Quantitative Economics 6 (2015), 153–187 1759-7331/20150153

Explaining the size distribution of cities: Extreme economies

Marcus Berliant
Department of Economics, Washington University, St. Louis

Hiroki Watanabe
Department of Economics and Finance, Lamar University

The empirical regularity known as Zipf's law or the rank-size rule has motivated development of a theoretical literature to explain it. We examine the assumptions on consumer behavior, particularly about their inability to insure against the city-level productivity shocks, implicitly used in this literature. With either self-insurance or insurance markets, and either an arbitrarily small cost of moving or the assumption that consumers do not perfectly observe the shocks to firms' technologies, the agents will never move. Even without these frictions, our analysis yields another equilibrium with insurance where consumers never move. Thus, insurance is a substitute for movement. We propose an alternative class of models, involving extreme risk against which consumers will not insure. Instead, they will move, generating a Fréchet distribution of city sizes that is empirically competitive with other models.

Keywords. Zipf's law, Gibrat's law, size distribution of cities, extreme value theory.

JEL classification. R12.

1. Introduction and motivation

A small industry has developed that seeks to provide a theory to explain a singular but robust stylized fact in urban growth: the size distribution of cities. Zipf's law or the rank-size rule, as applied to the size distribution of cities, states that for any country, the rank of a city according to population (for example, New York is ranked number one in the United States) multiplied by its population is constant. Thus, Los Angeles has half the

Marcus Berliant: berliant@artsci.wustl.edu
Hiroki Watanabe: watanabe.wustl@gmail.com
The first author thanks the Kyoto Institute of Economic Research at Kyoto University for their kind hospitality during the writing of the first draft of this paper. The second author thanks the Center for Research in Economics and Strategy (CRES) at the Olin School of Business at Washington University in St. Louis for financial support. Both authors thank Orazio Attanasio, three anonymous referees, John Conlon, Michael Magill, Tomoya Mori, Matthias Wrede, seminar participants at the University of Tokyo, and the participants at the North American Regional Science Association International meetings in Savannah for helpful comments, and thank K. T. Soo for providing European data. The authors thank Gilles Duranton and Jan Eeckhout for incisive, detailed comments and for sharing their data with us, but retain responsibility for any errors.

DOI: 10.3982/QE42

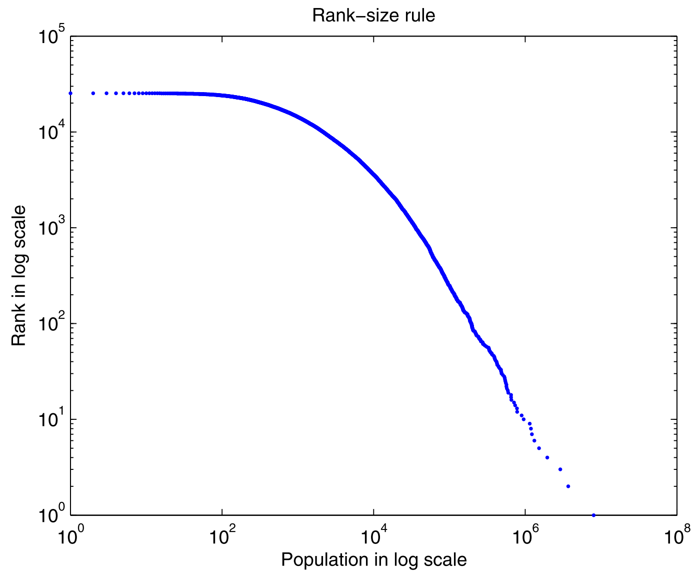


FIGURE 1. The rank-size rule. *Data source*: Census 2000.

population of New York, whereas Chicago has one-third the population of New York. This stylized fact holds across many countries and time periods (see Soo (2005)), but it is only one fact. In general, it is connected to Gibrat's law, stating that stochastic proportional growth tends to a log-normal distribution. The most compelling empirical work in this area shows that the size distribution of cities is log-normal (Eeckhout (2004)) when the data are not cut off at an arbitrary rank or population. For those unfamiliar with the empirics associated with this literature, we display in Figure 1 a graph of Eeckhout's data, consisting of more than 25,000 places from U.S. Census 2000. Since population on the horizontal axis and rank on the vertical axis are both plotted in log scales, the rank-size rule, taken literally, would say that the plot should be linear with slope -1. Deviations from the rule or law at the top and bottom of the size distribution are documented and discussed in the literature. See Gabaix and Ioannides (2004) for a fine survey of the entire area of research.

Further orientation with the data will prove useful so that the finer details of the distribution might be seen. The log–log plot is rather uninformative since very different distributions can appear similar because the majority of observations are bunched in the middle where there is little variation in the log–log scale. To that end, in Figure 2 we provide a graph of the empirical distribution function, whereas in Figure 3 we provide the density function.

Explanation of the stylized fact illustrated in these figures by a theory has long been an objective of urban economists; it is quite robust, but also very difficult to theorize about. Three recent articles, Eeckhout (2004), Duranton (2007), and Rossi-Hansberg and Wright (2007), have tackled this issue head on. The general methodology in this literature is as follows. A city is defined as a set of firms that receive a common technological shock to their production functions. Generally speaking, the shock is observed each period *before* the agents make their decisions. Consumers are freely mobile between cities.

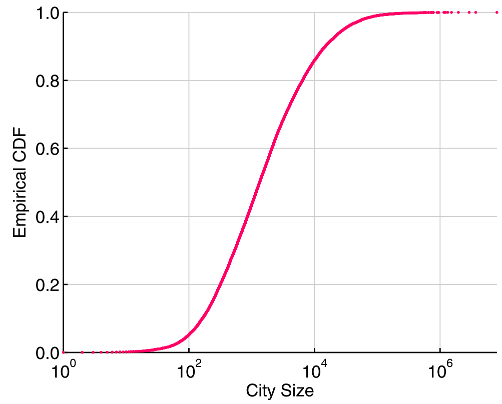


FIGURE 2. Empirical cumulative distribution function. *Data source*: Census 2000.

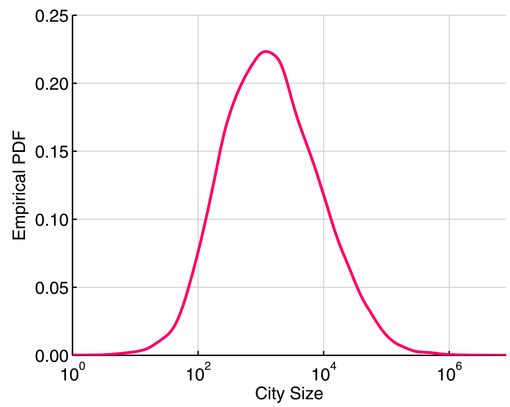


FIGURE 3. Empirical density function. *Data source*: Census 2000.

A model of demand and supply is formulated, generally relying on specific functional forms to obtain an analytical solution for equilibrium prices and quantities as a function of shock realizations. The key equation obtained from the models is the reduced form for the evolution of city population over time. Frequently (but not universally) this equation yields stochastic proportional growth for each city's population, where the stochastic component is derived from the city-specific technology shock. Then Gibrat's law is applied. The log-normal distribution matches Zipf's law well for the upper portion of the distribution.

The contribution of our work is as follows. First, we propose a new stochastic model of technological innovation in cities under perfect competition, giving rise in the limit to a generalized extreme value distribution of city sizes in aggregate, where the Fisher–Tippett theorem replaces the central limit theorem and Gibrat's law in a natural way. This model and its implications are robust against the introduction of self-insurance or insurance into the framework. The other models are generally not robust to the introduction of self-insurance or insurance, as we illustrate formally for one example from the literature in Appendix A. Our model is empirically competitive with other models of the size distribution of cities. In particular, the error in the estimate is very close to Eeckhout's (2004) for the log-normal distribution in the data on places, but better than Eeckhout's (2004) for the metropolitan statistical area (MSA) data.

All of these models, including ours, feature uncertainty that affects consumers through the budget constraint only. By self-insurance, we mean that consumers have an integrated budget constraint over time and know the distribution of future realizations of the random variables. Thus, they can smooth consumption. Since the consumers are risk averse, insurance or self-insurance is a substitute for migration. In the theoretical models, since moving is a discrete choice, partial insurance is never chosen in equilibrium by an individual agent (see the last subsection of Appendix A).

The paper is organized as follows. First, in Section 2, we propose a new type of model to explain the size distribution of cities, and implement it empirically. Only in Section 3 shall we discuss in detail the related literature that attempts to refine the stylized fact, namely the rank-size rule, and explain it. Then we shall raise specific objections, involving insurance or self-insurance against city-level risk, to these models. Section 4 discusses our conclusions and directions for future work. In Appendix A, we introduce Eeckhout's (2004) model and modify it to make the objections raised in Section 3 formal for a specific example, whereas in Appendix B, we examine our model with positive transport costs. Replication files are available in a supplementary file on the journal website, http://qeconomics.org/supp/42/code_and_data.zip.

2. Modeling the size distribution of cities

2.1 *A model*

2.1.1 *The basic model and its equilibrium* This model is loosely based on Duranton (2007), but in the context of perfect competition instead of monopolistic competition. It can also be viewed as a slice of a larger model that would include both our model and the model of Eaton and Kortum (2002). Our model adds labor and consumer mobility, whereas their model has them locationally fixed. In contrast with the other models in the literature, there is *economy-wide* risk in addition to *city-level* risk. But this in itself is not sufficient to generate consumer movement. For example, if all cities faced correlated shocks at each time, consumers could still insure against this risk by smoothing their consumption through borrowing and saving. Thus, we employ a more extreme form of aggregate risk.

Time is discrete and all consumers are infinitely lived. Assume that there are many cities (indexed by $i = 1, \dots, m$) and many industries, each producing one consumption commodity (indexed by $j = 1, \dots, n$). All commodities are freely mobile. The production function for commodity j in city i at time t is given by

$$y_{ijt} = A_{ijt} \cdot l_{ijt},$$

where y_{ijt} is the output of commodity j in city i at time t, and l_{ijt} is labor input.[1] The random variable $A_{ijt} \in \mathbb{R}_{++}$ will be discussed in detail shortly. Suppose that each con-

[1]The assumption of Starrett's (1978) spatial impossibility theorem that is violated by this model is the assumption of location-independent production sets.

sumer supplies one unit of labor inelastically and that the total number of consumers as well as total labor supply is given by N. We justify the assumption of perfect competition by implicitly assuming that there is a large number of firms in each city capable of producing a commodity using a constant returns technology, but all experiencing the same citywide technology shock.

In each time period t, each city i receives a random draw for its productivity in producing commodity j, namely A_{ijt}. Since we will be using the Fisher–Tippett limit theorem from extreme value theory rather than the central limit theorem, there is no requirement that these random variables be independent. It is assumed that with probability 1, the random draws for two industries at time t for city i are not both maximal among all cities for these given industries. In equilibrium, only the cities with the highest draw of the random variable for some industry will have employees and population. (Alternatively, we could simply classify cities exogenously by industry, and assume that a city in an industry receives only a draw for that industry.) Extensions that imply several cities produce in equilibrium will be discussed shortly, but first we must explain the basic model.

The wage rate for the (freely mobile) population of consumers is given by $w(t)$. In equilibrium, it will be the same across industries.

