STK4600

Solutions to Ex #6

1b

discharged: utskrevet

I: BLU estimator:

$$\hat{\beta} = \sum_s y_i / \sum_s x_i$$

$$\Rightarrow \hat{\beta} = \frac{6336}{2457} = \underline{2{,}579}$$

1c

II: BLU estimators

$$\hat{\beta}_1 = \bar{y}_s - \hat{\beta}_2 \bar{x}_s \quad \text{and} \quad \hat{\beta}_2 = \frac{\sum (x_i - \bar{x}_s) y_i}{\sum (x_i - \bar{x}_s)^2}$$

$\bar{y}_s = 633{,}6 \quad \bar{x}_s = 245{,}7$

$$\hat{\beta}_2 = \frac{1664371{,}8}{756486{,}1} = \underline{2{,}200}$$

and $\underline{\hat{\beta}_1 = 93{,}06}$

1d $\hat{T}_0 = N\bar{y}_s = 393 \times 633{,}6 = \underline{249005}$

$$\hat{T}_R = \hat{R} \cdot t_x = \hat{\beta} \cdot N\bar{x} = \cancel{278421} = \frac{\sum_s y_i}{\sum_s x_i} \cdot 393 \times 274{,}7 = \underline{278395}$$

Regression model:

$$\hat{T}_{pred} = \sum_{i \in s} y_i + \sum_{i \notin s} (\hat{\beta}_1 + \hat{\beta}_2 x_i)$$

$$= N\bar{y}_s + \hat{\beta}_2 (t_x - N\bar{x}_s)$$

$$= 249005 + 2.200(N \cdot 274{,}7 - N \cdot 245{,}7)$$

$$= 249005 + 2.200 \times N \times 29 = \underline{274078}$$

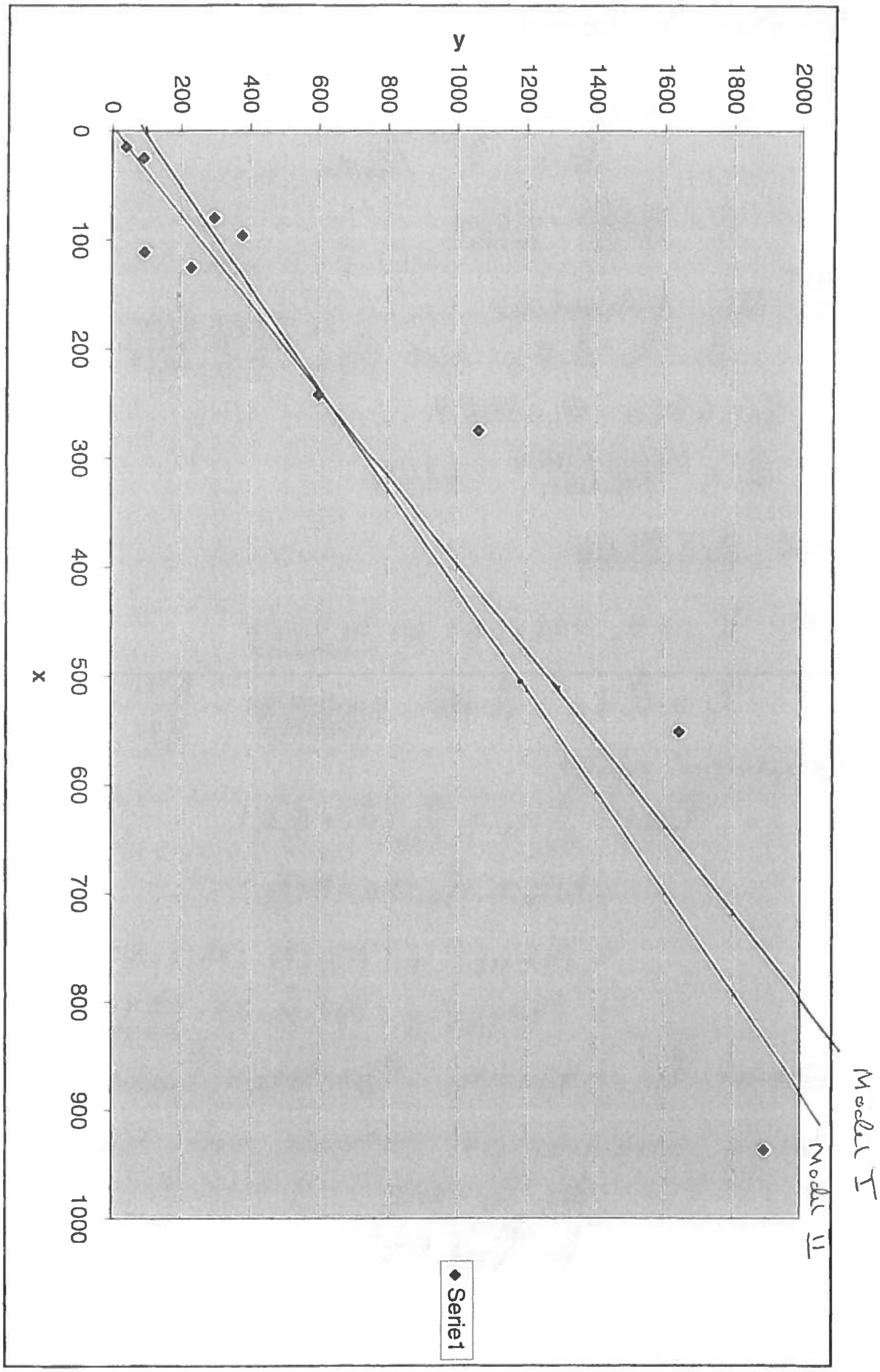
Errors: $\hat{T}_0$: -71154 $\quad \hat{T}_R$: -41764, $\hat{T}_{pred}$: -46081

