

IDE Discussion Papers are preliminary materials circulated  
to stimulate discussions and critical comments

IDE DISCUSSION PAPER No. 253

**The Size Distribution of All  
Cambodian Establishments**

Kiyoyasu TANAKA\*  
Naomi HATSUKANO

September 2010

**Abstract**

This paper presents empirical evidence on the size distribution of all Cambodian establishments in the nonfarm sector for 2009. Small- and large-scale establishments account for the largest share of employment, pointing to a “missing middle” that is commonly observed in developing countries. The analysis provides little evidence for Zipf’s law because Cambodian industry is characterized by a more dense mass of small establishments than the Zipf distribution would predict.

**Keywords:** Size distribution, establishments, Zipf’s law, Cambodia

**JEL classification:** L11, O17, O53

---

\* Corresponding author: Kiyoyasu Tanaka; address: Economic Integration Studies Group, Inter-Disciplinary Studies Center, Institute of Developing Economies, 3-2-2 Wakaba, Mihama-ku, Chiba-shi, Chiba 261-8545, Japan; e-mail: kiyoyasu\_tanaka@ide.go.jp

The Institute of Developing Economies (IDE) is a semigovernmental, nonpartisan, nonprofit research institute, founded in 1958. The Institute merged with the Japan External Trade Organization (JETRO) on July 1, 1998. The Institute conducts basic and comprehensive studies on economic and related affairs in all developing countries and regions, including Asia, the Middle East, Africa, Latin America, Oceania, and Eastern Europe.

---

The views expressed in this publication are those of the author(s). Publication does not imply endorsement by the Institute of Developing Economies of any of the views expressed within.

---

**INSTITUTE OF DEVELOPING ECONOMIES (IDE), JETRO**  
**3-2-2, WAKABA, MIHAMA-KU, CHIBA-SHI**  
**CHIBA 261-8545, JAPAN**

©2010 by Institute of Developing Economies, JETRO

No part of this publication may be reproduced without the prior permission of the IDE-JETRO.

# The Size Distribution of All Cambodian Establishments

Kiyoyasu TANAKA<sup>†§</sup>

*Inter-Disciplinary Studies Center, Institute of Developing Economies, Japan*

Naomi HATSUKANO

*Area Studies Center, Institute of Developing Economies, Japan*

September 2010

---

**Abstract:** This paper presents empirical evidence on the size distribution of all Cambodian establishments in the nonfarm sector for 2009. Small- and large-scale establishments account for the largest share of employment, pointing to a “missing middle” that is commonly observed in developing countries. The analysis provides little evidence for Zipf’s law because Cambodian industry is characterized by a more dense mass of small establishments than the Zipf distribution would predict.

**JEL Classification:** L11, O17, O53

**Keywords:** Size distribution, establishments, Zipf’s law, Cambodia

---

---

<sup>†</sup> Corresponding author: Kiyoyasu Tanaka; address: Economic Integration Studies Group, Inter-Disciplinary Studies Center, Institute of Developing Economies, 3-2-2 Wakaba, Mihama-ku, Chiba-shi, Chiba 261-8545, Japan; e-mail: kiyoyasu\_tanaka@ide.go.jp

<sup>§</sup> We thank Fumihiko Nishi and Net Kim for data assistance on the Establishment Listing for 2009 in Cambodia.

## 1. Introduction

This paper presents empirical evidence on the entire size distribution of business establishments in least developing countries. We use the first comprehensive *Establishment Listing* in Cambodia for 2009, which surveyed the economic activities of virtually all establishments in all areas of Cambodia. The data are used to describe the size distribution with respect to number and employment of establishments according to the fine disaggregation of employment sizes. Then, we examine the relationship between the rank and size of establishments to examine the validity of Zipf's law.

The size distribution of business firms has received considerable attention for its striking empirical regularity.<sup>1</sup> Axtell (2001) found that the size distribution of tax-paying firms in the U.S. is well described by a Pareto distribution with a shape parameter of 1. In other words, the probability that a firm has more than  $L$  workers is proportional to  $1/L$ . A similar pattern was also found for a large sample of European firms by Fujiwara et al. (2004).<sup>2</sup> Consequently, the regularity has yielded various theoretical explanations in which the underlying model of firm dynamics could generate an observed pattern of firm size distributions (Simon and Bonini, 1958; Luttmer, 2007; Rossi-Hansberg and Wright, 2007).

