Discussion Paper No. 009

Statistical Theory of Rank Size Rule Regression under Pareto Distribution

Yoshihiko Nishiyama
and
Susumu Osada

January, 2004

21COE
Interfaces for Advanced Economic Analysis
Kyoto University

Statistical Theory of Rank Size Rule Regression under Pareto Distribution[1.]

Yoshihiko Nishiyama
and
Susumu Osada[2]
Kyoto Institute of Economic Reseach
Kyoto University
Sakyo Kyoto 606-8501 Japan

Abstract

Letting $S_{(i)}$ be i-th largest city in a country, it is often observed that $\log S_{(i)} \approx \alpha_0 + \alpha_1 \log i$ for some $\alpha_0 > 0$ and $\alpha_1 < 0$. It is called rank size rule when $\alpha_1 = -1$. This relationship has been examined by means of OLS estimation and t test in the literature. However, since $S_{(i)}$ is heteroskedastic and autocorrelated, t statistics do not have standard distribution. Indeed we show $t \xrightarrow{p} \infty$ as the sample size increases. The purpose of this paper is to obtain statistical properties of OLS estimator of the rank size rule regression and distribution of t-statistics under Pareto distribution, and further to propose more efficient estimation procedures.

[1]This research was partially supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grant-in-Aid for 21st Century COE Program.

[2]The second author is a Postdoctoral Fellow of the 21st Century COE program.

1 Introduction

After pioneering work on city size distribution by Auerbach (1913) and Zipf (1949), many researchers have investigated a wide range of settlement systems. Zipf's main result called Zipf law is the following. Let S denote a random varialbe representing city size measured by its population, then for large x,

$$P(S \geqq x) = A/x$$

for some $A > 0$ or Pareto distribution with unit exponent. This is closely related to so-called rank size rule of city size data. Let $S_i, i = 1, \cdots, n$ be population of cities in a country, and $S_{(i)}$ be its order statisitcs satisfying $S_{(1)} \geqq S_{(2)} \geqq \cdots \geqq S_{(n)}$, then we often observe that

$$\log S_{(i)} \approx \alpha_0 + \alpha_1 \log i, \ i = 1, \cdots, n, \tag{1.1}$$

where $\alpha_0 > 0, \alpha_1 < 0$. This relationship is called the rank size rule when $\alpha_1 = -1$. When Zipf law holds, rank size rule follows approximately. Regarding (1.1) as a regression model, many researchers have estimated α_0 and α_1 by ordinary least squares (OLS) method and implemented t test for $\alpha_1 = -1$.

One of the most important papers in this field is Rosen and Resnick (1980). They examined city size distribution of the 50 largest administrative urban areas in 44 countries. The simple mean of the estimate for α_1 was 1.14 and in 32 countries exceeded unity and they concluded validity of the urban rank-size rule appears to be open to question. Soo (2002) also made an international comparison using updated data of 73 countries.

Data availability is an important issue in the estimation of (1.1). As Rosen and Resnick (1980) pointed out, many countries do not provide city size data which cover all the population, but that of some larger cities, which restricts the data. They show how sensitive the estimation results are when the cut off point is changed. This point is supported also by Malecki (1980) in terms of the US Midwest settlement system in 1940 and 1970 and Guerin-Pace (1995) for French settlement data between 1831 and 1991.

A series of Alperovich's study investigated the validity of the rule by several tests. Alperovich (1984) estimated the coefficients of 15 countries based on the OLS and examined the rank size rule by t tests. He concluded that the distribution pattern of most countries did not support the rank-size rule. See also Alperovich (1988, 1992) for his original methods.

There are some other approaches of studying city size distribution using different models or estimators. Hsing (1990) proposed to use Box-Cox transformation model. Cameron (1980) applied a truncated regression model with non-Pareto error distributions such as log-normal and extreme value distributions. Dobkins and Ioannides

(2000) recommended to use not the OLS estimation but the maximum likelihood estimator by Hill (1975), and applied it to the US data. Soo (2002) also examined the urban rank size rule comparing OLS and Hill estimators.

Rank size rule itself is a simple description of city size data, which does not tell us how and why it occurs, but it appears to be associated with economic development. Berry (1961) examined the city size distribution pattern for 37 countries to find a clear relationship between the pattern and the level of economic development. Parr (1985) also examined the same relationship in 12 countries over a period of seventy years. Recently Gabaix (1999) and Duranton (2002) constructed economic models which can describe this relationship. We also refer Mori, Nishikimi and Smith (2003) related with this issue.

In a purely statistical aspect, estimation of tail exponent for heavy tail distributions has been a research topic of interest from 1970s. Pareto distribution is one of the simplest distributions with a heavy tail. In the case when the Pareto assumption is true, the parameter(s) can be estimated by means of maximum likelihood principle efficiently. Since the pioneering work by Hill (1970, 1975), statisticians have proposed estimators for semiparametric models departing from the Pareto, where they parametrize the distribution only partly in the tail. Hill's estimator is a conditional maximum likelihood estimator (MLE) using only some larger observations. Dekkers et. al. (1989) extends Hill' estimator to allow for the parameter space to be R^1. A simple estimator is proposed in Haan and Resnick (1980) with the asymptotic distribution under stable distribution. Csorgo et.al. (1985) consider a kernel estimator which includes the Hill' and Haan and Resnick estimators as special cases. Beirland et.al., based on quantile-quantile plots, estimate the tail index by a right-side weighted least squares method, while Hosking et.al. (1985) apply the method of probability-weighted moments. There exist a number of other estimators proposed in the literature. However, as pointed out in Rootzen and Tejvidi (1997), these methods are not robust when the underlying distribution is not a member of generalized Pareto (GP) models. Feuerverger and Hall (1999) propose robust ML and LS estimators against departures form the GP assumption. We refer Csorgo and Viharos (1998) for detailed overview of the literature.

It is certainly important and interesting from statistical point of view to develop such new statistical tools to invetigate the rank size rule, but we also need to know the performance of OLS estimation based on (1.1) at least to evaluate the previous studies which used this method. The purpose of this paper is to derive the exact and approximate properties of the OLS estimator and t test statistics for the rank size rule null, or $\alpha_1 = -1$. We obtain the bias and variance of the estimator assuming S_i are independently and identically distributed (iid). Further we show t statistic does not have t distribution unlike standard classical linear regression theory because $S_{(i)}$ are in fact autocorrelated and heteroskedastic under the iid assumption. Since suggested by Zipf, it is often assumed that S_i have Pareto distribution. Under this assumption, we can show

$$E(\log S_{(i)}) = \alpha_0 + \alpha_1 \log i,\ \ i = 1, \cdots, n$$

does not strictly hold for $^\forall \alpha_0, \alpha_1$ in small sample, but it does approximately for large n and i. It yields bias of the OLS estimator for α_1.

The following section shows exact and approximate expressions for $E(\log S_{(i)})$ and $V(\log S_{(i)})$ then derive the bias and variance of the OLS estimator for (1.1). Then we present Monte Carlo results on the distribution of t value for the estimator which is far away from t distribution. We further show t explodes asymptotically, indicating t test is not applicable to test the null of $\alpha_1 = -1$. Section 3 proposes more efficient estimators, while Section 4 gives empirical results from Japanese city size data of Metropolitan Employment Area (MEA). Section 5 is conclusion.

2 OLS estimation of the rank size rule regression

2.1 Bias and variance of the OLS estimator

Let $F_X(x)$ and $f_X(x)$ denote the cumulative distribution function and density function of a random variable X. Assume $S_i, i = 1, \cdots,\ n$ are iid from the distribution function

$$F_S(x) = 1 - \left(\frac{1}{x}\right)^\beta \quad (\beta > 0, x \geqq 1), \tag{2.1}$$

or the density

$$f_S(x) = \frac{\beta}{x^{\beta+1}}. \tag{2.2}$$

Before giving a justification to (1.1), we prove the following Lemma on some moments of $\log S_{(i)}$, which is also used to evaluate the variance of the OLS estimator.

Lemma 1 When $\{S_1, \cdots, S_n\}$ is a random sample from (2.1), and $\left\{S_{(1)}, \cdots, S_{(n)}\right\}$ is the order statistics satisfying $S_{(1)} \geqq \cdots \geqq S_{(n)}$, we have

(a) $E(\log S_{(i)}) = \sum_{k=1}^{n-i+1} \frac{1}{\beta(n-k+1)} = \frac{1}{\beta}\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{i}\right)$

(b) $V(\log S_{(i)}) = \sum_{k=1}^{n-i+1} \frac{1}{\beta^2(n-k+1)^2} = \frac{1}{\beta^2}\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{i^2}\right)$

(c) $Cov(\log S_{(i)}, \log S_{(j)}) = V\left[\log S_{(j)}\right] = \frac{1}{\beta^2}\left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \cdots + \frac{1}{j^2}\right), \quad (i < j)$

Proof

Because

$$F_{\log S}(x) = P(\log S \leqq x) = P(S \leqq e^x) = 1 - e^{-\beta x}, \tag{2.3}$$

density of $\log S_i$ is

$$f_{\log S}(x) = \beta e^{-\beta x}$$

which is a density of exponential distribution with parameter β. Therefore, $\log S_{(i)}$ can be regarded as the i-th largest value of a random sample of size n from (2.3).

