A study on firm size distribution of the service sector and manufacturing sector

Sang Sup Cho
Department of Digital Technology Management, Hoseo University, Asan, South Korea

Abstract
Purpose – This study aims to estimate the firm size distributions that belong to the service sector and manufacturing sector in Korea.

Design/methodology/approach – When estimating the firm size distribution, the author considers the following two major factors. First, the firm size distribution can have a gamma distribution rather than traditional accepted distributions such as Pareto distribution or log-normal distribution. In particular, industry-specific enterprises can have different size distributions of the type of gamma distribution. Second, the firm size distribution that is applied to this study’s data set should reflect a number of factors. For example, estimating mixture gamma distribution for firm size distribution should be required and compared, because the total amount of configuration data is composed of small businesses, medium-sized and large companies.

Findings – Using 8,230 number of firm data in 2013, the author estimates mixture gamma distribution for the firm size.

Originality/value – From the comparison, empirical results are found for the following characteristics of core firm size distribution: first, the firm size distribution of the manufacturing sector has a longer tail than firm size distribution of the service sector. Second, the manufacturing firm size distribution dominates the entire country firm size distribution. Third, one factor among the three factors that make up the mixed gamma firm size distribution is described for 99 per cent of the firm size distributions. From the estimated firm size distributions of the service sector and manufacturing sector in Korea, the author simply implies the strategy and policy implications for the start-up firm.

Keywords Firm size distribution, Pareto distribution, Log-normal distribution, Gamma distribution, Mixture gamma distribution, Business start-ups

Paper type Research paper

1. Introduction
Empirical research on firm size distribution has been conflicting. Traditionally, firm size distribution is known to log-normal distribution (Mansfield, 1962). However, when viewing the recent empirical analysis through a broader range of data relating to the firm size, firm size distribution is reported to be out of the log-normal
distribution (Evans, 1987 and Hall, 1987). A Pareto distribution has been proposed as another alternative firm size distribution for the conventional view point (Axtell, 2001).

The correct identification of firm size distribution is an important starting point in a company’s strategy and corporate policy and is a theoretical standard for enterprise growth theory of the firm size distribution as well. Theoretical models for firm size distribution may result from considering two main factors which depend on study objectives and business environment (Rossi-Hansberg and Wright, 2007). Two typical models of firm size distribution are presented according to above two considering factors.

Factors affecting the size distribution are features of the industry, technology development, financial structure, etc. Therefore, firm size distribution is influenced by these factors. However, it is important not only to correctly identify the causal relationship between firm size distribution and some of the determinants but also to determine a particular industry firm size distribution in advance. The reason is that the nature of firm size distribution in a particular industry is important information to show where to start the initial activities among industries.

This study is to compare the features with respect to firm size distribution shown in two industrial firms around the manufacturing and service activities in Korea. As mentioned earlier, the existing empirical analyzes can refer to the total firm distribution (Mansfield, 1962) or a number of studies to estimate the firm size distribution over a certain size (Axtell, 2001). In this study, the following two points of view are used to provide new results. First, we should refine research approach on firm size distribution. I conduct an empirical analysis of different size firm distributions in the industry. Second, total firm size distribution is that they should estimate the mixed gamma distribution because they are composed of small businesses, medium-sized and large enterprises.

To reflect the two aspects described in the above paragraph, I investigate that the firm size distribution may represent the distribution of the gamma probability form. Next, to analyze firm data in this study, I adapt a more appropriate form of the mixture gamma distribution rather than simple gamma distribution. This study’ contributions of the existing research are as follows: first, the empirical analysis of existing firm size distribution is from the assumption that it is a form of the log-normal distribution or Pareto distribution. In this study, I derive the firm size distribution as the gamma distribution and analyze gamma empirical distribution. Second, this study presents the majority of firm data suitable for mixed gamma distribution and finds out estimates for mixture gamma distribution composed of three data sources, namely, small businesses and medium-sized and large enterprises.

The study order is as follows. Section II investigates theoretically that the firm size distribution is the gamma distribution forms. It is then briefly described with respect to the techniques and methods for estimating the gamma mixture distribution. In Section III, I estimate the gamma mixture distribution based on 2013 data on firm size distribution and compare the two firm size distributions for different industries. Brief summary of the empirical results of this study is given in Section IV.
2. Theoretical model and empirical method

2.1 Theoretical model

I am to explain why gamma distribution for firm size distribution is important in this chapter. As already mentioned in the introduction, the existing research has been intensively discussed with respect to whether the firm size distribution log-normal distribution or a Pareto distribution. The firm size distribution has been made in consideration of the firm environment or the research objective. For the application of Pareto distribution to the appropriate firm size distribution, as illustration has been made to derive the importance of the general assumptions such that firms face the technology shock or capital accumulation process, and for a log-normal distribution, it has been made in order to introduce the home for general business environment (Rossi-Hansberg and Wright, 2007).