The firm size distribution in industrial countries has been widely examined, but there has been limited systematic analysis of the complete distribution of firms in developing countries. Using aggregate measures of employment distribution by firm size, Tybout (2000) describes a distinctive feature of manufacturing firms in developing countries as a bimodal structure in which a number of small firms and a handful of large

---

<sup>1</sup> For Zipf's law in the size distribution of cities, see Gabaix (1999) and Eeckhout (2004).

<sup>2</sup> Using a limited sample of relatively large firms in the U.S., Stanley et al. (1995) found that the size distribution of U.S. firms is approximately lognormally distributed.

firms account for a substantial share of employment. The observed feature is commonly attributed to high transaction costs due to imperfections in product and factor markets in developing countries (Nugent and Nabli, 1992; Sleuwaegen and Goedhuys, 2002). However, prior findings relied primarily on a limited sample of firms in a specific sector, making it difficult to illustrate the entire distribution of all business entities as well as to assess the empirical validity of Zipf's law. This paper is distinctive in that we examine the entire population of Cambodian establishments in an environment with arguably serious imperfections in market mechanisms.

## 2. Data

The empirical analysis uses a unique dataset on Cambodian establishments. The data are obtained from *Establishment Listing* in Cambodia for 2009 (EL2009) by the National Institute of Statistics.<sup>3</sup> The survey defines an establishment as a unit of economic entity managed by a single ownership in a single physical location with some durable facilities. EL2009 covers all the establishments that were in operation in the entire territory of Cambodia as of February 2009, except for individual proprietorships in agriculture, forestry and fishery. The survey information includes location, employment, and industrial category at the establishment level.<sup>4</sup> This dataset is novel in that it provides data on a fundamental unit of economic entity for the entire nonfarm private and public sectors.

Table 1 lists the size distribution of Cambodian establishments in the nonfarm

---

<sup>3</sup> See details at <http://www.nis.gov.kh/index.php/statistics/surveys/el2009>.

<sup>4</sup> The ownership types of establishments are classified into individual proprietorship, partnership and cooperative, private and public company, foreign company, state-owned enterprise, and non-governmental organization (NGO). We exclude NGOs from the analysis.

sector by tabulating the number and share of establishments and employment over size classes. The sample has 375,854 establishments with 1,455,526 workers in Cambodia for 2009. Small-scale establishments with less than 10 workers account for over 90% of all establishments in number, with nearly a 60% share of nationwide employment. Medium-scale establishments between 10–99 workers constitute 3.2% by number and 16.8% by employment. By contrast, large-scale establishments with 100 workers or more represent only 0.18% by number, but account for a quarter of employment. As shown by Tybout (2000) for other developing countries, the size distribution of Cambodian establishments exhibits a “missing middle” in which medium-scale establishments are underrepresented in the Cambodian economy.<sup>5</sup>

**[Table 1]**

**3. Zipf’s law**

It is of great interest to examine whether Zipf’s law holds for the observed size distribution of all Cambodian establishments. For a set of establishments  $i = 1, \dots, N$ , let  $S(i)$  denote the size of an establishment  $i$ , as measured by the total number of persons engaged in economic activity, including owners and/or family members. Assume that  $S(i)$  is a discrete random variable following a Pareto distribution. Then, the Pareto distribution of the size variable,  $S(i)$ , is defined by:

$$f(S(i)|\alpha, s_0) = \frac{\alpha s_0^\alpha}{S(i)^{\alpha+1}}, \quad S(i) \geq s_0, \alpha > 0$$

$$F(S(i)|\alpha, s_0) = 1 - \left(\frac{s_0}{S(i)}\right)^\alpha, \quad S(i) \geq s_0, \alpha > 0$$

where  $f(\cdot)$  is the probability density function and  $F(\cdot)$  is the cumulative density function.  $s_0$  is the lower bound of the establishment size.  $\alpha$  is a positive parameter that shapes the

---

<sup>5</sup> Rauch (1991) explains size dualism as a distributional consequence of entrepreneurial talent in an economy.

dispersion of the Pareto distribution. Zipf's law is a special case of the Pareto distribution with  $\alpha = 1$ .

