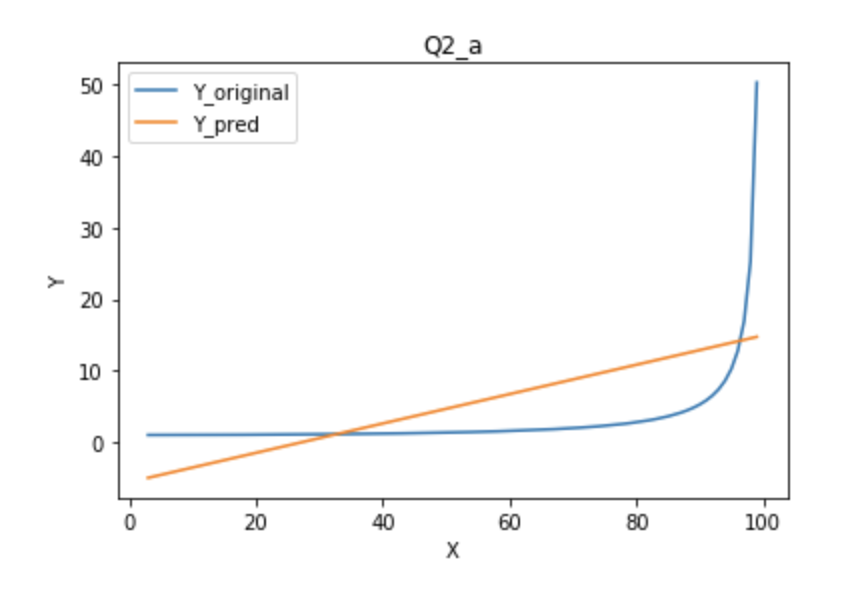
**Question – 2**

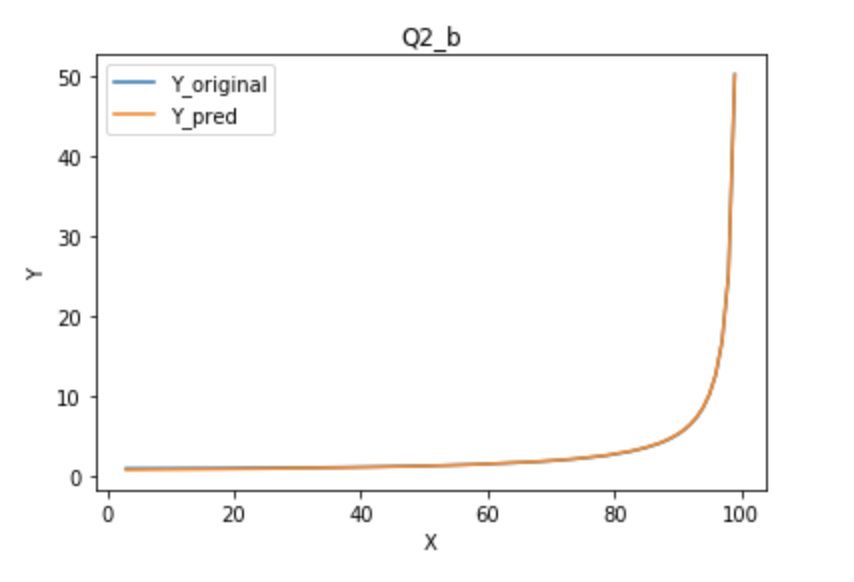
**Part(a)**



SSE: 21795.2971512

MAPE: 167.349490875

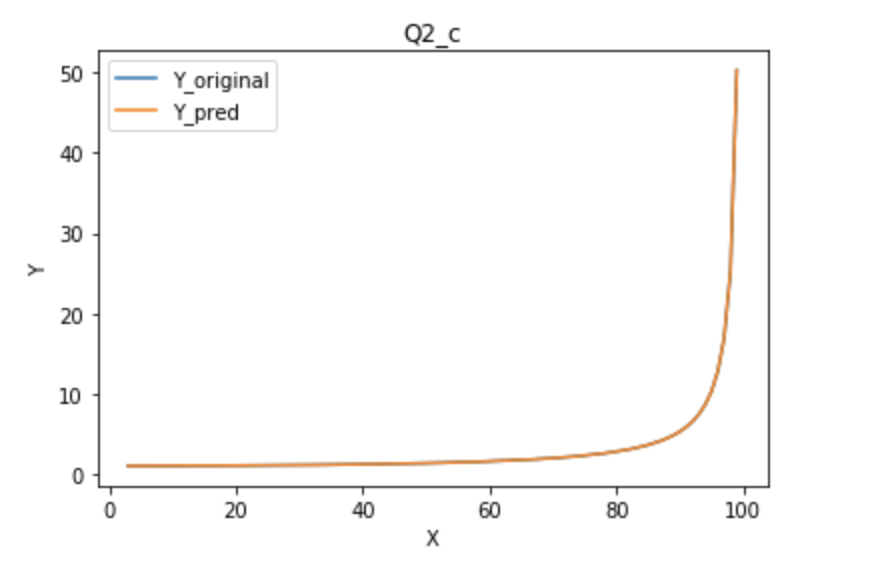
**Part (b)**



SSE: 0.738531163301

MAPE: 2.53111398665

**Part (c)**



SSE: 1.23090292921e-07

MAPE: 0.000806031498868

Function used – 1/(1-(X/100)^2) (squaring X/100 term)

**Question – 4**

**Part (a)**

Coef of FG% = 165.62434772

Coef of TRB = 0.68659874

Mean value of FG% = 0.45

Mean value of TRB = 44

Mean value of PTS = 105

Contribution of TRB = (44\*0.68)/105 = 0.2849

TRB contributes to 28% of points prediction

SSE = 219.92094440752467

MAPE = 1.9262952101991986

**Part (b)**

Coef of FG% = 162.50582185

Coef of TRB = 2.73594995

Coef of ORB = -2.11125195

Coef of DRB = -1.98843192

Contribution of TRB = (44\*2.73)/105 = 1.144

TRB contributes to 114% of points prediction

This happens because ORB and DRB have negative contribution.

SSE = 219.42900790556914

MAPE = 1.9123847664110871

