Lab 7: Estimating Inequality

Econometrics Beyond Ordinary Least Squares

Patrick Mokre

WS 2020/2021

Parametric Estimation

- Econometrics is about probability distributions.

$$y_i = \alpha + x_i'\beta + \epsilon_i \; ; \; \epsilon_i \sim N(0, \sigma_\epsilon^2) \Leftrightarrow y_i \sim N(\alpha + x_i'\beta, \sigma_\epsilon^2) \tag{1}$$

$$Z \sim N(\mu, \sigma^2) \Rightarrow E[Z] = \mu \tag{2}$$

- A Gaussian Normal distribution has **two parameters**, location/mean μ and variance/scale σ^2.
- The parameters can be consistently estimated by $(1/N)\sum_i^N y_i$ and $(1/N)\sum_i^N (y_i - \bar{y})^2$.

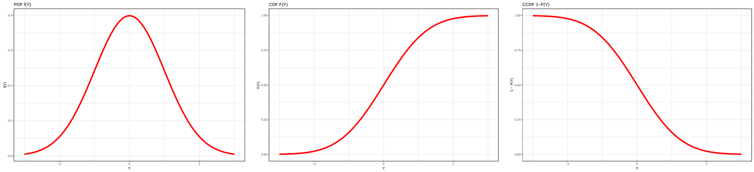
Distributions

- A parametric distribution can be sufficiently described by its **parameters**.
- A distribution can be represented as a **probability density function** $f(Y)$, **cumulative density function** $F(Y) = \int f(Y)dY$ or **complementary cumulative density function** $1 - F(Y)$.

$$f(Y) = P(y_i = Y) \tag{3}$$

$$F(Y) = P(y_i < Y) \tag{4}$$

$$1 - F(Y) = P(y_i > Y) \tag{5}$$



(a) (b) (c)

Figure 1: Representations of a Standard Normal Distribution N(0,1)

Parameter Estimation

Empirical CDF estimation:

Glivenko-Cantelli theorem: The empirical CDF, ie. the **share of observations below treshold a** converges to the true CDF with increasing sample size.

$$\tilde{F}_n(Y) = \frac{1}{n}\sum_{i}^{N} 1_{-\inf,a}(x_i) \tag{6}$$

$$max_a \mid \tilde{F}_n(a) - F(a) \mid \rightarrow 0 \tag{7}$$

Moment Condition Estimation

$$\frac{1}{n}\sum_{i}^{N} y_i \rightarrow \mu \tag{8}$$

$$\frac{1}{n}\sum_{i}^{N} (y_i - \bar{y}_i)^2 \rightarrow \sigma^2 \tag{9}$$

Maximum Likelihood Estimation

- $L(Y \mid \theta)$ gives the **likelihood** of Y to be observed if θ is the true parameter vector.
- Compare two potential θs, the one with the higher likelihood is the one with which the data agrees better -> do the for "all" potential values, find $\hat{\theta}_{ML}$.
- The likelihood is equivalent to the **product of the PDFs** $f(Y \mid \theta)$.

$$L(Y \mid \theta) = \prod_{i}^{N} L(y_i \mid \theta) = \prod_{i}^{N} f(y_i \mid \theta) \qquad (10)$$

- Logarithmic function is monotonous, ie. $y > x \Leftrightarrow log(y) > log(x)$: log-likelihood allows to find a better fit.
- Keep in mind: $log(a \times b) = log(a) + log(b)$.

