BIOS 6643 Fall 2021
Homework 1
Ian Arriaga MacKenzie

Question 1

I have read and agree to the homework expectations.

Question 2

(a)

Performing a linear regression of the change-score on the intercept yields the mean: -19.54 (95% CI: -26.63, -12.44; p-value: 8.43E-6).

(b)

The t-test for a mean of 0 yields the same results.

(c)

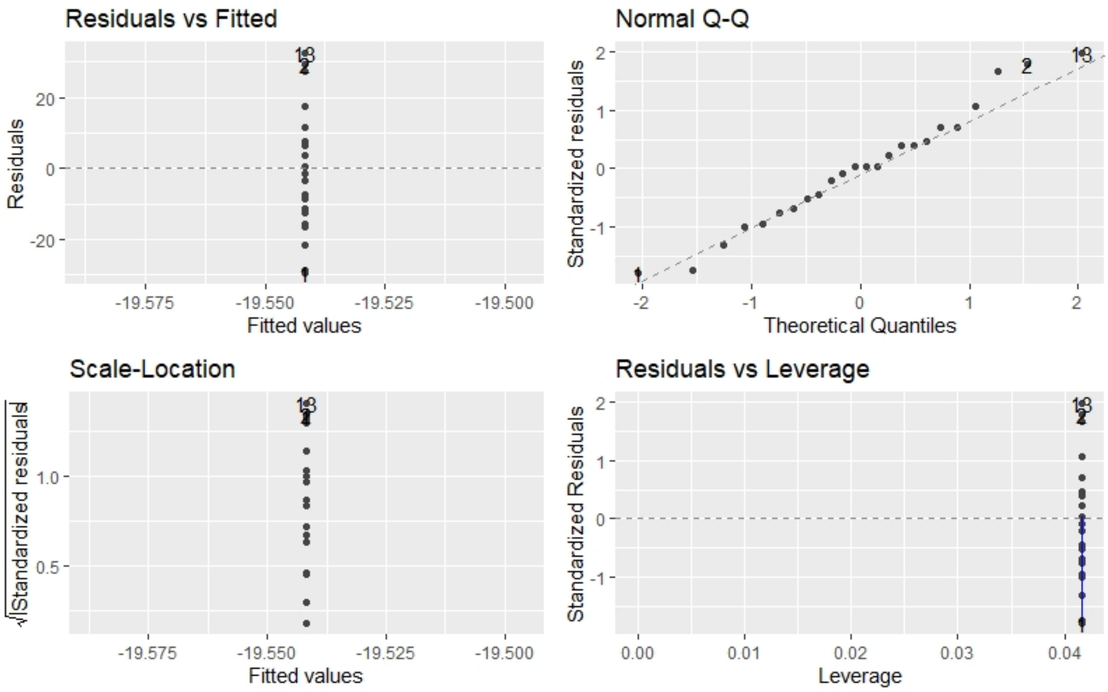
The results for the baseline-as-covariate model is:

