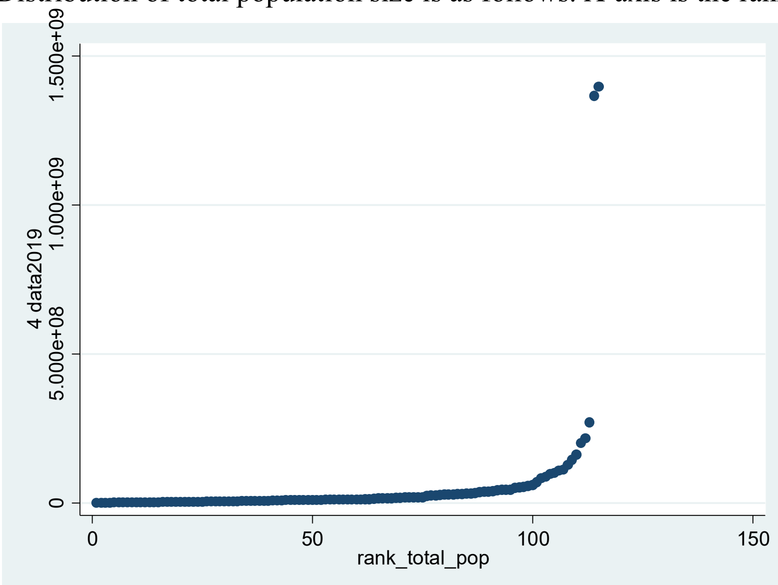
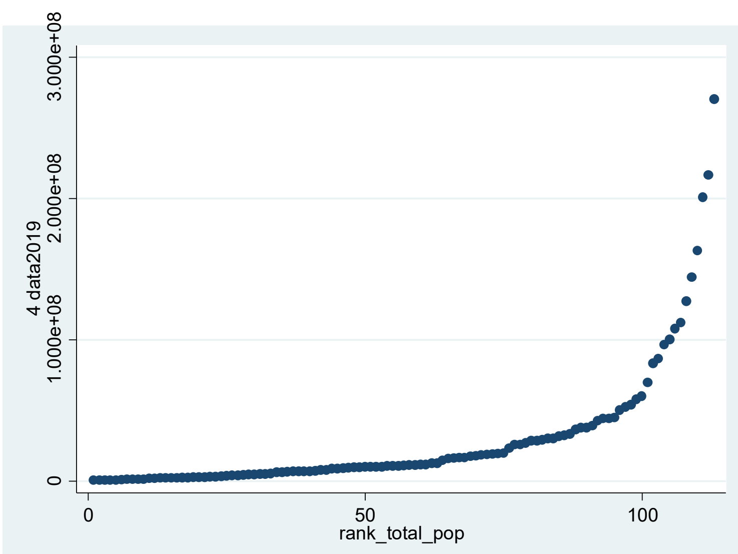
Use data_eviews - simp.dta. Perform structural break tests in Eviews 9.

(1) Choosing cutoffs
115 countries in total in the original WBES. We focus on total population in a country which is ranked in an ascending order. Rank_total_pop =1 for the country with the smallest population (Malta). Rank_total_pop =115 for the country with the largest population (China).

Distribution of total population size is as follows. X-axis is the rank and y-axis is the total population:




Rank = 114 and 115 are India and China, obviously outliers. If we remove China and India, we get:




The distribution looks exponential. Running the following regression and conduct structural break tests on the slope and the intercept:

Log(total_pop) = a*Rank_total_pop + constant

Use Bai-Perron test to identify the structural breaks in the distribution of country population size. Bai-Perron testidentifies multiple unknown break points in the whole sample. H0: no break points. H1: N break points. Set maximal N to be 5. Significance level = 10%.

Eviews code:

```
import(link) D:\Research\WEBS\Geo_data\city_category\WBdata\data_eviews - simp.dta @freq U 1 @smpl @all
series ln_pop = log(total_pop)
{%equation}.ls ln_pop c rank_total_pop
{%equation}.multibreak(method=glob, trim=5, size=10, heterr) c rank_total_pop
```

Results:

Multiple breakpoint tests
Bai-Perron tests of 1 to M globally determined breaks
Date: 03/13/21 Time: 23:36
Sample: 1 115
Included observations: 115
Breaking variables: C RANK_TOTAL_POP
Break test options: Trimming 0.05, Max. breaks 5, Sig. level 0.10
Allow heterogeneous error distributions across breaks

| | |
|---|---|
| Sequential F-statistic determined breaks: | 5 |
| Significant F-statistic largest breaks: | 5 |
| UDmax determined breaks: | 1 |
| WDmax determined breaks: | 5 |

