Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds

SUB-LOGNORMAL SIZE DISTRIBUTION OF HOSPITALS – AN AGENT-BASED APPROACH AND EMPIRICAL STUDY

Baojun Gao

School of Economics and Management
Wuhan University
Wuhan, Hubei Province 430072 P.R. CHINA

Wai Kin (Victor) Chan

Department of Industrial and Systems Engineering
Rensselaer Polytechnic Institute
Troy, NY 12180 USA

ABSTRACT

This paper studies the size distribution of hospitals and its underlying generative mechanisms. Based on the U.S. hospital data, we find that the size distribution is sub-lognormal (a leptokurtic distribution more skewed than normal but less skewed than lognormal). This distribution is different from those of firms and cities. We develop an agent-based simulation model to simulate the preference behavior of patients and the service processes of hospitals. The model can produce a sub-lognormal size distribution similar to the U.S. hospital size distribution. Sensitivity analysis shows that the patients' preference behavior and search distance are two key factors for the emergence of the sub-lognormal size distribution.

1 INTRODUCTION

Many human systems are found to have a power law or lognormal distribution on their sizes or degrees. Among many examples are actor collaboration network, power grid, Internet (Barabasi and Albert 1999, Newman 2001), city sizes, firms, and frequency of words in any human language (Axtell 2001, Eeckhout 2004, Newman 2005).

The size distribution of firms and cities has been extensively studied in the last several decades, for it is the basis for modeling the growth and evolution dynamics of those systems. Back to 1931, Gibrat established a well-known proposition called the proportionate growth law or Gibrat's Law, which states that the growth rate of size is independent of the size itself (Sutton 1997). According to this law, the proportionate growth process gives rise to the lognormal distribution of size. A number of studies have shown that this law holds true for cities and firms (Simon and Bonini 1958, Hart and Oulton 1996, Eeckhout 2004). Many empirical studies have also reported that cities and firms size distributions satisfy the power law (Axtell 2001, Rozenfeld et al. 2011). Regardless of whether the size distribution is lognormal or power law, there is sufficient evidence to convince one that the size distributions of both firms and cities are highly skewed to the right.

From a service point of view, both firms and cities provide certain types of services to people (e.g., a product or a living environment). Hospitals are also one type of service systems providing healthcare services to people.

Would hospitals exhibit a size distribution similar to those of firms and cities? If not, what are the factors that make the hospital service system special? These questions will be studied in this paper. We first conduct an empirical study to show that the size distribution of hospitals is sub-lognormal as defined in the following.

Definition 1 *A* sub-lognormal distribution *is a leptokurtic continuous probability distribution that is more skewed than a normal distribution but less skewed than a lognormal distribution.*

Roughly speaking, a sub-lognormal distribution lies in-between the normal and lognormal distribution. Both the lognormal and sub-lognormal distributions are more skewed than the normal distribution. The differences between the lognormal and sub-lognormal distributions can be seen from the logarithm of the distributions. The logarithm of the lognormal distribution is normal, which is symmetric. The logarithm of a sub-lognormal distribution is left skewed. Therefore, the sub-lognormal distribution is less skewed than the lognormal distribution. In addition, to be qualified as sub-lognormal, the distribution must also be leptokurtic.

Motivated by the empirical results, we then construct an agent-based simulation model to explain the empirical results, and more importantly to reproduce the underlying mechanisms that give rise to the sub-lognormal size distribution. The agent-based method has been used for studying size distribution of firms and cities (Axtell 1999, Gulden and Hammond 2012). What makes our model different from existing ones is the integration of ABS and queuing models to mimic the service processes of hospitals. With this model, we will perform sensitivity analysis to examine the effects of various potential factors that could influence the size distribution of hospitals.

Chan (2008) introduces the incorporation of queuing models into agent-based models. Classic queuing models usually treat customers as non-autonomous objects who simply come, join, and leave the system. In other words, customers lack human-like behaviors. Agent-based simulation provides a solution to these limitations. In an agent-based model, agents are objects capable of reproducing various behaviors (Chan et al. 2010). The agent-based method allows us to investigate and test the underlying mechanisms giving rise to the size distribution.

To our best knowledge, unlike firms and cities the literature on the size distribution of hospitals is unseen. Most of the studies on hospital sizes focus on the relationship between service quality and size, and are mostly reported on the Internet rather than on journal publications (Blizzard 2004).

The rest of the paper is organized as follows. Section 2 analyzes the data for the sizes of U.S. hospitals. Section 3 introduces the agent-based model for investigating the underlying mechanisms of hospital size distribution. Section 4 presents the simulation results and sensitivity analyses. Section 5 offers a conclusion.