However, there is no change in SSE in this case. Thus addition of new variables ORB and DRB does not affect the error. We further calculated the correlation between different variables -

ORB, DRB are highly correlated with TRB

Thus, we can say that the independent variables are multicollinear. This causes the weights to change erratically.

But the error remains constant**.**

**Part (c)**

Part (i) :

SSE: 139.466218

MAPE: 1.529443

Part (ii) :

SSE: 199.142764

MAPE: 1.861052

Part (iii) :

SSE: 680.809330

MAPE: 3.927207

**Part (d)**

In order to ensure that least square estimators are best linear unbiased estimator for a population, OLS assumes that the standard deviation of error terms will be constant (Homoscedastic).

As seen from the plots, variance of residuals increases with the value of predicted PTS. This violates the assumptions of OLS technique. Thus, for the given data, OLS is not the best linear unbiased estimator.

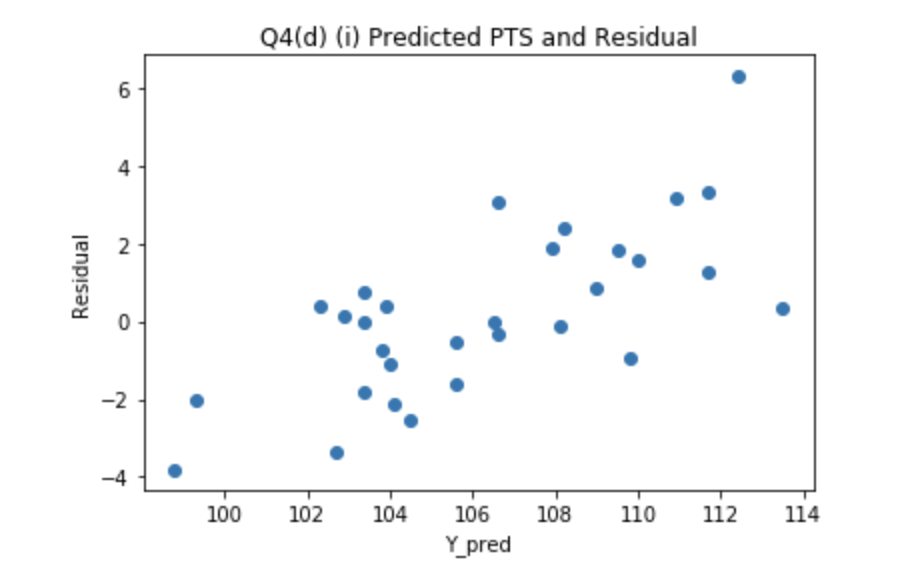
Further, we calculated the variance of residuals in all 3 cases.