Putting $Y_i = \beta i \left\{\log S_{(i)} - \log S_{(i+1)}\right\}, i = 1, \cdots, n$ with $\log S_{(n+1)} = 0$, standard order statistic theory for exponential distribution gives Y_i's are iid standard exponential random variables. Noting

$$\log S_{(i)} = \frac{1}{\beta}\left(\frac{Y_n}{n} + \frac{Y_{n-1}}{n-1} + \cdots + \frac{Y_i}{i}\right), \tag{2.4}$$

and

$$E(Y_i) = 1, V(Y_i) = 1, Cov(Y_i, Y_j) = 0 \ \ (i \neq j),$$

we immediately obtain (a), (b) and (c).

■

From Lemma 1(a), we obtain the following proposition justfying the rank size rule regression (1.1).

Proposition 1

$$E(\log S_{(i)}) = \frac{\log n}{\beta} - \frac{1}{\beta}\log i + O\left(\frac{1}{n} + \frac{1}{i}\right) \text{ as } n \to \infty, \ i \to \infty$$

Proof

Noting $\sum_{k=1}^{n} \frac{1}{k} = \gamma + \log n + O\left(\frac{1}{n}\right)$ where $\gamma \approx 0.577$ is the Euler's constant (see e.g. Hardy and Wright(1988)), we have, due to Lemma 1(a),

$$E(\log S_{(i)}) = \sum_{k=1}^{n}\frac{1}{k} - \sum_{k=1}^{i-1}\frac{1}{k} = \log n - \log i + O\left(\frac{1}{n} + \frac{1}{i}\right),$$

the second equality using $|\log i - \log(i-1)| \leqq \frac{1}{i-1}$.

■

Proposition 1 implies that approximation of $E(\log S_{(i)}) \approx \frac{\log n}{\beta} - \frac{1}{\beta}\log i$ is justified when n and i are large.

We next focus on the bias and variance of the OLS estimator for (1.1). Letting $(\hat{\alpha}_0, \hat{\alpha}_1)$ be the OLS estimator for (α_0, α_1) of (1.1), we have

$$\begin{aligned}\begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{bmatrix} &= \begin{bmatrix} n & \sum \log i \\ \sum \log i & \sum \log^2 i \end{bmatrix}^{-1} \begin{bmatrix} \sum \log S_{(i)} \\ \sum \log i \log S_{(i)} \end{bmatrix} \\ &= \frac{1}{n\sum \log^2 i - \left(\sum \log i\right)^2} \begin{bmatrix} \sum \log^2 i \sum \log S_{(i)} - \sum \log i \sum \log i \log S_{(i)} \\ -\sum \log i \sum \log S_{(i)} + n \sum \log i \log S_{(i)} \end{bmatrix},\end{aligned}$$

where and hereafter we drop summation range if not ambiguous.

We use the following Lemma to obtain approximate expressions of bias and variance of $\hat{\alpha}_1$.

Lemma 2

(a) $\sum \log i = n\log n - n + \frac{1}{2}\log n + O(1)$.

(b) $\sum \log^2 i = n\log^2 n - 2n\log n + 2n + \frac{1}{2}\log^2 n - \log n + O(1)$.

(c) $\sum \log i \left(\frac{1}{n} + \cdots + \frac{1}{i}\right) = n\log n - 2n + \frac{1}{4}\log^2 n + O(\log n)$.

(d) $\sum \frac{\log i}{i} = \frac{\log^2 n}{2} + o\left(\log^2 n\right)$.

(e) $\sum \frac{\log^2 i}{i} = \frac{\log^3 n}{3} + o\left(\log^3 n\right)$.

(f) $\sum \frac{\log^2 i}{i^2} = O(1)$.

Proof

(a) is a standard result used to prove Stirling's formula. (d), (e), (f) are easily obtained by their integral approximation. To prove (b), write

$$\sum_{i=1}^{n} \log^2 i = n\log^2 n - \sum_{i=1}^{n-1} i\left\{\log^2(i+1) - \log^2 i\right\},$$

and

$$\begin{aligned}\sum_{i=1}^{n-1} i\left\{\log^2(i+1) - \log^2 i\right\} &= \sum_{i=1}^{n-1} i\log\left(1+\frac{1}{i}\right)\left\{\log\left(1+\frac{1}{i}\right) + 2\log i\right\} \\ &= \sum_{i=1}^{n-1} i\log^2\left(1+\frac{1}{i}\right) + 2\sum_{i=1}^{n-1} i\log\left(1+\frac{1}{i}\right)\log i \\ &= (A) + (B).\end{aligned}$$

Because

$$\log\left(1+\frac{1}{i}\right)=\frac{1}{i}-\frac{1}{2i^2}+\frac{1}{3i^3}+O\left(\frac{1}{i^4}\right),$$

$$(A)=\sum_{i=1}^{n-1} i\left(\frac{1}{i}-\frac{1}{2i^2}+\frac{1}{3i^3}+O\left(\frac{1}{i^4}\right)\right)^2=\log n+O\left(1\right),$$

and

$$\begin{aligned}(B)&=2\sum_{i=1}^{n-1} i\left(\frac{1}{i}-\frac{1}{2i^2}+\frac{1}{3i^3}+O\left(\frac{1}{i^4}\right)\right)\log i\\&=2\left(n\log n-n+\frac{1}{2}\log n+O\left(1\right)\right)-\left(\frac{1}{2}\log^2 n+O\left(1\right)\right)+O\left(1\right)\\&=2n\log n-2n-\frac{1}{2}\log^2 n+\log n+o\left(1\right),\end{aligned}$$

using (f), we have (b).

To prove (c), write

$$\begin{aligned}\sum\log i\left(\frac{1}{n}+\cdots+\frac{1}{i}\right)&=\sum_{i=1}^{n}\log i\left(\sum_{k=i}^{n}\frac{1}{k}\right)\\&=\sum_{k=1}^{n}\frac{1}{k}\sum_{i=1}^{k}\log i.\end{aligned}$$

Applying (a), the right hand side equals to

$$\begin{aligned}&\sum_{k=1}^{n}\frac{1}{k}\left(k\log k-k+\frac{1}{2}\log k+O\left(1\right)\right)\\&=\sum_{k=1}^{n}\left\{\log k-1+\frac{\log k}{2k}+O\left(\frac{1}{k}\right)\right\}\\&=n\log n-2n+\frac{\log^2 n}{4}+O\left(\log n\right)\end{aligned}$$

because of (a) and (d).

■

The following proposition provides exact and approximate expressions for the bias and variance of the OLS estimator.

Proposition 2

$$
\begin{aligned}
&(a)\quad E\left(\hat{\alpha}_1\right)-\left(-\tfrac{1}{\beta}\right)=\frac{n\sum\log i\left(\frac{1}{n}+\cdots+\frac{1}{i}\right)-n\sum\log i+n\sum\log^2 i-\left(\sum\log i\right)^2}{\beta\left\{n\sum\log^2 i-\left(\sum\log i\right)^2\right\}}\\
&\qquad\qquad\qquad\qquad\;=\frac{C_n}{\beta},\qquad C_n=-\frac{\log n}{4n}+o\left(\frac{\log n}{n}\right)\\
&(b)\quad V\left(\hat{\alpha}_1\right)=\frac{1}{\beta^2\left\{n\sum\log^2 i-\left(\sum\log i\right)^2\right\}^2}\sum_{j=1}^{n}\left(\frac{n}{j}\sum_{k=1}^{j}\log k-\sum\log i\right)^2\\
&\qquad\qquad\quad=\frac{D_n}{\beta^2},\qquad D_n=\frac{2}{n}+\frac{\log^3 n}{3n^2}+O\left(\frac{\log^2 n}{n^2}\right).\\
&(c)\quad E\left(\frac{\hat{\alpha}_0}{\log n}\right)=\frac{1}{\beta}+o\left(\frac{1}{\log n}\right).\\
&(d)\quad V\left(\frac{\hat{\alpha}_0}{\log n}\right)=O\left(\frac{\log^4 n}{n}\right).
\end{aligned}
$$

where C_n and D_n are constants depending only on n.

