PY538 - HW 1

Where is problem 2?

8.5/10

Efe Yigitbasi
(Dated: February 12, 2015)

1) a)

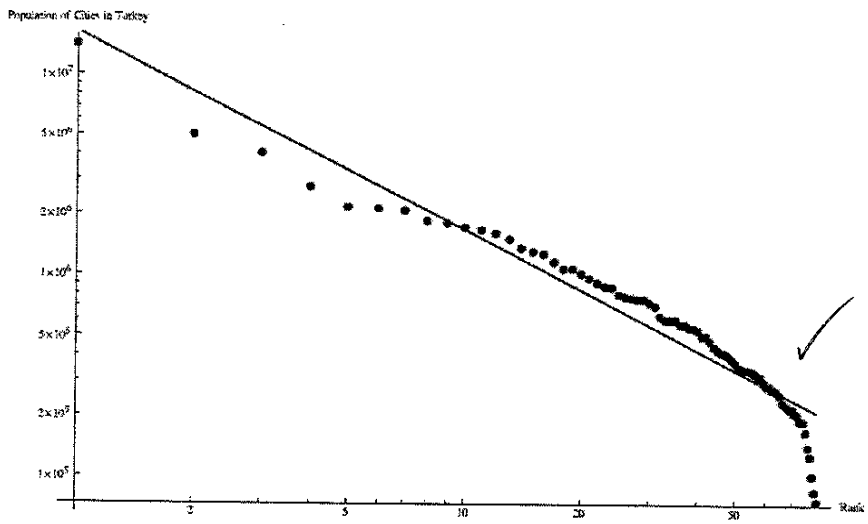


FIG. 1: Zipf Plot for cities of Turkey.

The plot is not really linear for both the upper tail and the lower tail. For the middle part it look like the exponent is $b = -0.99$. One explanation for the odd behavior of the tails might be internal migration which is a serious problem in Turkey. The basic model for Zipf's Law assumes a growth rate proportional to size, basically dominated by birth rates. However due to economical reasons the migration from smallest cities to the largest ones is not something negligible. Interesting...

b)

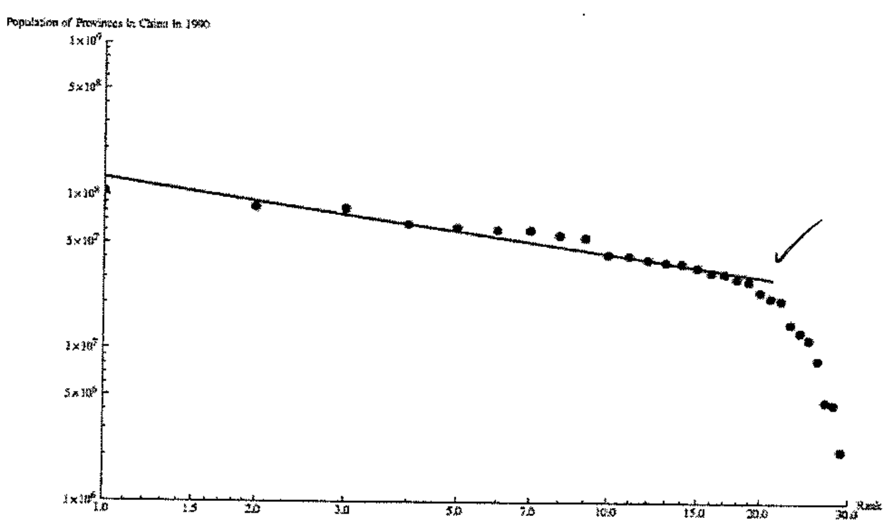


FIG. 2: Zipf Plot for provinces of China (1990). $b = -0.48$.

c)

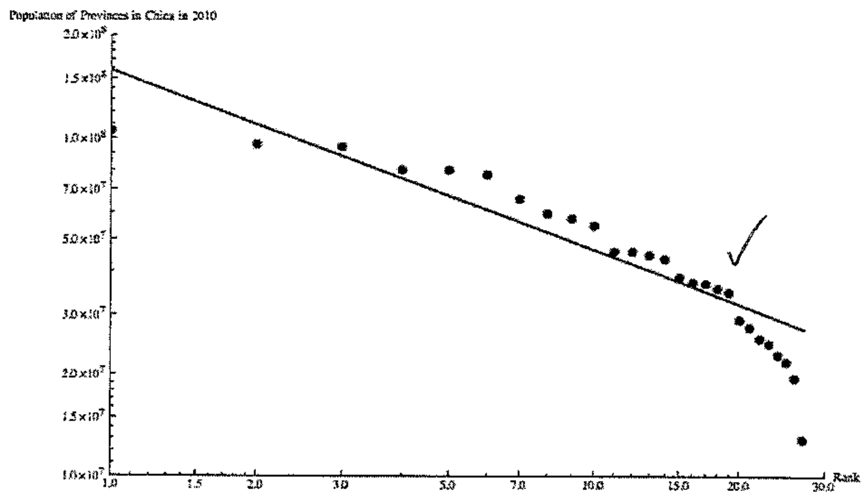


FIG. 3: Zipf Plot for provinces of China (2010). $b = -0.53$:

Assuming a finite birth rate that is roughly constant for each city, assuming a constant city founding rate for creation of new cities (α); the rank of a city (R) is proportional to the time it's founded multiplied by the creation constant (αt_0). Also the ratio of the population of a city (n) to the total population (N) remains a constant. Then,

$$R = \alpha t_0 \tag{1}$$

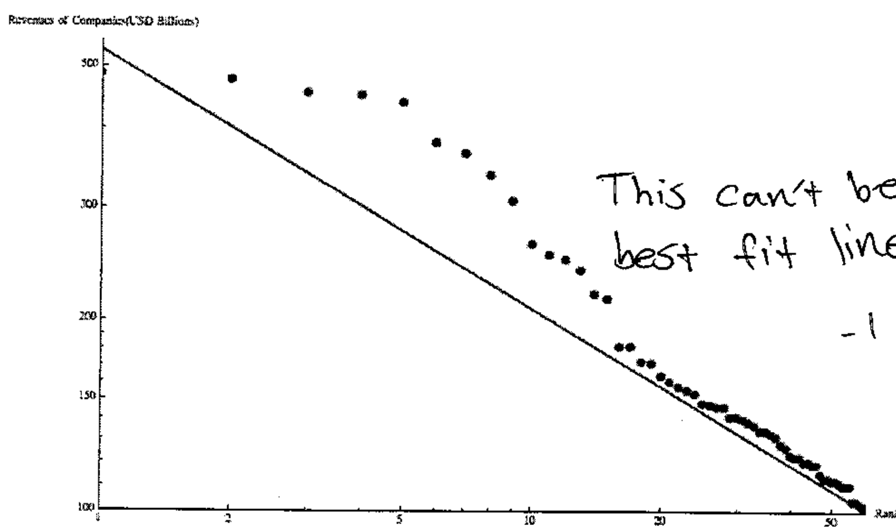
This is great, Efe

$$\frac{n}{N} = \frac{1}{t_0} \tag{2}$$

+1

$$n = \frac{\alpha N}{R} \tag{3}$$

So the size of a city is inversely proportional to its rank.
3)



This can't be the best fit line...

-1

FIG. 4: Zipf Plot for largest companies in the world sorted by revenue. $b = -0.41$.

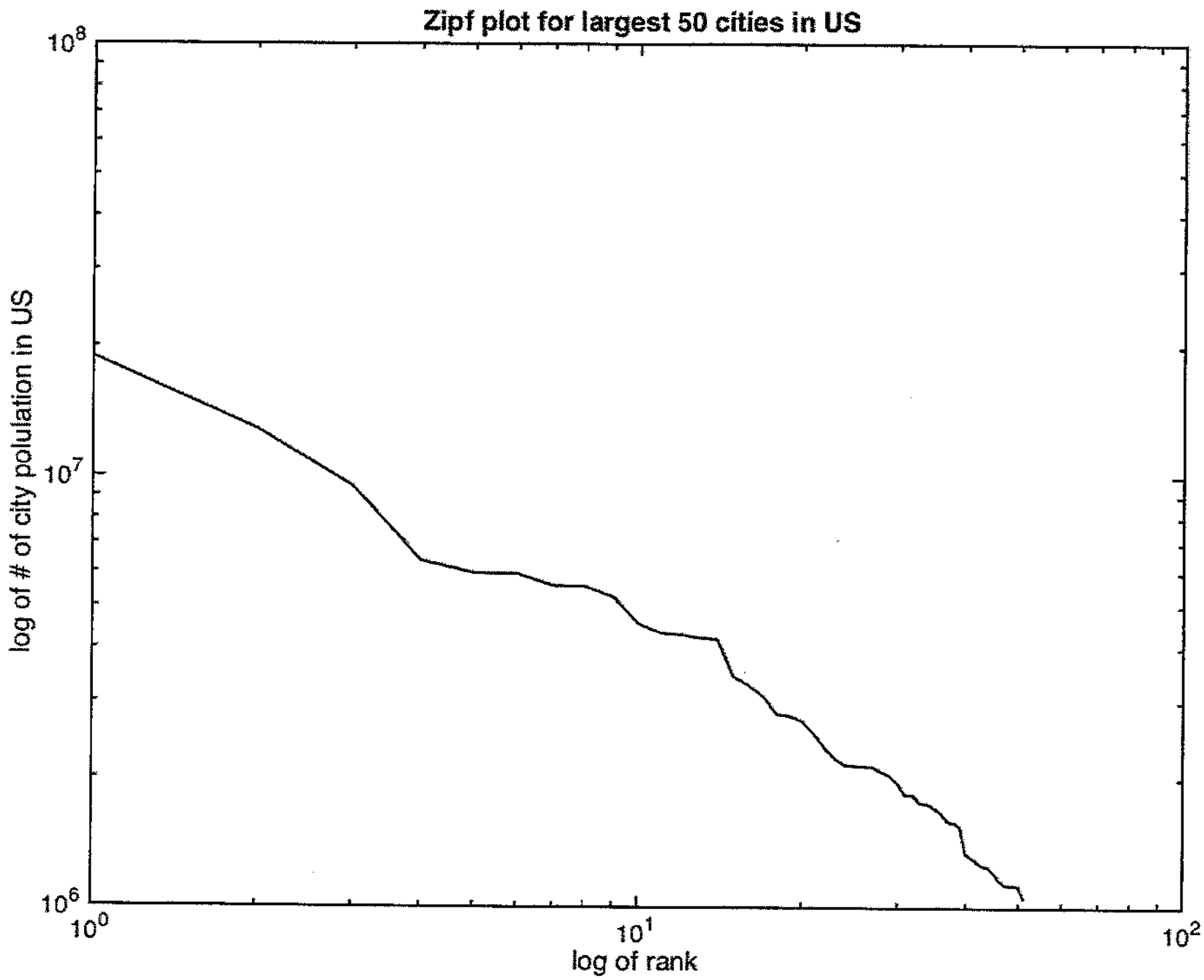
fair interpretation of "firm size"

Tianchi Chen

1. data for US cities.

10/10

(a)

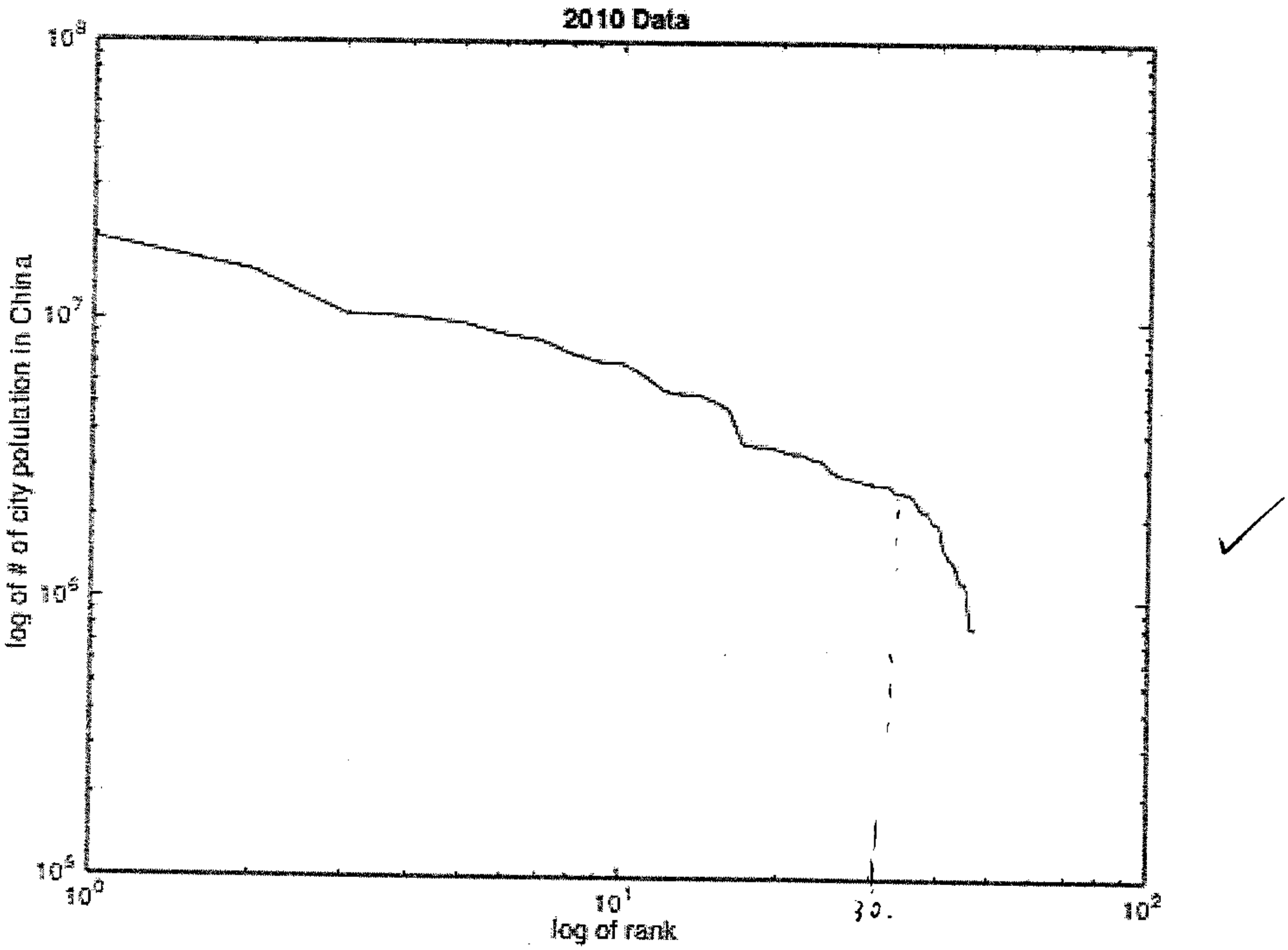


since I get the zipf plot roughly linear, then we can speculate, that:

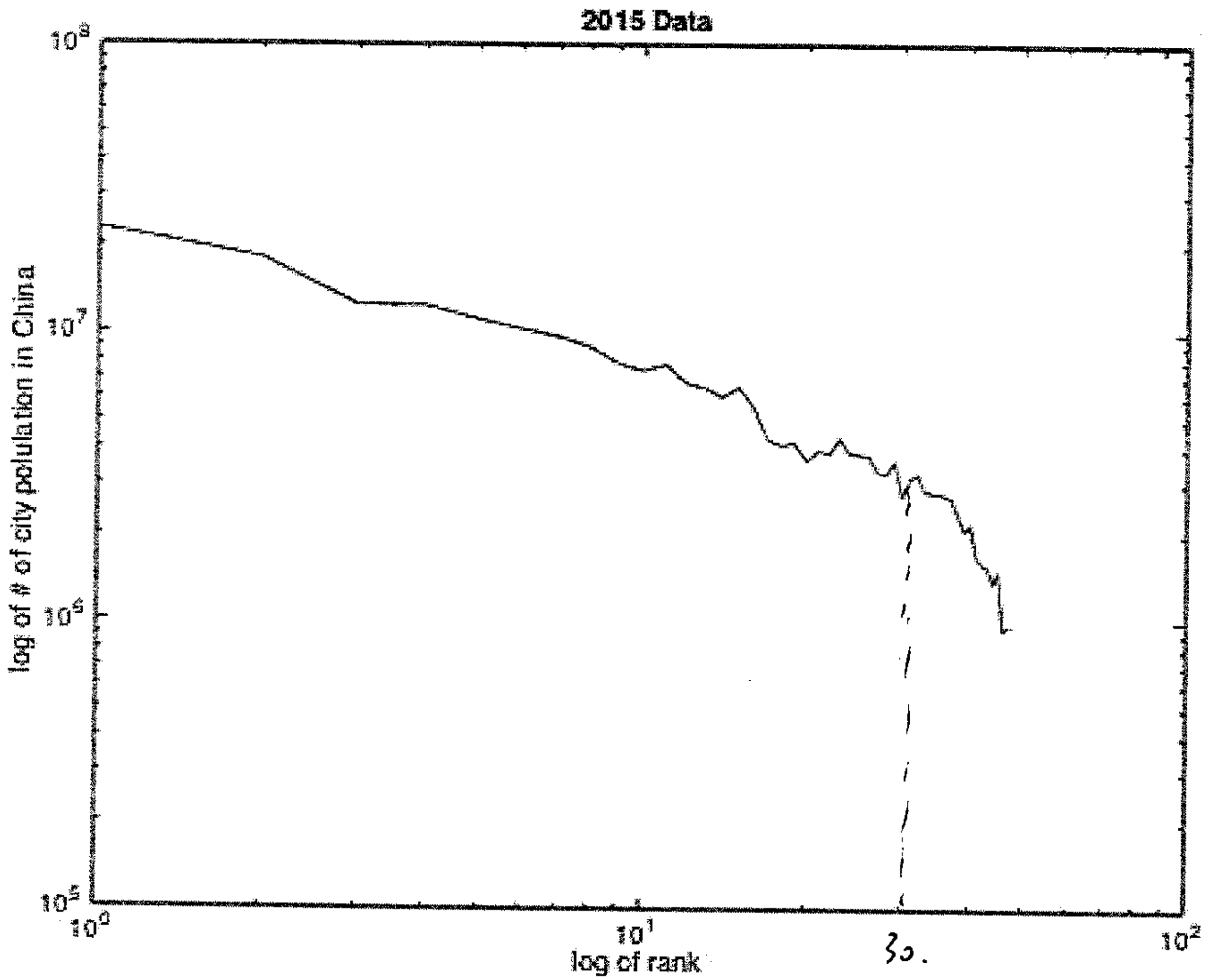
① at each time step a person is born in a city.

② All cities have approximately the same birth rate.

③ with very small probability a person creates a new city (b). data for China.



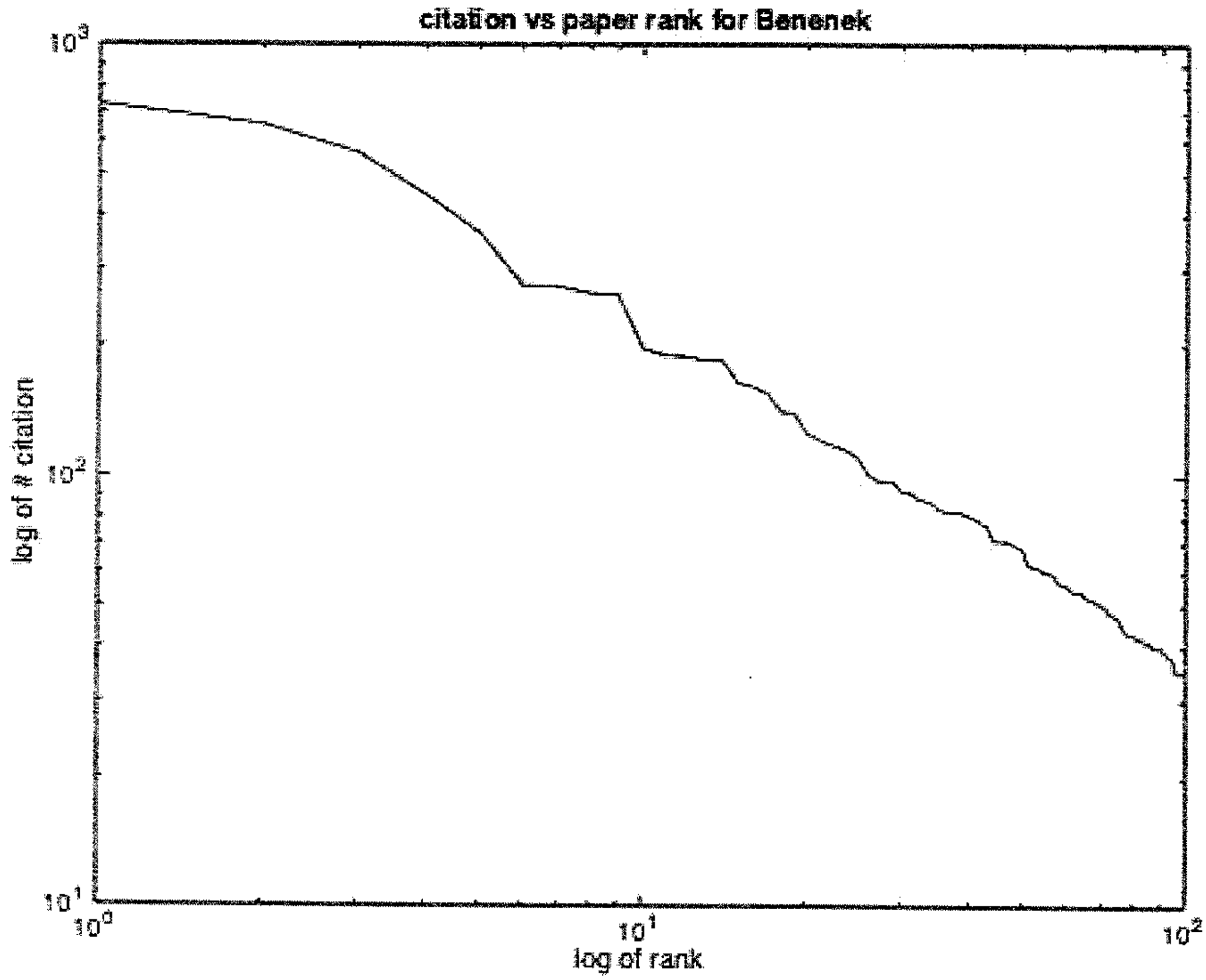
linear data for the top 30 largest cities.
after 30 is not linear probably because in China people tend to move from ~~country~~ urban area to cities. that explain why for ~~top~~ 30 cities. data is linear.



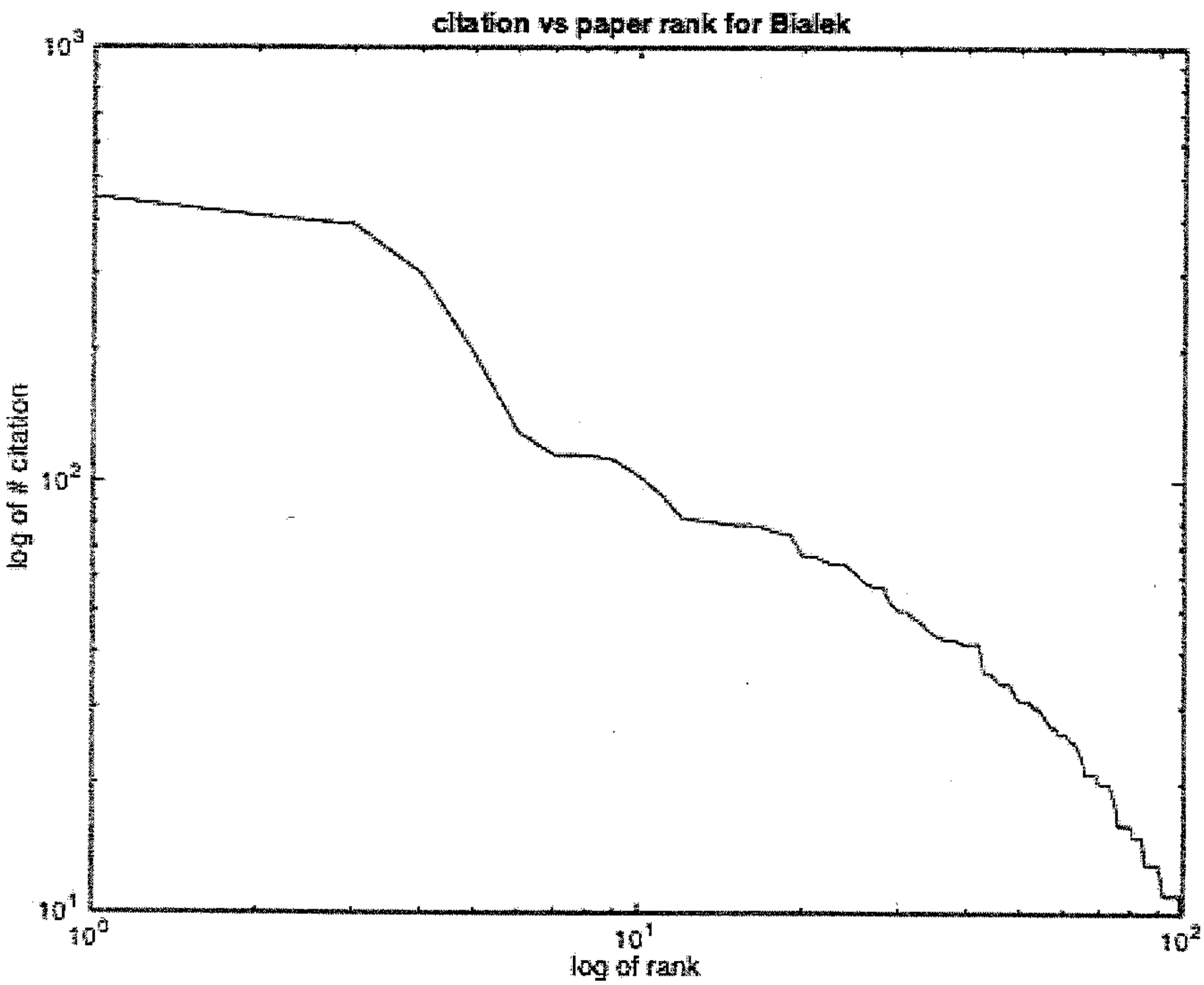
same explanation with 2010 data.

2. zipf plot of 5 different authors.