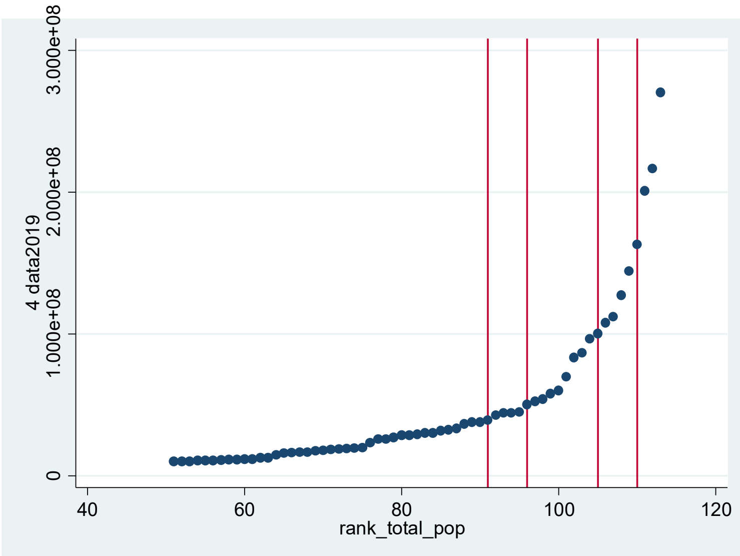
| Breaks | F-statistic | Scaled F-statistic | Weighted F-statistic | Critical Value |
|---|---|---|---|---|
| 1 * | 6.753357 | 13.50671 | 13.50671 | 11.02 |
| 2 * | 6.614795 | 13.22959 | 13.91127 | 10.48 |
| 3 * | 4.879671 | 9.759342 | 11.19125 | 9.61 |
| 4 * | 5.309670 | 10.61934 | 13.01726 | 8.99 |
| 5 * | 5.658020 | 11.31604 | 14.67091 | 8.50 |

| | | | |
|---|---|---|---|
| UDMax statistic* | 13.50671 | UDMax critical value** | 11.69 |
| WDMax statistic* | 14.67091 | WDMax critical value** | 12.33 |

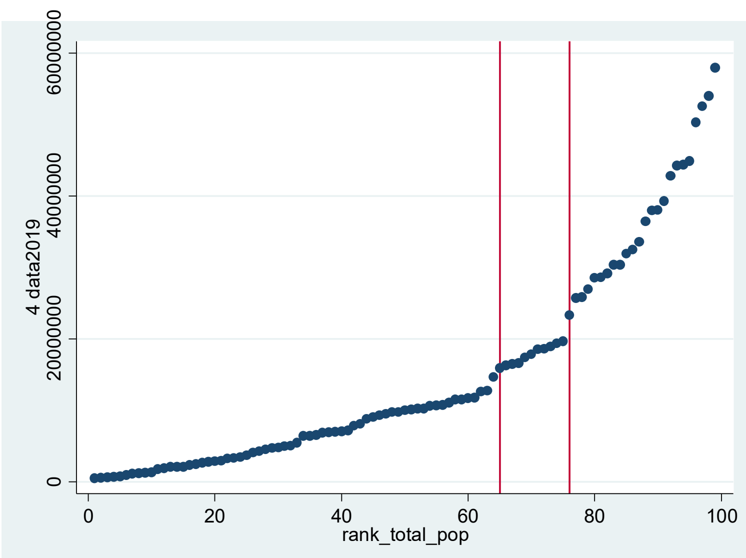
* Significant at the 0.10 level.
** Bai-Perron (Econometric Journal, 2003) critical values.

Estimated break dates:
1: 90
2: 91, 96
3: 91, 96, 107
4: 91, 96, 105, 110
5: 34, 91, 96, 105, 110

Bai-Perron robustly identifies breaks at rank = 91 (just reaching 40 million population), 96 (just reaching 50 million population), 105 (just reaching 100 million population), 110 (about to reach 200 million population), as follows. 96 seems to be a better cutoff than 91 based on the following graph. So we keep rank = 96, 105, 110 as structural breaks.




It is hard to identify robust structural breaks below rank =95 in the whole sample. Take the subsample of rank 1-95. Judging from the graph, the break can be rank = 65 (15 million) or rank = 76 (20 million). Chow tests on these two known breaks show that rank =76 gives more significant statistics than rank =65. Choose rank = 76 as the break point.




Eviews code:

```
{%equation}.ls ln_pop c rank_total_pop  (note: for rank = 1-95 only)
{%equation}.chow 76
{%equation}.chow 65
```

Combining the above structural break tests and the distribution of country population, we get:
-Class 1: Rank 1-75, country population <20 million
-Class 2: Rank 76-95, country population 20-50 million
-Class 3: Rank 96-104, country population 50-100 million

-Class 4: Rank 105-110: country population 100-200 million
-Class 5: Rank 111-113: country population > 200 million
-Class 6: Rank 114 and 115, China and India, country population >1 billion

We include more cities in the sample in larger countries. How to determine the number of cities? Think about the modal country populations for each category:

1. <20 million: modal = 10 million
2. 20-50 million: modal = 25 million (X2.5)
3. 50-100 million: modal = 75 million (X3)
4. 100-200 million: modal = 150 million (X2)
5. 200-1000 million: modal = 600 million (X4)
6. >1000 million: modal = 1200 million (X2)

The modal population goes up by 120 across the six categories. If in the base category we choose 1 city, category 6 will have 12 cities. This implies:

1. country population <20 million: include top 1 city
2. country population 20-50 million: include top 2 cities
3. country population 50-100 million: include top 3 cities
4. country population 100-200 million: include top 5 cities
5. country population 200-1000 million: include top 6 cities
6. country population >1 billion: include top 12 cities