Pareto firm size distribution has the following probability density function:

\[ f(x) \propto x^{-(1+\alpha)} \]  

But, bright scholar Gibrat had criticized that firm size distribution is Pareto probability density function for the top firm size and had presented the probability density function of the log-normal distribution form for a medium-size firm.

\[ f(x) = \frac{1}{x \sqrt{2\pi\sigma^2}} \exp\left(\frac{- \log(x/x_0)^2}{2\sigma^2}\right), \]  

where \(x_0\) is the mean of distribution.

In compliance with firm size distribution, the Pareto distribution, or a log-normal distribution of the equations (1) and (2) for the firm distribution comprises two problems. First, Pareto distribution and the log-normal distribution is a distribution firm size for a certain level or size. Second, there is a need for a theoretical basis to represent the whole size of firm. To solve the problems of these mentioned business size distributions in the present study, I apply a thermodynamic model that is utilized in physioeconomics. When describing economic activity of economic entity reflects, the role of the interaction or exchange action is emphasized among the economic actors (Angle, 1986). I want to show that I can apply the gamma distribution forms to a firm size distribution in following section.

According to the model of Patriarca et al. (2013), I define firm size distribution as the gamma distribution as follows. Let \(N\) be the companies present on the economy, and they exchange their products. Firms produce their product on the condition of \(x_i \geq 0\). After exchange of the product, the product amount is determined as follows:

\[ x_i \rightarrow x_i - \Delta x_{ij}, \]  
\[ x_j \rightarrow x_j - \Delta x_{ji}, \]

Then, a change in \(\Delta x_{ij}\) can be defined as follows:

\[ \Delta x_{ij} = (1 - \epsilon)x_i + \epsilon x_j. \]
Here, a random variable $\varepsilon$ is equal to the position $[0, 1]$. Normal stationary distribution in this case is determined as follows:

$$f(x) = \beta \exp(-\beta x).$$  \hfill (6)

The above equation (6) is therefore called the Boltzmann–Gibbs energy distribution, which is referred to in thermodynamics.

I extend the above basic Boltzmann–Gibbs energy distribution in the general business environment to the multidimensional space. If $d > 2$, the standard Boltzmann–Gibbs energy distribution is transformed into the following gamma distribution. That is:

$$f(x, \alpha, \beta) = \frac{1}{\beta^d \Gamma(\alpha)} \exp\left(-\frac{x}{\beta}\right),$$  \hfill (7)

$$\alpha = \frac{d}{2},$$  \hfill (8)

$$\beta = \frac{\bar{x}}{\alpha}.$$  \hfill (9)

As briefly described above, if a firm meets regularly the assumptions of environment, a firm size distribution will have a gamma distribution, such as equation (7).

Given these analysis data, mixing gamma distribution for firm size distribution is more appropriate than gamma distribution. Macro data reflect the number of microscopic material properties in most cases. In this study, data on the firm data are a mixture of small- and medium-sized businesses and large firms. Therefore, it is more useful for mixed gamma distribution forming multiple configurations factors than the gamma distribution consisting of one configuration factor.

From a simple gamma distribution, let the random variable $x_i$ be otherwise defined as follows: $x_i$ is extracted from different groups composed of the entire population:

$$X = \{x_i\}, \quad 1 < i < N.$$  \hfill (10)

Mixed gamma distribution may be represented as a combination of the probability of $j$ several groups. That is:

$$P(x_i|\theta) = \sum_{i=1}^{j} \pi_i f(x_i|\theta),$$  \hfill (11)

where $\theta$ is the parameter of gamma distribution and $\sum \pi_i = 1$ is satisfied. Finally, this study is to estimate the gamma mixed distribution shown in equation (11) to determine the size distribution company.