As is standard in this literature, the utility function of a consumer at time t is given by

$$u(t) = \sum_{j=1}^{n} \frac{1}{n} c_j(t)^{\gamma},$$

where $c_j(t)$ is the consumption of commodity j by a consumer at time t and $\gamma \in (0, 1)$. Let $p_j(t)$ be the price of commodity j at time t. Assuming that commodities are freely transportable, a consumer's budget constraint at time t is

$$\sum_{j=1}^{n} p_j(t) \cdot c_j(t) = w(t).$$

Let $\lambda(t)$ be the Lagrange multiplier associated with the budget constraint in the consumer optimization problem. Standard calculations yield demand for commodity j at time t for a single consumer $d_j(t)$:

$$d_j(t) = \left(\frac{\gamma}{-\lambda(t) n \cdot p_j(t)} \right)^{1/(1-\gamma)}.$$

Aggregate demand is given by

$$N \cdot d_j(t) = N \left(\frac{\gamma}{-\lambda(t) n \cdot p_j(t)} \right)^{1/(1-\gamma)}.$$

To reduce notation, for $j = 1, \dots, n$, define i^* to be the city with $A_{i^* jt} = \max_{1 \le i \le m, 0 \le t' \le t} A_{ijt'}$.

Profit optimization yields, for each t,

$$p_j(t) \cdot A_{i^*jt} = w(t).$$

Here we are assuming total recall, in that the best technology from the past is remembered, so new technologies are not used unless they are better than all the old ones. Also, only the best technology in industry j survives, where the best is across all cities and previous time periods. This assumption is made for convenience. We discuss it more below.

Hence

$$p_j(t) = \frac{w(t)}{A_{i^*jt}}. \tag{1}$$

In other words, even though wage is constant across occupied cities, output price varies inversely with the production shock. Consumption commodity market clearance requires, for each t,

$$l_{i^*jt} \cdot A_{i^*jt} = N \cdot d_j(t) = N\left(\frac{\gamma}{-\lambda(t)n \cdot p_j(t)}\right)^{1/(1-\gamma)}. \tag{2}$$

This is the key equation for our analysis.

Labor market clearance requires, for each t,

$$\sum_{j=1}^{n} l_{i^*jt} = N. \tag{3}$$

Setting the constant to be

$$\kappa(t) = N\left(\frac{\gamma}{-\lambda(t) \cdot n \cdot w(t)}\right)^{1/(1-\gamma)},$$

and using (1) and (2), we obtain

$$l_{i^*jt} \cdot (A_{i^*jt})^{\gamma/(\gamma-1)} = \kappa(t).$$

Hence

$$l_{i^*jt} = \kappa(t) \cdot (A_{i^*jt})^{\gamma/(1-\gamma)}. \tag{4}$$

Since $\gamma < 1$, labor usage l_{i^*jt} and the shock A_{i^*jt} are positively correlated. Notice that cities that do not have an industry with the largest shock in that industry at time t are empty.

Existence of an equilibrium is not an issue here, since the equilibrium prices and quantities can be solved analytically. For example, at $t = 1$, setting $p_1(1) = 1$, then $w(1) = A_{i^*11}$, $p_j(1) = A_{i^*1t}/A_{i^*jt}$, $\lambda(1) = -\frac{\gamma}{nA_{i^*11}}(\sum_{j=1}^{n} A_{i^*j1}^{\gamma/(1-\gamma)})^{1-\gamma}$, $l_{i^*j1} = N(\frac{\gamma}{-\lambda(1)nA_{i^*11}})^{1/(1-\gamma)} A_{i^*j1}^{\gamma/(1-\gamma)}$, and so forth. Thus, equilibrium is also unique.

The original work on the asymptotic distribution of maxima drawn from a distribution is due to Fisher and Tippett (1928). Modern, more general treatments are given in

Coles (2001) and Embrechts, Kluppelberg, and Mikosch (1997). We shall return to a discussion of extreme value theory momentarily, but first we will draw the implications for our analysis.

The bottom line from this literature is that A_{i^*jt} has an asymptotic distribution, known as the generalized extreme value (GEV) distribution, of the form

$$F_{\mathrm{GEV}}(x) = \begin{cases} \exp\left\{-\left[1+\xi\cdot\left(\dfrac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, & \text{when } \xi \neq 0, \\ \exp\left\{-\exp\left[-\left(\dfrac{x-\mu}{\sigma}\right)\right]\right\}, & \text{when } \xi = 0. \end{cases}$$

Notice that there are three free parameters to be estimated here, namely μ, σ, and ξ. Also notice that to use rank as the left hand side variable in the regression, one simply computes $1 - F_{\mathrm{GEV}}(x)$. But from a pragmatic point of view, it is easier to use $\ln(F_{\mathrm{GEV}}(x))$ as the left hand side variable.

If there are no upper or lower bounds on the distribution, then $\xi = 0$ and the distribution is Gumbel. If there is an upper bound on the distribution, then $\xi < 0$ and the distribution is reverse Weibull. If there is a lower bound on the distribution, for example, 0 in our case, then $\xi > 0$ and the distribution is Fréchet.

Substituting (4),

$$\ln\bigl(F(l)\bigr) = \begin{cases} -\left[1+\xi\cdot\left(\dfrac{\left(\dfrac{l}{\kappa(t)}\right)^{(1-\gamma)/\gamma} - \mu}{\sigma}\right)\right]^{-1/\xi}, & \text{when } \xi \neq 0, \\ -\exp\left[\dfrac{-\left(\left(\dfrac{l}{\kappa(t)}\right)^{(1-\gamma)/\gamma} - \mu\right)}{\sigma}\right], & \text{when } \xi = 0. \end{cases} \tag{5}$$

Notice that if we use cross section data, then t and hence $\kappa(t)$ is constant. Thus, in addition to the three standard parameters for the GEV distribution of A_{i^*jt} (namely μ, σ, and ξ), for the distribution of l_{i^*jt} there are two additional parameters, namely κ and γ, that arise from our economic model.

Now that the basic model is fully developed, we can discuss why, unlike other models in this literature, consumers will not want to insure against this risk. Instead, they will move. If only a small percentage of cities produce at any time, then insurance would cost only slightly less than the wage, so the consumers might as well move and receive the wage in each period. For example, to keep things simple, suppose that there are 100 industries (or consumption commodities) and 100 cities in each industry (that is, each city is capable of producing only one commodity). Then there is only one city producing in each industry at each given time, and 100 cities out of 10,000 producing in each given time. As time plays out, as long as some consumers are willing to move, each of the cities producing at a given time will eventually be replaced by another in the industry. The city using old technology has zero wage and no production. So if some workers do not move, their average wage tends to 1% of the expected new wage with time. Under symmetry of cities in an industry, actuarially fair insurance would cost 99% of the expected new wage. In other words, if workers move, they will receive the wage next period, but if they insure, they will receive 1% of the wage next period. The only way workers will not move is if they all agree to use old, frozen technology in each industry, and collude so that none will move for a higher wage. In contrast, we assume competitive behavior.

This is the main idea motivating our specific model. Next, we discuss extensions and the intuition behind why this idea is robust.

2.1.2 *Extensions of the basic model and further implications* In fact, what we have presented is an extreme example. All that is needed to induce consumers to reject insurance and move is that the probability of unemployment next period is greater than zero if they do not move. To obtain stronger results, for example, the GEV distribution, stronger assumptions are required. Thus, there are many models like this in which consumers will not take up insurance, but that do not require such strong assumptions. We provide a simple model that is tractable.

We claim that the choice of insurance or moving is essentially a bang–bang phenomenon, not only in this model, but in other models of stochastic growth belonging to the literature that will be surveyed in Section 3. That is, generically one or the other will be better for consumers, so in equilibrium they will not coexist. Moreover, in equilibrium there will be no partial insurance. To see this, notice first that utilities are not state-dependent, so the state only directly affects budget constraints. Second, the decision to move is a discrete one: Either all of a moving cost or none of it is incurred by a particular consumer. If competition forces insurance to be priced competitively, implying both that consumer cost is proportional to price and that it is actuarially fair, then risk averse consumers will always want to fully insure or move, facing no uncertainty in equilibrium. The consumers must consider whether the moving cost or the cost of full insurance is cheaper. Generically these exogenous parameters are unequal, so only one or the other will be observed in equilibrium. Partial insurance will not result unless there is some defect in insurance markets, but the random shock in this entire class of models is assumed to be observed by all agents. Generically, none would predict that consumers use partial insurance. We prove this more formally in Appendix A in the case where moving is costly and insurance is actuarially fair.

Given the structure of the model and the other models in the literature, it is much more natural to introduce a market imperfection in the labor market: labor heterogeneity and adverse selection, moral hazard, or search frictions, for example. This new source of uncertainty or asymmetric information requires an additional dimension for states of nature beyond the states we have specified for production shocks. It leads to a different form of a distortion or market imperfection than that in this literature, since, for example, labor supply might be distorted. Although individual labor supply is inelastic in our basic model, it is elastic, for example, in Eeckhout (2004); see Appendix A. Elastic labor supply could easily be put into our model in an additively separable way at the cost of further notation. The consequences of a distortion in the labor market would be very different from the introduction of an exogenous mobility cost that varies between zero and infinity, as described in the previous paragraph. Full or partial insurance would have to be defined over states of the world associated with a new source of uncertainty or asymmetric information related to the labor market, in contrast with the one already in the model that is related to production shocks.

Returning to our basic model, the consumers still might want to insure against aggregate wage volatility (namely movement in $w(t)$ over time) by saving and borrowing to smooth consumption, but their spatial distribution is still as we have laid out.

Returning now to our assumptions and extreme value theory, the original theory of Fisher and Tippett presumed that, fixing j, the random variables, A_{ijt} in our case, were independent and identically distributed (i.i.d.) across i and t. Of course, in our context this makes little sense. *In general, the city with the best technology for some good j at a particular time t is more likely to innovate and produce a better technology for the next period than an arbitrary city.* Moreover, it is possible that cities nearby are more likely to innovate than an arbitrary city. Fortunately, much progress has been made in extreme value theory since 1928. The modern versions of the Fisher–Tippett theorem, as given by Coles (2001, Theorem 5.1) and Embrechts, Kluppelberg, and Mikosch (1997, Theorem 4.4.1), allow some dependence. Specifically, what is required is that the sequence of random variables be stationary and that a form of asymptotic independence (as blocks of random variables become farther apart in time) hold.[2] *Since temporal (as well as spatial) correlation is allowed, the model can explain the persistence of an industry in a given city over time.* For example, the assumption that the process is stationary imposes some symmetry on the spatial correlation, in that the influence of neighbors on the productivity of one reference location is the same, independent of the reference location. However, we note that even the modern versions of the Fisher–Tippett theorem we have cited give only sufficient conditions for convergence to the GEV distribution. There are yet further generalizations to nonstationary processes; see Coles (2001, Chapter 6) for example. So asymmetries in space, implying that the process is not stationary, can still lead to the GEV distribution.

Returning to the case of i.i.d. technology draws, an implication of extreme value theory (Embrechts, Kluppelberg, and Mikosch (1997, Chapter 5.4)) is that the time between new record draws of technology in an industry grow in a roughly exponential fashion with the passage of time. This implication of the theory might not hold in more general settings, for example, nonstationary ones.

It is also important to note that the model and results can be extended to the case where more than one city in an industry produces. This could happen, for example, if there is transportation cost for consumption goods between cities, so a city with a high realization of productivity for a commodity, but not the highest, might serve a local market. It turns out that extreme value theory applies not only to the maximum of a sequence of random variables, but also to the upper order statistics. A detailed discussion of the results can be found in Embrechts, Kluppelberg, and Mikosch (1997, Section 4.2). These extensions of the model require a simulation approach, as the analytics are difficult. Specifically, the calculation of aggregate demand on the right hand side of equation

[2] An easy way to fit our structure into the theory is to fix an industry j and imagine that at each time t, there are m subperiods. A city i draws its random variable A_{ijt} in subperiod i of time t.

