Homework 5: Pareto and Kuznets on the Grand Tour

DingHuan, 3170102085

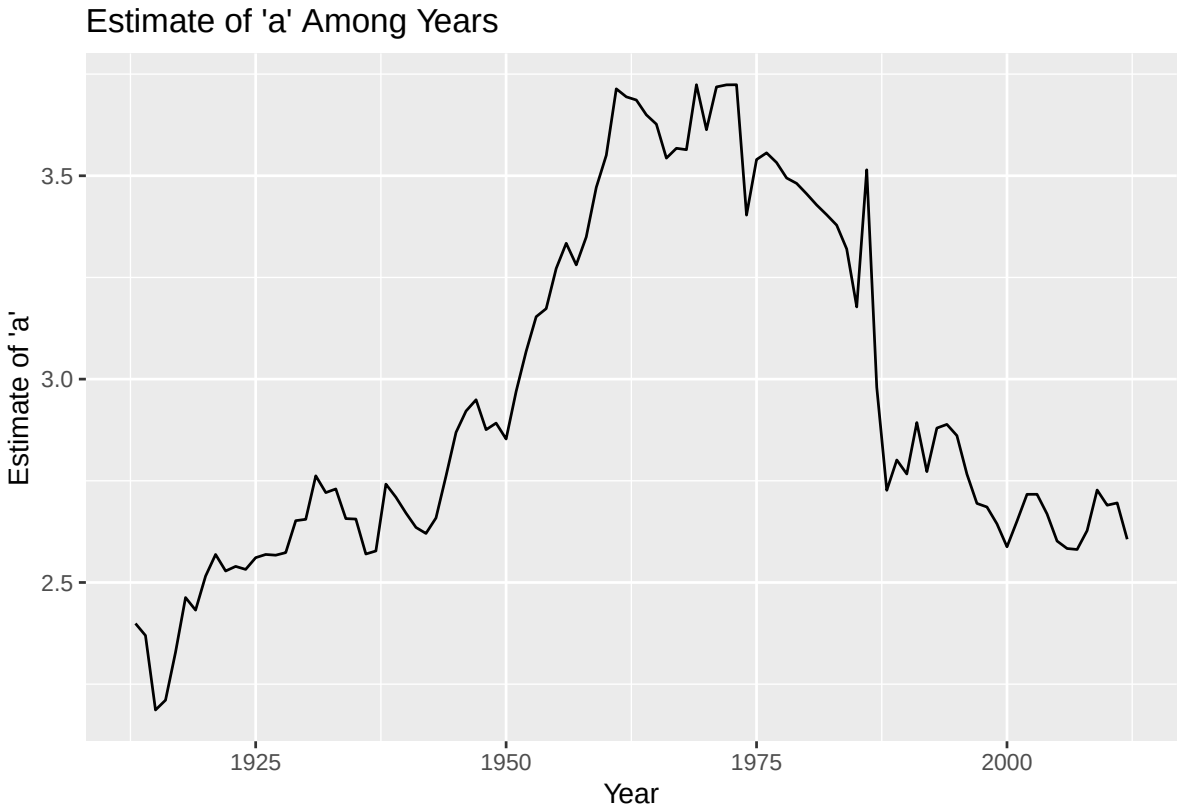
1.
```
percentile_ratio_discrepancies <- function(P99, P99.5, P99.9, a){
  p <- sum(((c(P99,P99.5,P99) / c(P99.9,P99.9,P99.5))^(1-a) - c(10,5,2))^2)
  return(p)
}
percentile_ratio_discrepancies(P99=1e6, P99.5=2e6, P99.9=1e7, a=2)
```

```
## [1] 0
```

2.
```
exponent.multi_ratios_est <- function(P99, P99.5, P99.9){
  prd <- function(a) percentile_ratio_discrepancies(P99, P99.5, P99.9, a)
  a <- 1 - log(10) / log(P99/P99.9)
  res <- nlm(prd, a)
  return(res$estimate)
}
exponent.multi_ratios_est(P99=1e6, P99.5=2e6, P99.9=1e7)
```