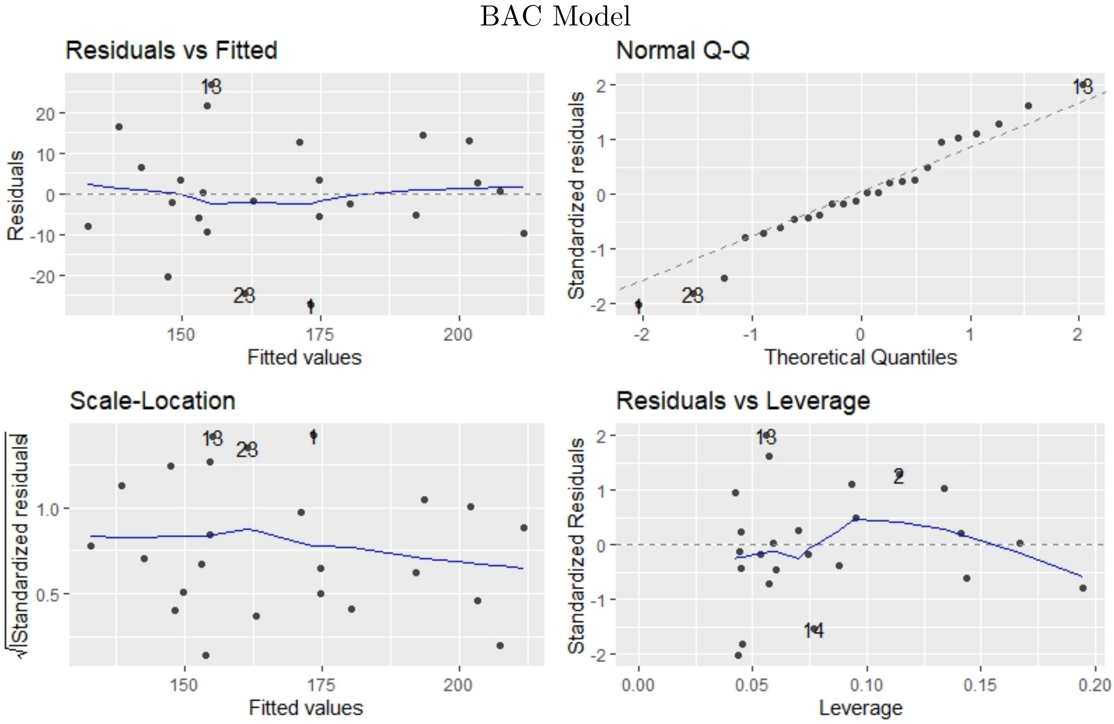
	Estimate	2.5% CI	97.5% CI	P-value
Intercept	37.15	2.85	71.45	0.035
Baseline	0.69	0.51	0.87	5.39E-8

(d)

The change score model is just the t-test for a mean of 0, whereas the baseline-as-covariate model is a true linear model with a slope over the parameter of interest. We can see this from the residuals as well:

CS Model





(e)

The results for the hybrid model are:

	Estimate	2.5% CI	97.5% CI	P-value
Intercept	37.15	2.85	71.45	0.035
Baseline	-0.30	-0.48	-0.12	0.00213

Which can be expressed as:

$$Y_{ChangeScore} = \beta_0 + \beta_1 \cdot X_{Baseline}$$

$$X_{Baseline} = \frac{Y_{ChangeScore} - \beta_0}{\beta_1}$$

The Hybrid and BAC models are the same in that they have the same intercept β_0, but differing slopes on the baseline. The hypothesis of interest is:

$$H_0 : \beta_1 = 0 \qquad\qquad H_1 : \beta_1 \neq 0$$

(f)

I tried to fit a linear mixed effects model, but it kept regressing to the original Hybrid linear model as there were no repeated measures in the data. We return a singular fit matrix, and the estimates are the same as above.

Question 3

(a)

$$E[\epsilon_t] = E\left[\sum_{j=0}^{k} \phi^j Z_{t-j}\right] = \sum_{j=0}^{k} \phi^j \left(E\left[Z_{t-j}\right]\right) = \sum_{j=0}^{k} \phi^j \left(0\right) = 0$$

(b)

$$Cov[\epsilon_t, \epsilon_{t+h}] = E[\epsilon_t \epsilon_{t+h}] - E[\epsilon_t]E[\epsilon_{t+h}] = E[\epsilon_t \epsilon_{t+h}]$$

$$\epsilon_{t+h} = \phi\epsilon_t + Z_{t+h} \Rightarrow E[\epsilon_t \epsilon_{t+h}] = E[\epsilon_t\left(\phi\epsilon_t + Z_{t+h}\right)] = E[\phi\epsilon_t^2 + Z_{t+h}\epsilon_t] = \phi E[\epsilon_t^2] + E[Z_{t+h}\epsilon_t]$$