2 EMPIRICAL STUDY

In this section, we analyze the size distribution of U.S. hospitals. The dataset is obtained from (American Hospital Directory 2012). The size of a hospital is measured by patient days. Patient days are used for accounting purposes by healthcare facilities and healthcare planners. One patient day is "*a unit of time during which the services of the institution or facility are used by a patient*" (Medical Dictionary 2013). For example, 100 patients in a hospital for one day is equivalent to 100 patient days. Patient days can be viewed as a scale measure of a hospital in both facility size (such as the number of beds) and facility utilization. This also makes patient days a better (more informative) size measure than the number of beds. There are also other measures for size; for example, in the business and economic literature, the size of a firm can be measured by asset, sales, revenue, or number of employees. For hospitals, these statistics however are not always available to the public and can depend on the economic environment of the city, state, or country. The patient days on the other hand is more stable and less economic dependent.

After removing zero data points, the dataset contains data for 3454 hospitals across the US. Figure 1(a), (b), and (c) show the histogram of U.S. hospital sizes, the histogram of logarithmic sizes and the QQ-plot of the logarithmic sizes, respectively. Figure 1(a) shows that the size distribution is highly right skewed, which is an important characteristic of the lognormal distribution. However, a further investigation of the logarithmic sizes reveals unfavorable support for the lognormal distribution as the histogram of the logarithmic sizes deviates from a symmetric bell-shaped curve and skews to the left (see Figure 1(b)). The QQ-plot of logarithmic sizes Figure 1(c)) further shows that the sample quantiles of upper tail and lower tail are less than those of the theoretical quantiles. All these results indicate that the distribution of the logarithmic sizes is thin tailed and leptokurtic.

To formally test the characteristics of the U.S. hospital size distribution, we conduct the Shapiro-wilk normality test on the logarithmic size. The null hypothesis that the logarithmic size follows a normal distribution is rejected at the 0.001 significance level. The skewness and kurtosis of the logarithmic size are, respectively, -0.86 and 4.43, which indicates that the distribution of the logarithmic size is left skewed and leptokurtic. Therefore, the size distribution is a sub-lognormal distribution.

We next examine the data at the state level. Most of the states with more than 50 hospitals are found to have a similar characteristic as the U.S. data. Due to space limitations, we shall only present the data for California (CA), which is one of the states with a large number of hospitals. The histograms of sizes, logarithmic sizes, and Q-Q plots of CA are given in Figure 2(a)-(c). These figures share similar characteristics with the U.S. hospital data and the size distribution is also sub-lognormal.

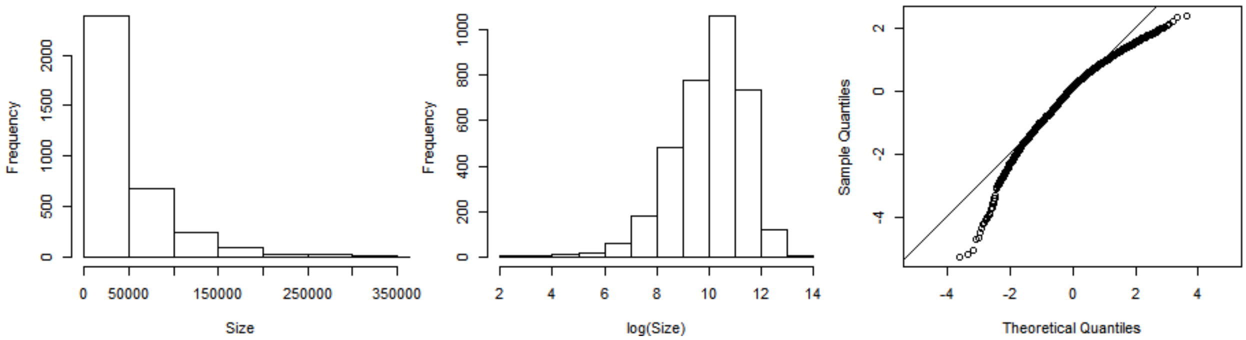

Figure 1: Hospitals Size Distribution in U.S.: (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes.

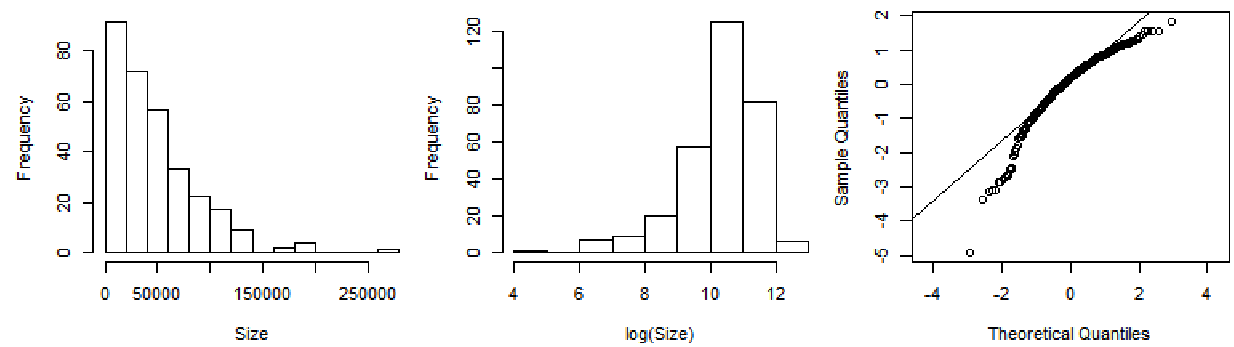

Figure 2: Hospitals Size Distribution in CA: (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes (The Shapiro-Wilk normality test on the logarithmic sizes rejects the null hypothesis with p-value<0.001. The skewness and kurtosis of logarithmic sizes are -0.84 and 3.87 respectively.)