(2) becomes difficult due to the endogeneity of market area. Our simulations appear in Appendix B.

A couple more remarks are in order. First, the role of having different industries j, as in the other models in the literature, is to generate a full distribution of limiting populations rather than just one realization of the asymptotic distribution of city populations. Second, in contrast with other models in the literature, the cities without the best technology for some industry at a given time have zero population, so they do not show up in the data because they are rural.

2.1.3 *Stochastic proportional growth* As a complement to our basic analysis of the model, it is interesting to see under what conditions our model will generate stochastic proportional growth in (occupied) city populations. To examine this, we must specialize and reinterpret slightly the stochastic part of our model, inspired by Eeckhout (2004, p. 1447). Suppose that a primitive productivity random variable is generated by the (autoregressive) AR(1) process

$$B_{ijt} = k \cdot B_{ij(t-1)} + \frac{1-\gamma}{\gamma} \cdot \varepsilon_{ij(t-1)}, \tag{6}$$

where $\varepsilon_{ij(t-1)}$ is i.i.d. with mean 0 and finite variance, and where $0 < k < 1$. For the purpose of approximation, we will be taking k close to 1. Then define the reduced form random variable by

$$A_{ijt} \equiv \exp(B_{ijt}).$$

Our previous analysis applies to this more specific model of A_{ijt}, for instance, the aforementioned Theorem 4.4.1 of Embrechts, Kluppelberg, and Mikosch (1997), along with all of the results in the subsections above. But with this additional structure, we can say more.

Consistent with our notation,

$$\text{for } j = 1, \dots, n, \text{ let } i^* \text{ be such that} \quad B_{i^*jt} = \max_{1 \le i \le m, 0 \le t' \le t} B_{ijt'}.$$

If the ε_{ijt} are small, we claim that

$$B_{i^*jt} \approx k \cdot B_{i^*j(t-1)} + \frac{1-\gamma}{\gamma} \cdot \varepsilon_{i^*j(t-1)}$$

in the sense that the distributions of the two sides of this expression viewed at time $t-1$ are close. The reasoning behind this approximation is as follows. Fix industry j. If $B_{i^*j(t-1)} \gg k \cdot B_{i'j(t-1)}$ for all $1 \le i \le m$, $0 \le t' \le t-1$, $i' \ne i^*$, then the city with the maximal draw remains the same between periods $t-1$ and t, so the approximation holds according to equation (6). If $B_{i^*j(t-1)} \approx k \cdot B_{i'j(t-1)}$ for some $1 \le i' \le m$, $0 \le t' \le t-1$, $i' \ne i^*$, then the distribution of $k \cdot B_{i^*j(t-1)} + \frac{1-\gamma}{\gamma} \cdot \varepsilon_{i^*j(t-1)}$ conditional on $B_{i^*j(t-1)}$ is close to the distribution of $k \cdot B_{i'j(t-1)} + \frac{1-\gamma}{\gamma} \cdot \varepsilon_{i'j(t-1)}$ conditional on $B_{i'j(t-1)}$, so the approximation holds.

Dividing equation (4) at time t by its value at time $t-1$ to the power k,

$$\frac{l_{i^*jt}}{[l_{i^*j(t-1)}]^k} = \frac{\kappa(t)}{[\kappa(t-1)]^k} \cdot \left(\frac{A_{i^*jt}}{[A_{i^*j(t-1)}]^k}\right)^{\gamma/(1-\gamma)} \quad \text{for each industry } j = 1, \dots, n. \tag{7}$$

Using (4) and (3),

$$\begin{aligned} N &= \sum_{j=1}^{n} l_{i^*jt} \\ &= \kappa(t) \cdot \sum_{j=1}^{n} (A_{i^*jt})^{\gamma/(1-\gamma)}. \end{aligned}$$

Now

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{n} (A_{i^*jt})^{\gamma/(1-\gamma)}}{n} = E\big[(A_{i^*jt})^{\gamma/(1-\gamma)}\big].$$

Hence for k close to 1,

$$\frac{\kappa(t)}{[\kappa(t-1)]^k} \approx \frac{E[(A_{i^*j(t-1)})^{\gamma/(1-\gamma)}]}{E[(A_{i^*jt})^{\gamma/(1-\gamma)}]} \approx 1.$$

Taking logarithms of both sides of equation (7),

$$\begin{aligned} \ln(l_{i^*jt}) &= k \cdot \ln(l_{i^*j(t-1)}) + \frac{\gamma}{1-\gamma} \cdot \frac{1-\gamma}{\gamma} \cdot \varepsilon_{i^*j(t-1)} \\ &= k \cdot \ln(l_{i^*j(t-1)}) + \varepsilon_{i^*j(t-1)} \\ &\approx \ln(l_{i^*j(t-1)}) + \varepsilon_{i^*j(t-1)}. \end{aligned}$$

This last equation is the form of stochastic proportional city population growth obtained in Eeckhout (2004).

The assumption that $k < 1$ is essential, in the sense that $k = 1$ yields Gibrat's law and a log-normal distribution for occupied cities. The assumption that $k < 1$ implies the asymptotic independence used for modern variants of the Fisher–Tippett theorem. In contrast, $k = 1$ implies some permanent path dependence. Another way to frame the arguments in this subsection is that the order of limits in k and t matters.

2.2 *Empirical implementation*

Notice that we are not overly concerned with identification of the five parameters in equation (5). In essence, the parameters are identified by the functional form itself. The economic interpretation of these variables is as follows. The three parameters of the GEV distribution, μ, σ, and ξ, are analogous to the mean and variance of the log-normal distribution estimated by Eeckhout or the regression coefficients estimated for Zipf's law using a log–log regression. They have no direct economic interpretation. Since γ and κ are derived from the model, they do have an economic interpretation. Standard calculations tell us that $\frac{1}{1-\gamma}$ is the elasticity of substitution for consumers between consumption commodities. The endogenous variable κ is more difficult to interpret, since it involves a number of endogenous variables as well as random variables. But equation (4) gives us the equilibrium relationship between the random variable representing productivity in an industry (exogenous) and employment in that industry (endogenous). So $\kappa(t)$ tells us equilibrium employment in an industry where one unit of labor produces one unit of consumption commodity.

We use the Census 2000 data set also used by Eeckhout. Table 1 gives the summary statistics for this data along with the metropolitan statistical area (MSA)-level data that we use later for comparison.[3]

As noted in the sources we cite for extreme value theory, the most common method of estimating extreme value distributions is to use maximum likelihood. The maximum likelihood estimator (MLE) does not yield the smallest Kolmogorov–Smirnov (KS) statistic in our data set. The KS statistic measures the maximum distance between a sample distribution and its estimate. As noted by Goldstein, Morris, and Yen (2004) in the context of social networks and later by Eeckhout (2009) in the context of the size distribution of cities, using a simple log–log regression can lead to serious statistical problems. The use of MLE and the KS statistic is preferred. It is interesting to note that both the literature on estimation of the GEV distribution and the literature on Zipf's law seem to be (independently) converging on MLE as the preferred method of estimation.

For purposes of comparison with Eeckhout (2004), we produce estimates using the log-normal (his) distribution and the generalized extreme value (our) distribution using equation (5), for both maximum likelihood estimation and minimization of the KS statistic (MinKS). We also report an estimate using the double Pareto log-normal distribution (DPL) from Giesen, Zimmermann, and Suedekum (2010) for comparison. Table 2 summarizes the estimation results. The results of maximum likelihood estimation for the log-normal distribution are identical to Eeckhout's. The rightmost columns contain the KS statistic, the log likelihood of the estimates (LogLH), the Akaike information criterion (AIC), and the Bayesian information criterion (BIC).

In the interest of full disclosure, we report both the MLE and MinKS estimates in Table 2. Notice that the MLE estimate implies a reverse Weibull distribution whereas MinKS estimates imply a Fréchet distribution. Since city sizes do not fall below zero, we expect the distribution to follow a Fréchet distribution. MLE predicts otherwise due to the large, uncensored data set containing places. The estimated Fréchet distribution under MinKS implies that the smallest place will have population 1.582, and two places actually fall below this size. Indeed, once we truncate the data to MSA's, MLE predicts a Fréchet distribution. So the reverse Weibull GEV distribution is driven by extremely small populations in the sample of places.

Of course, the comparison between log-normal and GEV is not quite fair. In general, the more parameters a distribution has, the better is its fit to data. There are only two

[3]For a definition of the spatial units used by the Census, see, for example, http://www.genesys-sampling.com/pages/Template2/site2/61/default.aspx.

Table 1. Summary statistics for U.S. data.

Unit	Sample Size	Mean	Variance	Median	Mode	Max	Min
Place	25,358	8.232E+03	4.677E+09	1,338	86	8,008,278	1
MSA	922	2.837E+05	9.490E+11	71,800.5	20,411	18,323,002	13,004

Table 2. Parameter estimates and related statistics—United States.

Unit	Distribution	Method	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\xi}$	$\widehat{\kappa}$	$\widehat{\gamma}$	KS	LogLH	AIC	BIC
Place	Log-normal	MLE	7.278	1.754				1.895E−02	−2.3477E+05	4.6955E+05	4.6957E+05
	GEV	MLE	1.410	0.3096	−2.902E−02	57.15	0.8827	8.638E−03	−2.3467E+05	4.6935E+05	4.6939E+05
	Log-normal	MinKS	7.249	1.738				1.336E−02	−2.3478E+05	4.6956E+05	4.6958E+05
	GEV	MinKS	1.592	0.6127	1.592	102.9	0.8100	6.970E−03	−2.3470E+05	4.6941E+05	4.6945E+05
MSA	Log-normal	MLE	11.46	1.190				9.426E−02	−1.203E+04	2.406E+04	2.407E+04
	GEV	MLE	4.295	2.192	0.6383	4552	0.6276	2.582E−02	−1.190E+04	2.382E+04	2.384E+04
	DPL	MLE								4.694E+05	4.695E+05

parameters in the log-normal distribution whereas there are five parameters in our distribution, and these parameters do not contain the parameters used for the log-normal distribution. In Table 2, we report the Akaike and Bayesian information criteria, that penalize distributions with more parameters. Smaller values for these criteria mean better performance. Those two statistics indicate that the log-normal and our distribution are still comparable when an adjustment for the number of parameters is made. In particular for the uncensored place data, *after penalizing each estimate for the number of parameters used*, where the penalty is larger for the GEV estimate, the error is quite similar, with GEV slightly ahead. The penalties are actually quite small relative to the log likelihood, since the data sets are so large. There is more divergence between log-normal and GEV in the error for the MSA estimates. Clearly, as Eeckhout (2004) points out, there are problems with truncation of this data. On the other hand, it seems quite odd to give places with just a few people in them the same weight as, say, New York City, in the data. Implicitly in the place data, all places have weight 1. In the MSA data, places above a certain population have weight 1, whereas all other weights are set to 0. There is likely a way to weight the data better than these extreme cases, but we do not attempt a formal theory of data weighting for these estimates.[4] So we do not completely discount the MSA estimates (as Eeckhout might), but rather await a less extreme data weighting scheme than the two standard ones. The truth probably lies somewhere in between. This appears to be an interesting topic for future research.