I show that firm size distribution through the entire theoretical model may represent a gamma distribution. If the configuration factor or space is $d \geq 2$, then I prefer mixed gamma distribution to general gamma distribution. I describe a method of estimating the simply mixed gamma distribution in the following section.
2.2 Empirical method

Firm size distribution in the present study is estimated from the parameter approach through a mixed gamma distribution. As already mentioned in the Introduction, controversy about the form of the firm size distribution is present. In this study, I will show that the most appropriate firm size distribution is mixed gamma distribution, because firms that belong to a particular industry are composed of large-, medium- and small-sized firms. In terms of characteristics of these firm data, estimating parameters of a mixed gamma distribution are useful, because they show the characteristics of the different groups.

Before discussing the statistical properties for the mixed gamma distribution, I will investigate the statistical properties and the estimator of the general gamma distribution. General gamma distribution is as follows:

\[ f(x, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{(\alpha - 1)} \exp\left(-\frac{x}{\beta}\right), \]  

(12)

where \( \Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \) is the gamma function. The usefulness of general gamma distribution is that it can be transformed to the various types of distributions such as the exponential distribution according to the parameter \( \alpha \) that determines the shape of the distribution. If the random variables are generated from the Gamma distribution, the mean and variance of random variables can be simply calculated by two parameters \((\alpha, \beta)\). That is:

\[ E(x_i) = \alpha \beta, \]  

(13)

\[ V(x_i) = \alpha \beta^2. \]  

(14)

Mixed gamma and general gamma distributions mentioned in the preceding paragraph are essentially the same (Johnson et al., 1997). So, first, I look at the statistical properties of the mixed gamma distribution and describe how to estimate the technical parameters of the mixed gamma distribution in empirical analysis. As described earlier, if gamma distribution is mixed with some factor, the general gamma distribution is as follows:

\[ f(x, \alpha, \beta, \pi) = \sum_{i=1}^{k} \pi_i x^{\alpha_i - 1} \frac{\alpha_i}{\beta_i^\alpha_i} \exp\left(-\frac{x}{\beta_i}\right), \]  

(15)

where \( \sum_{i=1}^{k} \pi_i = 1. \)

Statistical properties of the mixed gamma distribution equation (15) are as follows (Gharib, 1995). First, to simplify the discussion, assume that the mixed gamma distribution consists of two constituent factors. If a parameter for determining the shape of the mixed gamma mix is equal to \( \alpha_1 = \alpha_2 \), then the distribution is simplified in the form of an exponential distribution which is determined according to \( \beta_1 \) and \( \beta_2 \). Second, if mixed gamma distribution is limited to the case of the assumption, it is determined by a mixture of two Pareto distributions. Statistical properties of such a flexible mixture show that gamma distribution can be useful as an alternative to the different probability distributions.
This study applies Vegas-Sanchez-Ferrero et al.’s (2012) estimating method for mixed gamma distribution to estimate the firm size distribution. The expectation maximization method for estimating mixed gamma distribution is presented as follows. Preceding equation (15) can be simply defined as:

\[ f(x | \theta) = \prod_{i=1}^{k} f(x_i | \theta). \]  

(16)

Expectation maximization method for estimation is that when \( Z = \{z_i\}, \{1,2,\ldots,k\} \) is given a random variable, it maximizes the likelihood function. If \( \theta^n \) is the \( n \)th step maximization estimator of \( L(\theta | \theta^n, x) \), this iterative process is performed until the following condition is satisfied:

\[ \| \theta^{n+1} - \theta^n \| < t, \]  

(17)

where \( t \) refers to a very small threshold value. Numerically, when data \( X \) and unknown random variable \( Z \) are given, expected likelihood function for the random variable \( Z \) is as follows:

\[ L(\theta | \theta^n, x) = \sum_{i=1}^{n} \sum_{j=1}^{k} f(z_i = j | x_i, \theta^n)(\log f(x_i | \theta_j + \log \pi_j). \]  

(18)

To solve equation (18), I need the several steps of the estimation procedure resulting in the estimated parameters \( \hat{\theta} \) of mixed gamma distribution.

In Section 3, I will compare the estimates of the three parameters that determine the distribution of the gamma mixture distribution described above and will summarize key statistical characteristics of firm size distribution of two industries.

### 3. Empirical analysis

#### 3.1 Basic data

This study used survey data conducted by the Korea Corporate Innovation in Science and Technology Policy Institute in 2014. Although the purpose of this study is to clarify the various features of the innovation activity, it is useful in the present research in terms of direct and most extensive survey data. The number of employees that refers to the size of the firm has been used for the base year 2013.