In short, the U.S. hospitals data is sub-lognormal that exhibits the following characteristics; (1) the size distribution is right skewed; (2) the logarithmic size distribution is left skewed; and (3) the logarithmic size distribution is leptokurtic.

These results imply that the hospital size distribution is different from those of firms. In the business and economic literatures, many empirical studies show that the lognormal distribution (or a more skewed distribution like power law) is a good approximation for the size distribution of firms. For example, Cabral and Mata (2003), Angelini and Generale (2008) suggest that the logarithms of firm size of a given cohort is skewed to the right at the beginning, and gradually evolves towards a more symmetric distribution. However, for firms aged over 30, the distribution of log size is still far from symmetric. The total logarithms of firm size, in turn, is fairly stable over time, and somewhat skewed to the right.

3 GENERATIVE MODEL FOR HOSPITAL SIZE DISTRIBUTION

Why does the hospital service system differ from the service systems of firms and cities? In this section, we develop an agent-based model combined with queuing models and try to answer this question. The objective of the agent-based model is to simulate the underlying mechanisms and factors that give rise to the sub-lognormal size distribution observed in Section 2.

We assume that the hospital service system is a system with a number of hospitals providing healthcare services to a population of patients. Patients are able to choose their preferred hospitals so long as a certain number of constraints are not violated. From the supply and demand points of view, the possibility of losing patients (i.e., customers) due to long wait times (e.g., reaching the capacity) drives the hospitals to grow in size to respond to demand increase. On the other hand, hospitals may reduce their size if the demand is low to cut down cost and increase revenue. If the resource of a hospital cannot meet the current demand, the hospital will expand its size, and vice versa.

A central question is: *what will the hospitals size distribution be as a consequence of the interactions between patients and hospitals* (i.e., patients' choosing behaviors and hospitals' response to demand changes)?

To investigate the underlying mechanisms for the stylized facts of hospitals size distribution, we develop an agent-based model incorporating both the patients' choosing behavior and hospitals' response to demand changes. There are two kinds of agents: patients and hospitals. All agents are situated in a two-dimensional n x n lattice. Each cell has a hospital and each hospital consists of multiple service stations. For simplicity, we assume that all the service stations are homogeneous with identical exponential service time distribution. Let μ be the mean service time of a single service station and QS_i be the number of service stations in the i^{th} hospital. The patients are randomly distributed over the lattice with each cell containing a number of patients (see details in Section 4). We describe the behaviors of patients and hospitals in the next two subsections respectively.

3.1 The Patients

Patients visit hospitals according to a Poisson process, i.e., the intervals between two consecutive visits follow the exponential distribution. If it is the time to visit a hospital, the patients must decide which hospital they will visit. Patients can only visit the hospitals within a distance threshold, *Search-Radius*.

When making their decisions, patients mainly consider two factors: the waiting time and reputation of the hospitals. Patients prefer a high-reputation and short-waiting-time hospital over a low-reputation and long-waiting-time one.

Following the size definition in Section 2, we also use the *patient days* to measure the hospital size. In addition, patient days is also the indicator of reputation because a large patient days value tells the choice of the majority. We can assume that reputation is a monotonic increasing function of the majority choice. This assumption can be justified as studies have shown that the choice of popular products or service tends to minimize potential risk. For example, DeSarbo and colleagues (2002) argue that consumers prefer popular products because popularity represents a type of social cue, and following the social cue tends to reduce perceived risk. In a similar vein, the literature on herding suggests that it is sometimes optimal for consumers to ignore or not seek private information and to follow the crowd (e.g., Banerjee 1992; Bikhchandani, Hirshleifer, and Welch 1992).

We note that a high patient days value does not necessarily mean long waiting time if the hospital is a huge facility with sufficient beds. However, when the resources (e.g., beds) become insufficient, long waiting time would result.

With the above connection among patient days, reputation, and hospital size, we can say that the larger the size, the greater the reputation of the hospital will be. This size-reputation indicator propels the patients to choose larger size hospitals over small ones if everything else is equal.

The decision rules of how patients choose hospitals, *Patients-Choose-Hospital-Process* are defined as follows.

