36-402 Homework 4

Eu Jing Chua
eujingc

February 12, 2019

Question 1

$$
\begin{aligned}
Y &\approx cN^b && (1) \\
\log Y &\approx \log c + b \log N && (2) \\
\log \frac{Y}{N} &\approx \log c + (b-1) \log N && (3) \\
\log P &\approx \log c + (b-1) \log N && (4)
\end{aligned}
$$

Hence $\beta_0 = \log c$ and $\beta_1 = b - 1$.

Question 2

Q2 a)

If equation 1 were true, then using the linear model, we can find that $c = e^{\beta_0}$ and $b = \beta_1 + 1$

Q2 b)

Table 1: Coefficients of linear model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.79641	0.14558	60.42185	0
log(pop)	0.12418	0.01148	10.81692	0

Looking at the point estimate of the coefficient of log N, or `log(pop)`, we can see that the estimated exponent $b = 1.12 > 1$, which supports the idea of supra-linear scaling.

Q2 c)

Table 2: MSE (5-fold CV)

x
0.06679

Question 3

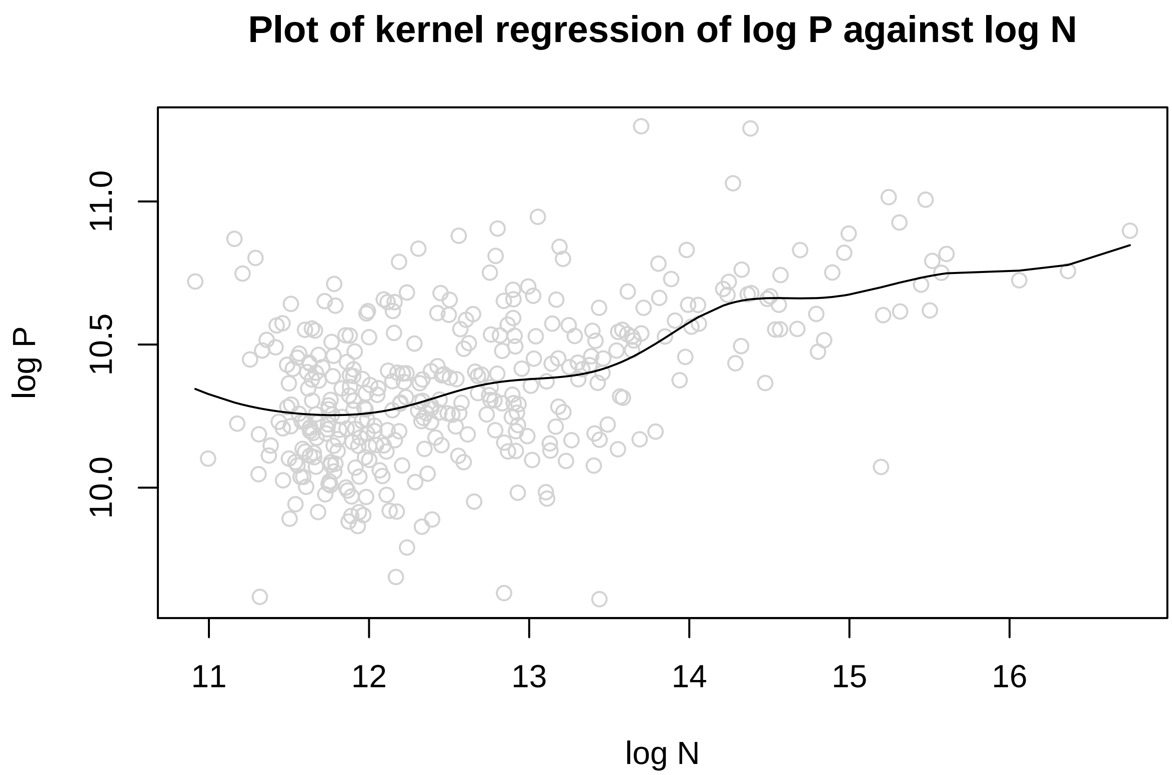


Table 3: MSE of Kernel Regression

x
0.05302

Question 4

Q4 a)

Table 4: Predicted per-capita GMP using linear model

	x
Cape Girardeau, MO / Jackson, IL	27354.54
Pittsburgh, PA	40895.54
Washington, DC	45172.97

Q4 b)

Table 5: Predicted per-capita GMP using kernel regression

	x
Cape Girardeau, MO / Jackson, IL	28747.15
Pittsburgh, PA	42680.45
Washington, DC	46009.34

Question 5

Q5 a)

Table 6: 92% C.I. for slopes of linear model

	lower	upper
(Intercept)	8.54472	9.05147
log(pop)	0.10393	0.14408

Table 7: 92% C.I. for predicted per-capital GMP

	lower	upper
Cape Girardeau, MO / Jackson, IL	26479.04	28244.61
Pittsburgh, PA	39071.14	42834.42
Washington, DC	42534.78	47998.21