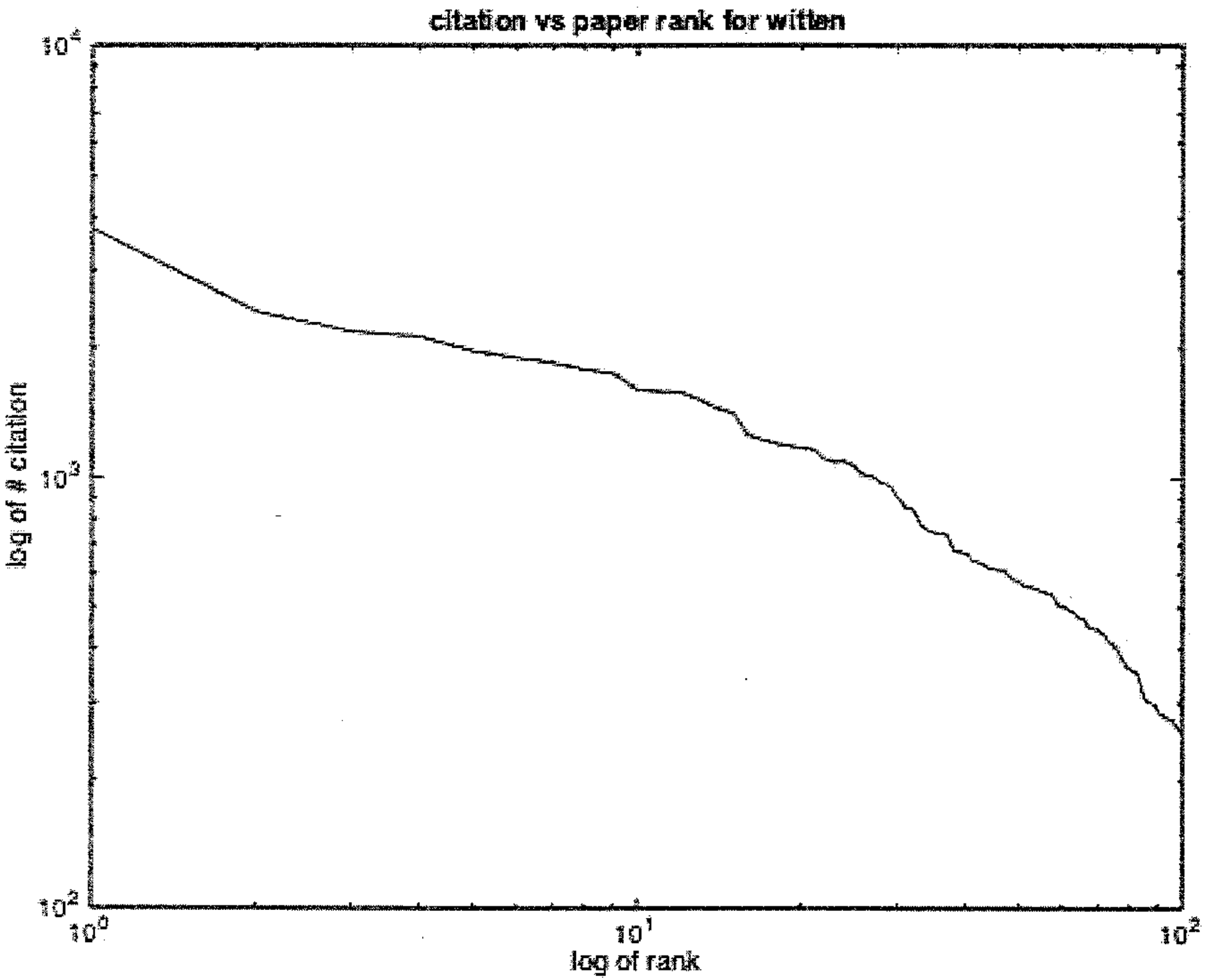
I got linear plot (almost) for all authors.

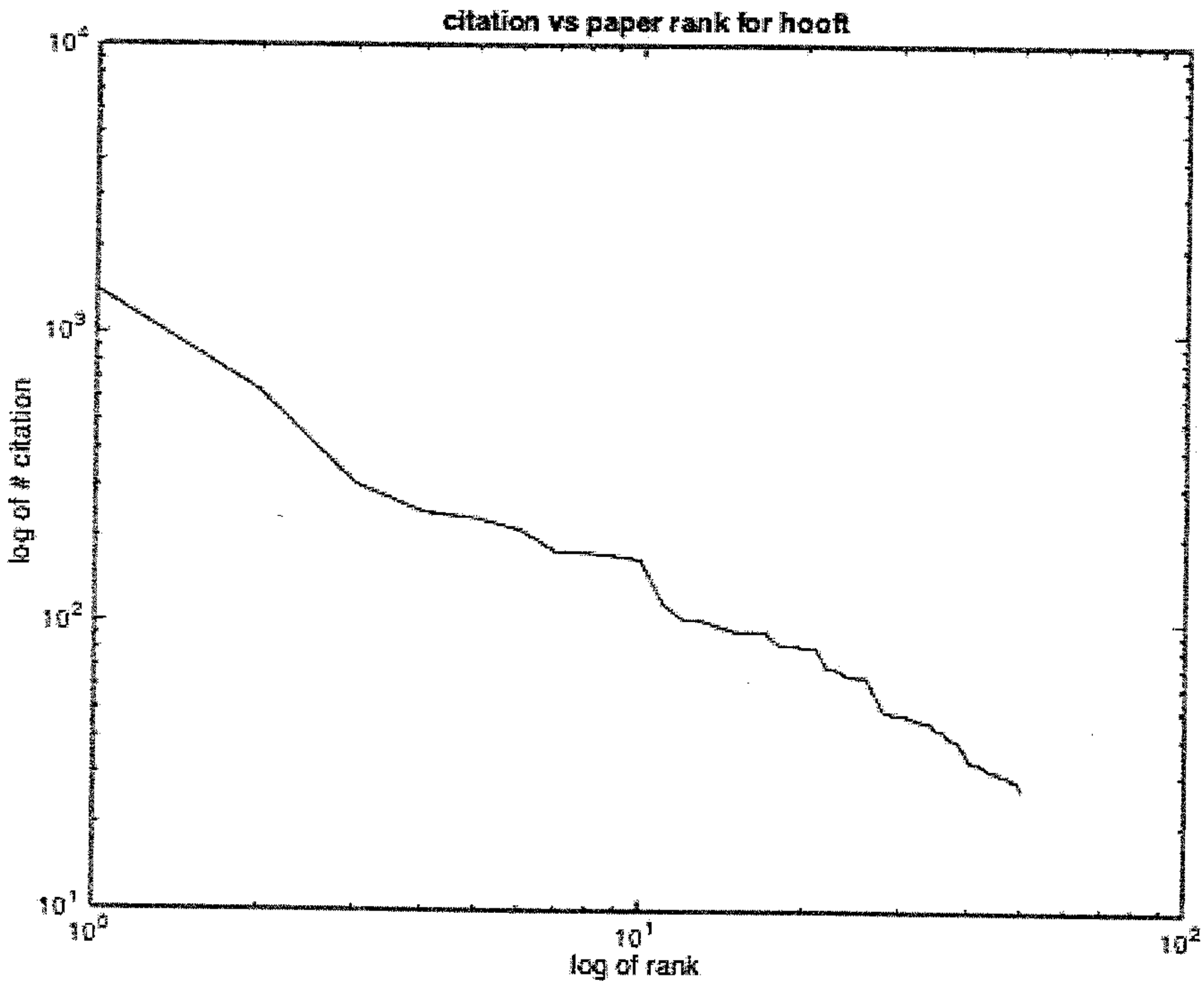


✓

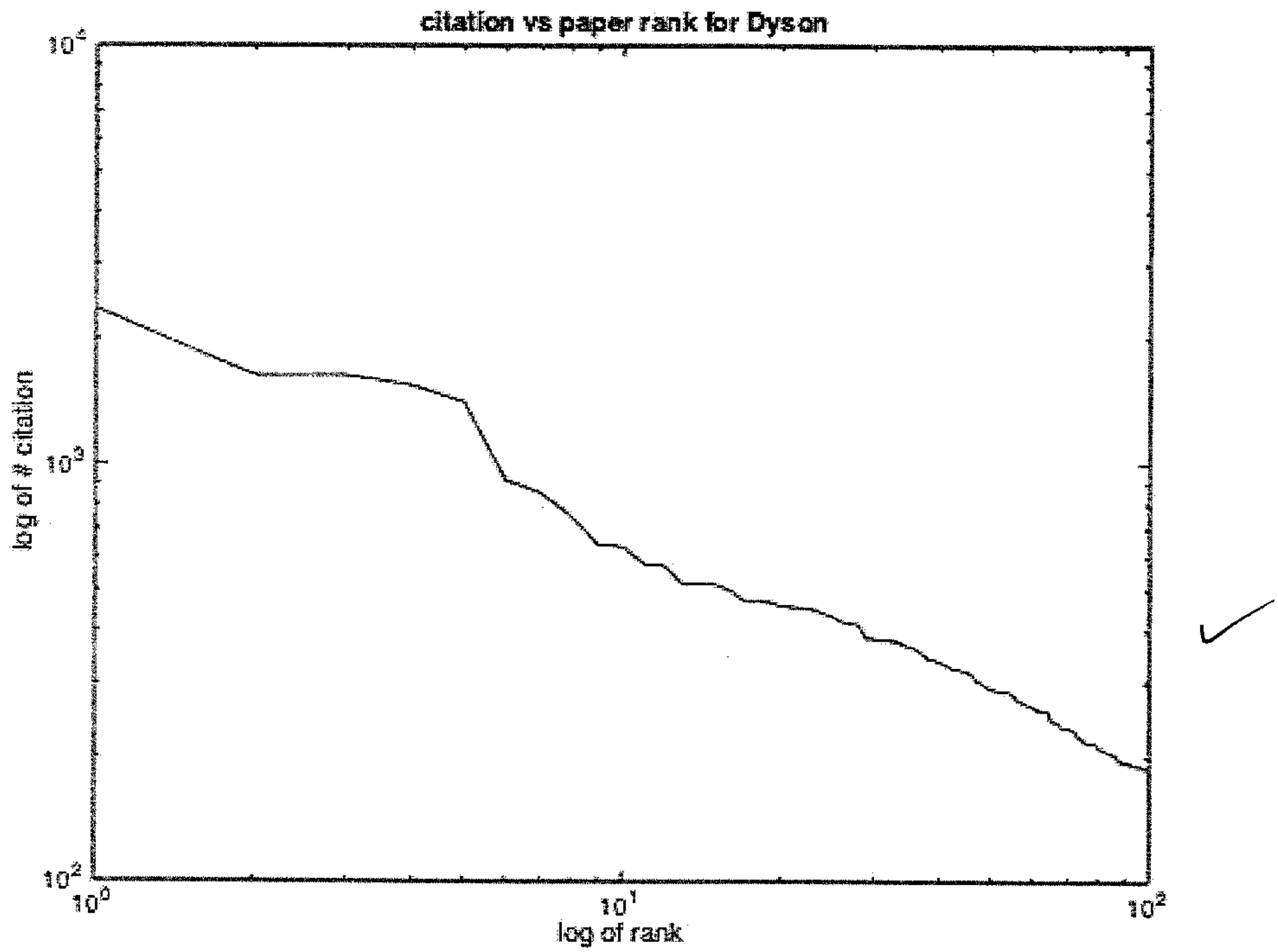


citation vs paper rank for Bialek
log of # citation
log of rank
10^3
10^2
10^1
10^0
10^1
10^2


citation vs paper rank for witten
log of # citation
log of rank
10^4
10^3
10^2
10^0
10^1
10^2



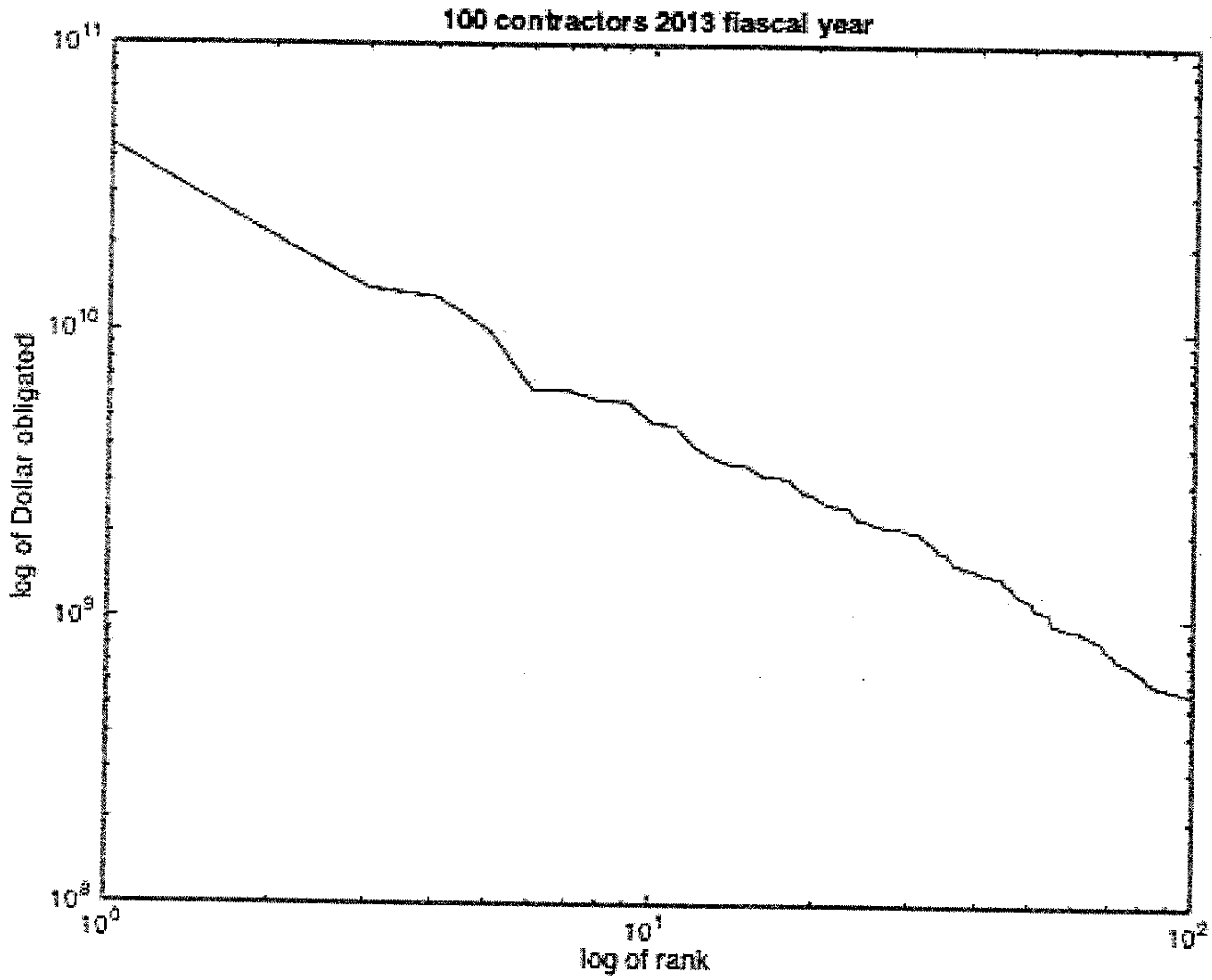
citation vs paper rank for hooft
log of # citation
log of rank


citation vs paper rank for Dyson
log of # citation
log of rank
10^4
10^3
10^2
10^0
10^1
10^2

3. I choose the firm size to be "Dollars Obligated".

it turns out that zipf plot for the top 100 contractors in 2013 fiascal year is Linear.