Patients-Choose-Hospital-Process

If there is at least one hospital available within the *Search-Radius* at time *t*,
  pick one available hospital with the probability Pa_i as defined in Eq.(1)
Else
  pick one hospital within the *Search-Radius* based the probability Pua_i as defined in Eq.(2)
Generate the next visit time

If there is at least one hospital available within the *Search-Radius* at time *t*, a patient will pick one from them with a probability proportional to the size ranks of the hospitals. Other than the ordinal ranking information, no other information is available to the patients. To pick a hospital, patients first sort all the available hospitals within the *Search-Radius* in ascending order of size and assign an ordinal number RS_i of size rank to each of them. The smallest available hospital will have a rank of one and the largest available hospital a rank of *K*, where *K* is the number of all available hospitals within the *Search-Radius* at time *t*. Specifically, the i^{th} available hospital for a patient will be selected by the patient with probability Pa_i:

$$Pa_i = \frac{RS_i}{\sum_{i=1}^{K} RS_i} \tag{1}$$

If there is no hospital available within the *Search-Radius* at time *t*, patients will make their choices by balancing the two decision factors, waiting time and hospital size. First, they still need to obtain the size rank ordinal number RS_i of all the hospitals within *Search-Radius*. Then, they sort all hospitals within *Search-Radius* in descending order based on their earliest available times. They will assign an ordinal number RW_i of waiting time rank to each of the hospitals. The hospital with the longest waiting time will have a rank one and the hospital with the shortest waiting time will get the largest waiting time rank number *L*, where *L* is the number of hospitals within the *Search-Radius* of a patient. Specifically, the i^{th} hospital within the *Search-Radius* of a patient will be selected by the patient with probability Pua_i:

$$Pua_i = a\frac{RW_i}{\sum_{i=1}^{L} RW_i} + (1-a)\frac{RS_i}{\sum_{i=1}^{L} RS_i} \tag{2}$$

where *a* is a number between 0 and 1 which is the weight of the waiting time (and size or reputation) used in the patients' decision making.

After choosing the preferred hospital, a patient needs to generate the time for the hospital next visit. Since the intervals between two consecutive hospital visits follow an exponential distribution, that time can be computed by adding an exponential random number to the current time *t*.

3.2 The Hospitals

Each hospital consists of several "standard" service stations that provide patients with medical services. Once a patient selects a hospital to visit, this selected hospital will have to either put him/her into the waiting queue or provide service immediately. This procedure is modeled by the *Service-Process* in our model and is described in the following.

The hospitals are somewhat adaptive. They will respond to the demand changes by adjusting their resources (i.e., number of service stations). They increase (or decrease) the number of service stations if the average waiting time is longer than a upper threshold (or shorter than a lower threshold). By adding more service stations, the average waiting time will drop, but the utilization will decrease as well. The hospitals try to keep an optimal balance between service quality (measured by average waiting time) and costs (measured by utilization rate). In our model, the hospitals adjust their sizes every T_{adj} time steps if needed. The adjustment rules and the adjustment process called *Adjust-Facilities-Process* are given below. It is followed by the definition of the variables and the explanation of this process.

Service-Process

```
Generate the exponential service time with mean μ.
If there is at least one idle service station in this hospital,
    Begin the service using one of the idle service stations.
Else
    Choose the earliest available service station and add this patient to the waiting queue.
Update the available time of the service station.
Update the earliest available time of the hospital.
```

Adjust-Facilities-Process

```
For each hospital
    If avg-wt_i > MAX_WT and exp-sum-svt_i > (QS_i + 1) * T_adj
        QS_i grow to int (exp-sum-svt_i / T_adj)
    If ut_r_i < L_UT or avg-wt_i < MIN_WT
        QS_i decreased by int (sum-spt_i / T_adj)
        If QS_i <1, set QS_i =1
    Reset avg-wt_i, sum-spt_i, exp-sum-svt_i, and ut_r_i to zero.
```

In the pseudo code above, *MAX_WT*, *MIN_WT* and *L_UT* are three constants specified at the beginning of the simulation. *MAX_WT* and *MIN_WT* refer to the upper bound and the lower bound of average waiting time, respectively. *L_UT* is the lower bound of the system utilization rate. QS_i, $avg\text{-}wt_i$, $sum\text{-}spt_i$ and ut_r_i are the number of service stations, the average waiting time, total spare time, and the utilization rate of the i^{th} hospital, respectively. $exp\text{-}sum\text{-}svt_i$ is the total service time (from the last adjustment till now plus the expected service time for those in queue). As at the end of this process, $avg\text{-}wt_i$, $sum\text{-}spt_i$, $exp\text{-}sum\text{-}svt_i$, and ut_r_i are reset to zero to prepare for the calculation in the next time step.

Due to the randomness in the arrival and service processes, the condition for adding service stations is augmented by the criterion "$exp\text{-}sum\text{-}svt_i > (QS_i + 1) * T_{adj}$. This is to add a safety margin on the number of service stations to handle variations in the arrival and service processes.

