q_{ni} = \exp(\mu_R \omega_n) = 1 \qquad \forall n \tag{24}$$

The implication of equation (24) is that the expected maximum welfare of every agent in the city –households and firms- is null, which represents a *brake even condition* $\omega_n = 0$ (because $\mu_R > 0$ or variance is not infinite). This means that all the welfare is captured through land rents by landowners. However, notice that a proportion of landowners reside in the city whose income is increased by rents. If we consider real estate property as one of the individuals' activities available, then rents naturally feed back into the real estate stakeholders' incomes; absentee landlords may be treated as exogenous agents in this system that collect their rents for use out of the system.

3.2 The welfare scale law

Consider a city with a given population *N,* composed by $N_h = aN$ household units and $N_f = eN$ firm units, all generically called agents. Each of these agents' maximum welfare is a random $\overline{\omega}_n = \omega_n + \varepsilon_n$ centered in zero, the brake even condition, but realizations of welfare is a random variable distributed IID Gumbel Type I. In order to define the total system's welfare we have to identify the composition of total population.

If we assume that the composition of the population is exogenous, that is the N individuals are exogenously identified, then total welfare is the sum of individuals' expected maximum welfare, which we know is $W_1 = 0$. In this case landowners capitalize the region's welfare and a measure of the total wealth is the spatial aggregate of rents. This assumption, however, requires that population of the city is exogenously selected from the universe of individuals by some unknown and undefined absentee agent, which is an implausible situation.

An alternative assumption is that a self-selecting process has drawn the N members of the city agents' population from the universal population of different households and firms, or in amore limited way, from the country's population of such agents. We follow this assumption hereafter. Consider that the city's population is endogenously chosen by an absentee pseudo-agent (say the invisible hand) from a large region (say the county), with the objective of maximizing the expected maximum welfare of the population; we call this the population selection process. This pseudo-agent chooses randomly one household from each the H population clusters, each one containing N_h households; the pseudo-agent performs this process a large number of tries. At each try the pseudo-agent observes the welfare realization of chosen household of each cluster and selects the household with largest welfare. The model of this process is simple if we remind that welfare variates are IID Gumbel Type I, then, the maximum expected welfare of any household chosen to live in the city by the population selection process is:

$$\overline{\omega}_H = \frac{1}{\mu_H} ln(\textstyle\sum_{h=1\ldots H} N_h \exp(\mu_H \omega_h)) + \varepsilon_H = \omega_H + \varepsilon_H \qquad (24)$$

with ε_H a Gumbel Type I variate. Additionally, the expected demography of households' types in the city population is:

$$\frac{\partial \omega_H}{\partial \omega_h} = \frac{N_h \exp(\mu_H \omega_h)}{\sum_{g=1\ldots H} N_g \exp(\mu_H \omega_g)} = P_h \qquad (25)$$

Similarly, firms' can also assumed to be selected by the same process, yielding the following maximum expected profit:

$$\bar{\pi}_F = \frac{1}{\mu_F} ln\left(\textstyle\sum_{f=1\ldots N_f} \exp(\mu_F \pi_f)\right) + \varepsilon_F = \pi_F + \varepsilon_F \qquad (26)$$

where ε_F is also Gumbel Type I variate and

$$\frac{\partial \pi_F}{\partial \pi_f} = \frac{N_f \exp(\mu_F \pi_f)}{\sum_{g=1\ldots F} N_f \exp(\mu_F \pi_f)} = P_f \qquad (27)$$

is the expected firmography of the city, i.e. the composition of firm types.

Some issues are worth noting about this population selection process.

i) The scale factors μ_H and μ_F define how diverse is the city's composition of households and firms, which in this model structurally depends on the clusters definition of households' socioeconomics and firms' business types. For example, if we consider the extreme case of only one cluster, then the allocation of population is certain: all households (firms) chosen must belong to that cluster; hence the scale factor is infinite and the population variance is null. Conversely, a large number of cluster increase variance. Therefore, specifying a larger number of clusters allows more diversity to be captured by scale factors.

ii) Two different scale factors were specified in equations 24 and 26, thus differentiating between diversity of population and firms, but they are necessarily related with the auction scale factor μ_R specified above. This theoretical relationship is a matter of further research.

iii) More importantly, the proposed population selection process describes a competition within the population to survive in the urban system, where the higher the welfare the higher the probability to survive in the competition for land, because it implies higher bids. Such probabilities are given by equations 25 and 27. Nevertheless, this selection criterion might be a matter subject to debate as a realistic assumption, because competition may occur on means different than just welfare. Therefore, this is a fundamental assumption of our model.