OK.

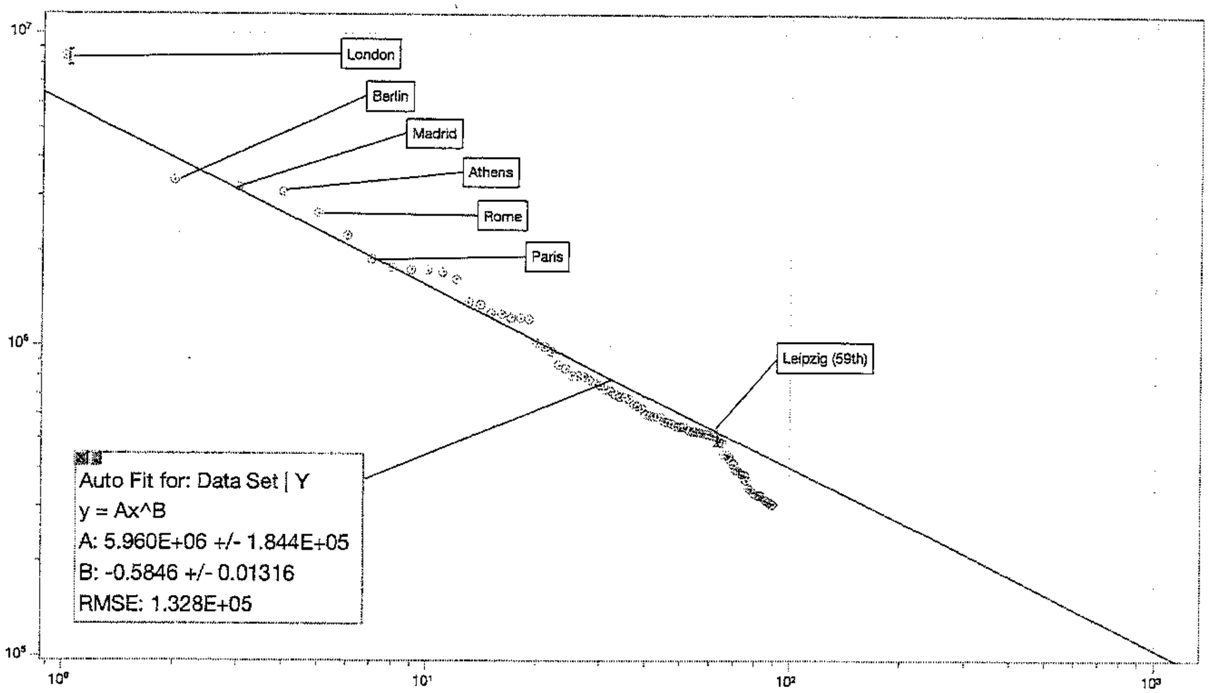


✓

Pun Chon Kit

7/10

1a) Instead of picking a country, I looked at the 90 most populated cities in the European Union. OK.



Graph 1. A log-log plot of population of the 90 most populated cities in the EU vs rank, data from year 2014 [1]

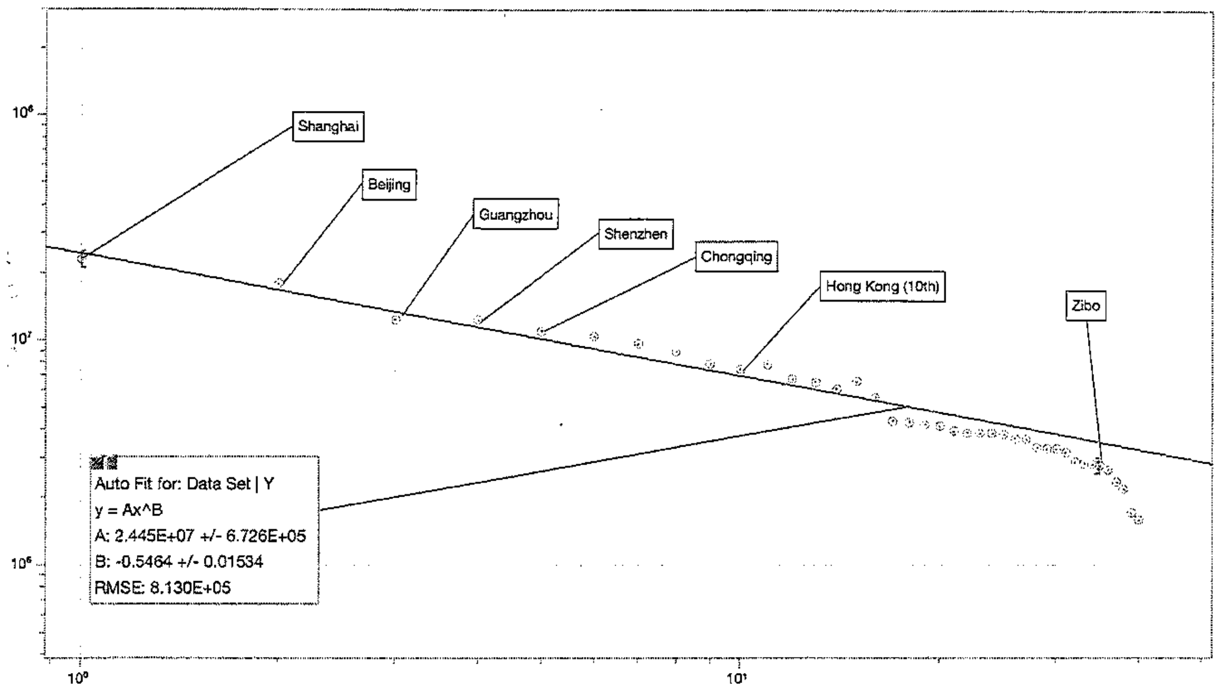
cities	Population (k)
1. London	8445
2. Berlin	3375
3. Madrid	3198
4. Athens	3089
5. Rome	2638
6. Paris	2243
7. Bucharest	1883
8. Hamburg	1770
9. Vienna	1741
10. Budapest	1735

Table 1. Top ten populated cities in the EU and their population (in thousands)

The plot has been pretty linear from Berlin (2^{nd}) to Leipzig (59^{th}). Beyond Leipzig, the data dip down and the data cease to be linear. The linear relationship shown in the data is known as Zipf law, which can be resulted from the presence of preferential attachment in the network. It makes sense since the larger the city, the more businesses and activities and thus attract more people. This results in a preferential migration of population to more populated cities. However, when population decreases to a certain point, the effect of preferential migration might not be that significant. For example below certain population limit, business activities might not vary much and thus cities below that limit will more or less attract the same number of people.

As seen from graph 1, London is obviously an outliner. This might be due to the vast distinction between the City and other EU cities in the financial aspect. London has been a very successful international city with a lot of financial activities going on (second to the New York city). This can be the reason that set London above other cities in the EU.

b)



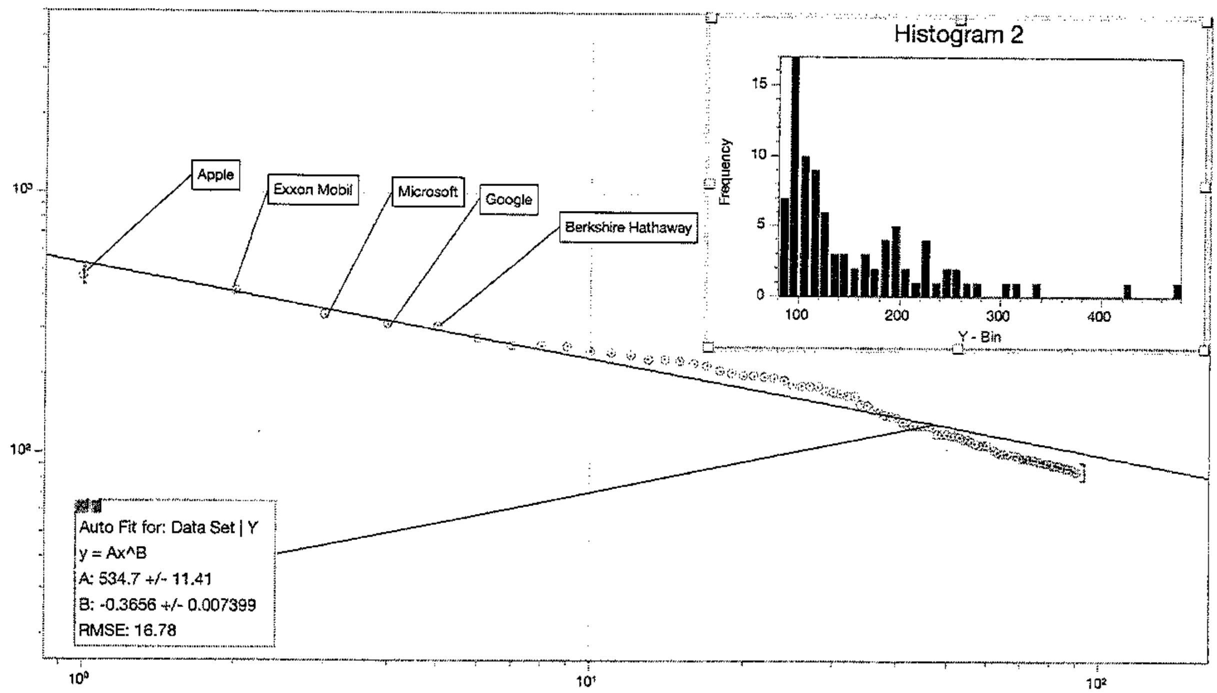
Graph 2. A log-log plot of population of the 40 most populated cities in China vs rank, data from year 2015 [2]

cities	Population (M)
1. Shanghai	23.0
2. Beijing	18.0
3. Guangzhou	12.38
4. Shenzhen	12.34
5. Chongqing	11.0
6. Wuhan	10.0
7. Tianjin	9.67
8. Foshan	8.91
9. Dongguan	7.86
10. Hong Kong	7.43

Table 1. Top ten populated cities in China and their population (in millions)

Similar to the EU, the data in China again shows Zipf law up to certain population (city of Zibo, 34th). There is difficulty in finding any data before the year 2000 online so I have not put it up here.

3.



Graph 3. A log-log plot of the largest firm by market capitalization (in billions USD) in the world vs rank. Histogram 2 showed the number of firms within ranges of market cap (bin size 10bn USD). Data from March 31st, 2014 [3]

Unlike population and citation, the market capitalization is much more volatile as stocks are traded in exchange. It is also interesting to see the same Zipf law describing the size of firm measured by market cap. I also made a histogram of number of firms within ranges of market cap. Note that the first data column is always not complete and the list did not include all the firms.

2) I cannot get to/ don't know how to use the web of science. Also I cannot get access to SCOPUS ....

Reference:

[1] TheGeographist: EU: 1000 largest cities by population
URL: https://thegeographist.wordpress.com/2014/04/12/eu-1000-largest-cities-by-population/

[2] Nationsonline: The most populated cities in China
URL: http://www.nationsonline.org/oneworld/china_cities.htm
* the data source quoted in this website is from the United Nations Department of Economic and Social Affairs. It seems that that data quoted from the UN website is from the year 2010, rather than the year 2015 data I used here. I have been quite sloppy in verifying source here....