Graphically, the estimates and data plots are shown in Figure 4.

In summary, estimates using the generalized extreme value distribution are quite competitive.

We ran simulations of our model with positive transportation cost. The results and discussion can be found in Appendix B. Related to this, Hsu, Mori, and Smith (2014) study a random growth model of the city-size distribution when city connections matter.

3. Relationship to the literature

3.1 *The older literature*

The innovative work of Gabaix (1999a, 1999b) is the source from which the modern literature on the size distribution of cities flows. This work uses an overlapping generations structure where consumers live for two periods. It is assumed that moving costs are so high that consumers can only choose their location (city) when they are young. This location decision is made after shocks to production and amenities are realized for that period, and are known to all. The consumer/workers cannot move again when old. The wages or income for the old in a city are never even specified, and it is simply assumed that the young make their decisions in a myopic manner. Moreover, the availability of

[4]As Eeckhout (2004, pp. 1434–1436) points out, the theoretical definition of a city as those firms that experience a common shock (perhaps assuming spatial independence of shocks) should drive the empirical unit used. In places with only a few inhabitants, it is difficult to see how to apply this definition. On the empirical side, Eeckhout (2004, Figure 7) finds some anomalies at the low end of the distribution. In our opinion, random models of city growth might not be appropriate for small places.

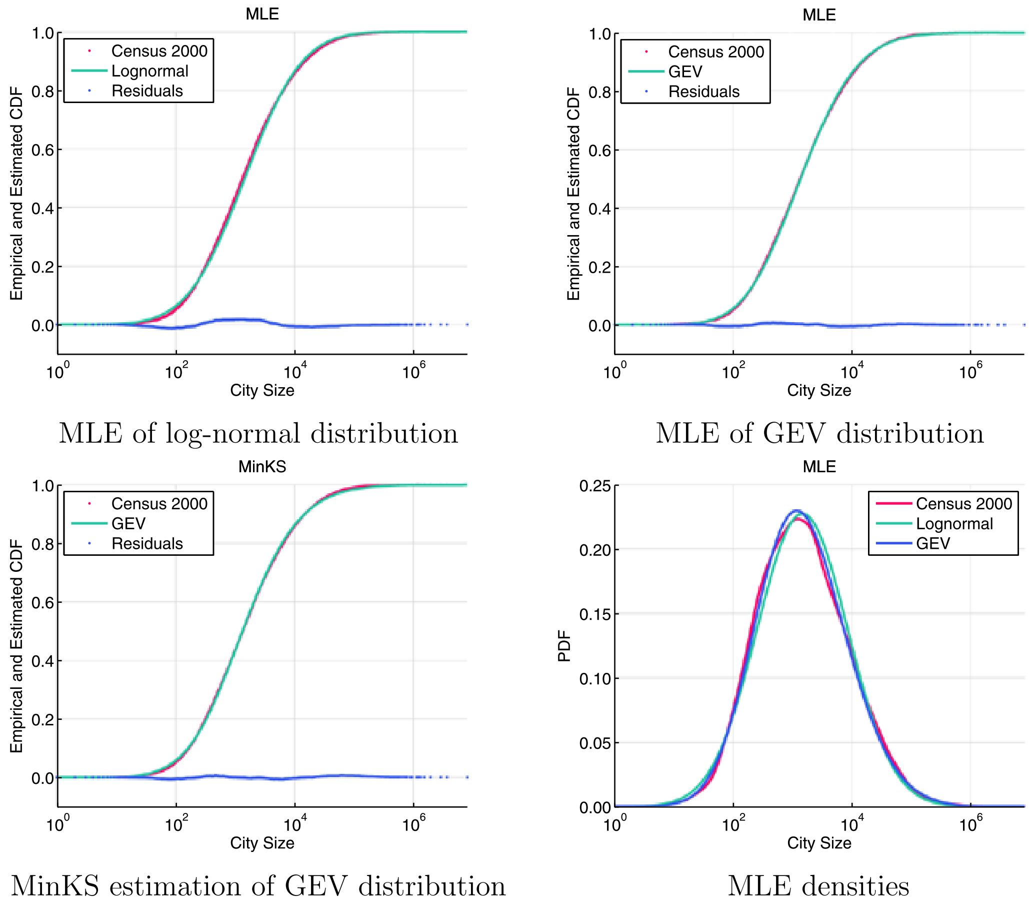


FIGURE 4. Empirical and estimated city-size distribution.

insurance or capital markets is never discussed, so it is unknown whether the young can hedge against uncertainty about their wage when they are old in the city they choose.

If the old people are immobile, why is this important? It is important because when the young make their decisions, they can anticipate what will happen when they are old and might change their minds about their location decisions when young. In other words, they will not behave myopically. Without myopia, insurance becomes important.

3.2 *Recent literature*

Chief among recent work are Rossi-Hansberg and Wright (2007), Duranton (2006), Eeckhout (2004), and Duranton (2007). We focus on the latter two.

Eeckhout's model has consumers who are infinitely lived, have foresight, and can move each period. There are technological shocks to production in each city in each time period. It is movement of the consumer/worker population in response to these shocks that generates Gibrat's law. The shocks generate changes in equilibrium wages, rents, and congestion across time and space that correspond to the consumer movements that equalize utility levels across space at each time. Eeckhout (2004, p. 1445) makes the following statement:

> "Moreover, because there is no aggregate uncertainty over different locations, and because capital markets are perfect, the location decision in each period depends only on the current period utility. The problem is therefore a static problem of maximizing current utility for a given population distribution, and the population distribution must be such that in all cities, the population $S_{i,t}$ equates utilities across cities."

Here we wish to make an important distinction between transfers of consumption across time, namely perfect capital markets, and across states, namely complete and perfect futures markets.

The actual consumer optimization problem in Eeckhout's model does not involve state-dependent assets or allow state-contingent transfers of income. If it were to allow this, as in a standard model of complete futures or insurance markets, then agents would never move. They would simply buy assets at the start of time that would pay them under a bad state in their city at a particular time, and such that they would pay under a good realization in their city. In other words, they would insure against the state of nature in their city. It is important to recognize that in this model there are two factors determining a worker/consumer's productivity, namely, the city-specific shock and the externality in production induced by total population in the city.

The basic model of Duranton (2007) has consumers maximizing an intertemporal utility function subject to an intertemporal budget constraint without facing uncertainty. However, once the detailed urban features are added (Duranton (2007, Section V)), the model looks similar to Eeckhout's at least in terms of the urban features. One simply needs some dependence of local prices (land rents or wages) on the state of nature. Then utility equalization implies that people will move depending on the state realization, but this movement disappears if one allows insurance.

There is not enough detail about the urban market in Duranton (2006, 2007) to make specific statements about how insurance would work, but the consumers in a city face uncertainty about employment due to the uncertainty about innovations in various industries, so similar insurance arguments should work if the details of the model are filled in.

Regarding contemporary developments in this literature, Behrens, Duranton, and Robert-Nicoud (2014) is a very interesting contribution that does not employ Gibrat's law to obtain Zipf's law. Using a static model, a number of stylized facts are matched. There is an asymmetric information/adverse selection component as well as a potentially insurable luck component in the model.

In general, we are inquiring whether moving or buying insurance is cheaper for the consumers in these models. Typically in these models, if moving costs are positive, it makes sense for consumers to stay put and insure.

3.3 *Criticism of the literature*

3.3.1 *How insurance reduces population movement* So how might this insurance occur in practice? Let us assume either that consumers cannot perfectly observe the technology shocks to cities or that moving has a small cost, or both.

- Self-insurance. Since consumers can transfer consumption across time and they know that shocks are i.i.d., then they can borrow or use their savings in bad times and save (or pay off their loans) in good, staying in the same city. In the literature, the intertemporal uncertainty faced by consumers does not show up in their objective function, whereas the possibility of self-insurance does not show up in the budget constraint. The earlier quote from Eeckhout seems to imply that this is allowed, but the formal statement of the consumer budget constraint makes it clear that it is not allowed. This type of insurance exploits the fact that for any given city, the shocks are i.i.d. over time. Empirically, the place to look for self-insurance is in the savings response to local employment shocks.

- Insurance markets. In all of these models, at each time, the state of nature (the random shock to each production function for each city) is known to all and is verifiable[5] before consumers make their decisions about consumption bundles and location. So this is a perfect setting for a viable insurance market. An insurance firm can step in or the continuum of consumers can simply pool resources in each period, smoothing their consumption without changing location, so it is independent of the state in their city. This type of insurance exploits the fact that at any given time, the shocks are i.i.d. across cities. Empirically, one place to look for insurance is a cross-country comparison of how varying benefits of unemployment insurance affect mobility in response to local employment shocks.

- Futures markets. Consumers formulate plans to sell labor, and buy consumption commodity and housing contingent on every possible state in every time period. There is no empirical complement. We mention this for completeness.

Given that for Gibrat's law to hold, the shocks to each city in each period must be "small" (see Eeckhout (2004, p. 1447)), it seems reasonable to think that insurance would yield higher consumer utility than movement if moving costs are at all significant or if consumers cannot observe shocks to firms perfectly, and, thus, face even a small amount of uncertainty in their optimization problems.

For models in the literature, consumers will choose to insure instead of move when insurance is available. A common feature of both the models in the literature and the model we have presented is the prediction that people will move and not insure. A major difference between our model and the balance of the literature is clear: An advantage of our model is that it can explain endogenously the lack of insurance, whereas the other models in the literature implicitly assume that such markets, namely insurance or self-insurance (saving and borrowing), do not exist. The empirical investigation of the use of insurance as a substitute for migration, especially when consumer heterogeneity is taken into account, seems quite interesting as a topic for *future* research.

As a preview, we present preliminary work. We compare U.S. data with analogous data for Belgium and Germany. For Germany and Belgium, we use data on municipalities, whereas for the United States, we use data on MSA's. Please note that all of these

[5] Thus, such models differ from models of human capital, for example, where verification is not a realistic assumption and, thus, insurance against fluctuations is not to be expected.

data are, therefore, truncated. For Europe, we use the data from Soo (2005), who obtains it from http://www.citypopulation.de/. We provide summary statistics for all three data sets in Table 3.

We present in Table 4 the estimates of the models for Germany and Belgium, to be used in conjunction with Table 2 for the MSA estimates for the United States. We note that for MLE, there is no general analog of t-statistics for each parameter estimate, just an overall measure of goodness of fit such as the KS statistic. The sample sizes are very different in Tables 2 and 4, resulting in very different log likelihoods as well as AIC and BIC statistics.

Likely insurance mechanisms are more developed and moving costs are higher for Europe compared with the United States. Thus, one would expect deviations from Zipf's law and log-normal models, but not from GEV, for Europe as compared to the United States. We examine AIC and BIC ratios of log-normal to GEV for each country in Table 5. What we find is that in three of the four cases, the GEV fit is better for Europe as opposed to the United States.

For a more complete analysis, it would be desirable to regress the ratios in Table 5 on proxies for moving cost and insurance mechanisms in each country for a larger sample of countries. That would give us a reading on which model performs better, but is beyond the scope of this paper.

3.3.2 *Possible objections to the criticism* *We emphasize that the criticism we make is a purely theoretical point concerning models in the literature.* Whether or not agents in the real world actually insure or self-insure against citywide risk is not relevant to the question at hand. Our point is that in the theoretical worlds of these models, insurance or self-insurance of the sort discussed in the previous subsection is implicitly excluded. The reasons are not given or, more importantly, included in the model. If these factors, such as asymmetric information, are included in the model to explain insurance market breakdown, other competing forces driving agglomeration can be important; see, for example, Berliant and Kung (2010), where it is shown that adverse selection alone can generate agglomeration. In other words, this criticism of the internal structure of the models, for example, when there is a nonzero moving cost, is that the consumers are not behaving rationally if they do not insure or self-insure.