From 24, the aggregate expected maximum welfare of a population with size N and an average household size a^{-1} (individuals per households) is defined as:

$$\overline{W} = aN \cdot \frac{1}{\mu_H} ln\left(\sum_{h=1\ldots H} N_h \exp(\mu_H \omega_h)\right) + \varepsilon_R = W + \varepsilon_R \qquad (28)$$

Replacing equation (24) yields:

$$W = aN \cdot \left(\frac{1}{\mu_H} ln(aN)\right) = \frac{a}{\mu_H} N \cdot ln(aN) \qquad (29)$$

Similarly, the aggregate expected maximum profit of firms with an average proportion of number of firms per individual denoted by b (obtained by dividing the employment rate by the average firms' size) is:

$$\overline{\Pi} = bN \cdot \frac{1}{\mu_F} ln\left(\sum_{f=1\ldots F} \exp\left(\mu_F \pi_f\right)\right) + \varepsilon_R \qquad (30)$$

$$\Pi = \frac{b}{\mu_F} N \cdot ln(bN) \qquad (31)$$

Additionally, we know that the variance of a Gumbel variate is defined as $\sigma = \frac{\pi}{\mu\sqrt{6}}$ or $\mu^{-1} \approx 0.78\,\sigma$ (see Annex).

Then, from equations 29 and 31 we define $c = \frac{\sqrt{6}}{\pi} \approx 0.78$ and $\tilde{\sigma} = c\sigma$ to obtain the following welfare to size relationship:

$$\boldsymbol{\overline{W} = \tilde{\sigma} a N\, ln(aN) + \varepsilon} \qquad (32)$$

which represents an entropy form for an *stochastic welfare's (and profit) scale law,* which is also distributed II Gumbel Type I.

This law reads:

> *total welfare and profit of a city scales with population, with a scale law that multiplies two diversity measures: population entropy and variance.*

4. Comments and implications

The scale law combines two measures of diversity in complex systems defined by Page (2011): the variance σ, which measures intra-type variability, and the entropy $Nln(N)$, which measure size and the inter-types distribution. Then, it follows that the combined measure obtained here represents an intriguing theoretical measure of diversity in urban systems.

Notice that for $\tilde{\sigma} = 1$ equation (28) defines a super-linear welfare increase with population, because $\ln(N) \geq 1$. In any other case, for this relationships to be super-linear we require $\tilde{\sigma} \ln(N) > 1$, which is attained for some very small population; for example if $\sigma = (1.2, 1.6, 2.0)$, the respective population conditions are as low as (2.3,1.8, 1.6) respectively.

Our scale law is: $\omega = \sigma N \cdot \ln(N)$, has an entropy form, which differs from the one assumed in empirical studies: $\omega' = N^{\beta}$. To compare these laws, consider that the empirical estimates of the β parameters associated to the welfare power law, with welfare measured as GDP, obtained by Bettencourt et al. (2008) are: 1.13 (Germany 2003), 1.15 (China 2002) and 1.26 (European Union 1999-2003). Comparing the entropy law and the power laws for $\beta = (1.1, 1.2, 1.3)$ yields the conclusion that the entropy law is quicker up to a size of (10^{18}, 10^{7}, 10^{3}) inhabitants, respectively. Since current megacities have sizes larger than 10^{3} but below 10^{7}, then we conclude that the entropy law is quicker than a power law for $\beta \leq 1.2$.

The scale parameter of the Gumbel distribution emerges as the natural explanation for the observed increasing returns to scale of our theoretical scale law, indicating that two cities similar in size may differ in their welfare output: the more the variance in the system's aggregate choice process, the lager the expected total welfare. This variance comes from our conjecture about the selection process of population and firms that survive in the city after the auction of the city land, which in turn depends on the diversity of households and firms in the larger region (or country). This tells us the following story: more diverse societies are more creative; produce higher number of opportunities for interaction among agents, hence the creation of welfare units is more likely to occur for a given population. In other words, a society composed by cloned humans (or with only one activity associated to only one type of firm) is expected to provide the least welfare among those of the same size, every thing else (including culture) being equal.