Using auxiliary x reduces error by:

for $\hat{T}_R$: 41%

for $\hat{T}_{pred}$: 35%

Ex#6 1a & b


y
2000
1800
1600
1400
1200
1000
800
600
400
200
0
x
0
100
200
300
400
500
600
700
800
900
1000
Serie1
Model I
Model II

2. $\hat{T}_{pred} = N\bar{Y}_s + \hat{\beta}_2 (t_x - N\bar{x}_s)$

$$\Rightarrow \text{Var}(\hat{T}_{pred} - T) = \text{Var}\Big(N\bar{Y}_s - n\bar{Y}_s + \hat{\beta}_2(t_x - N\bar{x}_s) \div \sum_{i \notin s} Y_i\Big)$$

$$= \text{Var}\{(N-n)\bar{Y}_s + (t_x - N\bar{x}_s)\hat{\beta}_2\} + (N-n)\sigma^2\}$$

Now: $\text{cov}(\bar{Y}_s, \hat{\beta}_2)$

$$= \frac{1}{\sum_s (x_i - \bar{x}_s)^2} \sum_{i \in s} (x_i - \bar{x}_s) \underbrace{\text{cov}(\bar{Y}_s, Y_i)}_{\frac{1}{n}\text{Var}(Y_i) = \frac{\sigma^2}{n}.}$$

$$= \frac{\sigma^2/n}{\sum_s (x_i - \bar{x}_s)^2} \underbrace{\sum (x_i - \bar{x}_s)}_{\overset{\shortparallel}{0}} = 0$$

$$\Rightarrow \text{Var}(\hat{T}_{pred} - T) = (N-n)\text{Var}(\bar{Y}_s) + (t_x - N\bar{x}_s)^2 \text{Var}(\hat{\beta}) + (N-n)\sigma^2$$

$$= (N-n)^2 \frac{\sigma^2}{n} + (N-n)\sigma^2 + N^2(\bar{x} - \bar{x}_s)^2 \cdot \frac{\sigma^2}{\sum_s (x_i - \bar{x}_s)^2}$$

$$= \frac{N^2}{n}\sigma^2 \left\{ \underbrace{\frac{(N-n)^2 + n(N-n)}{N^2}}_{\frac{N-n}{N} = 1 - \frac{n}{N}} + \frac{(\bar{x} - \bar{x}_s)^2}{\underbrace{\frac{1}{n}\sum_s (x_i - \bar{x}_s)^2}_{\frac{n-1}{n} \cdot s_x^2}} \right\} \quad \text{II}$$