To reduce the number of service stations, both the utilization rate and average time are considered. The amount to decrease is equal to the total spare time divided by T_{adj}. As the size must be greater than one, it will be reset to one if the amount of decrease will cause it to go lower than one.

The main procedure of the simulation model is given in the following.

4 SIMULATION RESULTS

4.1 Benchmark Case

We build the agent-based model using NetLogo (Wilensky 1999). All the parameters and their particular values are listed in Table 1. We simulate a system with 11×11 cells. Each cell is initially occupied by one hospital. The initial patient population at each cell is generated according to the lognormal distribution because it is well-known that the size distribution of cities (measured by population) is roughly a lognormal distribution (Eeckhout 2004). The initial size of hospitals are proportional to the population of local patients. Other parameter values are selected for simplicity and are listed in Table 1.

Main-Procedure

At time *t*, each patient and hospital runs independently.
For each patient
If it is time to visit a hospital, run the *Patients-Choose-Hospital* process.
For each hospital
If there are new arrivals, run the *Service-Process*
If time *t* is a multiple of T_{adj}, run the *Adjust-Facilities-process*.
Calculate waiting time, utilization rate and other queuing system indicators.
Advance the simulation clock to next time step.

Table 1: Parameters In the Agent-based Model.

Parameters	Value
Environment	11x11 lattice (121 cells)
Number of hospitals	121
Number of patients in $Cell_i$ (lognormal distribution)	Int (100×logN(0,1))
Initial size of hospitals	Int ($Num\text{-}Patient_{ij}$ /5)
μ, Mean service time (exponential distribution)	30
Mean visit interval of patients (exponential distribution)	100
Step size (time increment)	1
MAX_WT , upper bound of average waiting time	3
MIN_WT , lower bound of average waiting time	10
L_UT, the lower bound of system utilization rate	0.9
T_{adj} , the interval to adjust the hospital sizes	500
a, weight parameter in equation (2)	0.5
Search-Radius	2
Patients' size preference	larger

The simulation results are shown in Figure 3(a)-(c). It can been seen that the agent-based model can adequately reproduce the characteristics of the U.S. hospital data; the size distribution is highly skewed with its logarithmic skewed to the left and leptokurtic, i.e., a sub-lognormal distribution.

As we will see later, among all the parameters in the model, the patients' size preference and search radius are the key factors that allow the model to reproduce the characteristics of the U.S. hospitals size distribution. This simulation along with the parameters in Table 1 will be the benchmark case for comparison with other settings in the sensitivity analyses in Section 0.

4.2 Sensitivity Analysis

In this section, we conduct sensitivity analysis on the patients' size preference, search radius, initial beds distribution and environment size.

In Section 4.1, we assume that the patients prefer larger size hospitals. We now examine how this size preference affects the results. We study both neutral and opposite situations: What will happen if the patients have no preference to size? And what will happen if the patients prefer *smaller hospitals*? Figure 4 and Figure 5 give the results for the neutral size preference and smaller size preference, respectively. Statistical tests fail to reject the null hypotheses of lognormal distribution. Therefore, it is clear that size preferences play an important role in shaping the sub-lognormal size distribution.

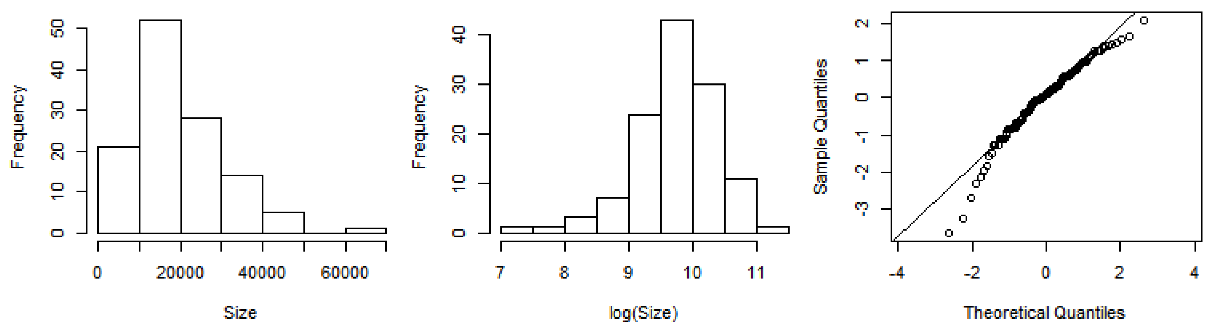

Figure 3: Simulation Results of Benchmark Case: (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes (The Shapiro-Wilk normality test on the logarithmic sizes rejects the null hypothesis with p-value<0.001. The skewness and kurtosis of logarithmic sizes are -0.98 and 4.70 respectively.).