Notice that the welfare scale law obtained distributes Gumbel. This point provides matter for empirical research to test this specific result and thus to support the validity of our model, which, by the way, uses extensively the asymptotic theorem for extreme distributions.

5. Comments

The paper proves social organizations follow an entropy scale law. This is the result of assuming that: there is a competitive selection process, all agents behave as rational beings facing stochastic information of alternative choices, and that the natural pdf for the choice process is the IID Gumbel Type I (or double exponential). If this holds, then the emerging total welfare of the society living in a city scales with its population size. This provides a theory that explains, in a very simple way, the empirical evidence supporting that welfare indicators scale with size (population), i.e. it explains increasing returns to scale. However, the entropy scale law theoretically obtained here differs from the power law observed empirically; it remains to test if the entropy fits well the available data.

The metaphor with organisms can be now told as follows. The welfare generated at one point in time feeds all agents -households and firms- in the society in the next period, as the nutrients do for life in organisms. Agents use these resources, including their time –as cells do in organisms- to maximize welfare production in the next period, producing goods and labor, relocating activities and setting new ones at the time when other activities die. A renovated system equilibrium emerges in the next period, with a flow of welfare produced by each agent that at the aggregate complies with the stochastic scale law conditional on the system diversity and population. In cities the interactions among agents occur in markets and in social organizations in a way that allow scientists to observe their intensity and productivity, for example by market prices. According to biologists, system-wide interactions between cells in organisms are also well known.

The connection between organisms and cities dynamics is simple: both share the assumption of an optimization behavior, e.g. to save energy, or more generally scarce resources. In the case of social systems the assumption of the utility maximizing behavior allows the optimization in a multiple dimensional set of resources, and also allows the explicit representation of constraints, including organizational (regulations) and resources (including time and material). Additionally, although it is implicit in the theory presented, this framework also assumes equilibrium conditions in multiple markets, for goods and for interactions between agents (agglomeration economies and social networks), which are already represented in some urban equilibrium models (see Martínez and Donoso, 2010).

There are several ways to complement and complete this theory; some of them has been mentioned above. For example, the extension to other extreme value pdf models; other assumptions to model the selection process of population and firms; clarify the relationship between different Gumbel scale factors, particularly in the auction and population selection processes; make explicit the relationship between cities with similar sizes but different cultures or levels of development. However, the above logit-based framework can provide answers to some of these issues.

Our theory does not explain economies of scale on physical resources (e.g. electricity or road networks sizes). Indeed this deserves future analysis, but this is a more direct consequence of the city area, which is in turn a consequence of non-constant location densities yield by the urban static equilibrium where agents agglomerate each other to reduce transport costs and to benefit from agglomeration economies. This argument questions the independency of the beta parameters presented in the introduction section, which raises the fundamental question of how many degrees of freedom describe the urban system? Thus, a direct research question in our theory is if it is possible to derive some (or all) of the other empirical scale laws from the theoretical welfare law proposed above.

In the above theory the role of the Gumbel distribution in the choice making process is essential. Based on theoretical arguments we have considered this assumption for all micro choices of all activities, but it is worth noting that this level of detail is not necessary to obtain the welfare scale law. In fact it is enough to assume that the land market (Section 3.1) performs auctions built from Gumbel bids of all participant agents. Hence, we conclude that we need fewer assumptions for the law to hold. On the other hand, our model provides a system wide consistency among individuals' choices on all their decisions.

ACKNOWLEDGEMENTS

This research was supported by FONDECYT 1110124 and by the Instituto de Sistemas Complejos de Ingeniería (ICM: P-05- 004-F, CONICYT: FBO16). I also thank G. West and L. Bettencourt for highly motivating comments.

REFERENCES

Bettencourt, L.M.A, Lobo, J., Helbing, D., Künhert, C., West, G. (2007). Growth, innovation, scaling, and the pace of life in cities. **PNAS 104**, 17, 7301-7306.

Bettencourt, L.M.A., Lobo, J. and West G.B. (2008). Why are large cities faster? Universal scaling and self-similarity in urban organization and dynamics. **The European Physical Journal B 63,** 285-293.