[3] Statista: The 100 largest companies in the world by market value in 2014
URL: http://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/

9/10

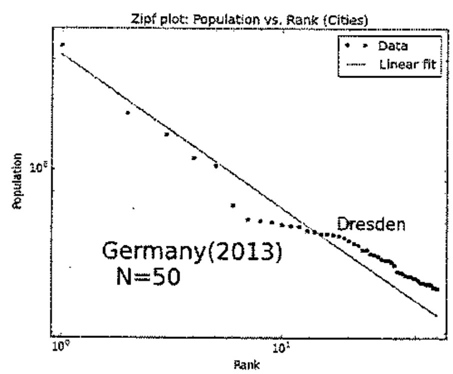
Econophysics
Problem Set 1

Sebastian Gemsheim

February 12, 2015

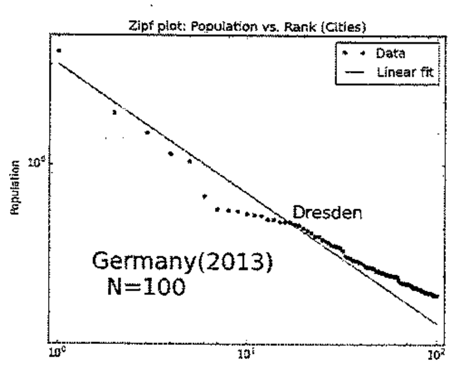
Problem 1

a) I took the data for Germany, my home country, from "Statistisches Bundesamt" (`http://www.destatis.de`) and their main database GENESIS (`https://www-genesis.destatis.de/genesis/online`). The data point of my home university (exchange student) is written explicitly.



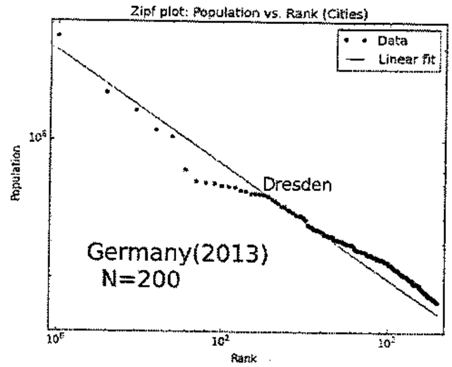
Parameter of fit on $\log(y) = a\log(x) + b$:

$$a = -0.658$$
$$b = 14.953$$
$$\sigma = \begin{pmatrix} 2.86 & -3.46 \\ -3.46 & 8.20 \end{pmatrix} \times 10^{-4}$$



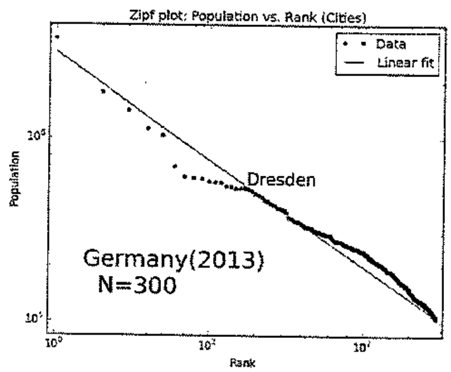
Parameter of fit on $\log(y) = a\log(x) + b$:

$$a = -0.613$$
$$b = 14.914$$
$$\sigma = \begin{pmatrix} 1.27 & -1.98 \\ -1.98 & 5.59 \end{pmatrix} \times 10^{-4}$$



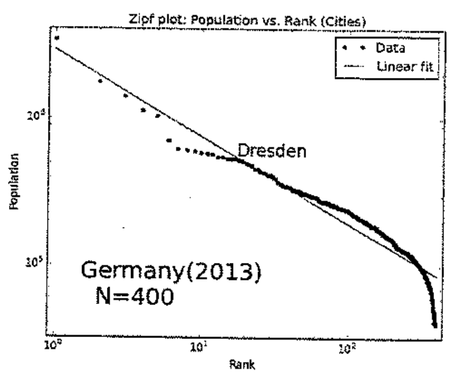
Parameter of fit on $\log(y) = a\log(x) + b$:

$$a = -0.585$$
$$b = 14.883$$
$$\sigma = \begin{pmatrix} 0.50 & -0.97 \\ -0.97 & 3.17 \end{pmatrix} \times 10^{-4}$$



Parameter of fit on $\log(y) = a\log(x) + b$:

$$a = -0.581$$
$$b = 14.878$$
$$\sigma = \begin{pmatrix} 0.28 & -0.58 \\ -0.58 & 2.06 \end{pmatrix} \times 10^{-4}$$

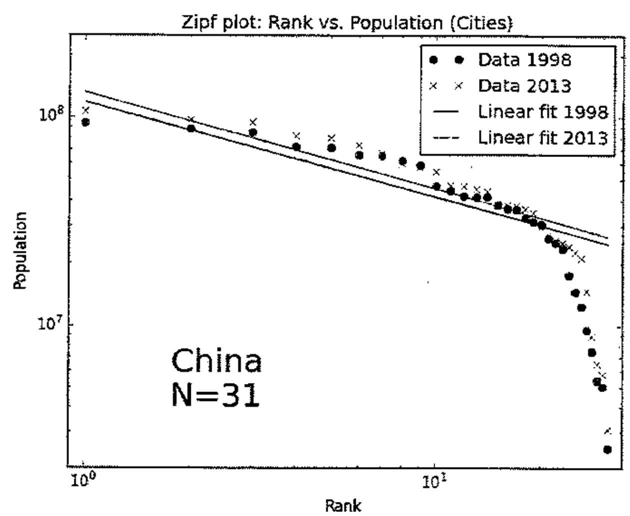


Parameter of fit on $\log(y) = a\log(x) + b$:

$$a = -0.588$$
$$b = 14.887$$
$$\sigma = \begin{pmatrix} 0.20 & -0.44 \\ -0.44 & 1.59 \end{pmatrix} \times 10^{-4}$$

The last plot for $N = 400$ shows that the plot ceases to be linear at rank ≈ 350. The range of validity of Zipf's law one order of magnitude higher than that of China (rank ≈ 30, see part b). I would guess, without proof, that this has historical reasons. The present-day area of Germany had many cultural and economical centers in the 18th and 19th century and was split to many territories, so that some of them grew independently. Therefore these are still major cities in Germany nowadays and the number of attracting regions might be higher than in other countries, for example China. Since Zipf's law requires a constant growth rate and is for the largest cities only valid up to a certain point, it seems that the growth rate changes for smaller cities/regions. Larger cities might be more attractive than smaller cities/rural regions and consequently those smaller cities loose citizens to larger cities which results in a different growth for both. One could guess, that after a certain point, in this case rank ≈ 350, the following range can be approximated by a linear fit as well but with a different set of parameters.

b) China for the years 1995 and 2013 (data taken from "National Bureau of Statistics of the People's Republic of China": `http://www.stats.gov.cn`)



The plot shows an overall growth in almost all major cities, but has a higher increase in the lower ranked cities/regions. This could be due to the introduction of the One-Child-Policy. On one hand it is more strictly enforced in densely populated urban areas and, on the other hand, it allowes couples in rural regions to have two children. Maybe this can explain the difference in the grwoth in regions with different populations.

OK.

Problem 2

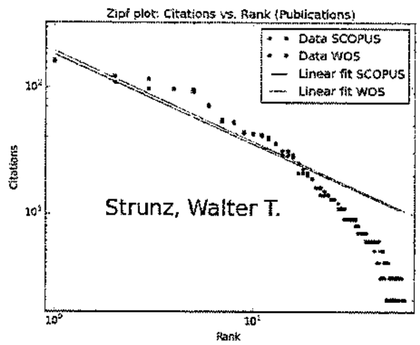


Figure 1: Prof. W. T. Strunz, who was supervisor of my Bachelor thesis.

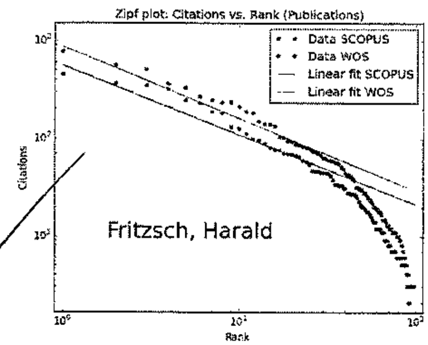


Figure 2: Fritzsch, Harald, who grew up in the same small German town and went to the same public high school as I did.

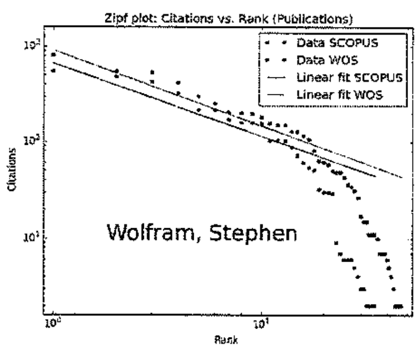


Figure 3: Wolfram, Stephen.

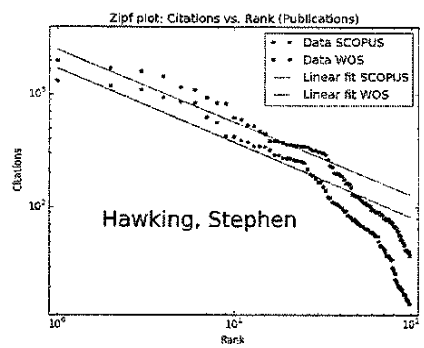


Figure 4: Hawking, Stephen.

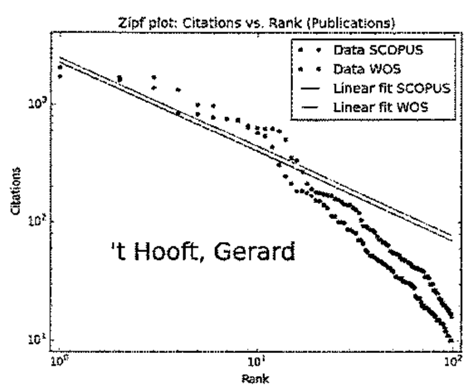


Figure 5: 't Hooft, Gerard.

I would guess that the rate of increase in the number of citation is approximately constant because the more citations a paper has the more scientists will be exposed to its title/reference. Consequently, assuming the fraction of scientists who know the publication and are willing to cite it is constant, the growth occurs with a constant factor. Again, this is only valid up to a certain point, which may depend on the scientific field of research.

Problem 3

Source of data: "2007 County Business Patterns and 2007 Economic Census" from `http://www.census.gov/econ/susb/data/susb2007.html`

Enterprise employment size	Number of firms
0-4	3705275
5-9	1060250
10-14	425914
15-19	218928
20-24	134254
25-29	89643
30-34	64753
35-39	47641
40-44	38221
45-49	29705
50-74	86364
75-99	41810
100-149	39316
150-199	18620
200-299	17780
300-399	8155
400-499	4715
500-749	6094
750-999	2970
1,000-1,499	2916
1,500-1,999	1542
2,000-2,499	942
2,500-4,999	1920
5,000-9,999	952
>10,000	975

Table 1: Combined data of 2007 U.S. Economic Census

Since the data is given in ranges of form size it is reasonable to represent the data in a histogram:

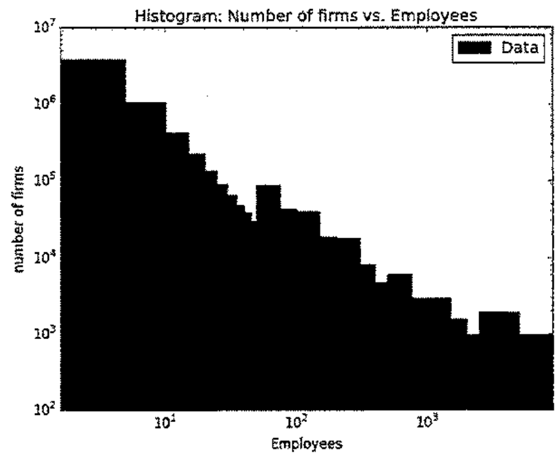


Figure 6: Histogram of the numbers of firms in dependency of the number of employees.

The bin sizes are not equidistant and the data points have to be normalized by their associated bin size in order to make it possible to apply Zipf's law. The size of the last bin is set to infinite because there is no information about the maximal value, respectively, the largest company. Therefore, the last data point is neglected in the following plot.

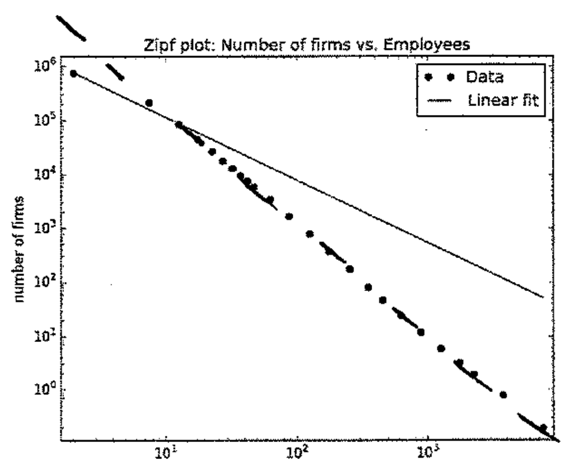


Figure 7: Plot of normalized data points.

It is obvious from the last figure that it would be better to treat companies without any employees as a special case since they seem not to apply to the normal behavior. Unfortunately, there is no distinction between firms with zero and a few employees. Consequently the first data point is also neglected and the remaining data follows Zipf's law as a good approximation.