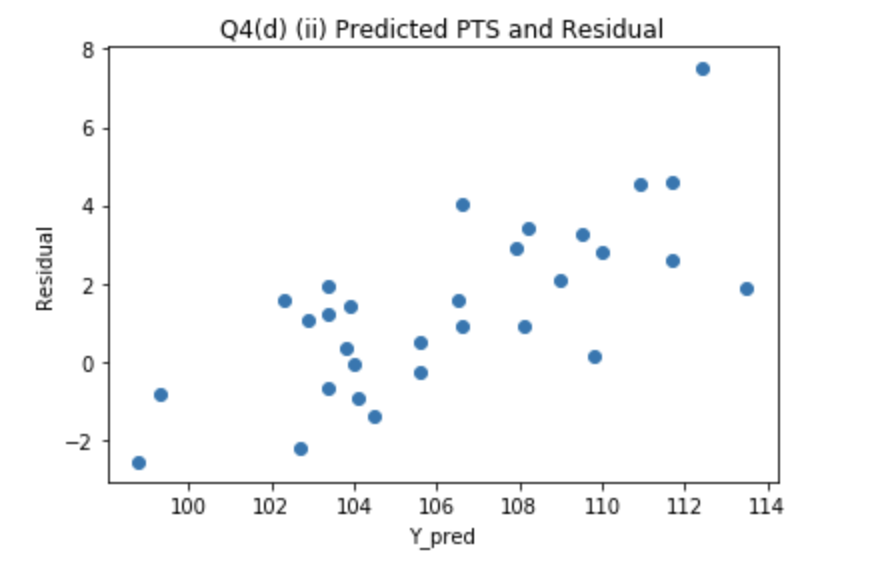
Variance of residuals on training 2017 data: 4.597430

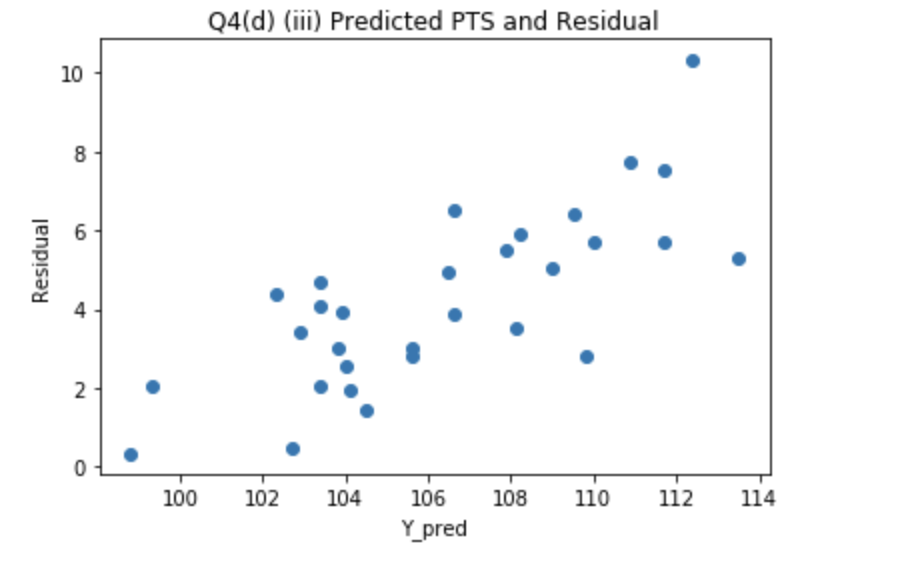
Variance of residuals training 2016-2017 data: 4.633787

Variance of residuals training 2010-2017 data: 4.809462

Variance of residuals increases as we are providing more data.