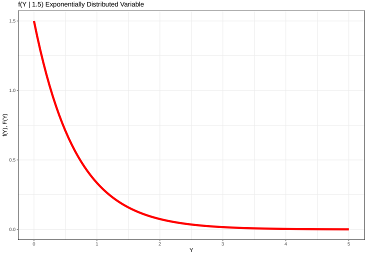
$$l(Y \mid \theta) = log(L(Y \mid \theta)) = \sum_{i}^{N} log(f(y_i \mid \theta)) \qquad (11)$$

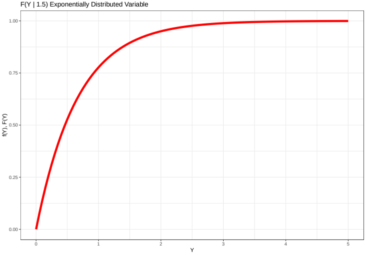
- Advantage: Pretty general method. Disadvantage: Estimating multiple parameters without analytical solution is pretty computationally intensive.

Example: Exponential Function

$$f(Y \mid \lambda) = \lambda \exp^{-\lambda y} \quad \forall y \geq 0 \tag{12}$$

$$F(Y \mid \lambda) = 1 - \exp^{-\lambda y} \tag{13}$$





Example: Estimating the Exponential Function Parameters

Maximum Likelihood

$$
\begin{aligned}
f(Y \mid \lambda) &= \lambda \exp^{-\lambda y} \quad \forall y \geq 0 \\
L(\lambda \mid Y) &= \prod_i^N f(y_i \mid \theta) = \prod_i^N \lambda \exp^{-\lambda y_i} \\
\hat{\lambda}_{ML} &: \frac{\partial log(L(\lambda \mid Y))}{\partial \lambda} = 0 \\
&= \frac{\partial \lambda^N \exp^{-\lambda \sum_i^N y_i}}{\lambda} = \frac{\partial log(N log(\lambda)) - \lambda \sum_i^N y_i}{\partial \lambda} \\
&= \frac{N}{\lambda} - \sum_i^N y_i \\
\hat{\lambda}_{ML} &= \frac{N}{\sum_i^N y_i} \qquad (14)
\end{aligned}
$$

Wealth: The Pareto Function

- **Vilfredo Pareto**: 20 % of Italian nobility own 80 % of the land.
- Insight: Inequality between persons **increases with relative wealth**.
- Vermeulen: Surveys are **less likely to even ask rich households**, rich households are more likely to **understate wealth** and to **reject participation** $\Rightarrow$ differential survey bias.
- Solution: Approximate distribution with available data, correct for bias.
- Generalization: $1 - F(w) = P(w_i \geq w) = 1 - F(w)1 - \frac{w_{min}}{w}^{\alpha} \quad \forall w \geq w_{min}$.
- Linearization: $log(1 - F(w)) = \alpha \times log(\frac{w_{min}}{W}) = \alpha log(w_{min}) - \alpha w$.

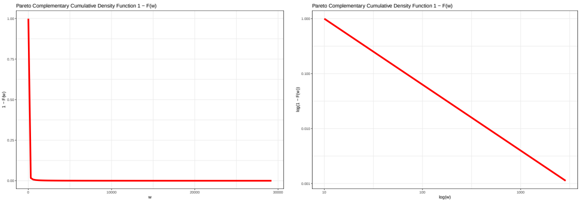
Estimating the Pareto Function

Estimation of α **by linear regression**:

- For all observations $w_i \geq w_{min}$ calculate empirical complementary cumulative density function $1 - \tilde{F}(w_i) = \frac{N(w_i)}{N(w)}$, regress on w_i, retrieve α as coefficient.
- **Note**: Rank i with $i = 1$ the richest household is proportional to CCDF. $1 - \tilde{F}(w) = \frac{i}{N} \propto i$.