Sample must have the property that

$\underline{\bar{x}_s = \bar{x}}$: Balanced sample

3. a) Suppose $\mu_i = E(Y_i) = \mu$

$$\mapsto \quad E(N\bar{Y}_s - T) = N\mu - E(T) = N\mu - N\mu = 0$$

b)
$$E(N\bar{Y}_s - T)$$

$$= \cancel{E(} N E(\bar{Y}_s) - \left(\sum_{i=1}^{N} \beta_1 + \beta_2 x_i\right)$$

$$= N\left\{\frac{1}{n}\sum_{i\in s}(\beta_1+\beta_2 x_i)\right\} - N\beta_1 - \beta_2 t_x \}$$

$$= N\beta_1 + \beta_2 N \bar{x}_s - N\beta_1 - \beta_2 t_x$$

$$= \underline{N\beta_2(\bar{x}_s - \bar{x})} = \text{prediction bias}$$

Bias is eliminated, $= 0$, $\underline{\text{if } \bar{x}_s = \bar{x}\,!}$

4) BLU predictor has prediction variance

$$\mathrm{Var}(\hat{T}_{pred} - T) = N^2 \cdot \frac{1-f}{n} \cdot \frac{\bar{x}_r \cdot \bar{x}}{\bar{x}_s} \sigma^2$$

Now: $\bar{x}_r = \dfrac{t_x - n\bar{x}_s}{N-n} = \dfrac{N}{N-n}\bar{x} - \dfrac{n}{N-n}\bar{x}_s$ and

$$\mathrm{Var}(\hat{T}_{pred} - T) = N^2 \frac{1-f}{n} \cdot \sigma^2 \frac{\bar{x}}{N-n} \cdot \frac{N\bar{x} - n\bar{x}_s}{\bar{x}_s}$$

is minimized when $\bar{x}_s$ is maximum.

i.e. when the sampling design selects the n units with largest x_i-values, with probability 1.

R-solution to nr.1 in Exercise 6

```
> x=c(15,25,80,96,111,125,242,275,551,937)
> x
 [1]  15  25  80  96 111 125 242 275 551 937
> y=c(41,92,297,377,95,231,601,1063,1645,1894)
> y
 [1]   41   92  297  377   95  231  601 1063 1645 1894
> mean(y)
[1] 633.6
> mean(x)
[1] 245.7
```

1a

```
>plot(x,y)
```

1b

```
> Discharges=data.frame(y,x)
> Discharges
        y     x
1       41    15
2       92    25
3       297   80
4       377   96
5       95    111
6       231   125
7       601   242
8       1063  275
9       1645  551
10      1894  937
```

Model I

```
> ratiomodel=lm(y/sqrt(x)~sqrt(x)-1)
> ratiomodel

Call:
lm(formula = y/sqrt(x) ~ sqrt(x) - 1)

Coefficients:
sqrt(x)
  2.579
```

1c

Model II

```
> linreg=lm(y~x,data=Discharges)
> linreg

Call:
lm(formula = y ~ x, data = Discharges)

Coefficients:
(Intercept)        x
     93.03      2.20
```

Plots on 1a, 1b and 1c

```
> plot(x,y)
> z=seq(0,900,10)
> yhat=93.03+2.2*z
> lines(z,yhat)
> yratio=2.579*z
> lines(z,yratio)
```

1d

```
> Tzero=393*mean(y)
> Tzero
[1] 249004.8
> Ratioest=mean(y)/mean(x)
> Ratioest
[1] 2.578755
> Tratio=Ratioest*393*274.7
> Tratio
[1] 278394.9

> Tpred=393*mean(y)+2.20*(393*274.7-393*mean(x))
> Tpred
[1] 274078.2
```