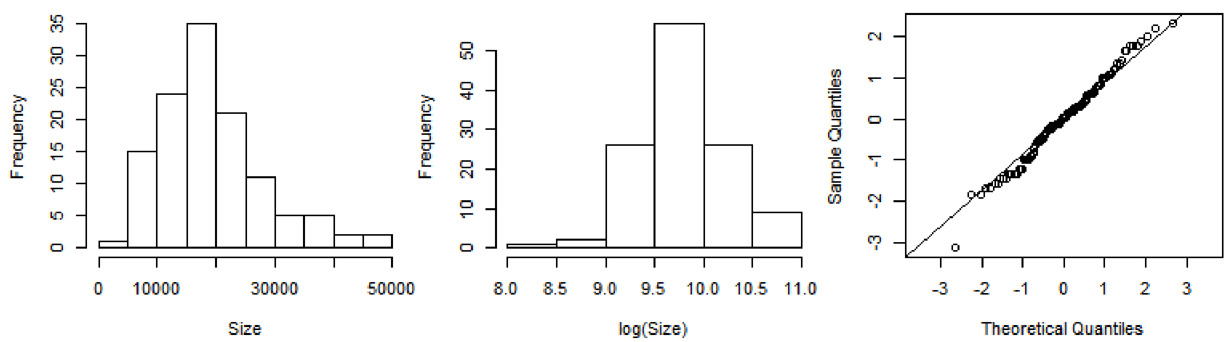

Figure 4: Neutral Size Preference: (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes (The Shapiro-Wilk normality test on the logarithmic sizes fails to reject the null hypothesis with p-value=0.83. The skewness and kurtosis of logarithmic sizes are -0.09 and 3.10 respectively.).

Next, we investigate whether the *Search-Radius* is a key factor for the sub-lognormal distribution. We test two scenarios: no search (i.e., a search radius 0) and global search (i.e., a search radius 5 as the environment is a 11×11 lattice). All other parameters are the same as in the benchmark case. The results for the no search scenario are presented in Figure 6. The size distribution is approximately lognormal as the logarithmic size is roughly normal. This lognormality result can be explained by the fact that a zero search radius completely eliminates the competition among hospitals (as patients can only visit the local hospital) and therefore, hospitals only need to grow their size to accommodate their local demands, which are initialized to lognormal. In the global search case, all hospitals compete for patients. The outcome is the emergence of a normal distribution for size (see Figure 7).

We now test the sensitivity of the initial size distribution of hospitals. We repeat the experiment by using three different initial size distributions: uniform, normal, and exponential. It turns out that regardless of the initial size distributions, a similar sub-lognormal distribution is obtained after a long period of time.

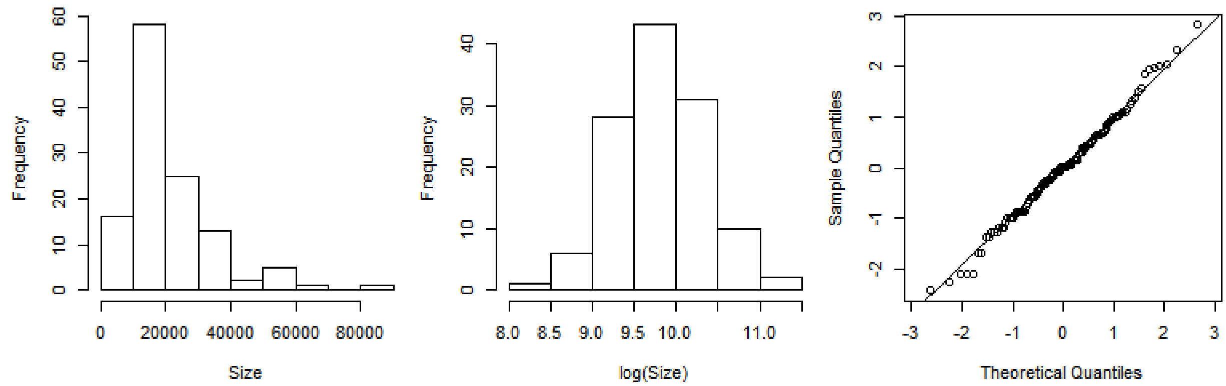

Figure 5: Smaller Size Preference: (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes (The Shapiro-Wilk normality test on the logarithmic sizes fails to reject the null hypothesis with *p*-value=0.49. The skewness and kurtosis of logarithmic sizes are -0.07 and 2.96 respectively.).