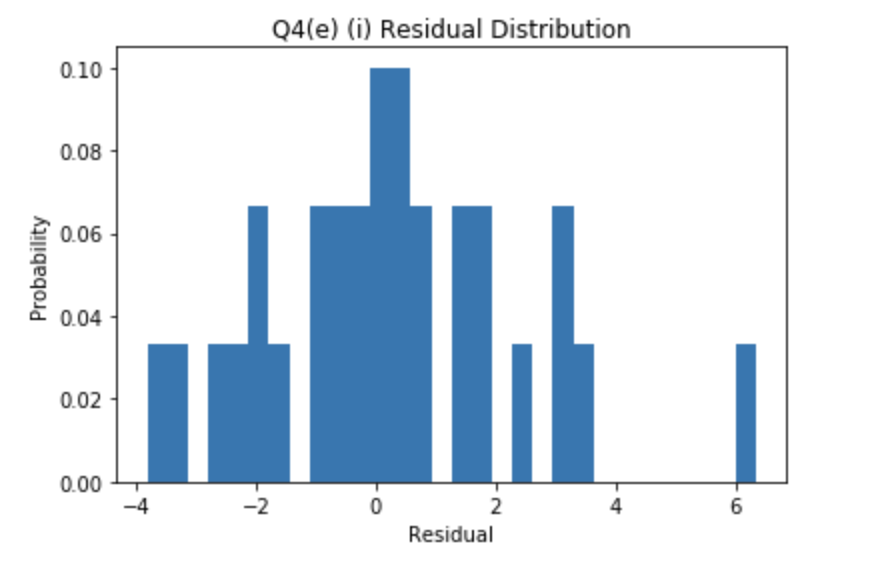
****

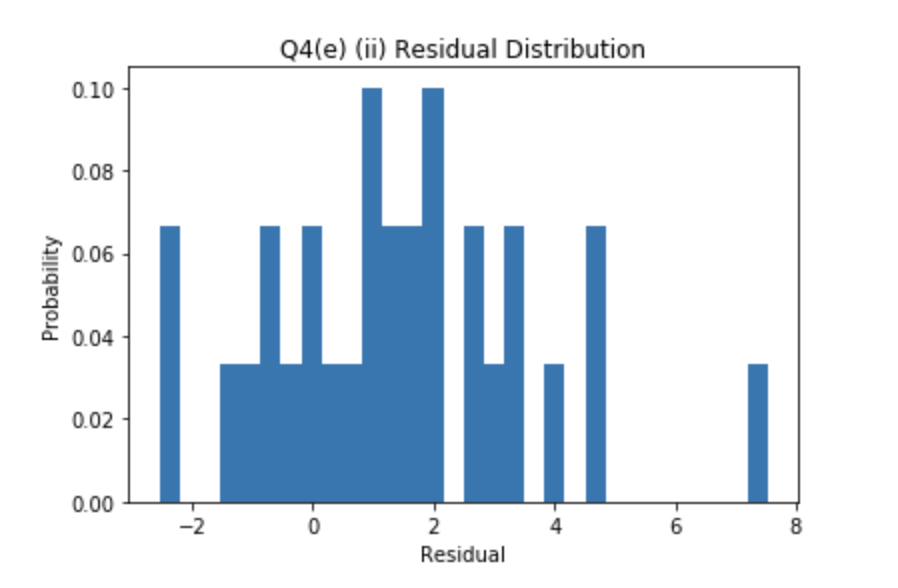
****

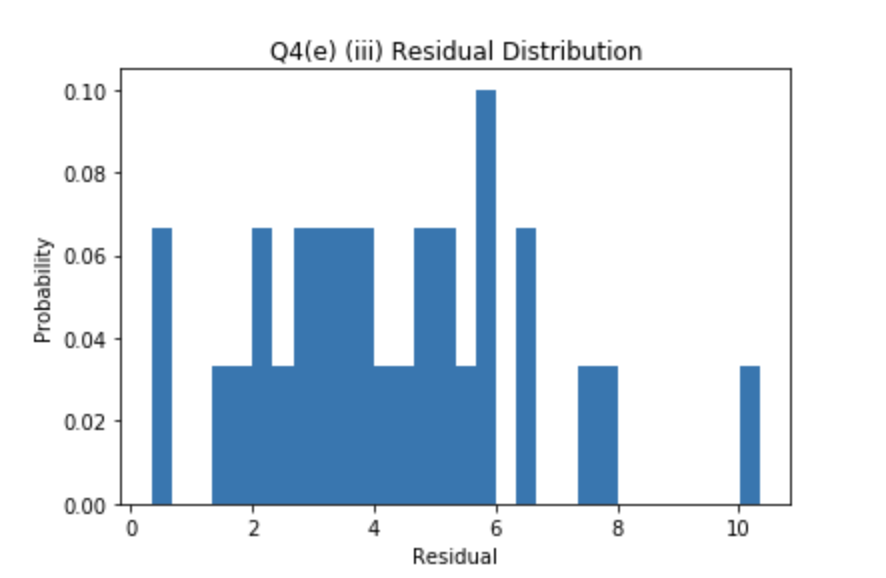
****

**Part (e)**

Bins = 30

****

****

****

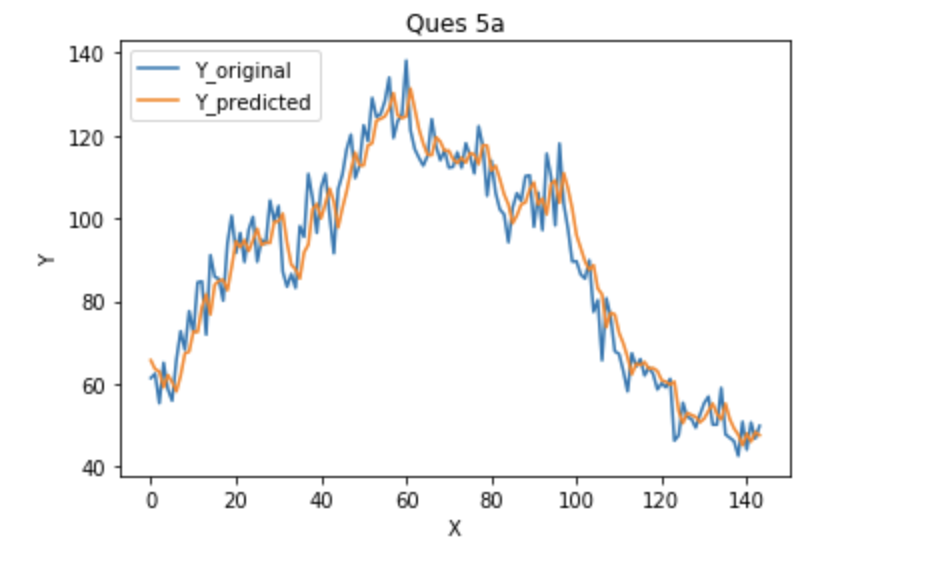
Our histogram of residuals suggests that the residuals are not normally distributed. The residuals have erratic probability distribution. There are many extreme outliers in between in all the 3 plots.

Normal distributions are very common in statistics. Theorems like Central Limit theorem result in normal distribution from linear combination of non normal distributions as well. Further, calculating the moments of normal distribution is quite easy. Transformation property of normal distribution also results in a normal distribution. Normal distribution also help in providing easy to interpret confidence intervals. This provides us an ability to estimate a parameter with a given confidence value. Thus, in linear regression, we assume that residuals are normally distributed( expectation of residual is 0 and variance is constant).

However, in our example, residuals does not come as normally distributed.