Proof

Using (2.4), write

$$
\hat{\alpha}_1=\frac{1}{\beta\left\{n\sum\log^2 i-\left(\sum\log i\right)^2\right\}}\left\{\sum_{j=1}^{n}\left(na_j-\sum_{i=1}^{n}\log i\right)Y_j\right\}\tag{2.5}
$$

where $a_j=\frac{1}{j}\sum\limits_{k=1}^{j}\log k$. Since $E\left(Y_j\right)=1, j=1,\cdots,n$,

$$
\begin{aligned}
E\left(\hat{\alpha}_1\right)&=\frac{1}{\beta\left\{n\sum\log^2 i-\left(\sum\log i\right)^2\right\}}\sum_{j=1}^{n}\left(na_j-\sum_{i=1}^{n}\log i\right)\\
&=\frac{1}{\beta\left\{n\sum\log^2 i-\left(\sum\log i\right)^2\right\}}\left(n\sum a_i-n\sum\log i\right).
\end{aligned}\tag{2.6}
$$

Noting

$$
\sum a_i=\sum\log i\left(\frac{1}{n}+\cdots+\frac{1}{i}\right),
$$

apply Lemma 2 (a),(b),(c) to obtain

$$
\begin{aligned}
E\left(\hat{\alpha}_1\right)&=\frac{\left(n^2\log n-2n^2+\frac{1}{4}n\log^2 n\right)-n\left(n\log n-n+\frac{1}{2}\log n\right)+o\left(n\log n\right)}{\beta\left\{n^2-\frac{1}{2}n\log^2 n+o\left(n\log n\right)\right\}}\\
&=\frac{-n^2+\frac{1}{4}n\log^2 n+O\left(n\log n\right)}{\beta\left\{n^2-\frac{1}{2}n\log^2 n+O\left(n\log n\right)\right\}}.
\end{aligned}
$$

Then (a) follows.

Next we prove (b). Due to (2.5), we have

$$V(\hat{\alpha}_1) = \frac{1}{\beta^2 \left\{ n \sum \log^2 i - \left(\sum \log i\right)^2 \right\}^2} \sum_{j=1}^{n} \left(n a_j - \sum \log i \right)^2 \tag{2.7}$$

since Y_j are iid with $V(Y_j) = 1$. Because of Lemma 2(a),

$$\begin{aligned} n a_j - \sum_{i=1}^{n} \log i &= \frac{n}{j} \sum_{k=1}^{j} \log k - \sum_{j=1}^{n} \log i \\ &= \left(n \log j + \frac{n \log j}{2j} \right) - \left(n \log n + \frac{1}{2} \log n \right) + O\left(\frac{n}{j} \right) \end{aligned}$$

so that

$$\begin{aligned} \sum_{j=1}^{n} \left(n a_j - \sum_{i=1}^{n} \log i \right)^2 &= n^2 \left(\sum \log^2 j + \sum \frac{\log^2 j}{j} + \sum \frac{\log^2 j}{4j^2} \right) \\ &\quad - 2n \left(n \log n + \frac{1}{2} \log n \right) \left(\sum \log j + \sum \frac{\log j}{2j} \right) \\ &\quad + n \left(n \log n + \frac{1}{2} \log n \right)^2 + O\left(n^2 \log n \sum \frac{1}{j} \right) \\ &= 2n^3 + \frac{1}{3} n^2 \log^3 n + O\left(n^2 \log^2 n \right) \end{aligned}$$

where we use Lemma 2 (b),(d),(e),(f) in the second equality. The denominator of (2.7) equals to $\beta^2 \left(n^2 + o\left(n^2\right) \right)^2$ due to Lemma 2 (a),(b). Therefore we obtain

$$V(\hat{\alpha}_1) = \frac{D_n}{\beta^2}, \ D_n = \frac{2}{n} + \frac{\log^3 n}{3n^2} + O\left(\frac{\log^2 n}{n^2} \right).$$

To prove (c) and (d), write using (2.4) and Lemma 2 (a), (b)

$$\begin{aligned} \frac{\hat{\alpha}_0}{\log n} &= \frac{\left(\sum \log^2 i\right) \sum Y_i - \left(\sum \log i\right) \left(\sum a_i Y_i\right)}{\beta \left\{ n \sum \log^2 i - \left(\sum \log i\right)^2 \right\} \log n} \\ &= \frac{\left\{ n \log^2 n - 2n \log n + 2n + O\left(\log^2 n\right) \right\} \sum Y_i - \left\{ n \log n - n + O\left(\log n\right) \right\} \sum a_i Y_i}{\beta \left(n^2 - \frac{1}{2} n \log^2 n + O\left(n \log n\right) \right) \log n} \end{aligned}$$

Dropping the smaller order terms, due to $E(Y_i) = V(Y_i) = 1$, we have

$$
\begin{aligned}
E\left(\frac{\hat{\alpha}_0}{\log n}\right) &\approx \frac{1}{\beta n^2 \log n}\left\{\left(n\log^2 n - 2n\log n + 2n\right)n - (n\log n - n)\sum a_i\right\} \\
&= \frac{1}{\beta} + o\left(\frac{1}{\log n}\right), \\
V\left(\frac{\hat{\alpha}_0}{\log n}\right) &\approx \frac{1}{\beta^2 n^4 \log^2 n}\sum\left\{\left(n\log^2 n - 2n\log n + 2n\right) - \left(n\log n - n + \frac{1}{2}\log n\right)a_i\right\}^2 \\
&= O\left(\frac{\log^4 n}{n}\right)
\end{aligned}
$$

because of

$$
\sum a_i = \sum \log i\left(\frac{1}{n} + \cdots + \frac{1}{i}\right) = n\log n - 2n + \frac{\log^2 n}{4} + O(\log n) \qquad (2.8)
$$

by Lemma 2(c), and

$$
\begin{aligned}
\sum a_i^2 &= \sum \frac{1}{i^2}\left(\sum_{k=1}^{i}\log k\right)^2 \\
&= \sum\left(\log^2 i - 2\log i + 1 + \frac{\log^2 i}{i} - \frac{\log i}{i} + \frac{\log^2 i}{4i^2} + O(1)\right) \\
&= n\log^2 n - 4n\log n + 4n + o(n) \qquad (2.9)
\end{aligned}
$$

by lemma 2(a), (b).

■

C_n and D_n are shown in Figure 1 and 2 respectively, and Table 1 tabulates them for some selected sample size. We observe both of them tend to zero as n increases while the variance $V(\hat{\alpha}_1)$ approximately equals to $2/(\beta^2 n)$, which, of course, is consistent with Proposition 2(b). In view of (2.6), we can easily construct an unbiased estimator, namely

$$
\bar{\alpha}_1 = \frac{n\sum(\log i)^2 - \left(\sum\log i\right)^2}{n\sum_{i=1}^{n}\log i\left(\frac{1}{n} + \cdots + \frac{1}{i} - 1\right)}\hat{\alpha}_1
$$

where the multiplicative constant depends only on n, independent of (nuisance) parameters. Because the constant is smaller than one in absolute value, this estimator not only eliminates the bias, but also reduces the variance.

Table 1. Bias and variance of OLS estimator
C_n=bias $\times\ \beta$, D_n=variance $\times\ \beta^2$

n	C_n	D_n
50	-0.0822	0.0410
100	-0.0534	0.0201
200	-0.0341	0.0099
500	-0.0183	0.0040

2.2 The distribution of t statistics

In testing the significance of coefficients of linear regression models, we implement t test. Suppose we would like to test the null of $\beta = \beta_0$ in the following regression model with nonstochastic regressor z_i,

$$y_i \sim N\left(\alpha + \beta z_i, \sigma^2\right),$$

where we assume y_i are mutually independent. Letting a, b be the OLS estimator of α, β,

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - a - bz_i)^2\,,\ Z' = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ z_1 & z_2 & z_3 & \cdots & z_n \end{bmatrix},$$

and

$$\hat{v}_b = s^2\left(Z'Z\right)^{-1},$$

we construct the corresponding t statistic

$$t_b = \frac{b - \beta_0}{\sqrt{\hat{v}_b}}.$$

Then t_b has t distribution with degree of freedom $(n-2)$ under the null. However, if the assumptions on the distribution of y do not hold, t_b does not have t distribution so that we will face size distortion. In our case, because $\log S_{(i)}$, the dependent variables, are not only normally distributed but also heteroskedastic and autocorrelated. We obtained the distribution of t statistics for α_1 in the regression (1.1) under the null of $\alpha_1 = -1$ by Monte Carlo simulation. Figure 3 and 4 show the histogram from 100,000 replications when $n = 100, 200$ respectively. The mean, variance, skewness and kurtosis are respectively -2.512, 171.0, 0.423, 1.012 when $n = 200$. Therefore they are obviously far from t distribution. Table 2 shows critical regions of two-sided test for different sizes calculated from the simulation which should be used in testing $\alpha_1 = -1$ instead of quitiles of t distribution.