Zipf's law can be analyzed by looking at the log of the rank plotted against the log of the size. Let  $R(i)$  denote the rank of the size of establishments,  $S(i)$ , which are sorted from largest to smallest. Because the rank is defined by  $R(i)/N = 1 - F(S(i))$ , the rank is expressed as:

$$R(i) = N \cdot \left( \frac{S_0}{S(i)} \right)^\alpha$$

Taking natural logarithms, we obtain  $\ln R(i) = \beta - \alpha \ln S(i)$ , where  $\beta = \ln N + \alpha \ln S_0$  is a constant term. By allowing for an error term in the deterministic specification, the rank-size relationship is specified as:

$$\ln R(i) = \beta - \alpha \ln S(i) + \varepsilon_i \quad (1)$$

An estimate for  $\alpha$  can be obtained by regressing  $\ln R(i)$  on  $\ln S(i)$  for the sample of Cambodian establishments. Consequently, an estimated coefficient of  $S(i)$  provides the basis for statistically investigating Zipf's law.

Figure 1 shows a scatter plot of  $\ln R(i)$  against  $\ln S(i)$  with a linear regression line estimated for the sample. The scatter plots in the range of 0–5 over log employment appear to fit well with the straight line, suggesting that the probability that an establishment has more than  $S$  employment is approximately proportional to  $1/S$ . However, the plots in the range of over 5 log employment start to deviate substantially from the linear regression line. Thus, the graphical representation of the rank-size relationship does not seem to support Zipf's law in the case of all Cambodian establishments.

**[Figure 1]**

While a visual inspection of Zipf's law is helpful to observe how well the plots

fit a linear regression line, a formal standard approach is to estimate the coefficient  $\alpha$  of  $\ln S(i)$  in equation (1) by Ordinary Least Squares (OLS).<sup>6</sup> Table 2 presents the OLS results with summary statistics of different samples used for the regression. Using the entire sample, an OLS estimate of  $\alpha$  is 1.33, which is significantly greater than one at the 1% significance level. The size distribution of all Cambodian establishments is associated with a larger number of small establishments and a smaller number of large establishments than the distribution predicted by a Pareto distribution with a shape parameter of 1. This finding seems to be consistent with the size distribution of U.S. establishments for 2000, as studied by Rossi-Hansberg and Wright (2007).

**[Table 2]**

To further investigate the validity of Zipf's law, the sample is split along various dimensions. First, establishments are separated into three classes by employment size. The estimated coefficients are 1.31 and 1.29 for small-scale establishments (1–9) and medium-scale establishments (10–99), respectively. As these estimates are significantly different from one, the size distribution for small and medium establishments does not appear to fit Zipf's law well. By contrast, an estimate of  $\alpha$  for large-scale establishments (100 workers or more) is still significantly different from one, but much closer to the value of 1.

Second, the sample is split by four major sectors in the Cambodian economy. The summary statistics show that average employment size per establishment for manufacturing (6.2) and education (13.5) is larger than that for wholesale and retail (2.2) and accommodation and food service (3.9). While the average employment size varies largely among these sectors, all the estimates of  $\alpha$  are significantly different from

---

<sup>6</sup> See Nitsch (2005) and Soo (2005) for empirical literature on Zipf's law for cities.



one, ranging from 1.30 to 1.38. Finally, the sample is broken down by ownership type: individual proprietorship, state-owned enterprise, private/public limited company, and foreign-owned company. As is the case for sectors, the average size of employment differs substantially by ownership. Nevertheless, the estimate for  $\alpha$  is significantly different from one for all the samples, with somewhat larger ranges than the results estimated across the sectors.

These results of the regression analysis suggest that the size distribution of Cambodian establishments does not provide strong evidence in favor of Zipf's law. Along different dimensions of the sample, the estimated coefficient  $\alpha$  is significantly larger than one, implying that Zipf's law is not likely to hold for various structures of Cambodian establishments. To further interpret the results, the size distribution in Cambodia can be characterized by a larger mass of relatively small establishments and a smaller mass of relatively large establishments than the pattern predicted by a Pareto distribution with a shape parameter of 1. This interpretation is consistent with the finding that the estimates of  $\alpha$  tend to be lower for the sample with larger average employment sizes.