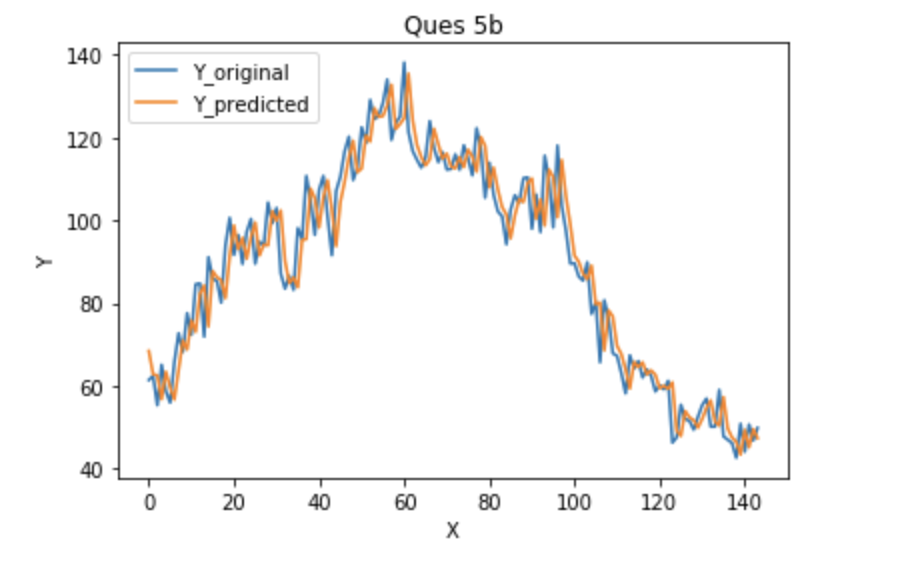
**Question – 5**

**Part(a)**

****

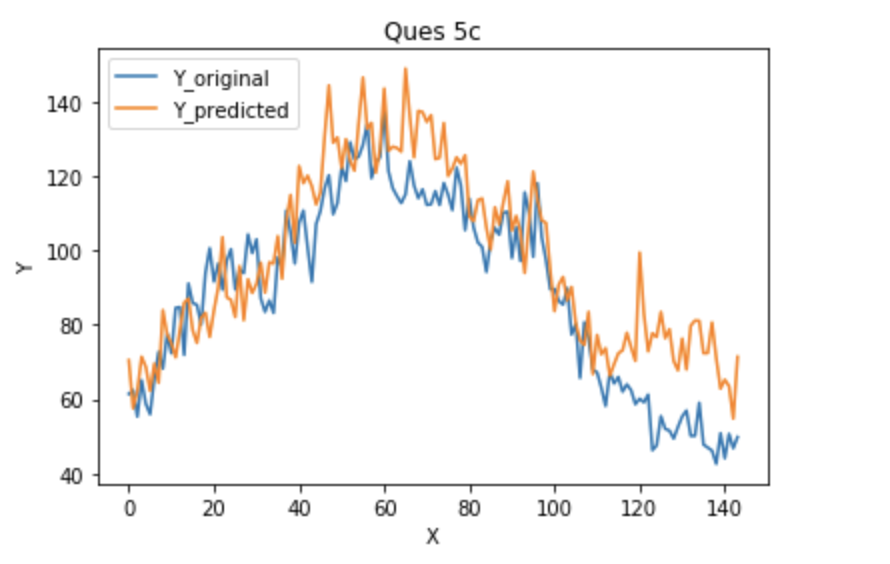
Average Error percent is 6.73759443495

**Part(b)**

****

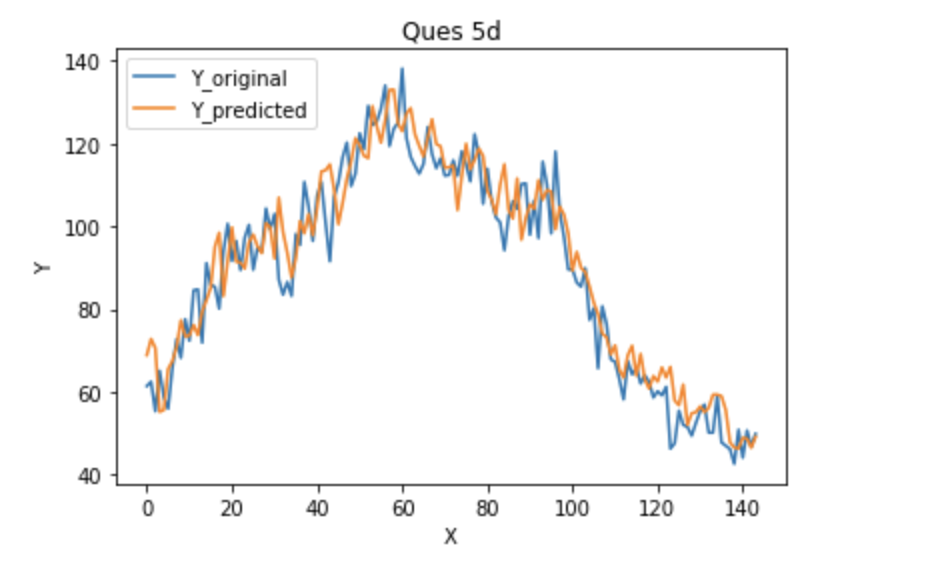
Average Error percent is 6.92667125341

**Part (c)**

****

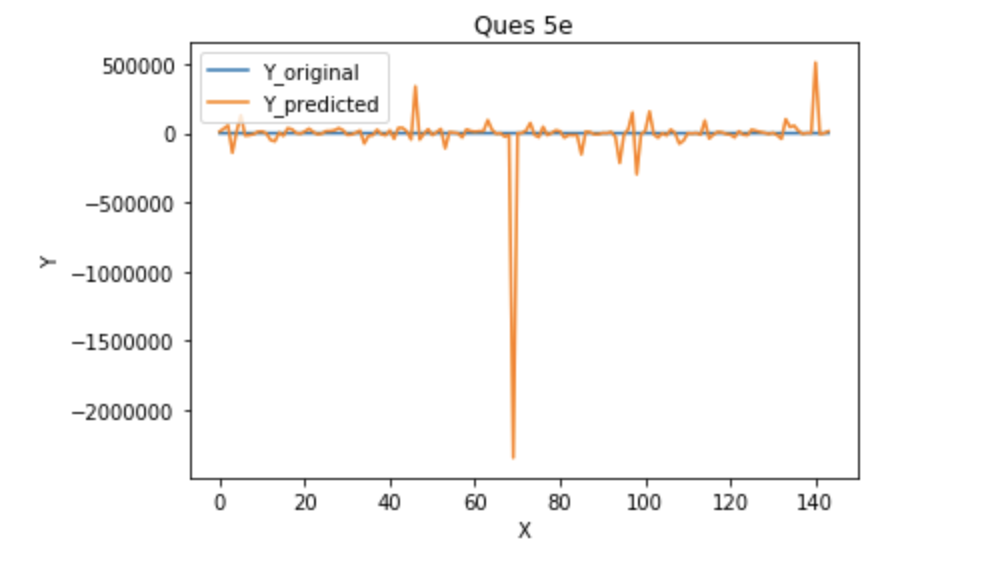