Next we present a discussion of why insurance market breakdown is not natural in the context of the models. Again, this is not meant to be a statement about the real world, but rather about whether the exclusion of an insurance option for consumers in the models makes sense.

The usual cause of a breakdown of insurance markets is adverse selection, represented, for example, by cream-skimming on the part of insurance companies. In the models discussed here, the state is assumed to be realized and observable to all before decisions are made in a given time period. So there is no issue of adverse selection. But one can easily imagine variations of these models that incorporate some form of information asymmetry. It would not be natural for, say, only consumers to know the shock to the local economy, since the technology shock really affects firms. If only firms knew the realization of the shock before making their decisions, then consumers could draw

TABLE 3. City population summary statistics for Germany, Belgium, and the United States.

Country	Germany	Belgium	U.S.
Year	1998	2000	2000
Mean	152,684.8	62,959.74	156,903.7
Standard error	19,194.33	7,372.42	15,141.98
Median	77,486	39,261	80,537
Standard deviation	286,632.5	61,239.92	391,062
Sample variance	8.22E+10	3.75E+09	1.53E+11
Minimum	20,425	24,791	50,052
Maximum	3,425,759	446,525	8,008,278
Count	223	69	667

TABLE 4. Parameter estimates and related statistics for Europe.

Country	Year	Distribution	Method	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\xi}$	$\widehat{\kappa}$	$\widehat{\gamma}$	KS	LogLH	AIC	BIC
Germany	1998	Log-normal	MLE	11.74	0.9865				0.1737	−2577	5158	5164
		GEV	MLE	2.242	1.159	1.081	3.179E+04	0.5085	0.05415	−2509	5027	5044
Belgium	2000	Log-normal	MLE	10.84	0.5697				0.2030	−806.3	1617	1594
		GEV	MLE	1.529	0.01775	0.3513	0.6682	0.9626	0.1064	−786.6	1583	1621

TABLE 5. Comparison of European and U.S. results.

	Germany	Belgium	U.S.
AIC log-normal/GEV	1.026	1.021	1.0101
BIC log-normal/GEV	1.024	0.983	1.0096

inferences from firm behavior or the consumers could self-insure or insure. It is not clear what hidden information or hidden action on the part of *consumers* would cause an insurance market breakdown in this context, given that the shock is to firms' technologies. It is natural to assume that amenities are observable.

One can imagine moral hazard at the city level with insurance markets, in that a city might try to claim a productivity level lower than the actual one so the residents can collect more insurance money. However, there are no local governments in the models in the literature to coordinate this, and the assumption is that local productivity is observable to all, including nonresidents of the city, when they make their location decisions.

Another objection that could be raised is the commitment required on the part of consumers. In fact, commitment to a plan or contract is a requirement of models that feature self-insurance, insurance, or futures markets generally. For example, a consumer might experience regret over the purchase of a long-term health insurance contract after the state of the world that tells him that he is healthy is realized or the insurance company might experience regret if the consumer turns out to be unhealthy, but they are committed to their contracts. In the models of the size distribution of cities, for example, one could begin the random process of technological change and at any point in time, allow insurance and commitment to begin. Then the population distribution will not change from that point on.

Self-insurance through borrowing and saving requires a long-term commitment to a plan. Insurance cooperatives or firms only require a one period commitment to stay in a city and work. The latter commitment problem can be solved with the following time line, which is a standard time line for insurance in the real world. First, people are in a city from last period. They make an insurance premium payment to the insurance company equal to the maximum possible income for a shock this period less the income workers received from work last period in the city. This will be "small" since the random shock is small, as explained in detail in Appendix A. Then they work and the shock for this period is realized (timing here is not important). Then any insurance payment is made from the pool to obtain the average income. After that, the next period begins.

This way, people cannot receive income and then move without sacrificing their insurance. Since in all equilibria the utility levels in every city in Eeckhout's model are the same, they must lose utility by moving and giving up insurance (the loss is their premium). Of course, one could then say that the insurance company could abscond with the money. But this stretches credulity.

One might easily object to even small moving costs or even a small amount of noise in consumer observations of shocks. Then what we present is another equilibrium, which yields exactly the same period by period utility as the equilibrium studied in this literature. This alternative equilibrium retains the initial distribution of consumers and does not generate Zipf's law.

Finally, there are costs associated with insurance contracts that, from the point of view of consumers, must be balanced against the cost of moving. Such costs involve lawyers and potentially complex transactions. Moreover, unemployment insurance might fulfill the role of explicit contracts. Self-insurance does not suffer from these problems. But credit constraints could limit self-insurance. In any case, insurance does not need to be perfect. If there is substitution between insurance and mobility, the type of mobility needed to generate the various empirical distributions of city size can be upset.

But we emphasize again that although these various insurance market imperfections can cause insurance market breakdown, their inclusion in a formal model is necessary to ensure that consumers behave rationally when they do not insure, and the consequences of their inclusion are far from obvious.

An alternative to insurance markets is self-insurance. Even if insurance markets are excluded from a model by assumption, for example, because they are not observed in the real world, self-insurance must also be excluded and this exclusion must be justified.

In Appendix A, we modify a model from the literature (Eeckhout (2004)) to include insurance (as well as moving cost) and to prove our claims formally. This represents an example. We conjecture that the other models in the literature can be modified in a similar fashion.

4. Conclusions

We are making several related points.

- First, when a model, markedly different from those found previously in the literature, is constructed to explain a specific empirical phenomenon, the microeconomic, structural assumptions about individual behavior and markets must make sense. Here, there is a rather obvious problem that self-insurance and insurance markets are assumed not to be functional. Models in the literature feature city-level risk, and it is generally possible to insure against such risk through many vehicles, barring asymmetric information. The latter does not arise naturally in these models, since consumers are assumed to know the state of nature before making their location and consumption decisions.
- With time in the model, it is even possible to insure against aggregate risk through borrowing and saving.
- However, it is much more difficult to insure against extreme aggregate risk, so we propose such a model. Our model begins with microfoundations and delivers a different functional form for the size distribution of cities than has been used in the literature.

In summary, we first propose a model based on primitive assumptions, not designed to match any particular stylized fact (like the rank size rule), but rather capturing the following theoretical notion: Insurance is allowed, but consumers will never use it, as it is very costly; instead, they move. The new model is based on extreme value theory and yields a functional form for the size distribution of cities different from the other models, and this prediction is empirically competitive with those in the literature. Then we advance a criticism of the literature based on the fact that a primitive assumption in previous work, that consumers cannot insure (either by borrowing and saving or by pooling resources) against the random productivity variable for each city that is observable to all. If insurance is allowed, there is another equilibrium of the model—retaining the initial distribution of consumers where there is never any migration. Instead, consumers insure against the risk, and the utility stream they obtain in this manner is the same as that in the equilibrium used in the literature. If there is any moving cost or residual uncertainty, the equilibrium used in the literature disappears.

Is insurance or self-insurance an important issue for the analysis of the size distribution of cities and city growth? The presence of insurance has no effect on our model, since it will never be taken up, and is simply prohibited in the other models in the literature. Thus, direct evidence regarding insurance or self-insurance is insufficient to distinguish between the models empirically. From the theoretical viewpoint, it makes no difference whether or not insurance is prohibited in our model, as the equilibrium is unchanged. But it makes a huge difference whether insurance is prohibited in other models, as the equilibrium with insurance and the equilibrium without insurance are vastly different. Other models from the literature that are modified to include insurance will not generate Zipf's law or Gibrat's law. It is in this sense that abstraction from consideration of insurance or self-insurance by other models in the literature is a first-order issue.

Future work includes testing further predictions of the model, for example, the wage and rent distributions when transport costs for consumption commodities are introduced, and applying the model in new (but appropriate) contexts, such as finance (see Gabaix, Gopikrishnan, Plerou, and Stanley (2003) for an application of Gibrat's law to finance) or crop abundance (see Halloy (1999) for an application of the log-normal distribution to crop abundance).

Application to the size distribution of firms is of interest; see, for example, Axtell (2001) in the context of Zipf's law or Gabaix (2011) more generally. Frequent churning might be expected more in firms than in cities. There are two issues with this idea. First, in an aspatial model, moving between firms is easy for workers, so our insurance critique will not apply to models using the log-normal or Pareto distributions, which therefore might be more appropriate. Second, we are using a competitive model since there is a continuum of firms in each city producing the same commodity and subject to the same productivity shock. The competitive assumption might not make as much sense in an aspatial model where productivity shocks are firm-specific, so only one firm has the state of the art production technology.

Finally, an interesting direction for future research is to merge our model with that of Eaton and Kortum (2002). Such a model would be very complicated. As an alternative to this approach, adding an iceberg transportation cost to our model, as we have done in simulations, seems more worthwhile.

Appendix A: A model from the literature modified to include insurance

A.1 *Notation*

We use the model of Eeckhout (2004) as the basis for the analysis because it is explicit about consumer behavior, in the form of an optimization problem, as well as endogenous urban variables, namely local wages and land rents.

The original model is specified as follows. For complete detail, see Eeckhout (2004, pp. 1445–1446). In general, there are a large number of cities and a continuum of identical consumers. Each city produces the same commodity using labor and a constant returns to scale technology. The production function is dependent on a citywide shock and on a positive agglomeration externality that is a function of city population. There is also a negative congestion externality that is a function of city population and that only affects consumers. On net, the random shocks to productivity cause some, but not all, population to move each period so as to equalize utility across cities in equilibrium.

Time is discrete and indexed by t. The set of cities is indexed by $i \in I$. Consumers are infinitely lived and identical. In city i at time t, consumption good is $c_{i,t}$, housing or land consumption is $h_{i,t}$, whereas leisure is $1 - l_{i,t}$ for labor supply $l_{i,t} \in [0, 1]$. Utility for a consumer in city i at time t is Cobb–Douglas:

$$u(c_{i,t}, h_{i,t}, l_{i,t}) = c_{i,t}^{\alpha} h_{i,t}^{\beta} (1 - l_{i,t})^{1-\alpha-\beta}$$

with $\alpha, \beta, \alpha + \beta \in (0, 1)$.

Production is constant returns to scale. The measure of population in city i at time t is $S_{i,t}$. Let $A_{i,t}$ be the technological productivity parameter of city i at time t. This parameter follows the law of motion

$$A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t}), \tag{8}$$

where $\sigma_{i,t}$ is the exogenous technological shock to city i at time t. It is assumed that $\sigma_{i,t}$ is i.i.d. with mean 0, symmetrically distributed, and satisfies $1 + \sigma_{i,t} > 0$. The positive local externality (spillover) function is given by $a_+(S_{i,t}) > 0$, where $a'_+(S_{i,t}) > 0$. The marginal product of a worker in city i at time t is given by

$$y_{i,t} = A_{i,t} a_+(S_{i,t}).$$

For prices, let the consumption good be numéraire, let the price of housing or land in city i at time t be $p_{i,t}$, and let the wage in city i at time t be $w_{i,t}$. The local negative externality or congestion function is given by $a_-(S_{i,t}) \in [0, 1]$, where $a'_-(S_{i,t}) < 0$. The optimization problem of a consumer in city i at time t is

$$\max_{\{c_{i,t}, h_{i,t}, l_{i,t}\}} c_{i,t}^{\alpha} h_{i,t}^{\beta} (1 - l_{i,t})^{1-\alpha-\beta}$$

subject to

$$c_{i,t} + p_{i,t} h_{i,t} \leq w_{i,t} L_{i,t},$$

where $w_{i,t} = A_{i,t} a_+(S_{i,t})$ and $L_{i,t} = a_-(S_{i,t}) l_{i,t}$. Total land or housing in a city is H.