$$Z_{t+h} \perp Z_{t-j} \Rightarrow E[Z_{t+h}\epsilon_t] = E\left[Z_{t+h}\sum_{j=0}^{k}\phi^j Z_{t-j}\right] = E[Z_{t+h}]E\left[\sum_{j=0}^{k}\phi^j Z_{t-j}\right] = 0$$

$$Var[e_t] = E[e_t^2] - E[e_t]^2 = E[e_t^2] \Rightarrow \phi E[e_t^2] = \phi Var[e_t]$$

$$Var[e_t] = Var\left[\sum_{j=0}^{k}\phi^j Z_{t-j}\right] = \sum_{j=0}^{k}\phi^{2j} Var[Z_{t-j}] = \sum_{j=0}^{k}\phi^{2j}\sigma^2 = \sigma^2 \underbrace{\sum_{j=0}^{k}\phi^{2j}}_{\text{Geometric Series}} = \frac{\sigma^2}{1-\phi^2}$$

$$Cov[\epsilon_t, \epsilon_{t+h}] = \frac{\phi\sigma^2}{1-\phi^2}$$

(c)

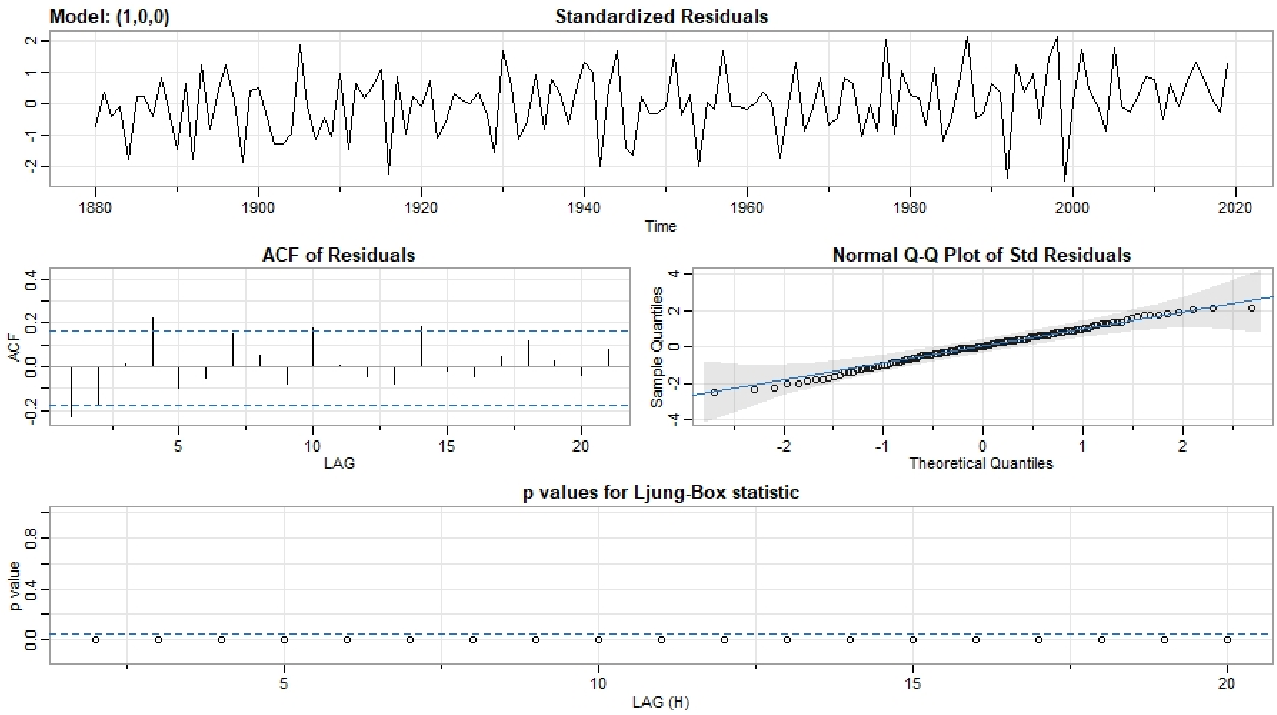
$$Corr[\epsilon_t, \epsilon_{t+h}] = \frac{Cov[\epsilon_t, \epsilon_{t+h}]}{\sigma_t \cdot \sigma_{t+h}} = \frac{Cov[\epsilon_t, \epsilon_{t+h}]}{\sqrt{Var[e_t]}\sqrt{Var[e_{t+h}]}} = \frac{\frac{\phi\sigma^2}{1-\phi^2}}{\sqrt{\frac{\sigma^2}{1-\phi^2}}\sqrt{\frac{\sigma^2}{1-\phi^2}}} = \phi$$