#### **4. Concluding remarks**

This paper employs the first comprehensive data on Cambodian establishments to characterize the detailed distribution of establishment sizes. Covering virtually all establishments in Cambodia, we examine the empirical validity of Zipf's law in the context of least developing countries. The descriptive analysis shows that small-scale and large-scale establishments account for a majority of the number and employment of establishments in the Cambodian economy. In contrast, mid-sized establishments are

underrepresented in the domestic industry, consistent with the “missing middle” that is commonly observed in a wide range of developing economies (Tybout, 2000). Additionally, the regression analysis provides considerable evidence against the strict validity of Zipf’s law. Instead, the estimated shape parameter is generally greater than one, indicating that the dispersion of establishment sizes is relatively small with a more dense mass of small establishments.

The large mass of small establishments in the size distribution could be a manifestation of substantial barriers to the growth of small- and medium-scale enterprises in the Cambodian economy (Sleuwaegen and Goedhuys, 2002). While monotonically increasing numbers of progressively smaller firms are also observed in developed countries as shown in Axtell (2001), an underrepresented share of mid-sized establishments in employment is a distinctive feature of Cambodian industry. Possible deterrent effects on the growth of small establishments would range from regulatory barriers (taxes, registration fees, and corruption) to financial constraints on external credit, demand constraints on mass production, and infrastructure obstacles to transportation. An empirical investigation of prominent barriers is a crucial step to understand why the size distribution of Cambodian establishments deviates from Zipf’s law. However, the issue of establishment size dynamics is left for future research.

## References

- Axtell, R. L., 2001. Zipf distribution of U.S. firm sizes. *Science* 293, 1818–1820.
- Eeckhout, J., 2004. Gibrat's law for (all) cities. *American Economic Review* 94, 1429–1451.
- Fujiwara, Y., Guilmi, C. D., Aoyama, H., Gallegati, M., Souma, W., 2004. Do Pareto-Zipf and Gibrat laws hold true? An analysis with European Firms. *Physica A: Statistical Mechanics and its Applications* 335, 197-216.
- Gabaix, X., 1999. Zipf's law for cities: an explanation. *Quarterly Journal of Economics* 11, 739–767.
- Luttmer, E. G. J., 2007. Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics* 122, 1103-1144.
- Nitsch, V., 2005. Zipf zipped. *Journal of Urban Economics* 57, 86-100.
- Nugent, J.B., Nabli, M.K., 1992. Development of financial markets and the size distribution of manufacturing establishments: international comparisons. *World Development* 20, 1489– 1499.
- Rauch, J.E., 1991. Modelling the informal sector formally. *Journal of Development Economics* 35, 33– 47.
- Rossi-Hansberg, E., Wright, M. L. J., 2007. Establishment size dynamics in the aggregate economy. *American Economic Review* 97, 1639-1666.
- Sleuwaegen, L., Goedhuys, M., 2002. Growth of firms in developing countries, evidence from Côte d'Ivoire. *Journal of Development Economics* 68, 117–135.
- Soo, K. T., 2005. Zipf's law for cities: a cross-country investigation. *Regional Science and Urban Economics* 35, 239-263.
- Stanley, M. H. R., Buldyrev, S. V., Havlin S., Mantegna R. N., Salinger M. A., Stanely, H. E., 1995. Zipf plots and the size distribution of firms. *Economics Letters* 49, 453–457.
- Simon, H. A., Bonini, C. P., 1958. The size distribution of business firms. *American Economic Review* 48, 607–617.
- Tybout, J.R., 2000. Manufacturing firms in developing countries: how well do they do and why? *Journal of Economic Literature* 38, 11– 44.

**Table 1: Distribution of Cambodian Nonfarm Establishments for 2009**

Size	Establishment		Employment	
	Number	Share (%)	Number	Share (%)
1	112,131	29.83	112,131	7.70
2	149,293	39.72	298,586	20.51
3	44,611	11.87	133,833	9.19
4	24,268	6.46	97,072	6.67
5	14,466	3.85	72,330	4.97
6	8,419	2.24	50,514	3.47
7	4,947	1.32	34,629	2.38
8	3,201	0.85	25,608	1.76
9	1,796	0.48	16,164	1.11
10-19	7,972	2.12	102,374	7.03
20-29	1,956	0.52	45,348	3.12
30-39	1,013	0.27	32,680	2.25
40-49	388	0.10	16,839	1.16
50-99	711	0.19	46,787	3.21
100 or more	682	0.18	370,631	25.46
Total	375,854	100	1,455,526	100

*Notes:* Size indicates the number of workers for each establishment; non-governmental organizations are excluded from the sample.