Using the first-order conditions from this optimization problem and market clearance, equilibrium (denoted by asterisks) in city i at time t as a function of population $S_{i,t}$ can be found:

$$p_{i,t}^{*} = \frac{\beta A_{i,t} a_+(S_{i,t}) a_-(S_{i,t}) S_{i,t}}{H},$$

$$w_{i,t}^{*} = A_{i,t} a_+(S_{i,t}),$$

$$c_{i,t}^{*} = \alpha A_{i,t} a_+(S_{i,t}) a_-(S_{i,t}),$$

$$h^*_{i,t} = \frac{H}{S_{i,t}},$$

$$l^*_{i,t} = \alpha + \beta.$$

The last equation in particular, indicating that labor supply is independent of population, is an artifact of the Cobb–Douglas specification.

Substituting back into the utility function, indirect equilibrium utility as a function of population $u^*(S_{i,t})$ can be written as

$$u^*(S_{i,t}) = \left[\alpha A_{i,t} \cdot a_+(S_{i,t})a_-(S_{i,t})\right]^\alpha S_{i,t}^{-\beta} H^\beta [1 - \alpha - \beta]^{1-\alpha-\beta}. \tag{9}$$

Under free mobility of consumers, indirect utility is equated across cities in each time period, determining their populations as a function of their productivity and their realized history of shocks, summarized by $A_{i,t}$. Instantaneous utility is constant over both time and location in equilibrium. Again using Eeckhout's notation, call this instantaneous utility level U.

A.2 *Insurance*

Let the discount factor be denoted by $\rho \in (0, 1]$. In correspondence with the assumption of complete capital markets, it is assumed that all consumers can borrow or lend at rate $\frac{1}{\rho} - 1$. The consumer optimization problem (at time 0) becomes

$$\max_{\{c_{i,t}, h_{i,t}, l_{i,t}\}} \sum_{t=1}^{\infty} \rho^t \cdot c_{i,t}^\alpha h_{i,t}^\beta (1 - l_{i,t})^{1-\alpha-\beta}$$

subject to

$$\sum_{t=1}^{\infty} \rho^t \cdot (c_{i,t} + p_{i,t} h_{i,t}) \le \sum_{t=1}^{\infty} \rho^t \cdot w_{i,t} L_{i,t}.$$

As stated by Eeckhout, the problem reduces to the one period optimization problem if there are no insurance or futures markets. Formally, there should be an expectation in the objective function and a requirement that the budget constraint hold for every state of nature. However, this is omitted in the literature since the problem is reduced to a static optimization problem where the state of nature is observed before consumers make their choices.

There are several important points to be made at this juncture. First, it is useful to imagine the consumers stepping back at $t = 0$ and making decisions about their cities of residence and their consumption bundles for the entire time stream of their infinite lives, contingent on state realizations at each time. Second, and more important, *it does not matter which interpretation of the model one employs.* Specifically, resources can be transferred across states of the world (at any given time) in one or more of several ways (insurance, self-insurance, or futures contracts). In the end, what a consumer is choosing is their residence and consumption bundle for every time and for every possible state of the world, optimizing utility subject to the budget constraint. The state of the world at time t affects the optimization problem through the prices, $p_{i,t}$ and $w_{i,t}$, and income (through $a_-(S_{i,t})$ and $L_{i,t}$) only. These variables depend on $A_{i,t}$ both directly and indirectly, the latter because $S_{i,t}$ depends on $A_{i,t}$ in equilibrium. The state of the world at time t does not enter into the consumer optimization problem otherwise. For example, it does not enter into the utility function. We could index these prices and incomes by the state of the world, but that would only serve to complicate notation.

As already mentioned, what will matter are only the lifetime choices of residence and consumption bundles, contingent on the state of the world in each period. The method used to actually implement them, via *transfers across states* in a time period as opposed to across time periods, does not matter; there are many possibilities. With complete futures markets, at time $t=0$ the consumers can sell their labor in every future time period and state, buying consumption good and housing in every future time period and state. With insurance markets, at $t=0$, the consumers can buy actuarially fair insurance against price and income changes. With self-insurance, they can commit to a plan of borrowing and saving under all possible scenarios, namely realizations of states in each time period.

To get the basic idea across, in the next subsection we show how insurance would work from the beginning when all cities have the same initial state (productivity) and population. This yields no movement at any time in equilibrium. In the next subsection, we discuss how to extend this so that insurance can begin from equilibrium of the model at any time t. From that time on, there is no consumer movement unless the insurance is switched off.

A.2.1 *Insurance when the initial state is the same for all cities* To illustrate the ideas behind insurance, we begin with an example where all cities start with the same state at time 0 and consumers insure from then on.

For notational purposes, let $\overline{S}$ be the mean population of cities, that is, $\overline{S}=\frac{\sum_{i\in I}S_{i,t}}{|I|}$, where $|I|$ is the cardinality of the set I. Let $A_0=A_{i,0}$ denote the common initial technology level for all the identical cities before the process begins. Let $S_{i,0}=\overline{S}$ for all cities i, so they all have the same initial population. We assume that

$$U=u^*(\overline{S})=\bigl[\alpha A_0\cdot a_+(\overline{S})a_-(\overline{S})\bigr]^\alpha\overline{S}^{-\beta}H^\beta[1-\alpha-\beta]^{1-\alpha-\beta}.$$

Thus, we assume for illustrative purposes that the initial configuration of shock A_0 and uniform population distribution $\overline{S}$ generates the instantaneous equilibrium utility. This is to get the idea across; in the next section, we will show how to start insurance from equilibrium at an arbitrary given time. In either case, no consumer movement will occur once insurance begins.

With insurance, self-insurance, or a futures market (or some combination of all three), we propose the following equilibrium solution for all cities i and times t:

$$\overline{p}_{i,t}=\frac{\beta A_0 a_+(\overline{S})a_-(\overline{S})\overline{S}}{H},$$

$$\overline{w}_{i,t}=A_0 a_+(\overline{S}),$$

$$\overline{c}_{i,t} = \alpha A_0 a_+(\overline{S}) a_-(\overline{S}),$$
$$\overline{h}_{i,t} = \frac{H}{\overline{S}},$$
$$\overline{l}_{i,t} = \alpha + \beta.$$

In other words, this is the allocation generated by a constant, over both time and state, allocation with a uniform distribution of consumers. By construction, it generates the same instantaneous utility stream for all consumers in all cities and in all times as both the initial distribution and the equilibrium studied by Eeckhout.

But how does this work in a pragmatic sense? Regarding futures markets, each consumer works the same hours, independent of state. If the state realization is good, that is, if the consumer is in city i at time 0 and $A_{i,t} > A_0$, income in excess of $A_0 a_+(\overline{S}) \cdot (\alpha + \beta)$ is paid to the market. If the state realization is bad, then the consumer receives income from the market, smoothing consumption. Under self-insurance, the consumer commits to a plan of saving income in a good state, and withdrawing from savings or borrowing in a bad state, thus smoothing consumption. The banks know that $E(A_{i,t}) = A_0$, so they are willing to lend. Under mutual insurance, the same type of idea, with commitment, has consumers who are in cities with good states at time t contributing to an insurance pool, and those in cities with bad states receiving payments from an insurance pool. If the number of cities is large, the law of large numbers implies that the mutual insurance pool is solvent.

It is interesting to note that the phenomenon we describe is something like another manifestation of Starrett's spatial impossibility theorem (see Mills (1967), Starrett (1978), Fujita (1986), and Fujita and Thisse (2002, Chapter 2.3)), though here markets are incomplete due to the presence of unpriced local externalities, both positive (a_+) and negative (a_-). In particular, we obtain a uniform distribution of economic activity, in spite of the violation of one of the hypotheses of the theorem, namely perfect and complete markets. It is well known (from these cites) that the hypotheses of Starrett's theorem are sufficient but not necessary for the conclusion, namely the lack of agglomeration.

In summary, the equilibrium time path of utility for every consumer is the same, and constant, under insurance and under the equilibrium that generates movement and eventually becomes log-normal. At the very least, a discussion of why the latter equilibrium is selected should be offered in the literature.

With any moving cost, the insurance or futures market equilibrium (the one denoted with bars) clearly dominates the path with asterisks, the one put forth in the literature. Given a choice between moving along the equilibrium path or insuring at $t = 0$, each consumer will individually choose to insure.

A second, and perhaps more reasonable, possibility is that consumers observe $A_{i,t}$ imperfectly when they make their location decisions each period. In that case as well, the consumers will insure rather than move, since they are risk averse. This can be seen in equation (9). When consumers cannot perfectly observe $A_{i,t}$ when optimizing, equilibrium expected utility will vary in proportion to $E(A_{i,t})^{\alpha}$.

A.2.2 *Insurance starting when the state is an equilibrium at a given time* The preceding subsection was provided to give intuition. However, it has drawbacks in terms of commitment on the part of consumers if they use mutual insurance at each given time, and on the part of banks and consumers at time 0 if the consumers use self-insurance. Moreover, there is a strong assumption that at time 0, A_0 is the same across cities, each city has the same population $\overline{S}$, and this combination produces the instantaneous equilibrium utility level. Here we discuss how to dispense with some of these assumptions.

Suppose that we start running the model without insurance, so that consumers are generally moving around, and stop it at some arbitrary time t. At this time, the instantaneous utility level of each consumer is, of course, U. Consider a consumer in city i and the possibility of self-insurance. At that point, the productivity parameter in the city is $A_{i,t}$, and everyone knows from equation (8) that for $t' > t$, $E(A_{i,t'}) = A_{i,t}$. So if the consumers in that city freeze their consumption bundle at whatever it is at that time, and commit to staying in that city and consuming that consumption bundle forever through a plan of borrowing and saving, they will obtain utility level U in each period. This exploits the law of large numbers over time.

Mutual insurance, exploiting the law of large numbers over space at a given time, is more interesting. Pick an arbitrary time t and freeze all the consumers in their equilibrium locations as well as their consumption bundles. All consumers obtain utility U in this situation at time t. Now consider what would happen if they maintain the same location and consumption bundle in time $t+1$. Given equation (8), the surplus or deficit in total wage payments for city i relative to the benchmark inherited from the previous period t is

$$\sigma_{i,t+1} \cdot A_{i,t} \cdot a_+(S_{i,t}) \cdot (\alpha+\beta) \cdot S_{i,t}. \tag{10}$$

Thus, to ensure that this system of mutual insurance across cities is solvent at time $t+1$, it is necessary that

$$\sum_{i\in I} \frac{1}{|I|} \cdot \sigma_{i,t+1} \cdot A_{i,t} \cdot a_+(S_{i,t}) \cdot (\alpha+\beta) \cdot S_{i,t} = 0.$$

Although this cannot be assured for finite $|I|$, we can see that as the number of cities $|I|$ tends to infinity, the limiting result is a consequence of a law of large numbers with weights given by $\frac{1}{|I|} \cdot A_{i,t} \cdot a_+(S_{i,t}) \cdot (\alpha+\beta) \cdot S_{i,t}$.