Average Error percent is 16.0782035018

**Part (d)**

****

Average Error percent is 7.5326783364

**Part (e)**

****

Average Error percent is 57369.7568659

**Part (f)**

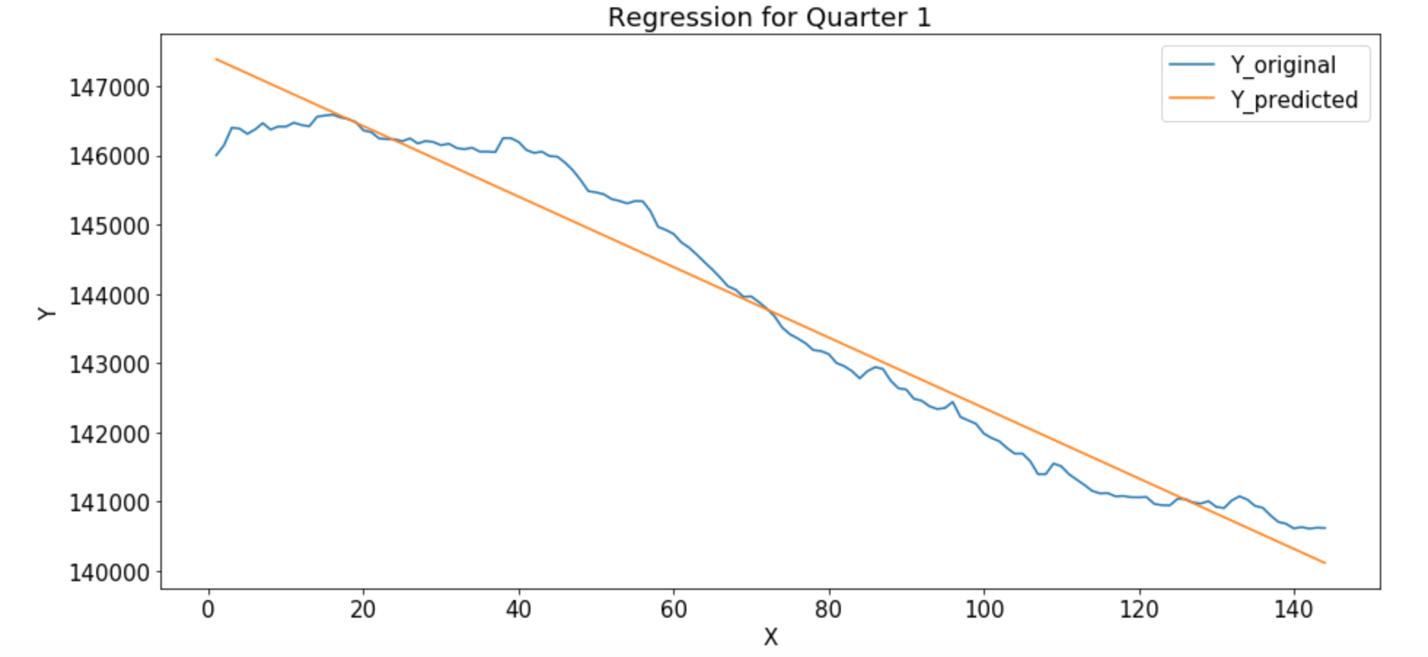
We conclude EWMA with alpha = 0.5 fits the data best.