Ben-Akiva, M., Lerman, S. (1985). **Discrete Choice Analysis: Theory and Application to Travel Demand**. MIT Press, Cambridge.

Domencic, T.A., McFadden, D. (1975). **Urban Travel Demand: A Behavioral Approach**. North-Holland.

Fisher, R.A., Trippett ,L.H.C. (1928). Limiting forms for the frequency distribution of the largest or smallest member of the sample. **Proc. Cambridge Phil. Soc.** 24, 180-190.

Gabaix, X. (1999). Zipf's law for cities: an explanation. **The Quaterly Journal of Economics 114**, 739-767.

Galambos, J. (1987). **The Asymptotic Theory of Extreme Order Statistics**, Wiley, New York.

Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aleatórie. **Ann. Math.** 423-453.

Leadbetter, M.R., Lindgren, G., Rootzén, H. (1983). **Extreme and related properties of random sequences and processes**. Springer-Verlag, New York.

Mattsson, L-G., Weibull, J., Lindberg, P. O. (2011). Extreme values, invariances and choice probabilities. Manuscript presented at the Kuhmo Nectar Conference on Transport Economics, Stockholm.

Martínez, F., Donoso, P. (2010). The MUSSA II Land Use Auction Equilibrium Model. In F. Pagliara, J. Preston, D. Simmonds (eds.), Residential Location Choice: Models and Applications, Advances in Spatial Science, Springe.

West, G.B. (1999). The origin of universal scaling law in biology. **Physica A**, 104-113.

West, G.B, Brown, J.H., Enquist, B.J. (1999). The fourth dimension of life: Fractal geometry and allometric scaling of organism. **Science 284**, 1667-1679.

West, G.B, Brown, J.H., Enquist, B.J. (2001). A general model of ontogenetic growth. **Nature 413,** 628-631

Annex 1: Individuals' behavior distributes Gumbel.

The rational individuals' choice-making process under stochastic shocks of information may be represented by the following nested optimization problem:

$$Max_{k\in K}\ (\omega_k + \varepsilon_k)$$
$$\text{with}\quad \omega_k + \varepsilon_k = Max_{l\varepsilon L}\ (\omega_{kl} + \varepsilon_{kl})$$

where ω_k and ε_k are the deterministic utility or benefit and the random utility shocks of option *k* and *K* the set of optional discrete choices. The utility of each option *k* is in turn built from considering a set *L* of draws from stochastic events whose utilities are represented by a systematic utility ω_{kl} and shock ε_{kl}. The Extremal Types Theorem, attributed to Fisher and Trippett (1928) and generalized by Gnedenko (1943), proves that the asymptotic (no-degenerative) distribution functions of ε_k belongs to the set of three types of extreme values distributions (df), no matter what are the df of the ε_{kl} terms (see also Leadbetter, et al. 1983; Galambos 1987).

In this paper, in the absence of information, we consider the case that ε_{kl} terms are normally distributed, which belong to the Type I, also called double exponential or Gumbel domain of attraction. Hence, it directly follows that the Gumbel is to the *maximum* operator as the normal distribution is, by the central limit theorem, to the *sum* operator. Therefore, the Gumbel df is a natural distribution for the classical individual's discrete choice problems with stochastic utilities, as recognized by Domencic and McFadden (1975).

If a variate ε is distributed Gumbel with parameters (η, μ) then

$$F(\varepsilon) = exp\left[-e^{-\mu(\varepsilon-\eta)}\right] \quad \text{and} \quad f(\varepsilon) = \mu e^{-\mu(\varepsilon-\eta)} exp\left[-e^{-\mu(\varepsilon-\eta)}\right]$$

where η is the mode and μ is a positive parameter. The variate mean is $\eta + \frac{\gamma}{\mu}$ with γ the Euler's constant (~0,577) and the standard deviation is $\sigma = \frac{\pi}{\mu\sqrt{6}}$ or $\frac{1}{\mu} \approx 0.7796\,\sigma$. Notice that in the paper I ignore γ as it is a constant not relevant in our context.