Since the support of the random variable $1+\sigma_{i,t}$ is contained in $(0,2)$, equation (8) implies that the size of $A_{i,t}$ at given time t can be bounded over i by $2^t A_{i,0}$. Since $A_{i,t}$ and $S_{i,t}$ are positively related, there is also a bound for $S_{i,t}$ and thus for the continuous function $a_+(S_{i,t})$ for fixed t over i. There is an extensive literature on the law of large numbers for sums of weighted random variables. Our framework would fit, for example, in Cabrera and Volodin (2005, Corollary 1).

Notice that there is no commitment required under mutual insurance beyond the next period. So it can be switched on and off as desired, with no consumer movement when it is on and movement when it is off. If insurance is carried on to period $t+2$, then expression (10) updated to time $t+2$ represents the change in the surplus or deficit in total wage payments for city i relative to time $t+1$, so solvency at time $t+2$ requires that these changes sum to zero across cities.

There is a subtle issue of commitment relevant to mutual insurance that is not as subtle for self-insurance. An interesting strategy for consumers is to insure for the first period, wait for uncertainty in the first period to be resolved, then pick a winner (a location where $A_{i,t+1} > A_{i,t}$), move there, and then reinsure. If a long-term commitment to insurance is required, then of course (similar to self-insurance) this strategy is not possible. Alternatively, if such commitment is not desirable, we could make an assumption about off the equilibrium path beliefs. That is, we could assume that if a consumer moves to a winning location, that person thinks that others will also follow this strategy, driving down instantaneous utility until it is again equalized across locations. Thus, this strategy does not do better than insuring in every period.

A.2.3 *Extensions: Partial insurance and moving cost* Before discussing partial insurance, it is useful at this point to make some remarks and to give some detail about equilibrium in Eeckhout's model with moving costs. In such a model, equilibria do not feature equalization of utility across locations, and this makes matters more complicated. We assume that a moving cost, if nonzero, is paid in terms of the numéraire consumption good. We call it $M \geq 0$. The main fact we need is that the explicit utility function is strongly monotonic in consumption good.

In this framework, we assume that before uncertainty is realized in a given period, each consumer must choose between a commitment to stay in their current location (and possibly insure against the uncertainty) or to move (*after* the realizations of all shocks that period are known to everyone).

First, consider the situation where the initial distribution generates the same instantaneous utility, independent of location. (With moving cost, this might not happen, but it is useful for thinking about the problem.) Then we will consider the situation where the initial distribution does not generate the same instantaneous utility.

The first claim we make is that under these conditions, for fixed positive M, every consumer will unilaterally decide to commit to stay and insure. The reason is that for Eeckhout's arguments to work, uncertainty must be arbitrarily small. This is made explicit by Eeckhout (2004, p. 1447), where $\varepsilon_{i,t}$ is written as an increasing function of $\sigma_{i,t}$, and asymptotic statements such as $\varepsilon_{i,t} \to 0$ are used to derive Gibrat's law. So for any fixed $M > 0$, we can find sufficiently small $\varepsilon_{i,t}$, and thus $\sigma_{i,t}$ (or its support), so that each consumer would find it more costly to move than to insure. Thus, the potential gain from moving will be less than the cost, namely M. If consumers have heterogeneous moving costs, as long as there is not an atom of consumers at zero moving cost, as the random variables representing the shocks become small, the measure of consumers who prefer to move rather than insure will tend to zero.

Clearly, in the original equilibrium with consumer movement, the measure of consumers who actually move between periods to equate utility across locations is relatively small. But given the last argument, no individual will want to be among the movers. And over time, the social cost of moving will add up.

Given that nobody chooses to move, we can next calculate the amount of insurance they will purchase. Since $S_{i,t+1} = S_{i,t}$, we can calculate from (9) and (8) that

$$Eu^*(A_{i,t+1} \mid A_{i,t}, S_{i,t}) = E\Big(\big[\alpha\big\{A_{i,t}(1+\delta\cdot\sigma_{i,t+1})\big\} \cdot a_+(S_{i,t})a_-(S_{i,t})\big]^\alpha [S_{i,t}]^{-\beta} H^\beta [1-\alpha-\beta]^{1-\alpha-\beta}\Big),$$

where $1-\delta$ is the percentage of insurance purchased. By concavity, $E\{(1+\delta\cdot\sigma_{i,t+1})^\alpha\}$ is maximized at $\delta = 0$, so everyone purchases full insurance. Partial insurance is not a feature of equilibrium.

Finally, consider the case where the initial indirect utility levels are not equalized across locations in equilibrium due to the moving cost. In fact, due to costly mobility, the utility difference in terms of numéraire cannot exceed M, for otherwise consumers would move to the location with higher gross utility (as it exactly compensates for the moving cost). But since consumers can look ahead, they realize that this is not just a one period difference in utility levels. In other words, after sinking moving cost this period, next period it is expected that the higher utility location will yield the same utility (for example, by fully insuring) without having to pay the moving cost. So more people will move there this period. The present discounted value of the utility difference will be at most the moving cost. So in the first period under consideration, people might move. But after that, the population distribution will be stable, and everyone in all locations will fully insure for the same reason as given above.

Appendix B: Positive transport costs

Before turning to the details of our simulations, we first summarize our conclusions. Table 6 reports the statistics from our simulations. Typically, in equilibrium, there will be more than one city producing the same commodity to serve nearby cities. Whereas both the log-normal and GEV distributions track the simulated data well, the results are inconclusive concerning which of the two distributions better explains the economy with multiple production sites. The KS statistic is small for both distributions; AIC and BIC are lower for the log-normal distribution when the iceberg transportation cost is large, but in the end the differences are small.

Now let us turn to the details of the simulations. We have 10 industries with 30 potential production sites each, so there are 300 potential city locations lined up on a circle. Transportation cost is of the iceberg form. In particular, we ship out 1.01, 1.05, or 1.1 units of commodity and an immediate neighbor on the circle receives 1 unit of it. Shipment cost grows with distance traveled. City residents purchase goods from the city that quotes the smallest delivered price within each industry. Everything else is the same as the basic model we have described.

The results from our simulations depart from the results from the basic model in two ways. As is the case for the basic model, high productivity reduces the mill price. However, it is not practical to serve the entire population from a single city just because that city is the most productive in the industry; delivered price grows with distance. Rather,

Table 6. Simulations with positive iceberg transport cost.

Iceberg	Productive Cities/Cities	Distribution	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\xi}$	$\widehat{\kappa}$	$\widehat{\gamma}$	LogLH	AIC	BIC	KS
1.01	78/300	Log-normal	−4.624	0.8466				−275.2	554.4	559.1	8.343E−2
1.01	78/300	GEV	4.845E−2	1.051E−2	−0.4621	1233	0.7967	−267.8	545.5	557.3	7.152E−2
1.02	107/300	Log-normal	−5.250	1.1704				−345.2	694.5	699.8	4.886E−2
1.02	107/300	GEV	4.240E−2	1.711E−2	−0.04938	27.60	0.7377	−346.1	702.2	715.5	5.477E−2
1.05	92/300	Log-normal	−5.301	1.3861				−307.5	618.9	623.9	1.062E−1
1.05	92/300	GEV	4.589E−2	1.491E−2	−0.1759	1839	0.8107	−310.7	631.4	644.0	1.028E−1
1.1	93/300	Log-normal	−5.345	1.4345				−319.1	642.2	647.4	8.516E−2
1.1	93/300	GEV	1.456E−1	3.837E−2	−0.3230	250.5	0.8530	−321.6	653.2	666.0	8.148E−2

a couple of cities will coexist within the same industry to serve the cities in close proximity to each of them. Also, surviving cities are not necessarily the most productive in the industry. It depends not only on a city's own productivity, but also on the size and productivity of close neighbors.

For the 300 locations, we mark an industry by $j = 1, \dots, 10$ and a location block of 10 cities by $i = 1, \dots, 30$. On a circle, we have city $(i, j) = (1, 1)$ located next to $(1, 2)$, next to $(1, 3)$, and so on. Then, next to $(1, 10)$ we have $(2, 1)$. As we travel along the circle in this way, we will eventually reach $(30, 10)$, whose next neighbor is $(1, 1)$, completing the circle. Figure 5 is a schematic representation of the arrangement of cities so that we can illustrate how the model works. A circle with 300 dots would be difficult to interpret, so instead we give the industry on the horizontal axis and the location block on the vertical axis.

Figure 5(a) represents the size of the random productivity draw A, where a larger dot represents a larger value of A. We shall use this single draw of the random variable for subsequent illustrations. In Figure 5(b) we represent the equilibrium city-size distribution for a transport cost of 1.01, where a larger dot means a larger equilibrium

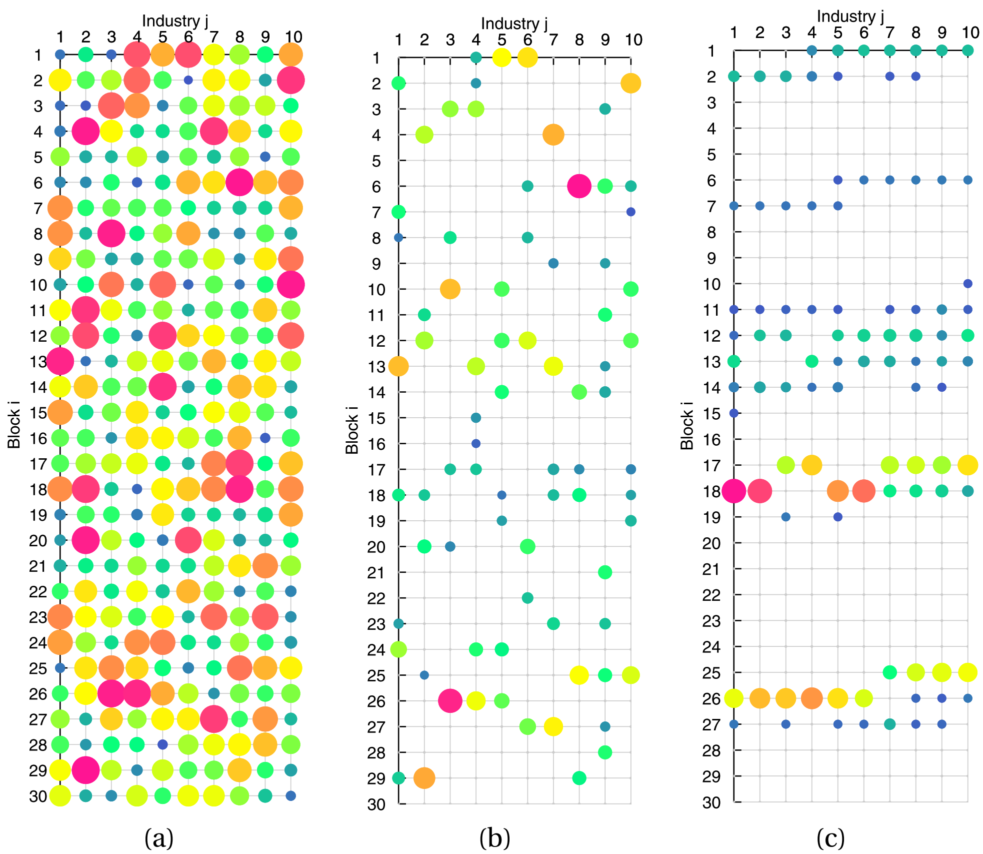


FIGURE 5. (a) Productivity parameter A. (b) City size ($\tau = 1.01$). (c) City size ($\tau = 1.1$).