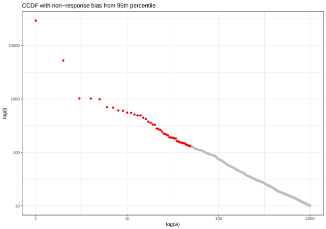
$$log(1 - \tilde{F}(w_i)) = \delta - \alpha log(w_i) + \epsilon_i \quad (15)$$

$$\tilde{F}(w_i) = \frac{\sum_{j:w_j > w_i} n(w_j)}{\sum_{k:w_k > w_{min}} n(w_k}$$

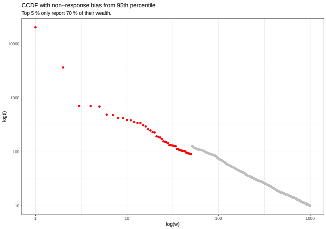


(a) (b)

Differential Bias



(a) Differential Non-Response



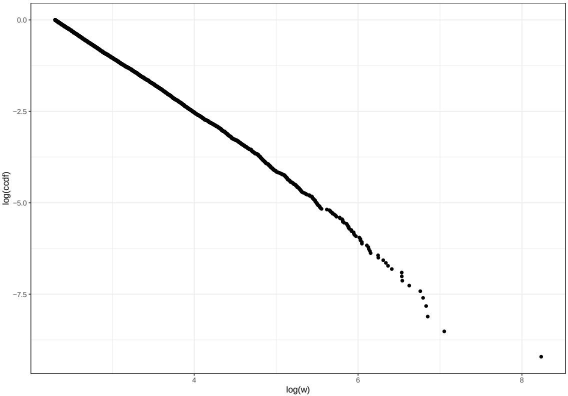
(b) Differential underreporting

Figure 3: Differential biases in wealth survey

Demonstration: Estimating Pareto's Alpha

Scale Pareto α gives the (increasing) inequality in the distribution: The lower α, the higher the inequality.

```
dta_tmp <- data.frame(w=EnvStats::rpareto(10000, location=10, shape=1.5
  arrange(desc(w)) %>% mutate(i=dplyr::row_number()) %>% mutate(ccdf=(i
dta_tmp %>% ggplot() + geom_point(aes(x=log(w), y=log(ccdf))) + theme_b
```



Demonstration: Estimating Pareto's Alpha 2

```
lm_tmp1 <- lm(log(ccdf) ~ log(w), data=dta_tmp)
lm_tmp1$coefficients %>% print()
```

```
## (Intercept)      log(w)
##    3.499471   -1.513731
```

```
lm_tmp2 <- lm(log(i) ~ log(w), data=dta_tmp)
lm_tmp2$coefficients %>% print()
```

```
## (Intercept)      log(w)
##   12.709811   -1.513731
```

Increasing Estimation Robustness

- Gabaix and Ibragimov (2011): In small samples, $log(i)$ produces **bias to leading rank**. $log(i-0.5)$ rather than $log(i)$ as approximation of $log(1-\bar{F}(w))$ is a primitive but effective remedy.
- Klass et.al. (2006), Vermeulen (2014): Add **Forbes list observations** to sample to determine top of the distribution.
- Chakraborty and Waltl (2018): **Conditional Median Regression** is more robust to outliers than ordinary linear regression.

$$
\begin{aligned}
log(1-\bar{F}(w)) &= \delta - \alpha w_i + \epsilon_i \\
log(i-0.5) &= \delta_{\tau=0.5} - \alpha_{\tau=0.5} w_i + \nu_i
\end{aligned} \tag{16}
$$

Demonstration: Small Sample Robustness

```
set.seed(1)
dta_tmp2 <-
  data.frame(w=EnvStats::rpareto(1000, location=10, shape=1.5)) %>%
  arrange(desc(w)) %>% mutate(i=dplyr::row_number()) %>%
  mutate(ccdf=(i/n()))
lm_tmp3 <- lm(log(i) ~ log(w), data=dta_tmp2)
lm_tmp3$coefficients %>% print()
```

```
## (Intercept)      log(w)
##   10.116195   -1.412089
```

```
lm_tmp4 <- lm(log(i-0.5) ~ log(w), data=dta_tmp2)
lm_tmp4$coefficients %>% print()
```

```
## (Intercept)      log(w)
##   10.170325   -1.431623
```

```
qr_tmp <- quantreg::rq(log((i-0.5)) ~ log(w), data=dta_tmp2)
qr_tmp$coefficients %>% print()
```

```
## (Intercept)      log(w)
##   10.313573   -1.480314
```