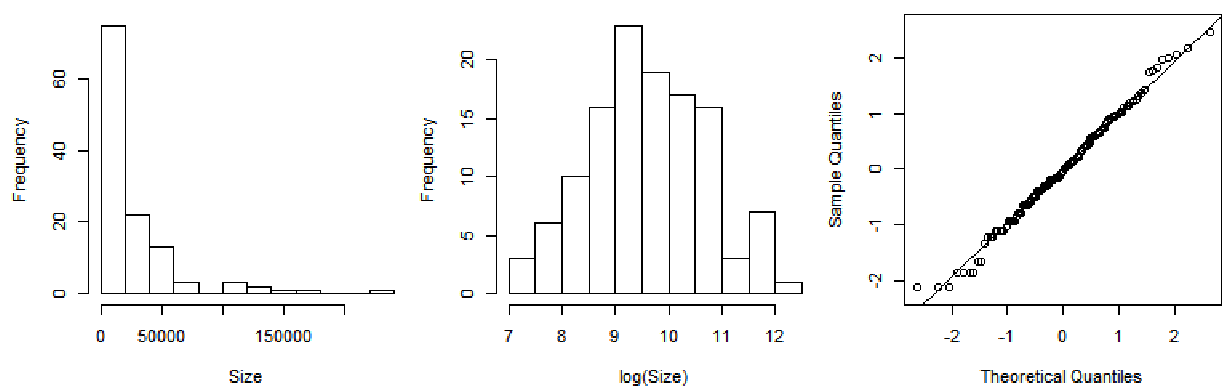

Figure 6: No Search Scenario (*Search-Radius* = 0): (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes (The Shapiro-Wilk normality test on the logarithmic sizes fails to reject the null hypothesis with *p*-value=0.67. The skewness and kurtosis of logarithmic sizes are 0.05 and 2.70 respectively.).

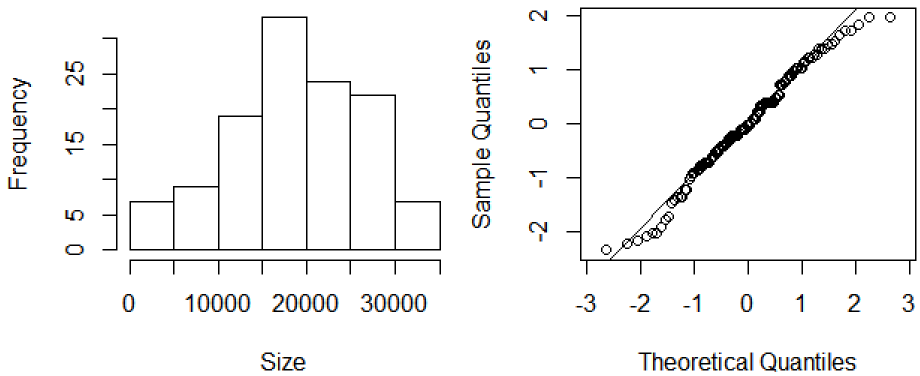

Figure 7: Global Search Scenario (*Search-Radius* = 5): (a) Histogram of Sizes, (b) Histogram of Logarithmic Sizes, (c) Q-Q Plot of Logarithmic Sizes (The Shapiro-Wilk normality test on the logarithmic sizes fails to reject the null hypothesis with p-value=0.14. The skewness and kurtosis of logarithmic sizes are -0.23 and 2.59 respectively.).

Another sensitivity question of interest is related to the size of the environment. It is well-known that the agent population of an agent-based model could significantly affect simulation results. This phenomenon is called the "finite size effects" (Alfi et al. 2009). To investigate whether this finite size effect exists in our model, we repeat the simulation at larger lattices: 15×15 and 21×21. A larger lattice means more agents and hospitals. Once again, the results from these two experiments are similar to what we found in the benchmark case. Therefore, the sub-lognormal size distribution is insensitive to the agent population.

All the sensitivity analysis results are summarized in Table 2. From these sensitivity analysis results, we can conclude that the model can reproduce the characteristics of the U.S. hospitals size distribution and that the two factors, size preference and search radius, are determining factors.

Table 2: Sensitivity Analysis Results. A check mark "√" indicates a size distribution matching the U.S. hospitals size distribution.

Parameters	Value	Results
Patients' Size Preference	Larger	√
	None	Lognormal
	Smaller	Lognormal
Search Radius	2,	√
	0	Lognormal
	5	Normal
Initial distribution of hospital sizes	Lognormal, Uniform, Normal and Exponential	√
Size of lattice	11×11, 15×15 and 21×21	√

5 CONCLUSION

Despite the public's great amount of interest in the size distribution of firms and cities, the size distribution of hospital, however, has not been rigorously studied. This paper analyzes the empirical data and found that the size distribution of hospital is sub-lognormal, which is different from those of firms and cities. An agent-based model equipped with queuing models mimicking the service processes of hospitals is developed and adequately reproduced the sub-lognormal size distribution. It is also found that the size preference and search distance are the two key factors for the emergence of the sub-lognormal distribution.

The hospital service system is different from other profit-oriented service systems, such as firms. Public service systems, such as hospitals, could be set up to provide high quality and timely services. The hospital service system is a queuing system taking feedback from customers. Therefore, customers' preferences are an important factor in determining the size distribution. Location is also a key factor influencing customers' decisions.

REFERENCES

Alfi, V., L. Pietronero and A. Zaccaria. 2009. "Self-Organization for the Stylized Facts and Finite-Size Effects in a Financial-Market Model". *Europhysics Letters* 86: 58003.