Table 2. Critical regions of t test by simulation

size	$n = 100$	$n = 200$
10%	$(-\infty, -17, 03], [16.14, \infty)$	$(-\infty, -22, 10], [20.38, \infty)$
5%	$(-\infty, -20, 17], [20.68, \infty)$	$(-\infty, -26, 37], [25.86, \infty)$
1%	$(-\infty, -27, 09], [31.08, \infty)$	$(-\infty, -35, 20], [38.27, \infty)$

We immediately know we face severe size distortion if we blindly apply t test for α_1, because its critical region is set to be around $(-\infty, -2], [2, \infty)$.

Moreover, we found in a simulation not reported here that t tends to become larger in magnitude as the sample size increases. This pheonomenon is caused by the fact that standard error of the regression tends to zero as $n \to \infty$, which is proved in the following proposition.

Proposition 3

Letting

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} \{\log S(i) - \hat{\alpha}_0 - \hat{\alpha}_1 \log i\}^2,$$

we have

$$(a) \quad E(s^2) = O\left(\frac{\log n}{n}\right).$$
$$(b) \quad s^2 \xrightarrow{p} 0 \text{ as } n \to \infty.$$

Proof

To prove (a), write

$$s^2 = \frac{1}{n-2}\sum \Big\{\log^2 S_{(i)} - \frac{\left(\sum \log^2 i\right)\left(\sum \log S_{(i)}\right)^2 - 2\left(\sum \log i\right)\left(\sum \log S_{(i)}\right)\left(\sum \log i \log S_{(i)}\right) + n\left(\sum \log i \log S_{(i)}\right)^2}{n\sum \log^2 i - \left(\sum \log i\right)^2}\Big\} \quad (2.10)$$

Because of (2.4), we have

$$\sum \log S_{(i)} = \sum \log S_i,$$
$$\sum \log^2 S_{(i)} = \sum \log^2 S_i,$$

and

$$\sum \log i \log S_{(i)} = \frac{1}{\beta}\sum a_i Y_i,$$

which leads to

$$E\left(\sum \log^2 S_{(i)}\right) = \frac{2n}{\beta^2} \tag{2.11}$$

$$E\left(\sum \log S_{(i)}\right)^2 = \frac{n(n+1)}{\beta^2} \tag{2.12}$$

$$E\left(\sum \log S_{(i)} \sum \log i \log S_{(i)}\right) = \frac{n+1}{\beta^2}\sum a_i \tag{2.13}$$

and

$$E\left(\sum \log i \log S_{(i)}\right)^2 = \frac{1}{\beta^2}\left\{\left(\sum a_i\right)^2 + \sum a_i^2\right\} \tag{2.14}$$

due to $E(Y_i) = 1, E(Y_i^2) = 2, E(Y_i, Y_j) = 1\,(i \neq j)$. (2.8)–(2.14) yield

$$E\left(s^2\right) = \frac{1}{(n-2)\beta^2}\left\{2n - \frac{2n^3 - n^2\log^2 n + O\left(n^2 \log n\right)}{n^2 - \frac{1}{2}n\log^2 n + O\left(\log^2 n\right)}\right\} = O\left(\frac{\log n}{n}\right).$$

It remains to show $V\left(s^2\right) \to 0$ to prove (b). Because

$$\frac{1}{n-2}\sum \log^2 S_{(i)} = \frac{1}{n-2}\sum \log^2 S \xrightarrow{p} E\left(\log^2 S\right) = \frac{2}{\beta^2},$$

it suffices to show

$$\frac{1}{n} y'X\left(X'X\right)^{-1}X'y \xrightarrow{p} \frac{2}{\beta^2}$$

where

$$X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \log 1 & \log 2 & & \log n \end{bmatrix},$$
$$y' = \left[\log S_{(1)}\ \ \log S_{(2)} \cdots \log S_{(n)}\right].$$

Writing

$$\begin{aligned} &\tfrac{1}{n} y'X\left(X'X\right)^{-1}X'y \\ &= \tfrac{1}{\beta n}\left[\log n, -1\right]X'y + \tfrac{1}{n}\left\{y'X\left(X'X\right)^{-1} - \tfrac{1}{\beta}\left[\log n, -1\right]\right\}X'y \\ &= (A) + (B), \end{aligned}$$

we have

$$(A) = \frac{1}{n\beta^2}\left[\log n, -1\right]\begin{bmatrix} \sum Y_i \\ \sum a_i Y_i \end{bmatrix} = \frac{1}{n\beta^2}\sum\left(\log n - a_i\right)Y_i$$

due to (2.4). Its expectation and variance are respectively

$$E[(A)] = \frac{1}{n\beta^2}\left(n\log n - \sum a_i\right) = \frac{2}{\beta^2} + o(1)$$

because of $E(Y_i) = 1$, (2.8), and

$$V[(A)] = \frac{1}{n^2\beta^4}\sum(\log n - a_i)^2 = \frac{1}{n^2\beta^4}\left(n\log^2 n - 2\log n\sum a_i + \sum a_i^2\right)$$
$$= \frac{1}{n^2\beta^4}\left\{4n + O\left(\log^3 n\right)\right\}$$

because of $Cov(Y_i, Y_j) = 0\,(i \neq j)$, $V(Y_i) = 1$, (2.8) and (2.9). Therefore

$$(A) \xrightarrow{p} \frac{2}{\beta^2}$$

by Chebyshev inequality.

Because

$$y'X - \frac{1}{\beta}[\log n,\ -1](X'X)$$
$$= \frac{1}{\beta}\left\{\sum(Y_i - 1) + \frac{1}{2}\log n + O(1),\ \sum a_i(Y_i - 1) + \frac{1}{2}\log^2 n + O(\log n)\right\}$$

using (2.4), we have

$$(B)' =$$
$$\frac{1}{\beta}y'X(X'X)^{-1}\begin{bmatrix}\frac{1}{\log n} & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}\log n & 0\\ 0 & 1\end{bmatrix}\frac{1}{\beta}\begin{bmatrix}\frac{1}{n}\sum(Y_i-1) + \frac{1}{2n}\log n + O\left(\frac{1}{n}\right)\\ \frac{1}{n}\sum a_i(Y_i-1) + \frac{1}{2n}\log^2 n + O\left(\frac{\log n}{n}\right)\end{bmatrix},$$

prime denoting transposition. Proposition 2 with Chebyshev inequality implies

$$\frac{1}{\beta}y'X(X'X)^{-1}\begin{bmatrix}\frac{1}{\log n} & 0\\ 0 & 1\end{bmatrix} \xrightarrow{p} \frac{1}{\beta^2}\begin{bmatrix}1\\ -1\end{bmatrix},$$

while

$$\begin{bmatrix}\log n & 0\\ 0 & 1\end{bmatrix}\frac{1}{\beta}\begin{bmatrix}\frac{1}{n}\sum(Y_i-1) + \frac{1}{2n}\log n + O\left(\frac{1}{n}\right)\\ \frac{1}{n}\sum a_i(Y_i-1) + \frac{1}{2n}\log^2 n + O\left(\frac{\log n}{n}\right)\end{bmatrix} \xrightarrow{p} \begin{bmatrix}0\\ 0\end{bmatrix}$$

because

$$E\left(\begin{bmatrix}\log n & 0\\ 0 & 1\end{bmatrix}\frac{1}{\beta}\begin{bmatrix}\frac{1}{n}\sum(Y_i-1) + \frac{1}{2n}\log n + O\left(\frac{1}{n}\right)\\ \frac{1}{n}\sum a_i(Y_i-1) + \frac{1}{2n}\log^2 n + O\left(\frac{\log n}{n}\right)\end{bmatrix}\right)$$
$$= \frac{1}{\beta}\begin{bmatrix}\frac{\log^2 n}{2n} + O\left(\frac{\log n}{n}\right)\\ \frac{\log^2 n}{2n} + O\left(\frac{\log n}{n}\right)\end{bmatrix} \to \begin{bmatrix}0\\ 0\end{bmatrix}$$

and

$$V\left(\begin{bmatrix}\log n & 0\\ 0 & 1\end{bmatrix}\frac{1}{\beta}\begin{bmatrix}\frac{1}{n}\sum\left(Y_i-1\right)+\frac{1}{2n}\log n+O\left(\frac{1}{n}\right)\\ \frac{1}{n}\sum a_i\left(Y_i-1\right)+\frac{1}{2n}\log^2 n+O\left(\frac{\log n}{n}\right)\end{bmatrix}\right)$$

$$=\frac{1}{\beta^2}V\begin{bmatrix}\frac{\log n}{n}\sum\left(Y_i-1\right)\\ \frac{1}{n}\sum a_i\left(Y_i-1\right)\end{bmatrix}=\frac{1}{\beta^2}\begin{bmatrix}\frac{\log^2 n}{n} & \frac{\log n}{n^2}\sum a_i\\ \frac{\log n}{n^2}\sum a_i & \frac{1}{n^2}\sum a_i^2\end{bmatrix}$$

$$=\frac{1}{\beta^2}\begin{bmatrix}\frac{\log^2 n}{n} & \frac{\log^2 n}{n}\sum a_i+O\left(\frac{\log n}{n}\right)\\ \frac{\log^2 n}{n}+O\left(\frac{\log n}{n}\right) & \frac{\log^2 n}{n}+O\left(\frac{\log n}{n}\right)\end{bmatrix}\to\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$$

because of (2.8), (2.9). Therefore $(B)\overset{p}{\to}0$ which completes the proof.