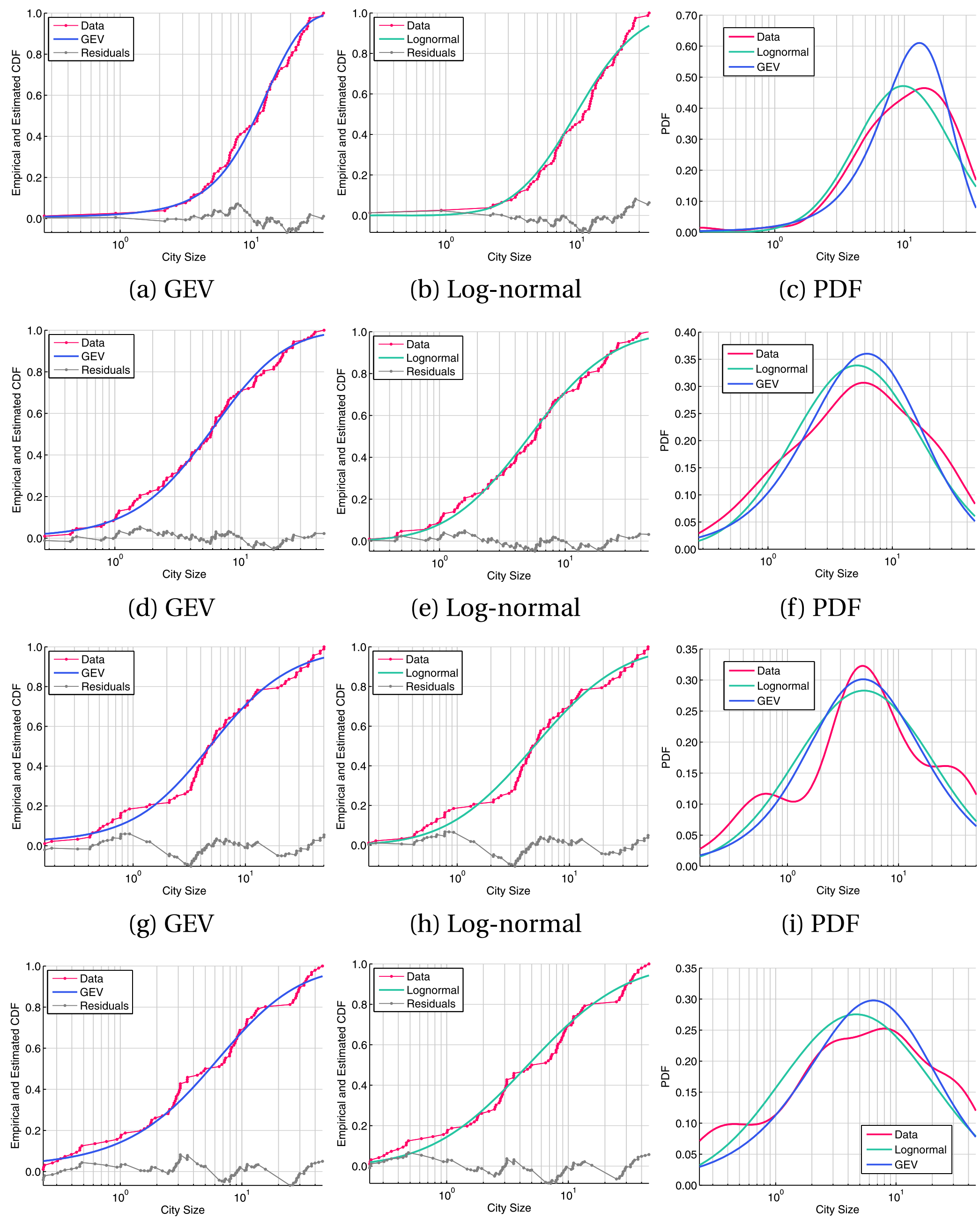
population and no dot means the area is rural. Such a small transport cost causes only a minor change from our basic model. Take industry $j = 8$ for example. The most productive city is $(i, j) = (6, 8)$. This would be the city that survives in the absence of any transport cost. City $(6, 8)$ still survives and produces most of commodity $j = 8$ when we throw in a minimal shipment fee. However, there are some other cities, such as $(25, 8)$, that also engage in production to serve local markets. City $(25, 8)$ undercuts the delivered price of city $(6, 8)$ for nearby cities. Indeed, shipment from $(6, 8)$ to $(25, 8)$ would require an impractical 189 steps.

As we raise transport cost, a qualitatively different city-size distribution emerges in equilibrium. *Own* productivity level becomes less pertinent and the size of *neighboring* cities becomes more influential in determining city size. This can be seen in Figure 5(c), where transport cost is raised to 1.1 units. Take industry 8 again. The most productive city $(6, 8)$ is still in the picture, but production is more intense in the less productive city $(17, 8)$ when transport cost is raised. City $(6, 8)$ is quite productive in isolation, but is surrounded by cities whose productivity is not exceptional. Consequently, its local market is small. On the other hand, city $(17, 8)$ is surrounded by productive cities. Since utility is concave, there is large demand for commodity 8 from these productive (and therefore, populous) neighbors. And this large demand will be fulfilled by the nearby city $(17, 8)$ rather than a faraway city $(6, 8)$ to ward off the increased shipping charge. City $(17, 8)$ grows to support local demand that, in turn, will create a large demand for goods other than 8 produced by its neighbors. As a result, high transportation cost creates snowballing clusters of cities, whose *average* productivity across the industries within the region is high, and eliminates cities of high productivity in geographic isolation. The equilibrium does not simply select the most productive cities as survivors.

The various theories we have summarized have zero transport costs. The simulations indicate that positive transport costs can generate a new force in city selection, namely a kind of local market effect illustrated in Figure 5(c).

Figure 6 and Table 6 summarize maximum likelihood estimation for the log-normal and GEV city-size distributions from these simulations. On the whole, both distributions fit the simulations well. Notice that the number of active cities is not monotonically increasing in transport cost. The reason is that some clusters of cities empty out as transport cost increases.

Table 6 reports the estimated parameters. The last four columns are the values of the log likelihood function (larger is better), Akaike and Bayesian information criteria (AIC and BIC), and the Kolmogorov–Smirnov (KS) statistic (smaller is better for all). Judging by the values of the KS statistic, the overall fit is not bad. Turning next to the comparison between the log-normal and GEV distributions, the simulated city-size distributions yield better log likelihood values for GEV when transport cost is small, but favor log-normal for AIC and BIC (GEV has more parameters). In the end, the fits are almost identical. In additional simulations not detailed here, we found no systematic relationship between transportation cost and how well either distribution fits the data.



(a) GEV (b) Log-normal (c) PDF

(d) GEV (e) Log-normal (f) PDF

(g) GEV (h) Log-normal (i) PDF

(j) GEV (k) Log-normal (l)

FIGURE 6. Maximum likelihood estimation with iceberg transport cost (from top row): 1.01, 1.02, 1.05, 1.1.

References

Axtell, R. L. (2001), "Zipf distribution of U.S. firm sizes." *Science*, 293, 1818–1820. [174]

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014), "Productive cities: Sorting, selection, and agglomeration." *Journal of Political Economy*, 122, 507–553. [168]

Berliant, M. and F.-c. Kung (2010), "Can information asymmetry cause stratification?" *Regional Science and Urban Economics*, 40, 196–209. [170]

Cabrera, M. O. and A. I. Volodin (2005), "Mean convergence theorems and weak laws of large numbers for weighted sums of random variables under a condition of weighted integrability." *Journal of Mathematical Analysis and Applications*, 305, 644–658. [179]

Coles, S. (2001), *An Introduction to Statistical Modelling of Extreme Values*. Springer-Verlag, London. [159, 161]

Duranton, G. (2006), "Some foundations for Zipf's law: Product proliferation and local spillovers." *Regional Science and Urban Economics*, 36 (4), 542–563. [167, 168]

Duranton, G. (2007), "Urban evolutions: The fast, the slow, and the still." *American Economic Review*, 97 (1), 197–221. [154, 156, 167, 168]

Eaton, J. and S. Kortum (2002), "Technology, geography, and trade." *Econometrica*, 70 (5), 1741–1779. [156, 174]

Eeckhout, J. (2004), "Gibrat's law for (all) cities." *American Economic Review*, 94 (5), 1429–1451. [154, 156, 160, 162, 163, 164, 166, 167, 168, 169, 173, 174, 180]

Eeckhout, J. (2009), "Gibrat's law for (all) cities: Reply." *American Economic Review*, 99 (4), 1676–1683. [164]

Embrechts, P., C. Kluppelberg, and T. Mikosch (1997), *Modelling Extremal Events*. Springer-Verlag, Berlin. [159, 161, 162]

Fisher, R. A. and L. H. C. Tippett (1928), "Limiting forms of the frequency by distribution of the largest or smallest members of a sample." *Proceedings of the Cambridge Philosophical Society*, 24, 180–190. [158]

Fujita, M. (1986), "Urban land use theory." In *Location Theory* (J. Lesourne and H. Sonnenschein, eds.), Harwood Academic Publishers, New York. [178]

Fujita, M. and J.-F. Thisse (2002), *Economics of Agglomeration*. Cambridge University Press, Cambridge. [178]

Gabaix, X. (1999a), "Zipf's law for cities: An explanation." *Quarterly Journal of Economics*, 114 (3), 739–767. [166]

Gabaix, X. (1999b), "Zipf's law and the growth of cities." *American Economic Review: Papers and Proceedings*, 89 (2), 129–132. [166]

Gabaix, X. (2011), "The granular origins of aggregate fluctuations." *Econometrica*, 79 (3), 733–772. [174]

Gabaix, X., P. Gopikrishnan, V. Plerou, and H. E. Stanley (2003), "A theory of power law distributions in financial market fluctuations." *Nature*, 423, 267–270. [174]

Gabaix, X. and Y. Ioannides (2004), "Evolution of city size distributions." In *Handbook of Regional and Urban Economics IV: Cities and Geography* (J. V. Henderson and J.-F. Thisse, eds.), North-Holland, Amsterdam. [154]

Giesen, K., A. Zimmermann, and J. Suedekum (2010), "The size distribution across all cities—Double Pareto lognormal strikes." *Journal of Urban Economics*, 68 (2), 129–137. [164]

Goldstein, M. L., S. A. Morris, and G. G. Yen (2004), "Problems with fitting to the power-law distribution." *The European Physical Journal B—Condensed Matter Physics*, 41 (2), 255–258. [164]

Halloy, S. R. P. (1999), "The dynamic contribution of new crops to the agricultural economy: Is it predictable?" In *Perspectives on New Crops and New Uses* (J. Janick, ed.), ASHS Press, Alexandria, VA. [174]

Hsu, W.-T., T. Mori, and T. E. Smith (2014), "Spatial patterns and size distributions of cities." Unpublished manuscript. [166]

Mills, E. S. (1967), "An aggregative model of resource allocation in a metropolitan area." *American Economic Review*, 57 (2), 197–210. [178]

Rossi-Hansberg, E. and M. L. J. Wright (2007), "Urban structure and growth." *Review of Economic Studies*, 74 (2), 597–624. [154, 167]

Soo, K. T. (2005), "Zipf's law for cities: A cross-country investigation." *Regional Science and Urban Economics*, 35 (3), 239–263. [154, 170]

Starrett, D. (1978), "Market allocations of location choice in a model with free mobility." *Journal of Economic Theory*, 17, 21–37. [156, 178]

Submitted December, 2009. Final version accepted March, 2014.