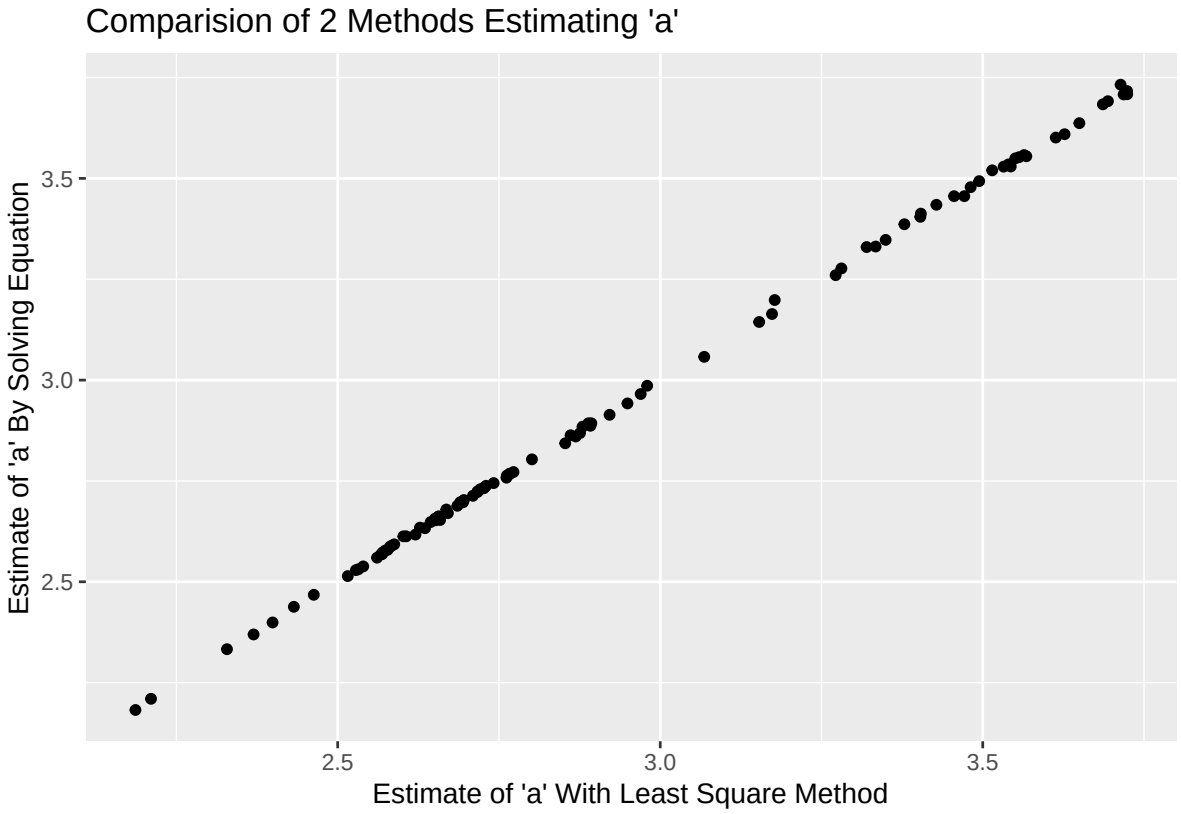
```
## [1] 2
```

3.
```
wtid.report <- read.csv("wtid-report.csv")
for (i in 1:nrow(wtid.report)){
  wtid.report[i,"a"] <- exponent.multi_ratios_est(
    wtid.report[i,"P99.income.threshold"],
    wtid.report[i,"P99.5.income.threshold"],
    wtid.report[i,"P99.9.income.threshold"])
}
wtid.report %>% ggplot(aes(x = Year, y = a)) +
  geom_line() +
  labs(x = "Year", y = "Estimate of \'a\'",
       title = "Estimate of \'a\' Among Years")
```



4. These two methods of estimating 'a' perform quit similar. The scatter-plot shows that in most cases, the results are very close, but the least square method requires more calculation with higher complexity. So, just calculating 'a' through equation (4) is a better choice in many cases.

```
wtid.report %>% mutate(a2 = 1-log(10)/
                         log(P99.income.threshold/P99.9.income.threshold)) %>%
  ggplot(aes(x = a, y = a2)) +
  geom_point() +
  labs(x = "Estimate of \'a\' With Least Square Method",
       y = "Estimate of \'a\' By Solving Equation",
       title = "Comparision of 2 Methods Estimating \'a\'")
```


Comparision of 2 Methods Estimating 'a'
Estimate of 'a' By Solving Equation
3.5
3.0
2.5
2.5
3.0
3.5
Estimate of 'a' With Least Square Method