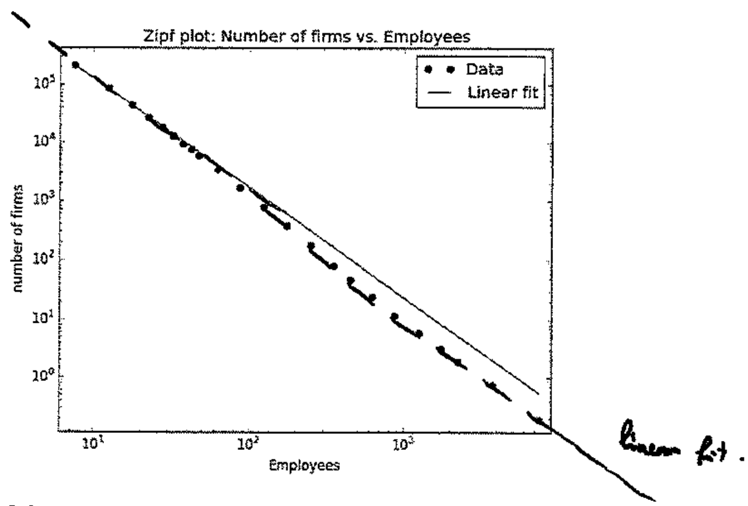


Figure 8: Zipf plot of data where the first and last data point are neglected since they have special behavior.

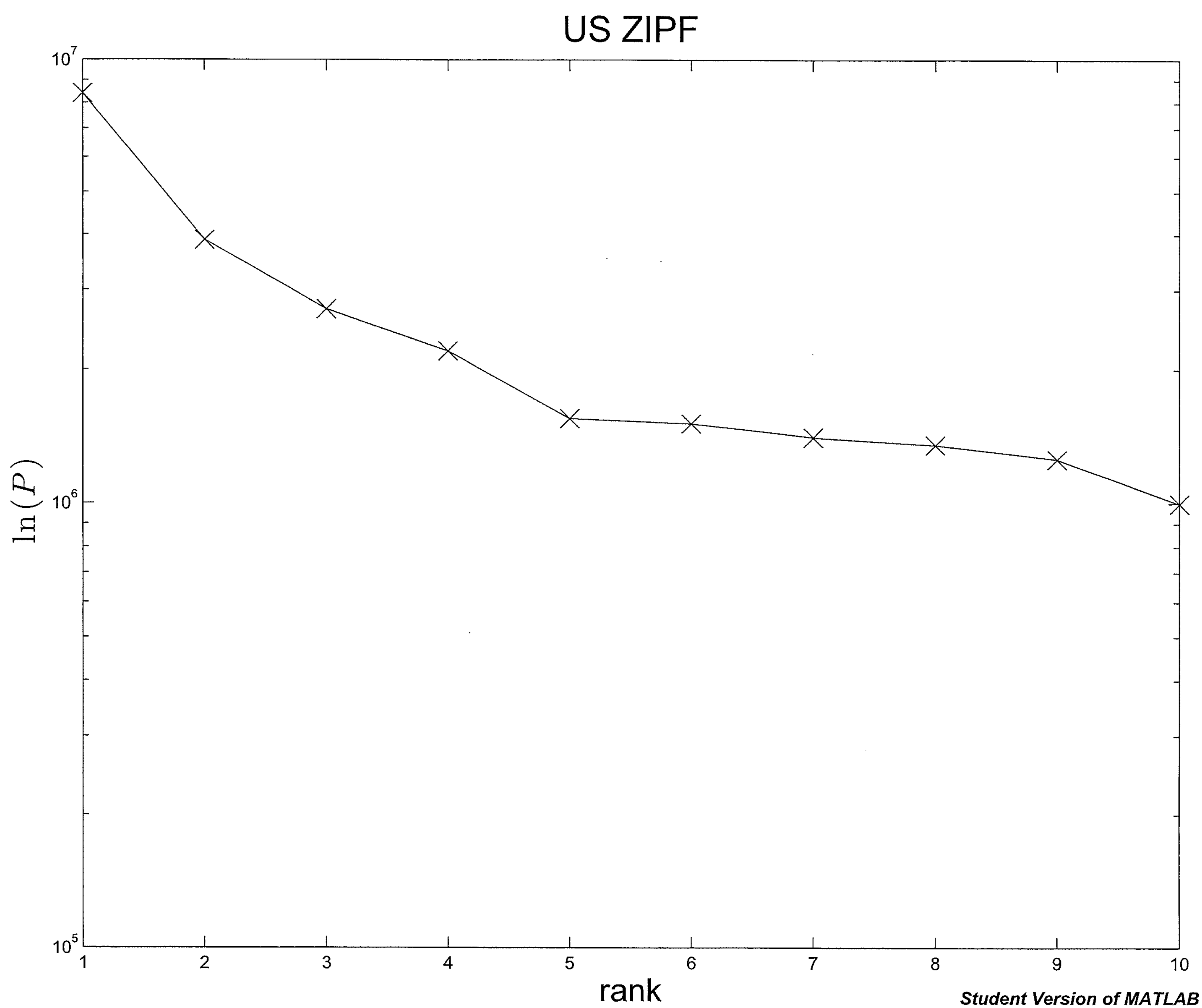
10/10

Jan Makkinje

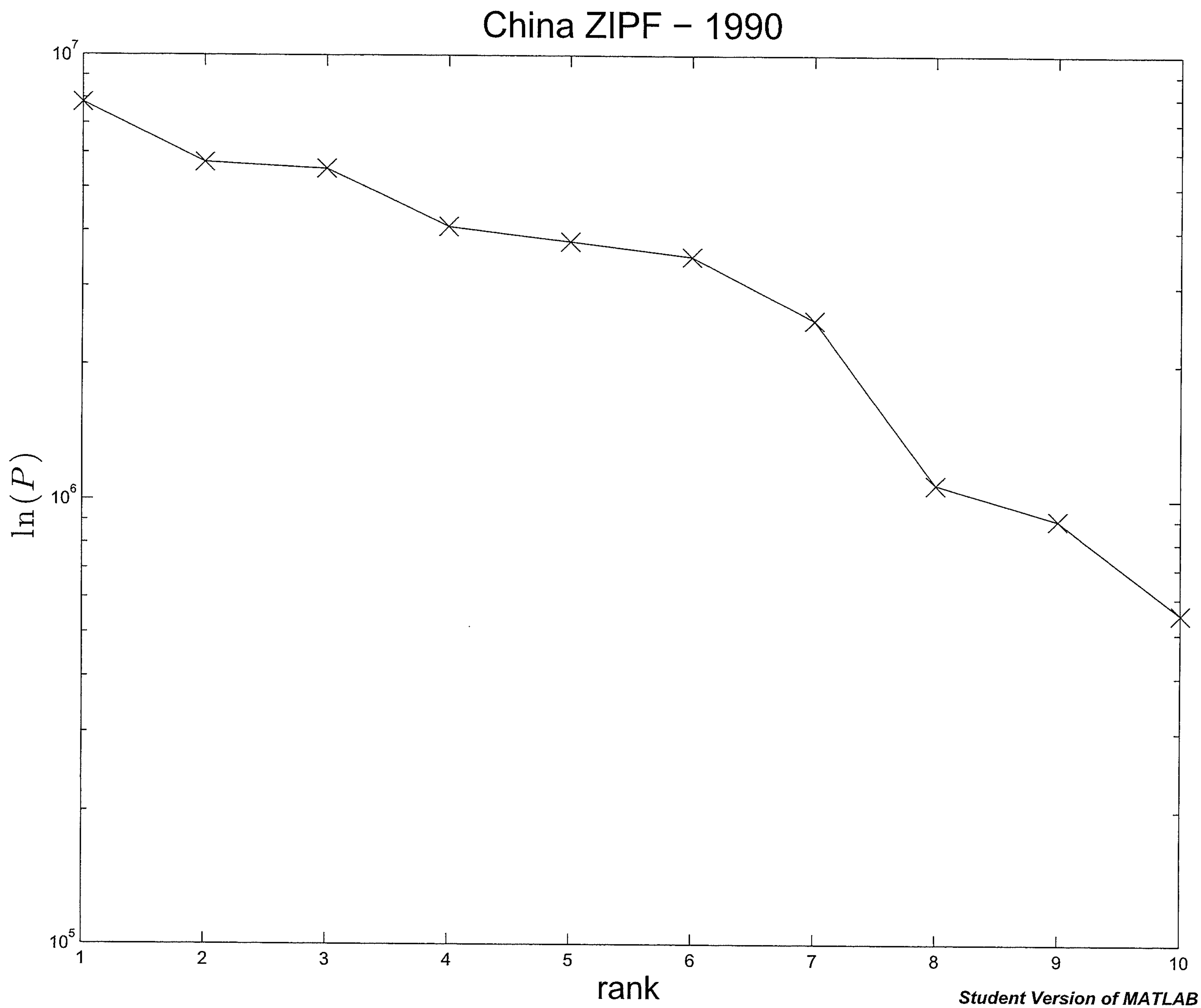
1). After leaving major cities the rural areas surrounding cities have the same lower population.

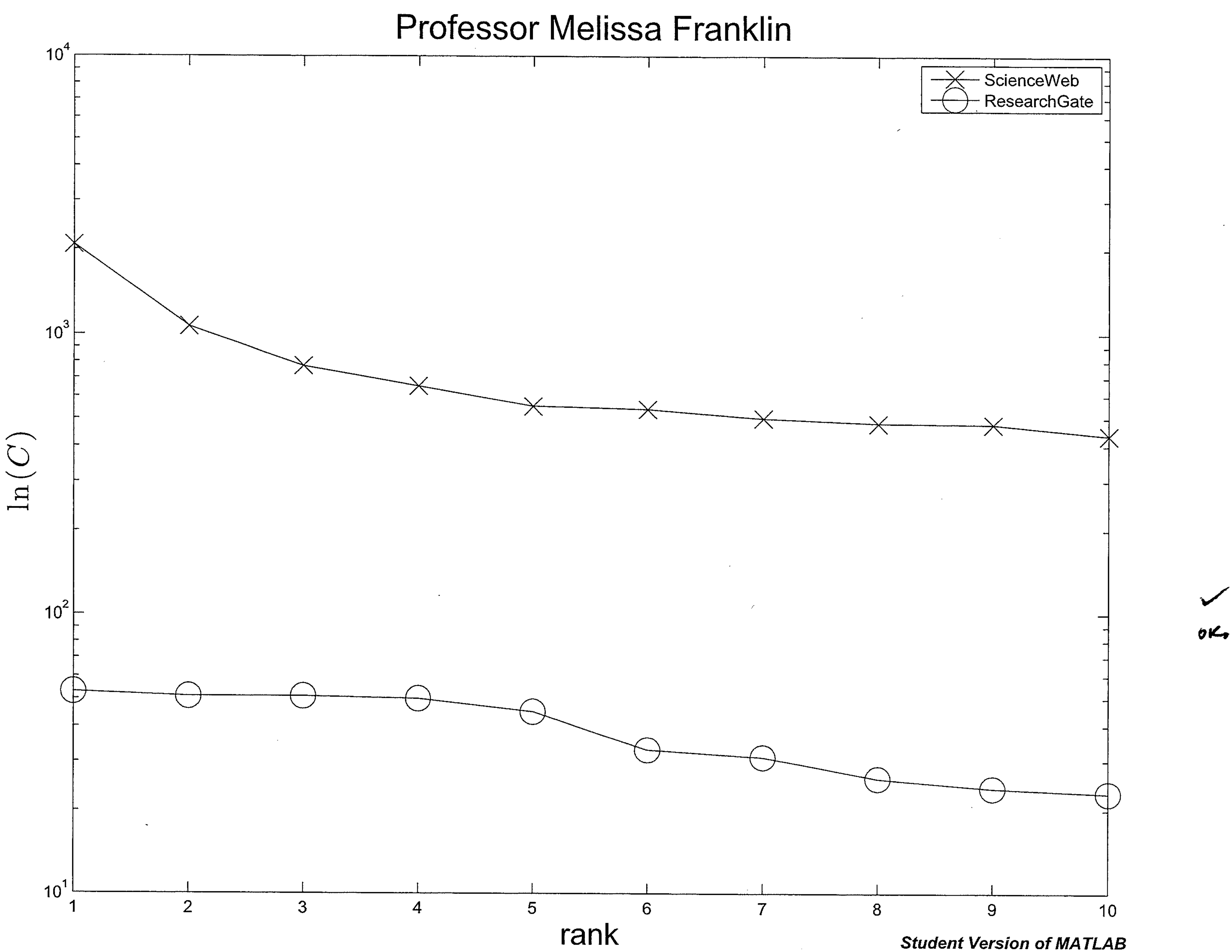
2). Approximately linear because of growth of career would be exponential. Especially when Professor is working for tenure.

JSKANDAR@BU.EDU.