Determine w_{min}

- Wealth distribution is a **mixed distribution**: Well-observed (bottom) majority, biased tail. w_0: Treshold value.
- Log-Log Linearity of Pareto distribution should hold only for Pareto tail.
- OLS/CQR **Root Mean Squared Error (RMSE)**: Measure of linearity/goodness-of fit. Minimize RMSE to estimate w_0.

```
set.seed(1)
dta_tmp3 <- data.frame(w=c(rexp(1000, 0.7),
                           EnvStats::rpareto(1000, 10, 1.5))) %>%
  arrange(desc(w)) %>% mutate(i=dplyr::row_number()) %>%
  mutate(ccdf=i/n())
dta_tmp3 %>% ggplot() + geom_point(aes(x=w, y=ccdf), size=2) +
  scale_x_log10() + scale_y_log10() + theme_bw()
```

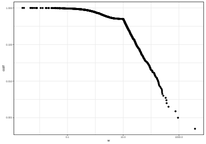
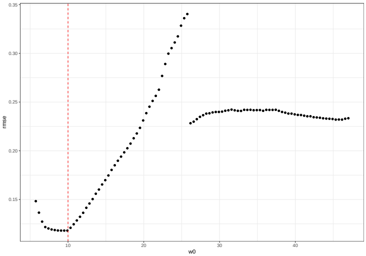


Figure 4: Mixed Distribution from Exponential (0.7) and Pareto (10, 1.5)

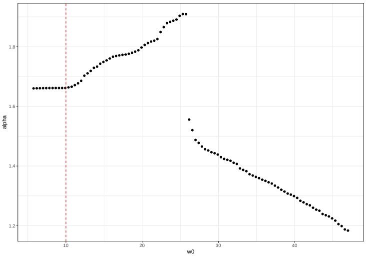
Demonstration: Determining w_0

```
for(i in 1:nw0){
  w0_tmp <- w0_est[i,1]
  dta_tmp4 %>%
    filter(w>w0_tmp) %>%
    quantreg::rq(log(i-0.5) ~ log(w),
                 data=.,
                 tau=0.5) -> qr_tmp
  w0_est[i,2] <- sqrt(mean((qr_tmp$residuals)^2))
  qr_tmp$coefficients[2] %>% abs() -> w0_est[i,3]
  rm(w0_tmp, qr_tmp)
}
rm(i)
```

Demonstration: Determining $w0$



(a) RMSE



(b) Alpha

Figure 5: Root Mean Squared Errors and Pareto's Alpha for different values w0.

Length of the Tail

- Differential Non-Repsonse implies a survey **under-estimates the number of rich households**.
- Cumulative Density Function $F(Y)$: Percentage of Observations with $y < Y$.
- Difference between two CDFs $F(y_1) - F(y_2)$: Percentage of observations between y_1 and $y2$.
- NUmber of observations in a distribution can be deducted from (1) number of observations between y_1 and y_2 and (b) the respective CDFs $F(y_1)$ and $F(y_2)$.

$$\sum_{i}^{N} 1(w_i >= w_{min}) = \frac{1 - F(w_0) - F(w_{min})}{F(w_0) - F(w_{min}} \sum_{i}^{N} 1(w_i \in [w_{min}, w_0])$$

$$= \frac{1 - F(w_0))}{F(w_0)} \sum_{i}^{N} 1(w_i \in [w_{min}, w_0]) \tag{17}$$

Simulating Tail Observations

$$w_i = w_{min} \left(\frac{\sum_i^N 1(w_i >= w_m in)}{\sum_j^N w_j > w_i} \right)^{1/\alpha} \quad (18)$$

Table 1: This is a caption

Col1	Col2	Col3