Some relevant properties are of common use in transport models (see Ben-Akiva and Lerman, 1985):

i) Linear transformation: If ε distributes Gumbel $(0, \mu)$, for any scalars α and W, $\alpha\varepsilon + W$ is G-distributed with parameters $(\alpha\eta + W, \mu/\alpha)$.

ii) Differences: If ε_1 and ε_2 are independent and identical Gumbel distributed (IIG) variates with parameters (η_1, μ) and (η_2, μ), then $\varepsilon = \varepsilon_1 - \varepsilon_2$ is logistically distributed:

$$F(\varepsilon) = \frac{1}{1 + e^{\mu(\eta_2 - \eta_1 - \varepsilon)}}$$

iii) Maximum: For ε_k , $k = 1, \dots, K$, variates distributed i.i.d Gumbel (η_k, μ), $\varepsilon^* = max_k\,(\varepsilon_k)$ is also distributed Gumbel $\left(\frac{1}{\mu} ln(\sum_k e^{\mu\eta_k}), \mu\right)$. This means that the Gumbel df is also the attractor of itself, or that the Gumbel df is closed under the maximization operator.

In what follows in this paper we use the extreme value distribution Type I, or Gumbel distribution as the attractor df. Nevertheless Type II may be relevant because it is the attractor for positive stochastic variates $(\varepsilon \geq 0)$ such as prices or willingness to pay variates. We also restrict our analysis for the simple case of identical and independently distributed (i.i.d) variates. It remains for further research to explore the cases of Type II and III df for the original variates, and the general case for on i.i.d variates.

Multi-level random utility model

To study multi-level choices in a random utility context, we consider McFadden (1978)'s Generalized Extreme Value (GEV) (see Ben-Akiva and Lerman, 1985). The GEV model was derived by McFadden directly from the expression of the random utility model (RUM), where:

$$P_n(i) = \int_{\varepsilon=-\infty}^{+\infty} F_i(V_i - V_1 + \varepsilon, \dots, V_k - V_1 + \varepsilon) d\varepsilon$$

is the individual n-th probability of choosing alternative i out of a choice set of k options. V_i is the non-random or systematic utility of the discrete option i , and F is the cumulative distribution of random disturbances $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)$, and Fi

The GEV model is obtained from assuming $F(\varepsilon) = exp[-G(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_k})]$, where function G is non-negative, unbounded from above, homogenous of degree $\mu > 0$ and the n-th partial derivative in non-negative if n is odd, and non-positive if n is even. Denoting G_i the partial derivative of G with respect to y_i the GEV model is:

$$P_n(i) = \frac{e^{V_i} G_i(e^{V_1}, e^{V_2}, \dots, e^{V_k})}{\mu\, G(e^{V_1}, e^{V_2}, \dots, e^{V_k})}$$

Additionally, the expected maximum utility of the GEV model is $V' = \frac{1}{\mu} \ln G(e^{V_1}, e^{V_2}, \dots, e^{V_k})$.

The GEV model is actually a large class of models, the most well known is the multinomial logit model obtained from $G = \sum_{i=1\dots k} {y_i}^{\mu}$ which yields:

$$P_n(i) = \frac{e^{\mu V_i}}{\sum_{j=1,..,k} e^{\mu V_k}}$$

Another well-known GEV model is the nested logit model, which is of interest in multilevel choice making analysis. Ben-Akiva and Lerman consider two choice levels, that can be regarded the macro scale choice (denoted by sub-index m) and the micro scale choice (denoted by d) and the following μ^m homogenous function:

$$G = \sum_{m=1}^{M} \left(\sum_{i \epsilon D_m} e^{\mu^d V_i} \right)^{\mu^m / \mu^d}$$

The switching signs condition for the derivatives of G is satisfied if $1 \geq \mu^m/\mu^d \geq 0$. Then, the GEV nested model is:

$$P_n(i) = \frac{e^{\mu^d V_i}}{\sum_{j \in D_m} e^{\mu^d V_j}} \cdot \frac{\left(\sum_{j \in D_m} e^{\mu^d V_j} \right)^{\frac{\mu^m}{\mu^d}}}{\sum_{m=1}^{M} \left(\sum_{j \in D_m} e^{\mu^d V_j} \right)^{\frac{\mu^m}{\mu^d}}} = \frac{e^{\mu^d V_i}}{e^{\mu^d V'_m}} \cdot \frac{e^{\mu^m V_m}}{\sum_{m=1}^{M} e^{\mu^m V'_m}}$$

where $V'_m = \frac{1}{\mu^d} \ln(\sum_{j \in D_m} e^{\mu^d V_j})$.