We only states the following asymptotic result without proof.

Proposition 4

$$\begin{bmatrix}\frac{\sqrt{n}}{\log n} & 0\\ 0 & \sqrt{n}\end{bmatrix}\begin{bmatrix}\hat{\alpha}_0\\ \hat{\alpha}_1\end{bmatrix}-\begin{bmatrix}\frac{1}{\beta}\\ -\frac{1}{\beta}\end{bmatrix}\overset{d}{\to}N\left(\begin{bmatrix}0\\ 0\end{bmatrix},\frac{2}{\beta^2}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\right).$$

We can straightforwardly prove this by Lindeberg-Feller central limit theorem and Cramer device based on the expressions written in terms of Y_j like (2.5) because Y_i are iid. We note the asymptotic covariance matrix is singular, and that $\hat{\alpha}_0$ is not $\sqrt{n}-$ consistent but $\frac{\sqrt{n}}{\log n}-$ consistent. Lemma 2 (a), (b), Proposition 3 and 4 imply the following result.

Proposition 5 For $t=\frac{\hat{\alpha}_1-\left(-\frac{1}{\beta}\right)}{\sqrt{s^2(X'X)^{-1}_{22}}}$,

$$t\overset{p}{\to}\infty \quad \text{as} \quad n\to\infty,$$

where $(X'X)^{-1}_{22}$ denotes the (2,2) element of $(X'X)^{-1}$, which equals to $n/\left\{n\sum\log^2 i-\left(\sum\log i\right)^2\right\}$.

The above proposition indicates the t value explodes asymptotically under the null of true parameter value. Therefore, when we would like to test a null hypothesis such as $\alpha_1=-1$, we know we should never use t test, but we should apply an asymptotic normality based test using

$$\frac{\hat{\alpha}_1-\alpha_1}{\sqrt{2\hat{\alpha}_1^2/n}}\overset{d}{\to}N\left(0,1\right)\ , \tag{2.15}$$

as recommended in e.g. Gabaix and Ioannides (2003). $1/\beta^2$ involved in the asymptotic variance is replaced by a consistent estimator $\hat{\alpha}_1^2$ under the null. In many application work, such as Rosen and Resnick (1980), Alperovich (1984) and Soo (2002), formal application of t test provides very large t values, leading to wrong conclusions.

3 More efficient estimation

We propose two methods of efficiency improvement in the estimation of (1.1). One is generalized least squares (GLS) method adjusting nonspherical disturbances, while the other is a trimmed least squares regression. The idea is that observing $V\left(\log S_{(i)}\right)$ is larger for smaller i, and also approximation (1.1) is worse for smaller i in view of Proposition 1, we can expect to improve the statistical propeties of the estimator by dropping some obsevations with smaller i, or larger observations.

3.1 GLS estimation

Putting $\Omega = V(y)$, GLS estimator for α_0, α_1 is simply

$$\begin{bmatrix} \tilde{\alpha}_0 \\ \tilde{\alpha}_1 \end{bmatrix} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}y, \tag{3.1}$$

where y and X are defined in the proof of Proposition 3.

Then its variance is

$$V\left(\begin{bmatrix} \tilde{\alpha}_0 \\ \tilde{\alpha}_1 \end{bmatrix}\right) = \left(X'\Omega^{-1}X\right)^{-1}. \tag{3.2}$$

Expressions for elements of Ω are in Lemma 1 (b),(c).

An interesting feature in (3.1) is that it is free from nuisance parameter unlike usual GLS estimation. Normally Ω involves some nuisance parameters and thus GLS estimation is infeasible, so that we will need to estimate Ω in the first step in practice. In view of the previous section, Ω itself involves an unknown parameter β, but it appears only as a multiplicative constant. Then, due to the form of (3.1), it cancels, so that (3.1) turns to be feasible. Table 3 compares the bias and variance of OLS and GLS estimators for some selected n.

Similarly to the OLS estimator, we can obtain the exact bias and variance of GLS estimator analogous to Proposition 2, which are in Figure 5 and 6. We do not present them explicitly because of their long and tedius expressions. GLS performs clearly better than OLS both in bias and variance. We give the following two remarks regarding this result in view of Table 3. Firstly, GLS procedure reduces not only the variance but also the bias. Secondly, we see the variance appoximately equals to $1/\left(\beta^2 n\right)$, which coincides with Cramer-Rao lower bound for $1/\beta$ in fact.

Table 3. Bias and variance of GLS estimator

n	bias$\times\beta$	variance$\times\beta^2$
50	-0.3681	0.0220
100	-0.2128	0.0105
200	-0.1217	0.0051
500	-0.0058	0.0020

Therefore, we anticipate GLS gives an efficient estimate, comparable with the maximum likelihood estimator (MLE).

3.2 Trimmed OLS and GLS

Proposition 1 and Lemma 1(a) imply that source of the bias of least squares estimators is the approximation error of $\frac{1}{n}+\cdots+\frac{1}{i}$ by $\log n-\log i$, and it is larger for smaller i. Then we conjecture the bias can be reduced by excluding observations with smaller i. Also Lemma 2(b) imply that variance of least squares estimators could become smaller if we trim observations with smaller i, though there should no doubt be trade-off between efficiency gain by exclusion of larger variance data points and efficiency loss due to the reduced sample size. Letting $(\hat{\alpha}_{0,k},\hat{\alpha}_{1,k})$and $(\tilde{\alpha}_{0,k},\tilde{\alpha}_{1,k})$ be OLS and GLS estimators from the subsample of $\left\{\log S_{(k+1)},\cdots,\log S_{(n)}\right\}$, where the larger k observations are excluded, we have similarly to Proposition 2,

$$
\begin{aligned}
E\left(\hat{\alpha}_{1,k}\right)-\left(-\frac{1}{\beta}\right) &= \frac{1}{\beta}\left\{\frac{(n-k)\sum_{i=k+1}^{n}\log i\left(\frac{1}{n}+\cdots+\frac{1}{i}-1\right)}{(n-k)\sum_{i=k+1}^{n}(\log i)^2-\left(\sum_{i=k+1}^{n}\log i\right)^2}+1\right\} \\
&= \frac{C_{n,k}}{\beta}
\end{aligned}
$$

and

$$
\begin{aligned}
&V\left(\begin{bmatrix}\hat{\alpha}_{0,k}\\ \hat{\alpha}_{1,k}\end{bmatrix}\right) \\
&= \begin{bmatrix} n-k & \sum_{i=k+1}^{n}\log i \\ \sum_{i=k+1}^{n}\log i & \sum_{i=k+1}^{n}(\log i)^2\end{bmatrix}^{-1} V\left(\begin{bmatrix}\sum_{i=k+1}^{n}\log S_{(i)} \\ \sum_{i=k+1}^{n}\log i\log S_{(i)}\end{bmatrix}\right)\begin{bmatrix} n-k & \sum_{i=k+1}^{n}\log i \\ \sum_{i=k+1}^{n}\log i & \sum_{i=k+1}^{n}(\log i)^2\end{bmatrix}^{-1} \\
&= \frac{D_{n,k}}{\beta^2}.
\end{aligned}
$$

$C_{n,k}$ and $D_{n,k}$ are constants determined only by n and k independent of unknown quantities. We omit the formulae of the variance for the GLS estimator. Table 4 shows these values with mean squared error (MSE) $=\beta^2\left(C_{n,k}^2+D_{n,k}\right)$ for $n=100$, $k=0,\cdots,20$. We find larger k yields smaller bias in both OLS and GLS, while variance of OLS estimator attains the minimum when $k=8$ as a result of the trade-off. GLS variance, on the other hand, increases with k thus there is no efficiency gain but only efficiency loss by decreased sample size.