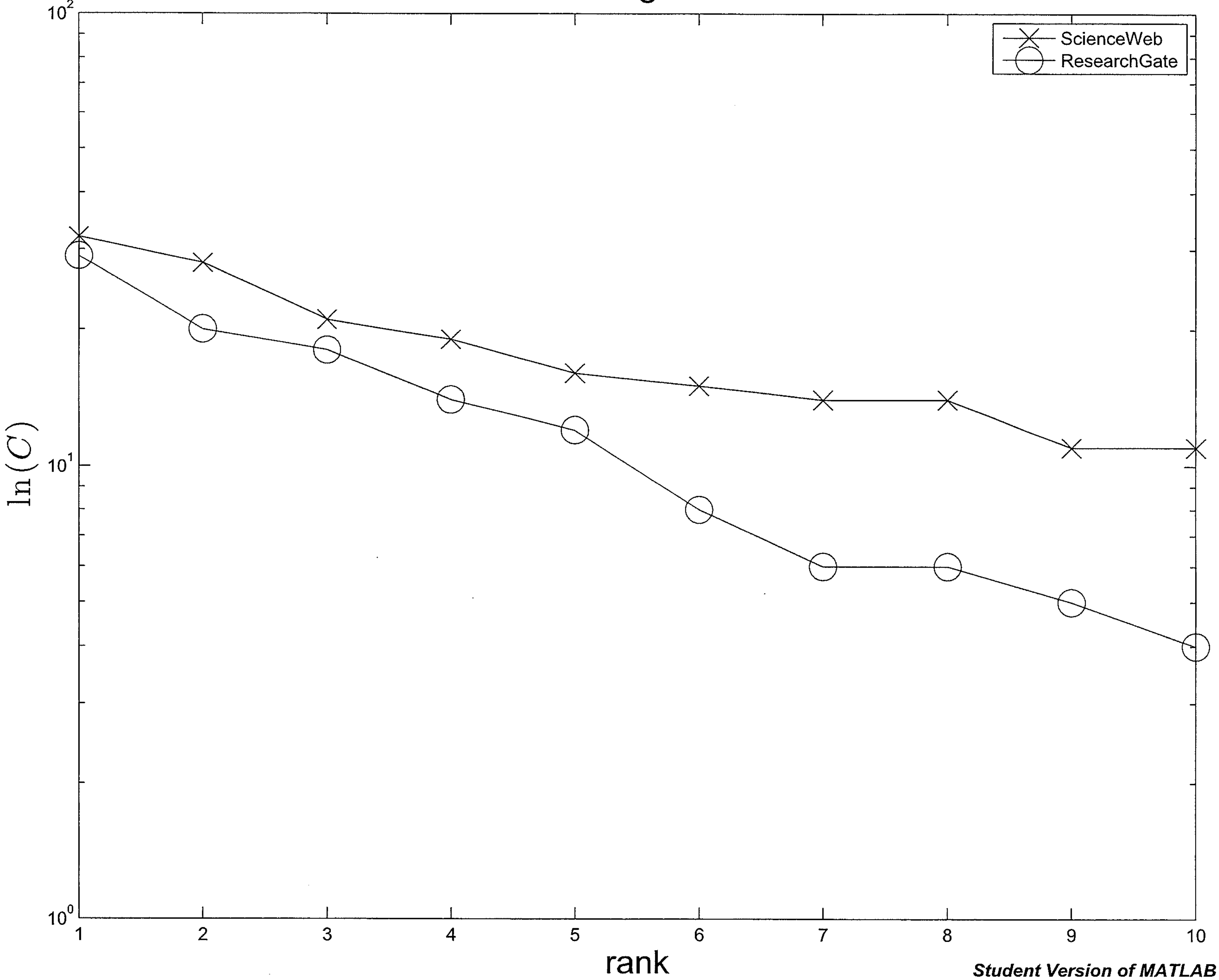

US ZIPF
10^7
10^6
10^5
$\ln(P)$
1
2
3
4
5
6
7
8
9
10
rank
Student Version of MATLAB

ok.

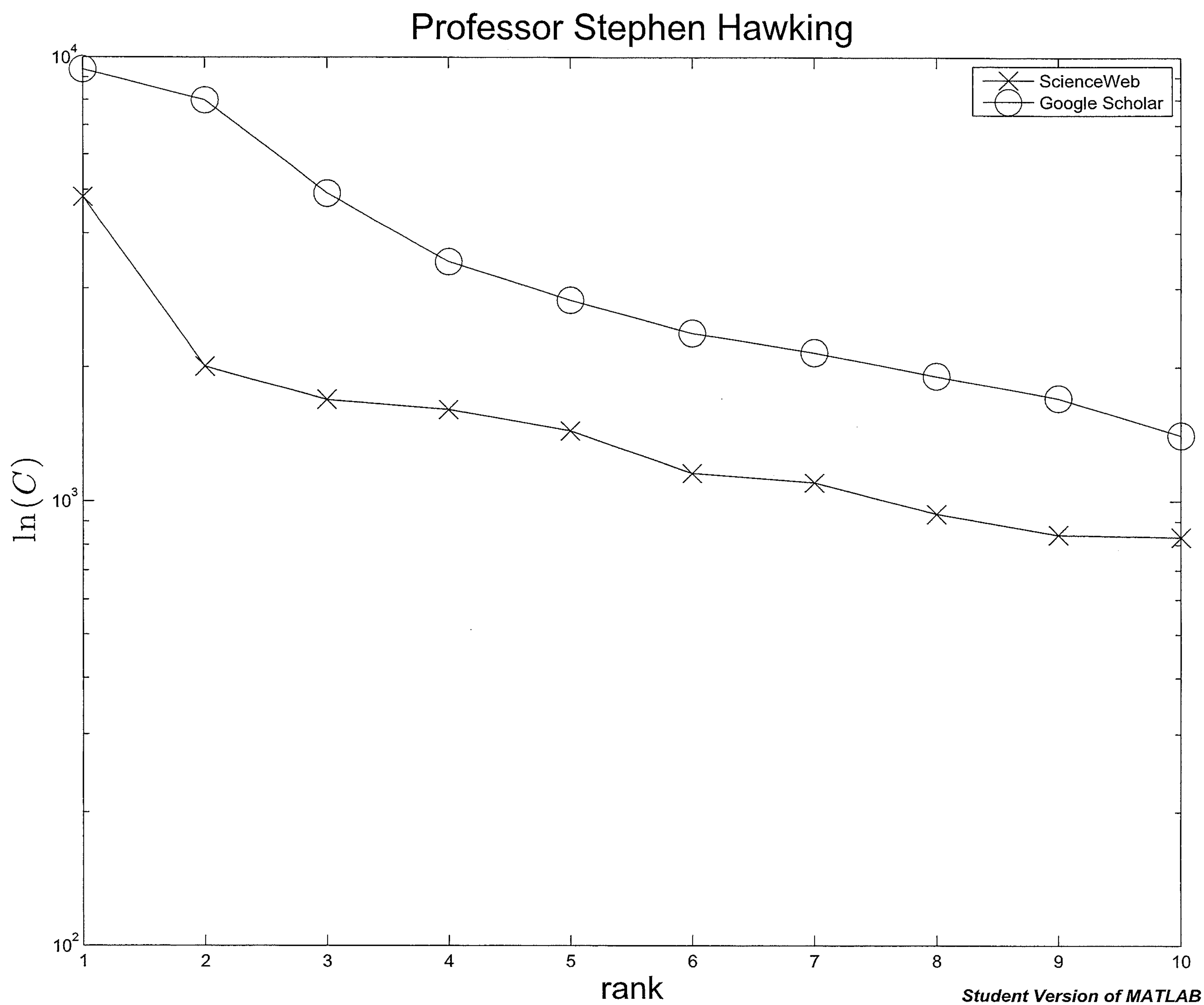

China ZIPF − 1990
10^7
10^6
10^5
ln(P)
1
2
3
4
5
6
7
8
9
10
rank
Student Version of MATLAB


Professor Melissa Franklin
ScienceWeb
ResearchGate
10^4
10^3
10^2
10^1
$\ln(C)$
1
2
3
4
5
6
7
8
9
10
rank
Student Version of MATLAB
OK

Professor George Zimmerman



ok.


Professor Stephen Hawking
ScienceWeb
Google Scholar
10^4
10^3
10^2
$\ln(C)$
1
2
3
4
5
6
7
8
9
10
rank
Student Version of MATLAB

10/10

Great job.

BOSTON UNIVERSITY

Division of Materials Science and Engineering

Econophysics

Homework 1

Alexandros Kyrtsos

Introduction

Zipf's Law for cities is an empirical law which states that the logarithm of the population of a city is proportional to the rank of the city. However, Zipf's Law can be generalized in other cases where a quantity is measured versus rank. For the Zipf's Law to apply, it is essential for all the cities in question to follow some proportional growth process.

Zipf's Law for cities

In order to validate Zipf's Law, we use Greece's 25 largest cities versus their rank as shown in figure 1. There is no linear relationship in these data except for a limited region of medium size cities. This is not surprising for Greece because it violates the most important condition of Zipf's Law which is the proportional growth of the cities in question. During the last decades, Greece experienced a huge internal immigration towards the largest cities. Interesting...

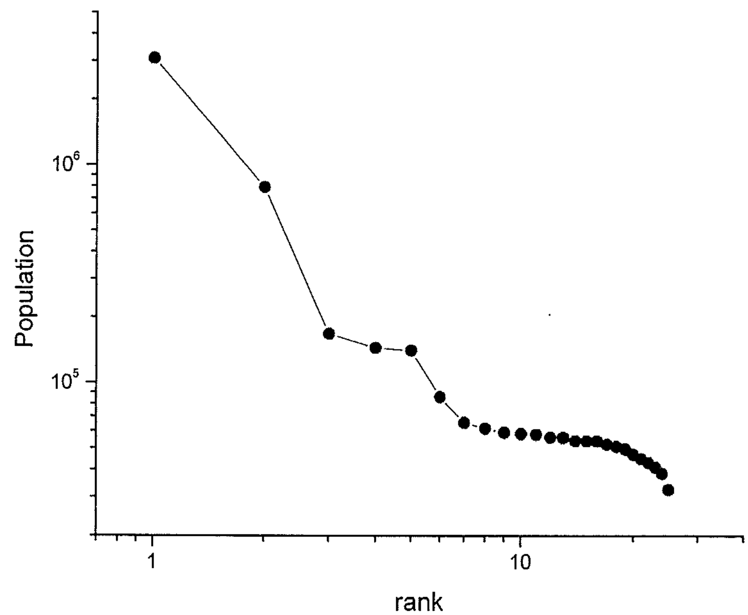


Figure 1: Zipf's plot for Greece's 25 largest cities [1]

Another sample could be the municipalities of Greece. Greece consists of 325 municipalities. Figure 2 shows all the municipalities. It can be seen that the Zipf's Law applies for the first 200 municipalities. After that the linearity no longer applies.

Finally, another sample is the cities of China. Figure 3 shows China's 500 largest cities. As it can be seen, this plot can be divided in two sub-regions which exhibit

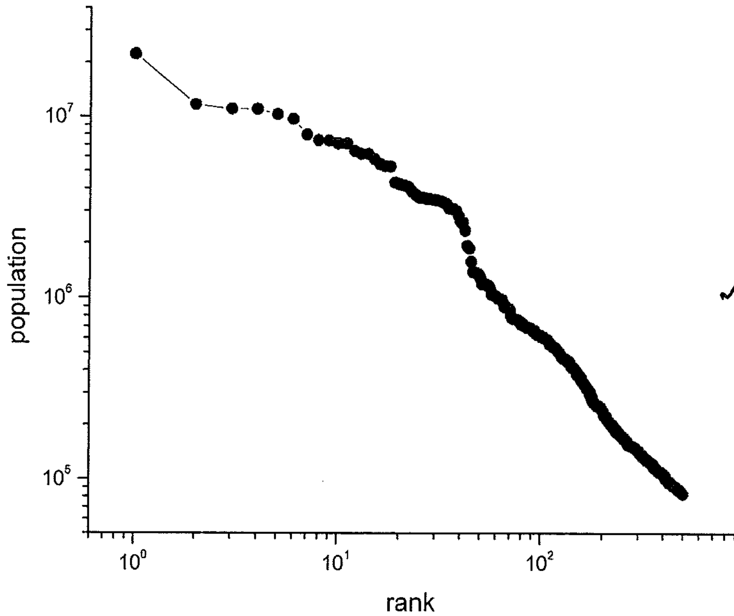


Figure 3: Zipf's plot for China's 500 largest cities [2]

Zipf's Law in finance

As an example of the Zipf's Law in finance we choose the capitalization of the top 100 stocks of the Greek Stock Exchange market. Except for the first 4, all the rest exhibit very good linearity, validating Zipf's Law.

The first 4 stocks consist of two multinational companies and two recently merged banks. Hence, their capitalization is not expected to be proportionally greater than the rest. However, for all the rest good linearity is achieved.

Market cap. is a fair interpretation of "firm size"

References

[1] http://www.statistics.gr/portal/page/portal/ESYE/PAGE-census2011.

[2] http://worldpopulationreview.com/countries/china-population/major-cities-in-china/.

[3] https://apps.webofknowledge.com.

[4] https://www.capital.gr.

[5] Y. Xinyue and X. Yichun, "Re-examination of Zipf's law and urban dynamic in China: a regional approach," *The Annals of Regional Science*, vol. 49, no. 1, pp. 135-156, 2012.

[6] X. Gabaix, "Zipf's Law for Cities: An Explanation," *The Quarterly Journal of Economics*, vol. 114, no. 3, pp. 739-767, 1999.

haichen

haichen

haichen

Haichen Zhan
U55327963.

File name: Microsoft Word - 1-12.doc

File size: 303395 chars

Job format: postscript

Submitted: 12:12:49 PM 02/11/2015

From system: IST-VDI-MUG-084.ad.bu.edu

On printer: mugar-ds-staple -> hspr1afp

Job ID: 2186047516

Sides: 2 (duplex)

Printing via MyPrint. See http://www.bu.edu/tech/help/myprint/ for info.