First, I present simple descriptive statistics for the data used in Table I. When I see the basic statistics in Table I, the average firm size belonging to service is higher than

| Industry  | Observation | Mean   | 95 (%) confidence | Minimum | Maximum |
|-----------|-------------|--------|-------------------|---------|---------|
| Service   | 4,155       | Arithmetic 129 | 120-139         | 10      | 7,488   |
|           |             | Geometric 52    | 50-54          |         |         |
|           |             | Harmonic 29     | 28-29          |         |         |
| Manufacturing | 4,075       | Arithmetic 109  | 98-120         | 10      | 12,000  |
|           |             | Geometric 43    | 41-44          |         |         |
|           |             | Harmonic 26     | 25-27          |         |         |

Table I. Basic statistics on firm size of service and manufacturing sectors in 2013 (unit: person)
that of manufacturing firm, and the maximum firm size of service sector is lower than that of the manufacture sector. Second, the difference between the two industry sectors for the arithmetic mean, the geometric mean and the harmonic mean may appear significantly different. These basic statistic results show that firm size distribution will be asymmetric.

To evaluate the visual differences in the size distribution of the two companies in the industry, I visualize the firm size distribution by estimating the nonparametric distribution method, which is shown in Figure 1. As shown in Figure 1, the size distributions of a firm in an industry are very similar but show a great difference in both tails of the firm size distribution. The type of distribution in the service industry shows a high probability density in the left tail of distribution, and distribution in the manufacturing shows very a long tail on the right. The following analytical section presents the estimation of mixed gamma distribution for the firm size distribution of the two industries, because the use of only visual characteristics to identify the characteristics of the two distributions is difficult.

3.2 Empirical results of firm size distribution

Prior to estimating the gamma mixed distribution, I estimate simple gamma distribution for the firm size distribution. Table II shows simple gamma distribution parameters estimated by dividing the data into three groups with respect to the classification of company size. Looking at the results of Table II, I can find the following statistical characteristics. First, parameters representing the distribution of the manufacturing firm size appear to be smaller than those of service firm size distribution.

![Figure 1. Firm size distribution in service and manufacturing industries](image)

| Industry   | $\alpha$ – estimate | $\beta$ – estimate |
|------------|----------------------|---------------------|
| Total      | 0.640                | 168.367             |
| Service    | 0.670                | 192.118             |
| Manufacture| 0.649                | 168.360             |

Table II. Comparison of firm size distribution estimates for a simple gamma distribution
Second, the estimated parameters for the whole firm size distribution are close to the estimated parameters of the manufacturing industry than to those of the services industry. From two characteristics that determine the industry firm size distributions, I can see the dominance of manufacturing firm size than the services firm size distribution.

The results of Table II start from the assumption that the firm size distribution determined from a single gamma distribution. The data for Korean companies to analyze the firm size distribution as discussed earlier show that the mixture gamma distribution is more suitable than the simple gamma distribution. Firm size distribution is formed by the various determinants rather than one determinant. In particular, the firm size data are data consisting of large and small businesses. Therefore, firm size distribution may be viewed as having at least three dimensions.

Table III shows the estimated parameters for mixture gamma distribution with respect to the mixture distribution obtained by dividing the firm data on firm size classification into three groups. Comparing the parameters estimated for mixing the gamma distribution for three industries, I may present particular features of firm size distributions. First, firm size distribution estimated from all three components are determined by the first component factor, and the first component factor dominates the other firm size distributions. Second, the shapes of parameters determining the form of firm size distribution of the manufacturing industry were smaller than those in the simple form of firm size distribution of the service industry. Third, whole industry distribution of firm size and manufacturing firm size distribution are very similar to each other.

Next, Figures 2 and 3 show the decomposition of the estimated mixed gamma distribution by three factors configuration. When looking at the two graphs, I cannot see a visual difference. But, it is necessary to keep in mind that there is very small difference among each gamma distribution from three factors.

4. Summary and implication
This study estimates the firm size distributions belonging to the service and manufacturing sectors. Especially, when I estimate a firm size distribution, I consider the following two import factors. First, firm size distribution may be specified to have a better availability to the gamma distribution than conventional Pareto distribution or log-normal distribution. In addition, industry-specific enterprises can have a different scale distribution of the gamma form. Second, I must estimate the firm size distribution

| Industry | Estimator | Component 1 | Component 2 | Component 3 |
|----------|-----------|-------------|-------------|-------------|
| Total    | weight    | 0.9497      | 0.0498      | 0.000049    |
|          | $\alpha$ – estimates | 1.0047      | 2.3952      | 286.990     |
|          | $\beta$ – estimates | 66.667      | 413.980     | 39.483      |
| Service  | Weight    | 0.9191      | 0.00753     | 0.0056      |
|          | $\alpha$ – estimates | 1.0281      | 3.3130      | 4.5720      |
|          | $\beta$ – estimates | 72.088      | 170.767     | 682.946     |
| Manufacturing | Weight | 0.9447 | 0.0498 | 0.000049 |
|          | $\alpha$ – estimates | 1.0086 | 2.3715 | 286.990 |
|          | $\beta$ – estimates | 66.651 | 413.980 | 39.483 |
as mixed gamma distribution, because it is composed of small businesses and medium-sized and large companies.