Table 4. Bias and variance of trimmed OLS and GLS estimators (n=100)

k	bias(OLS)	var(OLS)	MSE(OLS)	bias(GLS)	var(GLS)	MSE(GLS)
0	−0.05342	0.02006	0.02292	−0.02128	0.01055	0.01101
1	−0.03525	0.01723	0.01847	−0.01913	0.01061	0.01097
2	−0.02838	0.01628	0.01708	−0.01763	0.01068	0.01099
3	−0.02449	0.01578	0.01638	−0.01650	0.01077	0.01104
4	−0.02190	0.01549	0.01597	−0.01560	0.01086	0.01110
5	−0.02001	0.01531	0.01571	−0.01486	0.01096	0.01118
6	−0.01855	0.01520	0.01554	−0.01423	0.01106	0.01126
7	−0.01738	0.01514	0.01544	−0.01369	0.01117	0.01136
8	−0.01641	0.01511	0.01538	−0.01321	0.01128	0.01146
9	−0.01559	0.01511	0.01535	−0.01279	0.01140	0.01156
10	−0.01488	0.01514	0.01536	−0.01240	0.01152	0.01167
11	−0.01426	0.01518	0.01539	−0.01206	0.01164	0.01179
12	−0.01372	0.01524	0.01543	−0.01174	0.01177	0.01190
13	−0.01323	0.01532	0.01549	−0.01145	0.01190	0.01203
14	−0.01280	0.01540	0.01557	−0.01119	0.01203	0.01216
15	−0.01240	0.01550	0.01566	−0.01094	0.01217	0.01229
16	−0.01204	0.01561	0.01575	−0.01071	0.01231	0.01242
17	−0.01171	0.01572	0.01586	−0.01049	0.01245	0.01256
18	−0.01141	0.01585	0.01598	−0.01029	0.01260	0.01271
19	−0.01112	0.01598	0.01611	−0.01009	0.01275	0.01286
20	0.01086	0.01613	0.01624	−0.00991	0.01291	0.01301

Based on the above findings, we propose an optimal trimming rule by the minimum MSE principle. When $n = 100, k = 9$ gives the optimal trimming in OLS estimation, while in GLS estimation, $k = 1$ is the best choice. In OLS estimation, we attain about 33% efficiency gain. In GLS estimation, variance of $y' = \left[\log S_{(1)}, \cdots, \log S_{(n)}\right]$ is stabilized by Ω^{-1} (see (3.1)) so that we need to exclude much less observations than the OLS. We note the best trimming points depend only on n because MSE= $\beta^2 \left(C_{n,k}^2 + D_{n,k}\right)$ where $C_{n,k}$ and $D_{n,k}$ depend only on n and k. Table 5 gives the best trimming points for some n.

Table 5. Optimal trimming points

n	OLS	GLS
50	6	1
100	9	1
200	17	1
500	39	2

As easily expected, we should exclude more observations for larger sample size.

4 Empirical application

This section shows an empirical application of the theory provided in the previous sections. We implement OLS, GLS, trimmed OLS and trimmed GLS estimation for (1.1) using Japanese Metropolitan Employment Area (MEA) data developed by Kanemoto and Tokuoka (2001). We cut off cities with population smaller than 200,000[3], then the number of observation is 86. Optimal numbers of trimming when n=86 are computed similarly to Table 4 to be 8 for the OLS and 1 for the GLS. Therefore we have 78 and 85 observations in trimmed OLS and trimmed GLS estimation respectively. Table 6 reports the results only on α_1 because α_0 is of less importance and interest. The estimate of α_1 is mostly close to -1, but t test reject the null of $\alpha_1 = -1$ in all estimation methods at 5% size. However, as we pointed out in Propostion 5, t test does not work in this problem. We should trust the results from asymptotic normality based test of (2.15) instead shown in the fourth column, then we cannot reject the null of $\alpha_1 = -1$ in all four estimates, thus we may conlude the rank size rule applies in Japanese city size distribution. As the point estimate for α_1, trimmed GLS estimation should privide the best estimate in theory as far as the Pareto assumption is correct. Thus we believe $\alpha_1 = -1.049$ is the best point estimate for α_1.

Table 6. Inference in α_1 for Japanese city size data

Estimation method	Estimate	t value	normality based test statistic
OLS	–0.921	4.224	0.564
GLS	–1.051	–2.301	–0.451
trimmed OLS	–0.820	10.924	1.350
trimmed GLS	–1.049	–2.258	–0.408

5 Conclusion

We examined statistical properties of least squares estimators for rank size rule regression of city size under Pareto distribution. Standard method in empirical study of regional science has been OLS estimation and t test based on it. We obtained exact bias and variance of OLS estimator for the coefficient, where we found the bias is multiplicative and easy to correct by multiplying a constant depending only on n. By means of Monte Carlo simulation, we obtained distribution of t statistics, where we found t statistic does not have t distribution, and we will face a severe size distortion if we blindly implement t test. Moreover we proved t asymptotically explodes in fact.

We propose to apply GLS procedure because the explained variable is heteroskedastic and autocorrelated. Both of the bias and variance are significantly reduced and we believe the variance attains Cramer-Rao lower bound. As another

[3]As pointed out e.q. in Rosen and Resnick (1980), the results can be different if we change the cut off point. In this case we may consider Japanese city size is distributed as Pareto conditionally on $S \geqq 200,000$.

tool of efficiency gain, we propose a trimmed least squares method, which works well for OLS, but not so clearly effective for GLS. Obviously when we are sure of the Pareto assumption, GLS or MLE is the best, but when we are not so sure, OLS may have an advantage from robustness point of view, and we believe, trimmed OLS may have a good performance because $\log S_{(i)}$ should still have larger variance for smaller i even if the underlying distribution is not Pareto. Research toward this direction is currently under way.

Figure 1. Bias of OLS estimator

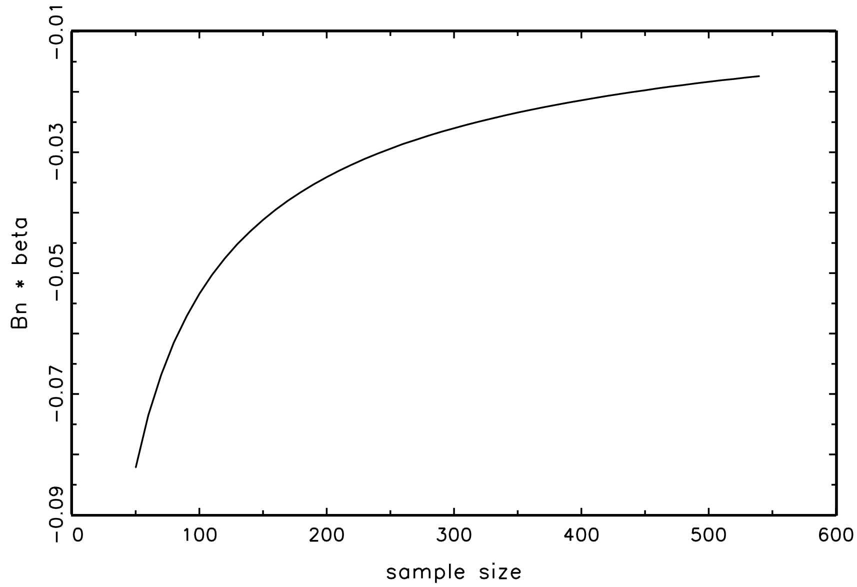


Figure 2. Variance of OLS estimator

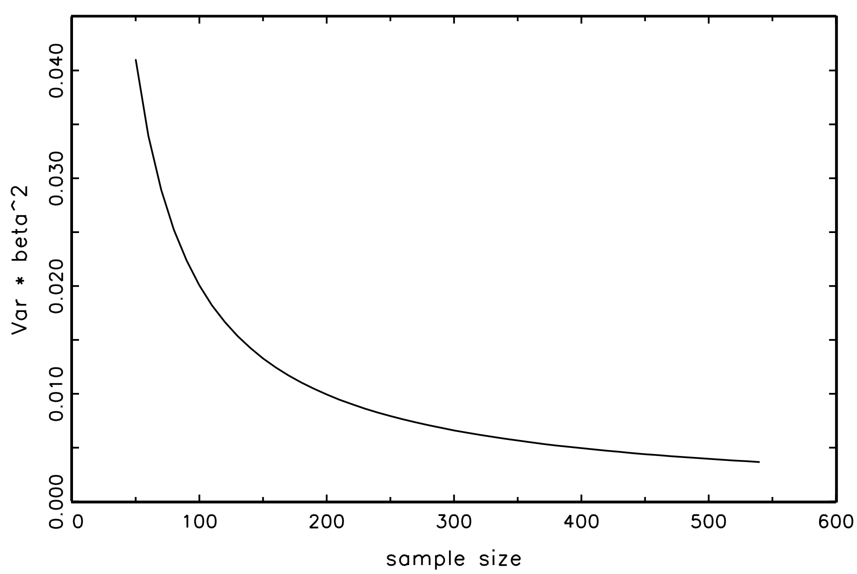


Figure 3. Null distribution of t statistics (n=100)