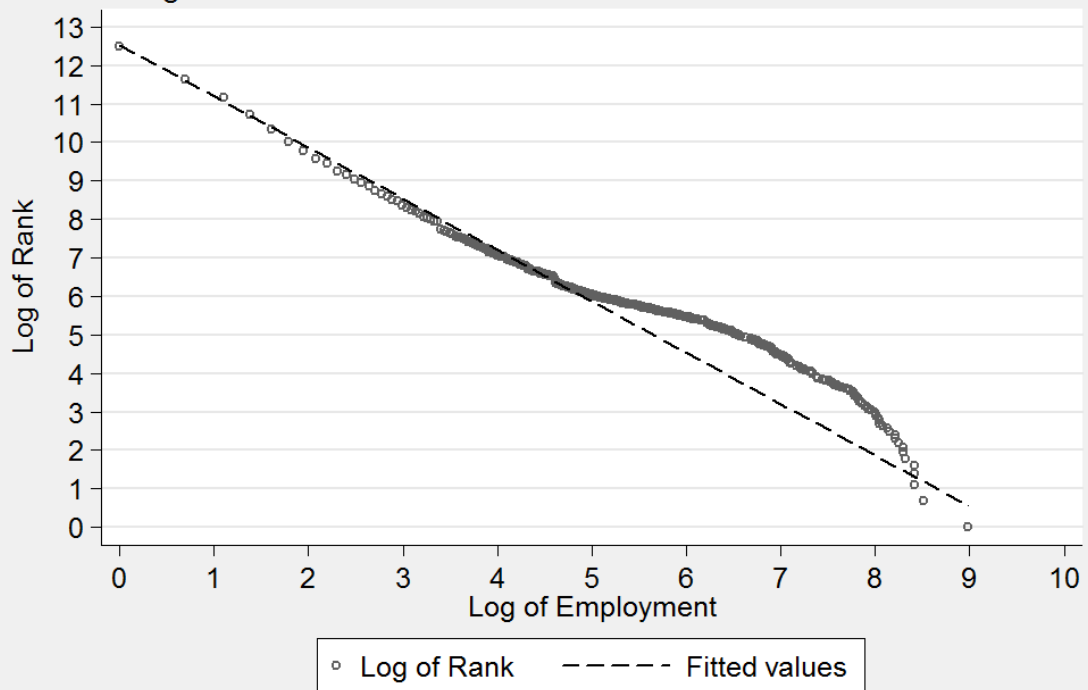
*Source:* Cambodian Establishment Listing 2009.

**Table 2: Regression Analysis of Zipf's Law**

Sample	Employment			OLS regression		
	Mean	S.D.	Obs.	$\alpha$	Std. err.	R <sup>2</sup>
All	3.9	41.3	375,884	1.33*	0.001	0.994
<b>Size</b>						
1-9	2.3	1.5	363,159	1.31*	0.0003	0.994
10-99	20.3	14.4	12,043	1.29*	0.001	0.997
100 or more	543.4	800.7	682	0.91*	0.022	0.926
<b>Sector</b>						
Manufacturing	6.2	83.8	84,629	1.30*	0.002	0.992
Wholesale/Retail	2.2	3.4	198,103	1.30*	0.0005	0.994
Accommodation/food service	3.9	8.3	29,225	1.37*	0.001	0.995
Education	13.5	29.3	9,020	1.38*	0.003	0.994
<b>Ownership</b>						
Individual proprietorship	2.8	21.2	358,182	1.32*	0.001	0.994
State-owned enterprise	13.9	33.0	8,690	1.37*	0.003	0.994
Private/public limited company	74.1	346.0	2,098	1.25*	0.006	0.987
Foreign-owned company	52.0	177.8	144	1.24*	0.027	0.983

*Notes:* \* indicates that the estimated coefficient is different from one at the 1% significance level; robust standard errors are reported.

Figure 1: Size Distribution of Cambodian Establishments in 2009



Source: Establishment Listing for 2009