(d)

Yes $\{e_t\}$ is a stationary process. The expectation is 0 and therefor constant. The autocovariance function is also finite, and so this is has weak-sense stationarity.

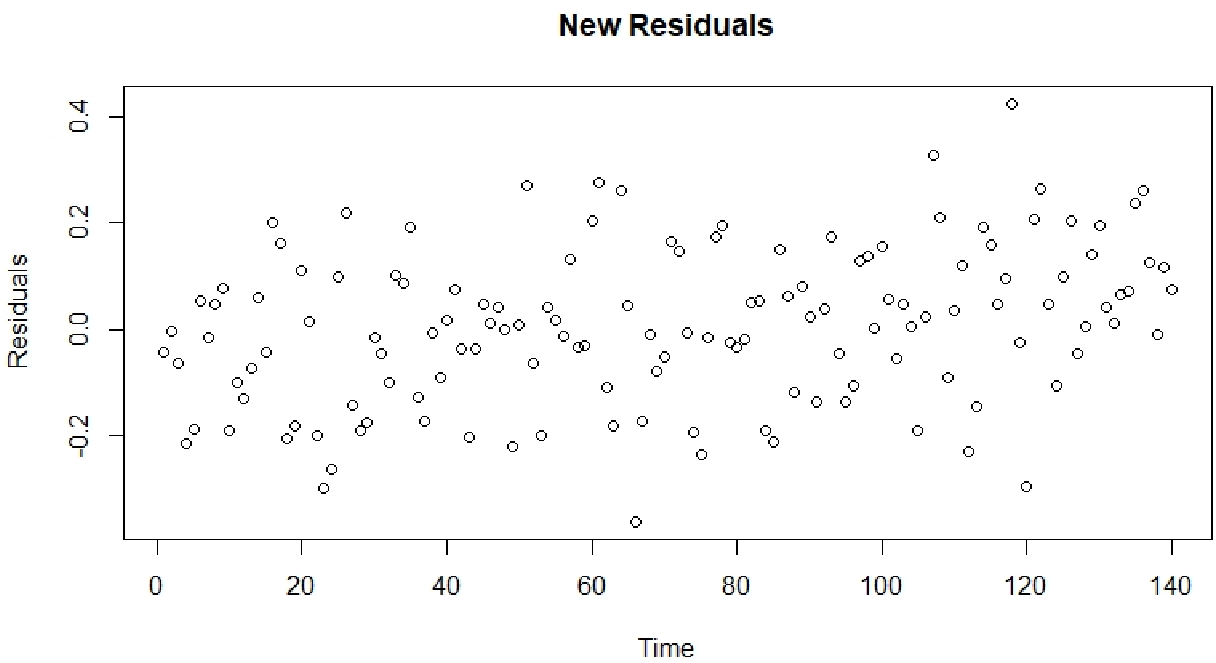
Question 4

(a)



Looking at the residuals we see roughly equal variance across the time series, however we do see a slight trend upward over time instead of a mean of 0 across the sample. The residuals are normal, however the autocorrelation function and Ljung-Box statistic show that our model could be fit better.

(b)



Here we see the residuals have centered around 0 more, and the upward trend in time has decreased slightly.