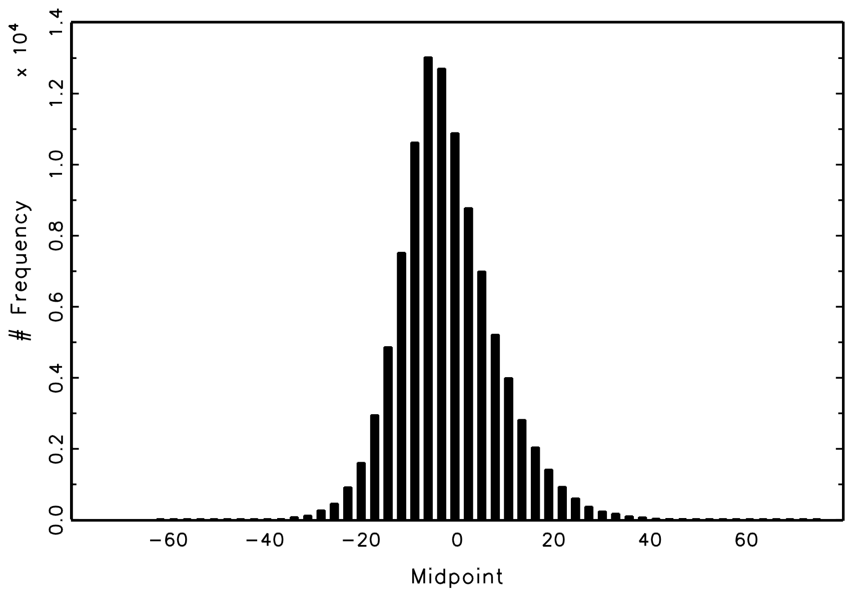


Figure 4. Null distribution of t statistics (n=200)

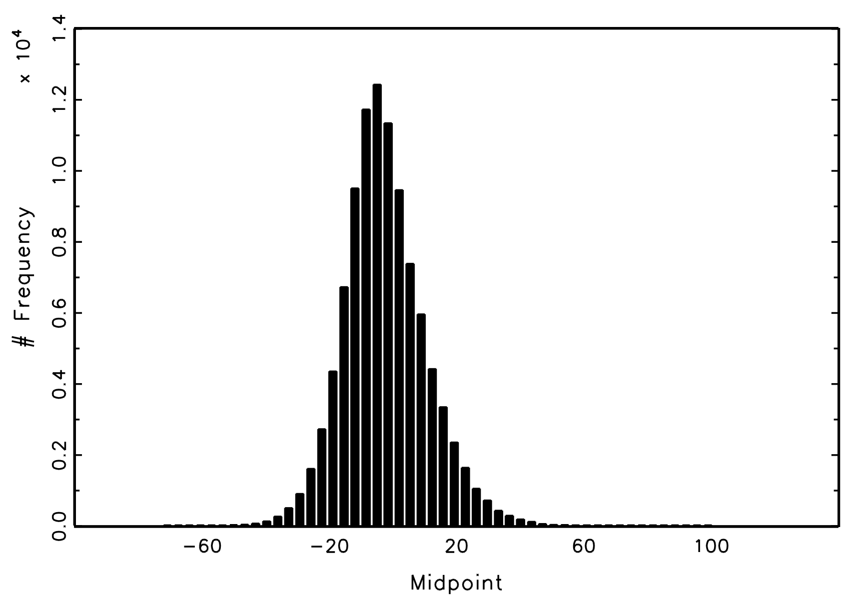


Figure 5. Bias of GLS estimator

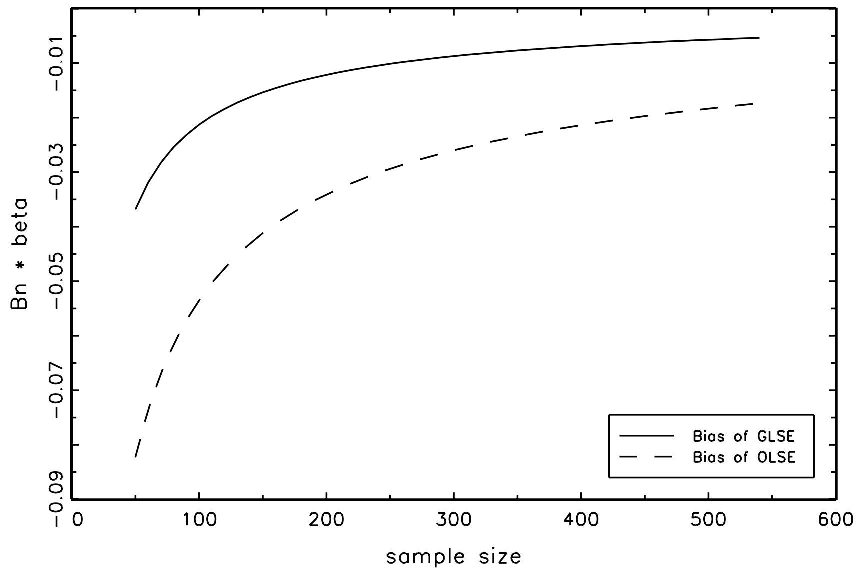


Figure 6. Variance of GLS estimator

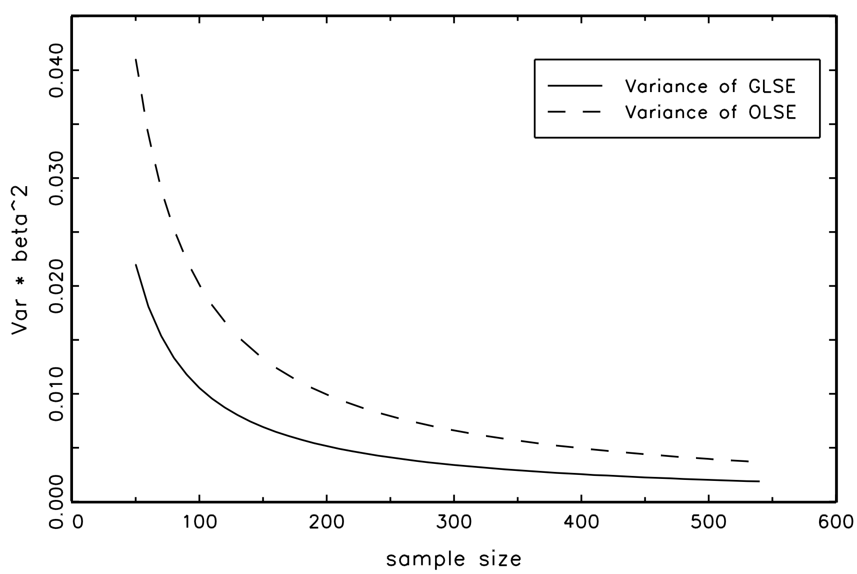


Figure 7. Bias of trimmed OLS and GLS estimators (n=100)

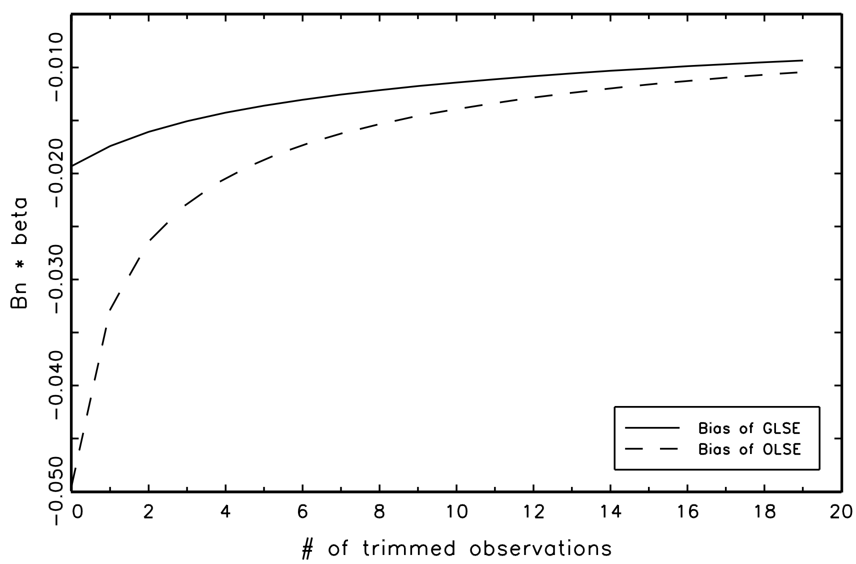


Figure 8. Variance of trimmed OLS and GLS estimators (n=100)