American Hospital Directory (2012). http://www.ahd.com/. [Accessed Junuarary 10, 2013].

Angelini, P. and A. Generale. 2008. "On the Evolution of Firm Size Distributions". *The American Economic Review* 98:426–438.

Axtell, R. L. 1999. "The Emergence of Firms in a Population of Agents: Local Increasing Returns, Unstable Nash Equilibria, and Power Law Size Distributions". CSED Working Paper No. 3. Brookings Institution. http://www.brookings.edu/research/reports/1999/06/technology-axtell [Accessed April 10, 2013].

Axtell, R. L. 2001. "Zipf Distribution of U.S. Firm Sizes". *Science* 293: 1818–1820.

Banerjee, A. V. 1992. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, 107:797–817.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100:992–1026.

Barabasi, A. L. and R. Albert. 1999. "Emergence of Scaling in Random Networks". *Science* 286: 509-512.

Blizzard, R. 2004. "Does Hospital Size Matter for Outpatient Test and Treatment Services?" http://www.gallup.com/poll/12601/does-hospital-size-matter-outpatient-test-treatment-services.aspx [Accessed February 20, 2013].

Cabral, L. M. B. and J. Mata. 2003. "On the Evolution of the Firm Size Distribution: Facts and Theory". *The American Economic Review* 93:1075–1090.

Chan, W. K. V. 2008. "An Analysis of Emerging Behaviors in Large-Scale Queueing-Based Service Systems Using Agent-Based Simulation." In *Proceedings of the 2008 Winter Simulation Conference*. Edited by S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson and J. W. Fowler, 872-878. Piscataway, New Jersy: Institute of Electrical and Electronics Engineers. , Inc.

Chan, W. K. V., Y. J. Son and C. M. Macal. 2010. "Agent-Based Simulation Tutorial - Simulation of Emergent Behavior and Differences between Agent-Based Simulation and Discrete-Event Simulation." In *Proceedings of the 2010 Winter Simulation Conference*. Edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan and E. Yücesan, 135-150. Piscataway, New Jersy: Institute of Electrical and Electronics Engineers. , Inc.

DeSarbo, W. S., J. Kim, S. C. Choi, and M. Spaulding. 2002. "A Gravity-Based Multidimensional Scaling Model for Deriving Spatial Structures Underlying Consumer Preference/Choice Judgments." *Journal of Consumer Research*, 29 :91–100.

Eeckhout, J. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 94:1429-1451.

Gulden, T. R. and R. A. Hammond. 2012. "Beyond Zipf: An Agent-Based Understanding of City Size." In *Agent-Based Models of Geographical Systems*, Edited by A. J. Heppenstall, A. T. Crooks, L. M. See, M. Batty, 677-704. New York: Springer Press.

Hart, P. E. and N. Oulton. 1996. "The Growth and Size of Firms." *Economic Journal* 106:1242-1252.

Medical Dictionary. 2013. "Patient Days." http://Medical-Dictionary.Thefreedictionary.Com/Patient+Day. [Accessed April 12, 2013].

Newman, M. E. J. 2001. "Scientific Collaboration Networks. I. Network Construction and Fundamental Results." *Physical Review E. Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics* 64: 016131-1.

Newman, M. E. J. 2005. "Power Laws, Pareto Distributions and Zipf's Law." *Contemporary Physics* 46: 323-51.

Rozenfeld, H. D., D. Rybski, X. Gabaix and H. A. Makse. 2011. "The Area and Population of Cities- New Insights from a Different Perspective on Cities." *American Economic Review* 101:2205-2225.

Simon, H. A. and C. P. Bonini. 1958. "The Size Distribution of Business Firms." *American Economic Review* 48:607-617.

Sutton, J. 1997. "Gibrat's Legacy." *Journal of Economic Literature* 35:40-59.

Wilensky, U. (1999) *Netlogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. available from: http://ccl.northwestern.edu/netlogo/ [cited October 10, 2010].

AUTHOR BIOGRAPHIES

BAOJUN GAO is an Associate Professor in the Economics and Management School at Wuhan University, P.R. China. He received his Ph.D. in Management Science from Xi'an Jiaotong University, P.R.China. His research interests include agent-based simulation and its applications in financial markets, marketing, innovation diffusion, service systems and electronic commerce. His email address is gaobj@whu.edu.cn.

WAI KIN (VICTOR) CHAN is an Associate Professor of the Department of Industrial and Systems Engineering at the Rensselaer Polytechnic Institute, Troy, NY. He holds a Ph.D. in industrial engineering and operations research from University of California, Berkeley. His research interests include discrete-event simulation, agent-based simulation, and their applications in social networks, service systems, transportation networks, energy markets, and manufacturing. He is a member of INFORMS, IIE, and IEEE. His e-mail address is chanw@rpi.edu