(c)

Fitting the AR(1) time series model, we get the covariates:

	ar1	xmean
Estimate	0.9484	0.1411
Std. Error	0.0286	0.1795

Which gives us the model:

$$x_t = 0.1411 - (0.1411 \cdot 0.9484) + 0.9484 \cdot x_{t-1} + w_t = 0.00728 + 0.9484 \cdot x_{t-1} + w_t$$

and the increase per decade would be:

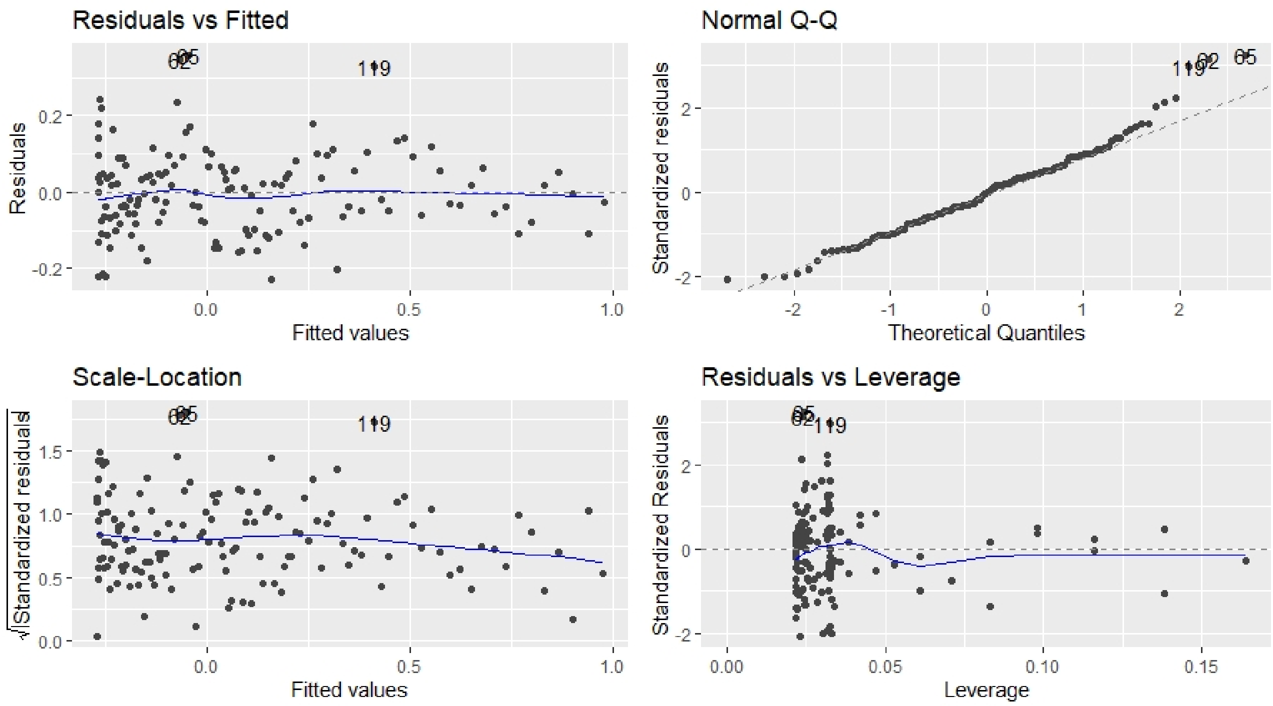
$$\frac{0.9484}{14 \text{ Decades}} = 0.0677 \text{ Degrees per Decade}$$

(d)

Fitting a linear model with a polynomial trend for time yields a slightly better fitting model by LL/AIC:

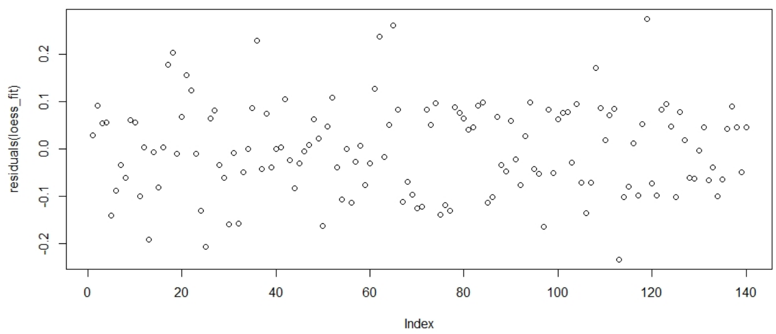
	Log Likelihood	AIC
AR(1)	96.05	-186.11
Polynomial Time	110.91	-209.83

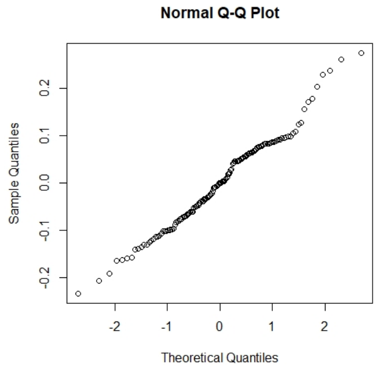
And examining the residuals yields no obvious problems aside from some slight heteroscedasticity:



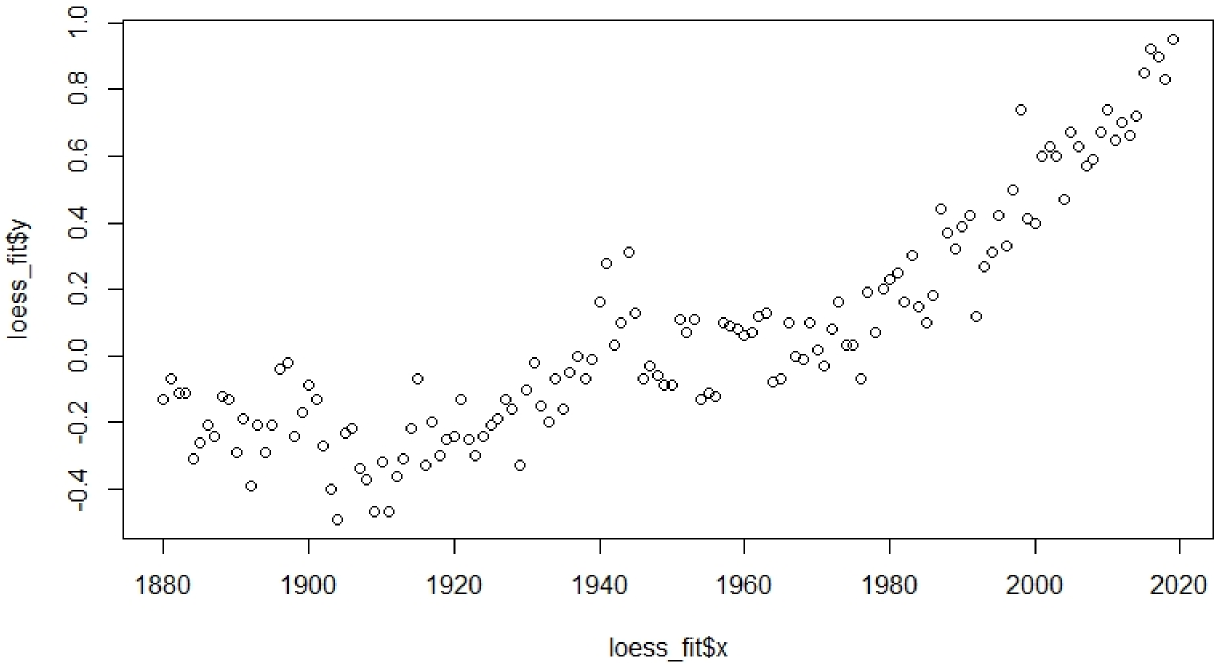
(e)

Fitting a LOESS model:





The residuals and Q-Q plot look good, and the fit seems to indicate an upward trend in time:



Based on the diagnostic plots, either the polynomial model or loess model would be adequate, however I would not use the AR(1) model for this data.