Q5 b)

Table 8: 92% C.I. for slopes of linear model

	lower	upper
(Intercept)	8.54648	9.04636
log(pop)	0.10459	0.14391

Table 9: 92% C.I. for predicted per-capital GMP

	lower	upper
Cape Girardeau, MO / Jackson, IL	26464.93	28238.13
Pittsburgh, PA	39225.32	42734.15
Washington, DC	42753.93	47907.81

Q5 c)

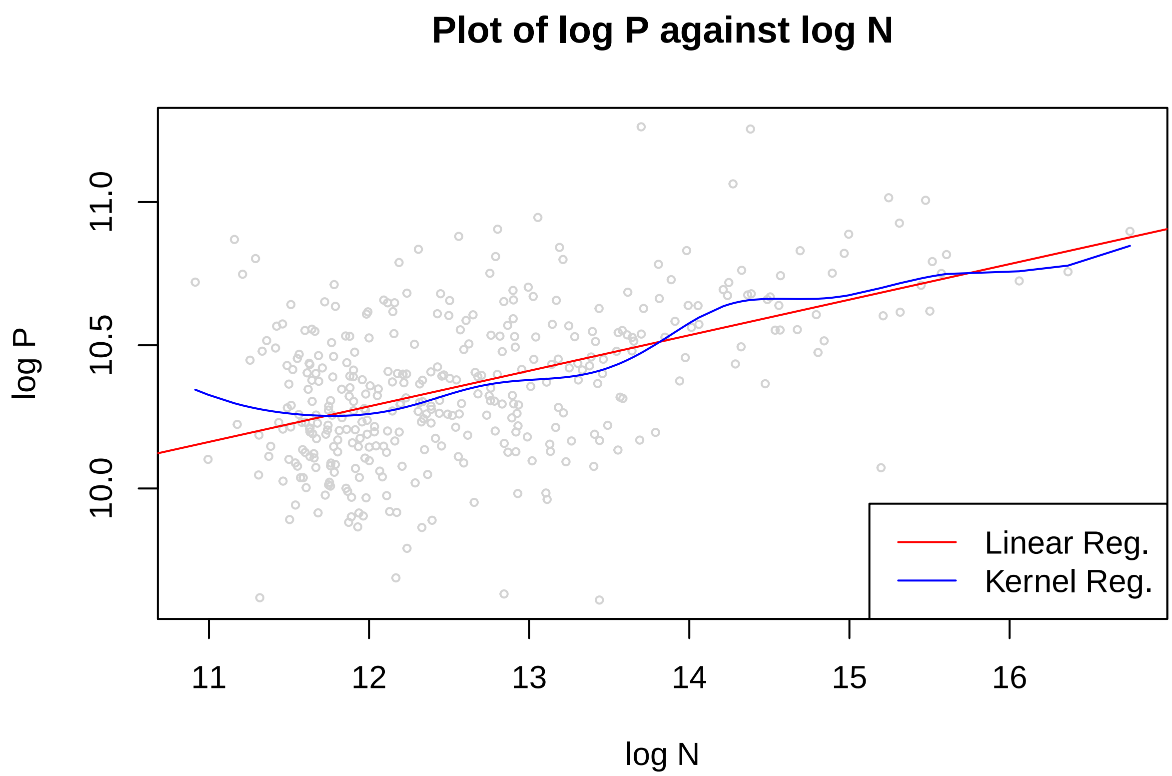
Table 10: 92% C.I. for predicted per-capital GMP