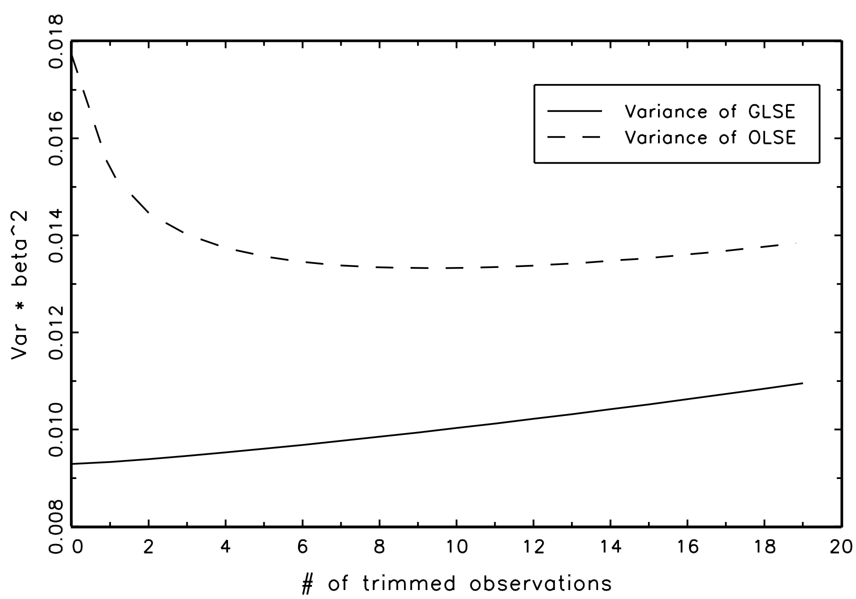
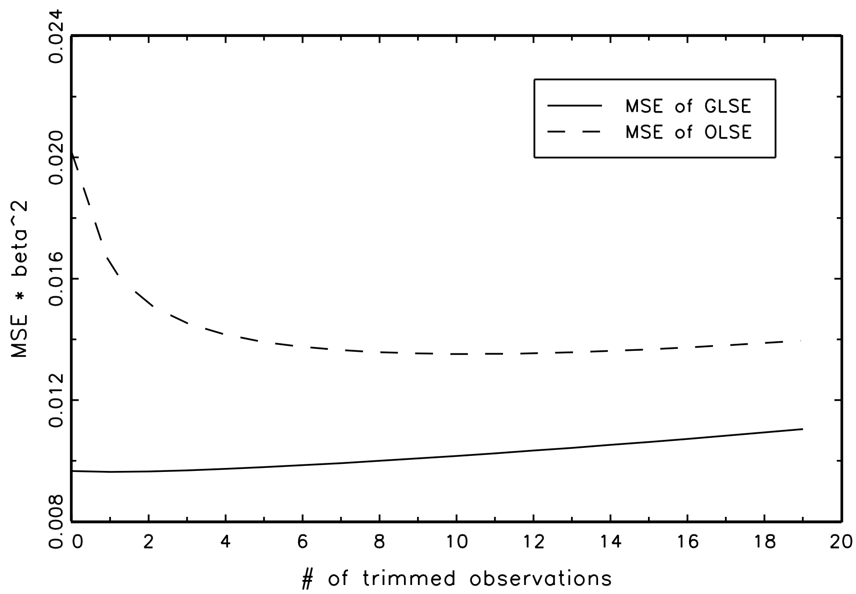


Figure 9. MSE of trimmed OLS and GLS estimators (n=100)



References

[1] Alperovich, G.A. (1984) "The Size Distribution of Cities: On the Empirical Validity of the Rank-Size Rule" Journal of Urban Economics, 16, 232-239.

[2] Alperovich, G.A. (1988) "A New Testing Procedure of Rank Size Distribution" Journal of Urban Economics, 23, 251-259.

[3] Alperovich, G.A. (1989) "The Distribution of City Size: A Sensitivity Analysis" Journal of Urban Economics, 25, 93-102.

[4] Alperovich, G.A. (1992) "Economic Development and Population Concentration" Economic Development and Cultural Change, 41, 63-73.

[5] Auerbach, F. (1913) "Das Gesetz der Bevolkerungskonzentration" Petermanns Geoghische Mitteilungen, 59, 74-76.

[6] Beirlant, J., G. Dierckx, Y. Goegebeur, and G. Matthys (2000) "Tail Index Estimation and an Exponential Regression Model," Extremes, 2, 177-200

[7] Beirlant, J., P. Vynckier, and J.L. Teugels (1996), "Tail Index Estimation, Pareto Quantile Plots, And Regression Diagnostics," Journal of the American Statistical Association, 91, 1659-1667.

[8] Berry, B.J.L. (1961) "City Size Distribution and Economic Development" Economic Development and Cultural Change, 9, 573-588.

[9] Cameron, T.A. (1990) "On-Stage Structural Models to Explain City Size" Journal of Urban Economics, 27, 294-307.

[10] Cheshire, P. (1999) "Trends in Sizes and Structures of Urban Areas" Regional Science and Urban Economics Vol.4, 1339-1373.

[11] Csorgo, S., P. Deheuvels, and D. Mason (1985) "Kernel Estimates of the Tail Index of a Distribution," Annals of Statistics, 13, 1050-1077.

[12] Csorgo, S. and L. Viharos (1998) "Estimating the Tail Index," in Asymptotic Methods in Probability and Statistics, et. by B. Szyszkowics, North-Holland, Amsterdam.

[13] De Haan, L. and S.I. Resnick (1980), "A Simple Asymptotic Estimates for the Index of a Stable Distribution," Journal of the Royal Statistical Society, B, 42, 82-88.

[14] Dekkers, A.L.M., J.H.J. Einmahl, and L. De Haan (1989) “A Moment Estimator for the Index of an Extreme Value Distribution,” Annals of Statistics, 17, 1833-1855.

[15] Dobkins, L.H. and Y.M. Ioannides (2000) ”Dynamic Evolution of the Size Distribution of US cities” J.M. huriot and J.F. Thesse (Eds.), Economics of Cities, Cambridge, cambridge University Press

[16] Duranton, G. (2002) ”City Size Distributions as a consequence of the growth process”, mimeo, London School of Economics.

[17] Feuerverger, A. and P. Hall (1999) “Estimationg a Tail Exponent by Modelling Departure from a Pareto Distribution,” Annals of Statistics, 27, 760-781.

[18] Gabaix, H. (1999) “Zipf’s Law for Cities : An Explanation” Quartarly Journal of Economics, 114, 739-767

[19] Gabaix, H. and Y.M. Ioannides (2003) “The Evolution of City Size Distributions”forthcoming in Handbook of Urban and Regional Economics, vol.4.

[20] Guerin-Pace, F. (1995) ”Rank-Size Distribution and the Process of Urban Growth” Urban Studies, 32, 551-62.

[21] Hardy, G.H. and E.M. Wright (1978) An Introduction to the Theory of Numbers, Springer.

[22] Hosking, J.R.M., J.R. Wallis, and E.F. Wood (1985) “Estimation of the Generalized Extreme-Value Distribution by the Method of Probability Moments,” Technometrics, 27, 251-261.

[23] Hsing, Y. (1990) ”A Note on Functional Forms and the Urban Size Distribution” Journal of Urban Economics, 27, 73-79.

[24] Johnson, N.L., S. Kotz and N. Balakrishnan (1995) Continuous Univariate Distributions, 2nd Ed., Wiley.

[25] Kanemoto, Y. and K. Tokuoka (2001) ”The proposal for the standard definition of the metropolitan area in Japan,” Discussion paper No.J-55, Center for International Research on the Japanese Economy, Tokyo University, Tokyo, Japan.

[26] Malecki, E.J. (1980) ”Growth and Change in the Analysis of Rank-Size Distributions, Empirical Findings” Environment and Planning A, 12, 41-52.

[27] Mori T., K. Nishikimi and T.E. Smith (2003) ”Some Empirical Regularities of Spacial Economies: A Relationship between Industrial Location and City Size.”

[28] Parr, J.B. (1985) "A Note on the Size Distribution of Cities over Time" Journal of Urban Economics, 18, 199-212.

[29] Pickands, J. III(1975) "Statistical Inference Using Extreme Order Statistics," Annals of statistics, 3, 119-131.

[30] Rosen, K.T. and M. Resnick (1980)"The Size distribution of Cities: An Explanation of the Pareto Low and Primacy", Journal of Urban Economics, 8, 165-186.

[31] Smith, T.E. (2003)"Notes on Rank-Size and NAS Estimation for the Pareto Mode", mimeo.

[32] Soo, K.T. (2002) "Zipf's Law for Cities: A Cross Country Investigation", mimeo, London School of Economics

[33] Teugels, J.L. (1981) "Limit Theorems on Order Statistics," Annals of Probability, 9, 868-880.

[34] Zipf, G.K. (1949) Human Behaviour and the Principle of Least Effort, An Introduction to Human Ecology, Cambridge, MA: Addison-Wesley