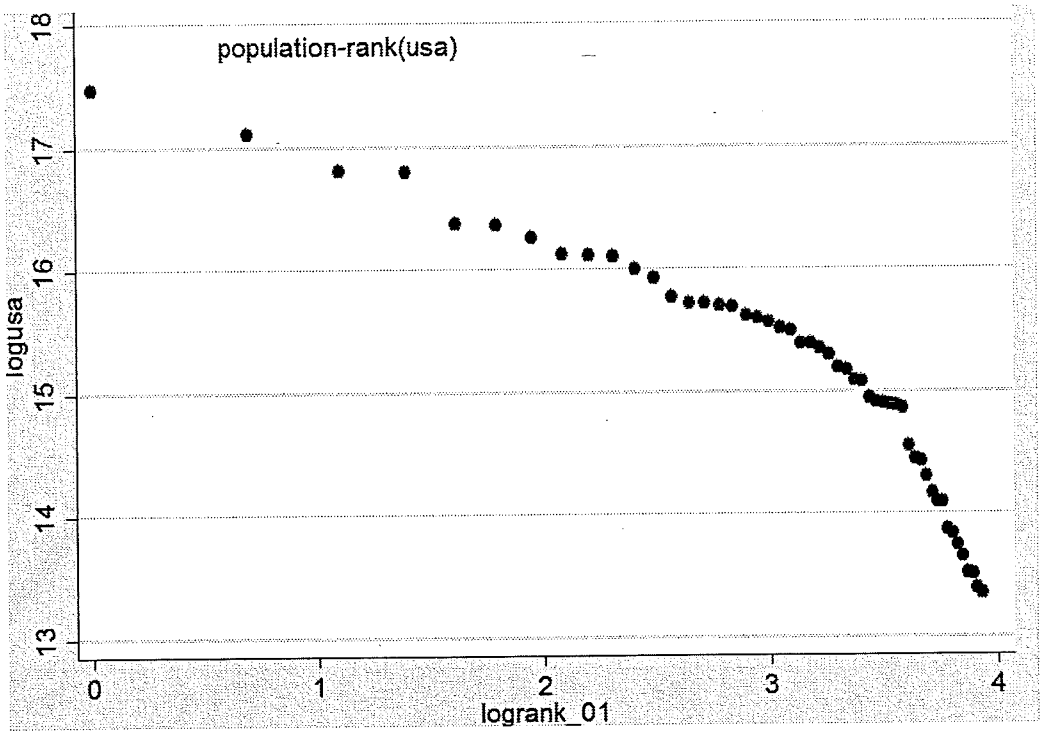
~~BNY MELLON online~~

BOSTON UNIVERSITY

Information Services & Technology
BU Common @ Mugar
771 Commonwealth Avenue
www.bu.edu/tech/help/myprint/

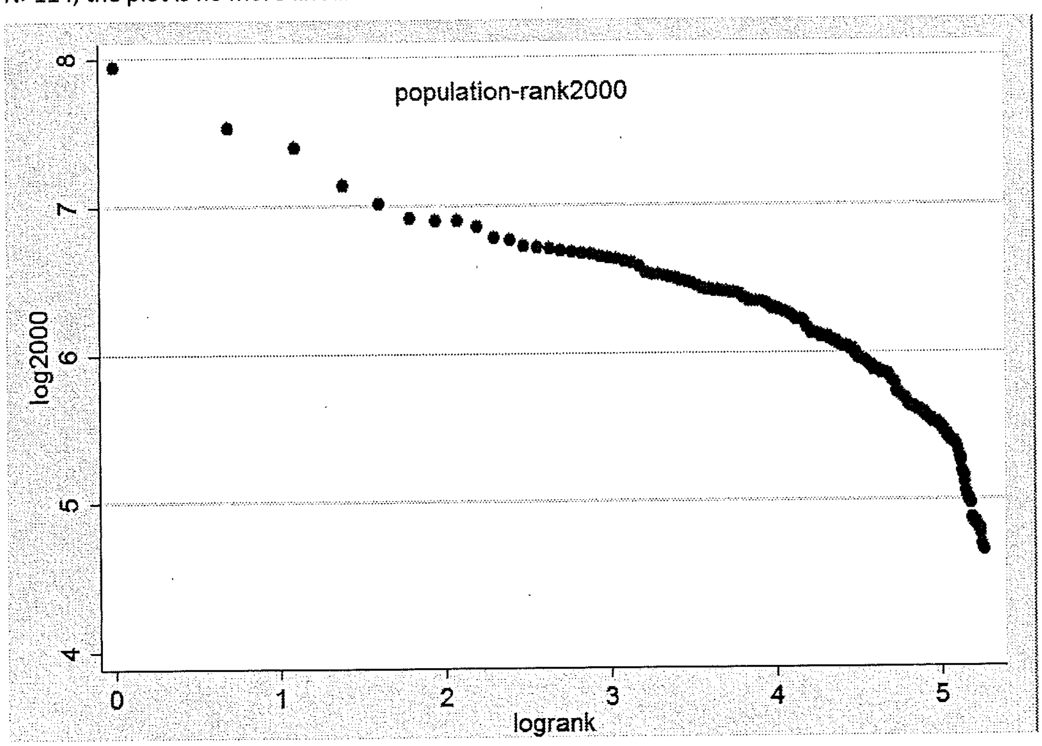
1. a

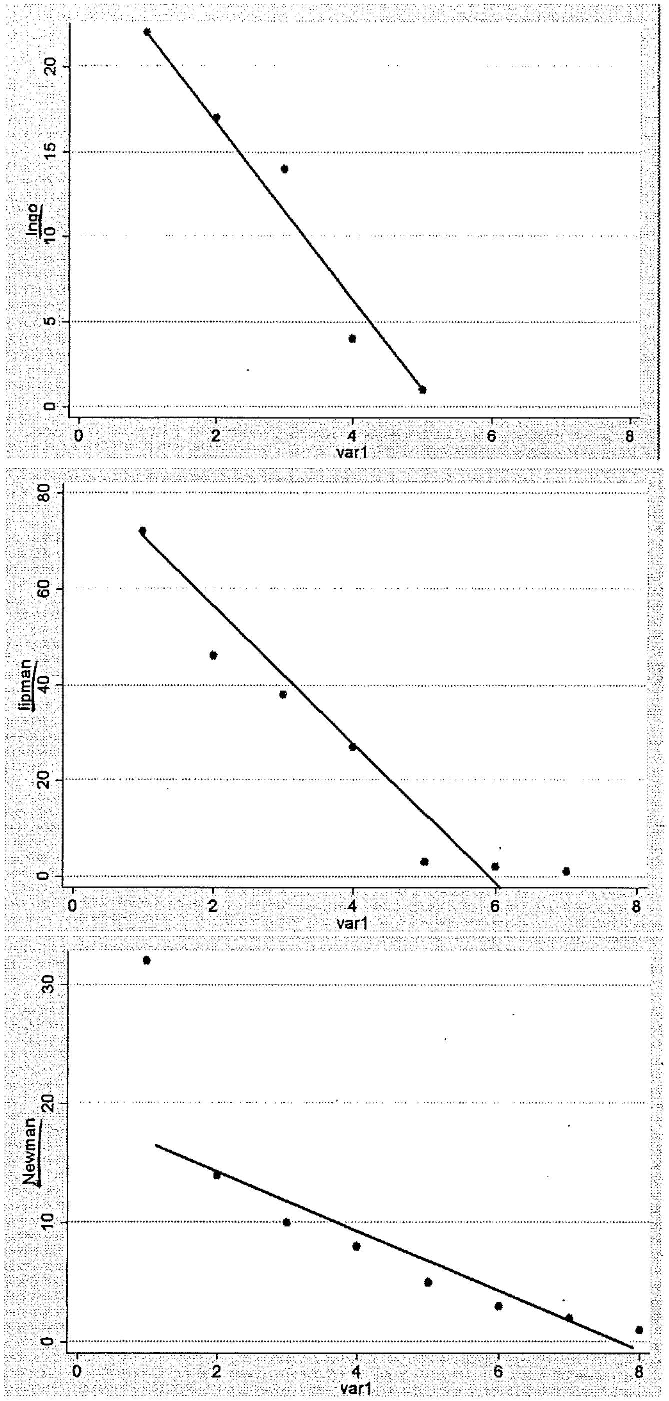
N>36, the plot is no more linear.



1. b

N>114, the plot is no more linear

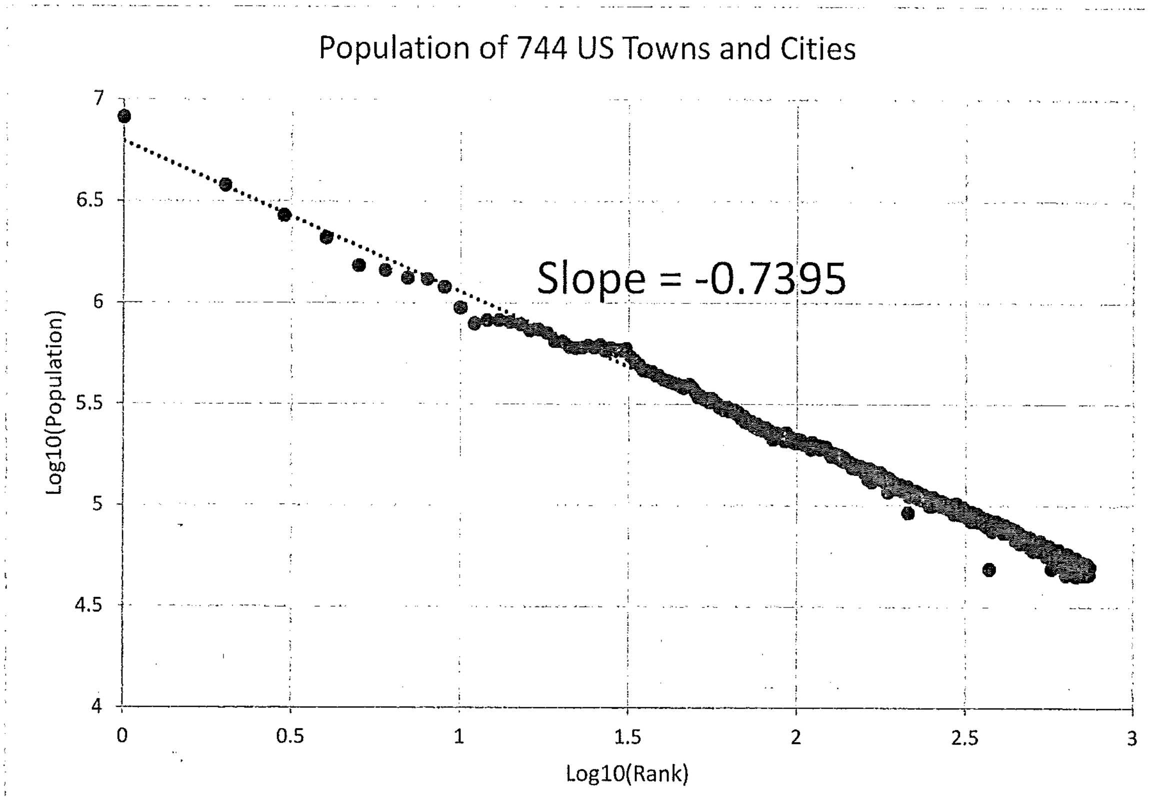




Ingo
var1
lipman
var1
Newman
var1

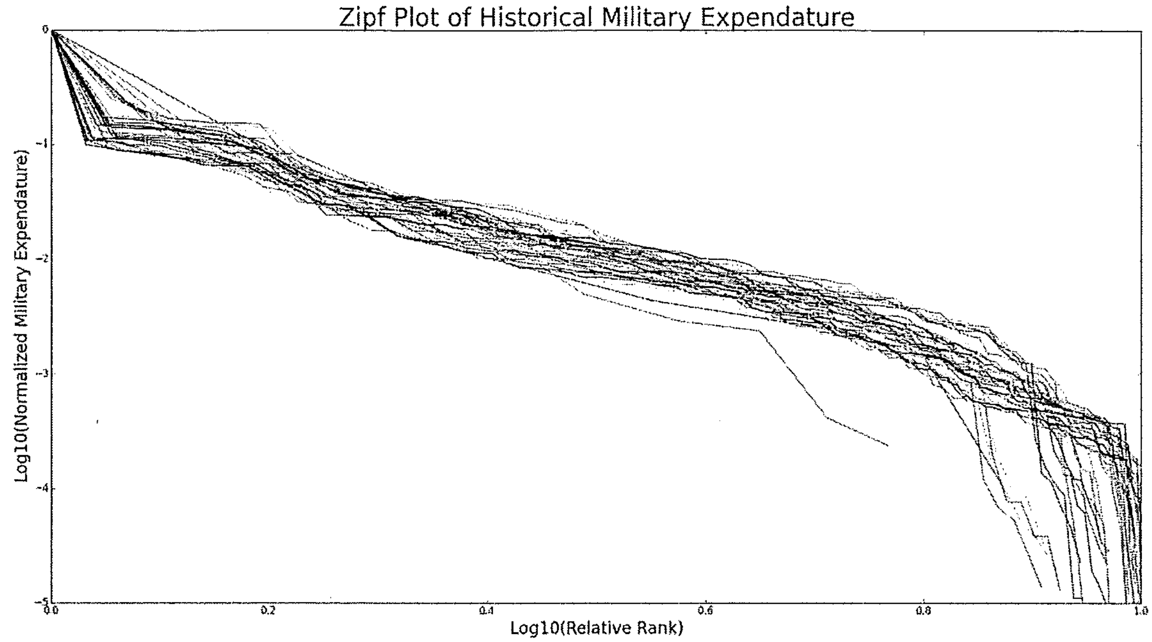
10/10

Jonathan Wurtz
Econophysics Homework I



Looks great

Source: US Census



Source: Stockholm International Peace Research Institute

Great

While we are on the subject of Zipf plots I was curious as to exactly how universal these sorts of trends are. Above is some data that I found for historical military expenditure for various countries since 1950. The data is normalized in two ways to show the universality: First, the expenditure is normalized to the top military spender of the year. Second, because the dataset is not complete (early years had less countries) I had to normalize the rank out of the number of countries in the dataset. This normalization process really flags the relevance of a zipf plot, for me, because it means there would be a different slope for different years (aka, different number of countries). This may also be an artifact of the finite data set.