Using 8,230 number of firm data in the year of 2013, I estimate mixture gamma distribution for the firm size. From the comparison, empirical results are found for the following characteristics of core firm size distribution: first, the firm size distribution of the manufacturing sector has a longer tail than the firm size distribution of the service sector. Second, the manufacturing firm size distribution dominates the entire country firm size distribution. Third, one factor among the three factors that make up the mixed gamma firm size distribution is described for 99 per cent of the firm size distributions. From estimated firm size distributions of the service sector and manufacturing sector in Korea.

Based on the results of this study, I can simply present the strategy and policy implications for the enterprise business. When I derive a corporate policy and strategy on corporate entrepreneurship, enterprise-wide distribution of the entire nation should be noted therefore drawn on whether the strategies and policies which target industrial sectors because most firm size distributions represent a manufacturing firm size

Figure 2. Firm size gamma distribution decomposition for the service sector

Figure 3. Firm size gamma distribution decomposition for the manufacturing sector
distribution. In situations where the foundation is active strategies and policies for corporate hospitality, recent establishment of the service should be based on firm size distribution belonging to the service sector. Second, the firm size distribution such as Pareto distribution or log-normal distribution that take references of many existing studies should be recognized in a special form of gamma distributions. It should be clear that the implications from Pareto distribution and log-normal distribution based on the partial distribution will be different from those of a gamma mixture distribution.

References
Angle, J. (1986), “The surplus theory of social stratification and the size distribution of personal wealth”, Social Forces, Vol. 65 No. 2, pp. 293-326.
Axtell, R.L. (2001), “Zipf distribution of US firm sizes”, Science, Vol. 293 No. 12, pp. 1818-1820.
Evans, D.S. (1987), “Tests of alternative theories of firm growth”, Journal of Political Economy, Vol. 95 No. 4, pp. 657-674.
Gharib, M. (1995), “Two characterizations of a gamma mixture distribution”, Bulletin of the Australian Mathematical Society, Vol. 52 No. 1, pp. 353-358.
Hall, B.H. (1987), “The relationship between firm size and firm growth in the United-States manufacturing sector”, Journal of Industrial Economics, Vol. 35 No. 4, pp. 583-606.
Johnson, N., Kotz, S. and Balakrishnan, N. (1997), Continuous Univariate Distributions, Wiley and Sons, Hoboken.
Mansfield, E. (1962), “Entry, Gibrat’s law, innovation, and the growth of firms”, American Economic Review, Vol. 52 No. 5, pp. 1023-1051.
Patriarca, M. and Chakrabarti, A. (2013), “Kinetic exchange model: from molecular physical to social science”, American Journal of Physics, Vol. 81 No. 1, pp. 618-623.
Rossi-Hansberg, E. and Wright, M. (2007), “Establishment size dynamics in the aggregate economy”, American Economic Review, Vol. 97 No. 5, pp. 1639-1666.
Vegas-Sanchez-Ferrero, G., Aja-Fernandez, S., Palencia, C. and Martin-Fernandez, M. (2012), “A generalized gamma mixture model for ultrasonic tissue characterization”, Computational and Mathematical Methods in Medicine, Vol. 2012 No. 4, pp. 1-25.

Further reading
Cabral, L. and Mata, J. (2003), “On the evolution of the firm size distribution: facts and theory”, American Economic Review, Vol. 93 No. 4, pp. 1075-1090.
Lawless, J. (1980), “Inference in the generalized gamma and log gamma distributions”, Technometrics, Vol. 22 No. 3, pp. 409-4019.
Luis, M. and Mata, J. (2003), “On the evolution of the firm size distribution: facts and theory”, American Economic Review, Vol. 93 No. 4, pp. 1075-1090.

Corresponding author
Sang Sup Cho can be contacted at: choss@hoseo.edu

For instructions on how to order reprints of this article, please visit our website:
www.emeraldkrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com