	lower	upper
Cape Girardeau, MO / Jackson, IL	21919.65	29903.66

	lower	upper
Pittsburgh, PA	36057.42	47499.89
Washington, DC	35140.43	51289.65

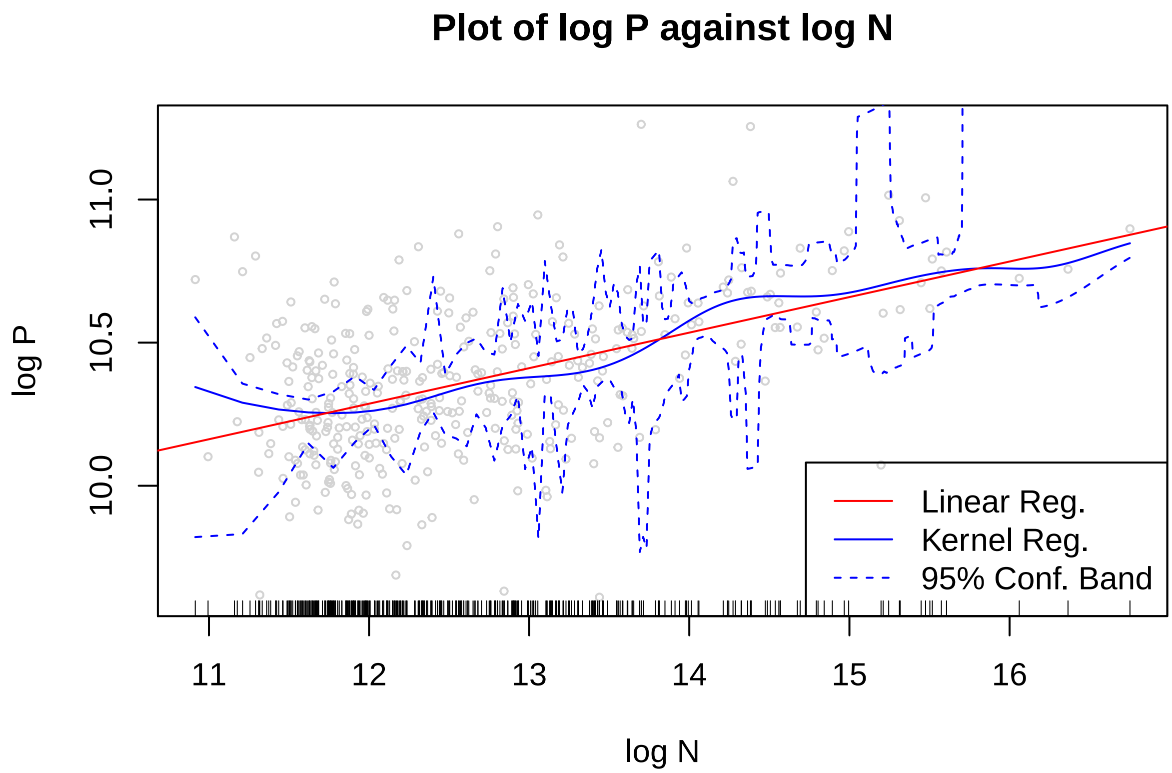
Question 6

Q6 a)



The curve estimated from the power law is linear in nature as expected, whereas the curve from the kernel regression seems to bend around more, roughly following the shape of the linear curve. By comparing their MSEs from cross-validation, we know that the kernel regression seems to predict better.

Q6 b)



The confidence band is has alot of variance, but follows the main kernel regression curve well. It can also be seen that the confidence band is narrower in the regions with more data, and wider in the regions with less data. It should be noted that the confidence band does contain the linear model from Problem 2.

Question 7

Q7 a)

Table 11: Predicted increase in log P with 10% increase in N

	x
Cape Girardeau, MO / Jackson, IL	0.01184
Pittsburgh, PA	0.01184
Washington, DC	0.01184

Q7 b)

Table 12: Predicted increase in log P with 10% increase in N

	x
Cape Girardeau, MO / Jackson, IL	-0.00613
Pittsburgh, PA	0.00031
Washington, DC	0.00943

Q7 c)

The non-parametric estimate of does not really support the idea of supra-linear scaling as it is possible for the predicted change in log P to be negative, depending on the starting value of log N. This contradicts the idea that increasing N leading to more-than-proportional increase in Y for all N.

Question 8

Q8 a)

Let $\log P' = \beta_0 + \beta_1 \log(1.1N)$, then

$$\begin{aligned} \log P' - \log P &= \beta_1 \log(1.1N) - \beta_1 \log N && (5) \\ &= \beta_1(\log 1.1 + \log N) - \beta_1 \log N && (6) \\ &= \beta_1 \log 1.1 && (7) \end{aligned}$$

Assuming the 92% C.I for β_1 is $(\beta_{1,L}, \beta_{1,U})$, then the 92% C.I for $\log P' - \log P = \log 1.1\beta_1$ would be $(\beta_{1,L} \log 1.1, \beta_{1,U} \log 1.1)$.

Q8 b)

Table 13: 92% C.I. for change in log P with 10% increase in N

lower	upper
-0.0668705	0.3139747
-0.0517113	0.2662718
-0.1126494	0.2739786

Q8 c)

The confidence intervals do not really support the idea of supra-linear scaling, as their bounds extend into the negative. This means that under 92% confidence, it is still possible for P to increase less-than-proportionally for some N, contradicting the idea of supra-linear scaling.

Q9

Based on the analysis so far, the situation seems more ambiguous as to whether the idea of supra-linear scaling holds. When comparing the supra-linear model against the non-parametric regression, we see that the fits are close, with the non-paramettric regression having slightly better prediction based on its MSE from cross-valiation. However, when we obtain the confidence bands via bootstrapping, we find that the supra-linear model is actually well contained within, a good sign as there were much less assumptions when using the non-parametric regression but they still ended up being similar.

However, the analysis done above for estimating the change in `log P` also provides another story where these two models give rise to different conclusions, with the non-parametric model contradicting the supra-linear model. Hence, the situation is still ambiguous as to whether the supra-linear model is well-supported by this data.