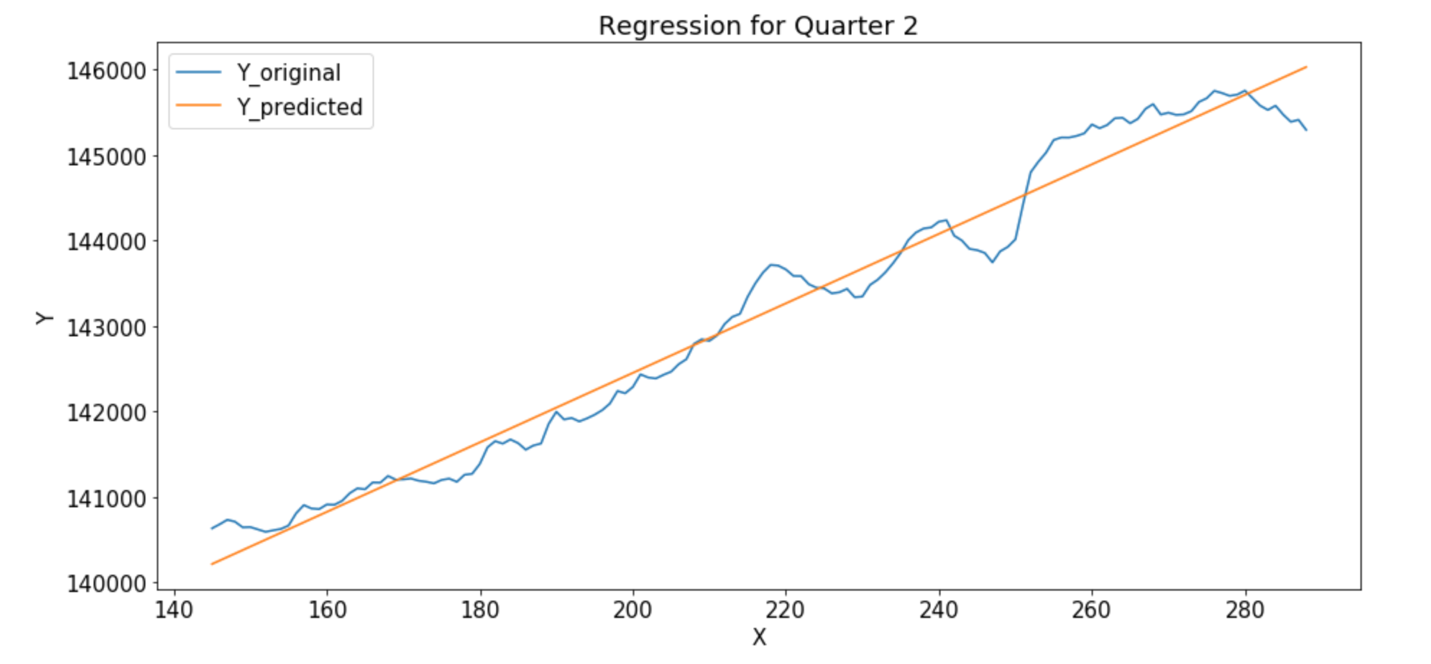
Linear regression also fits data well.

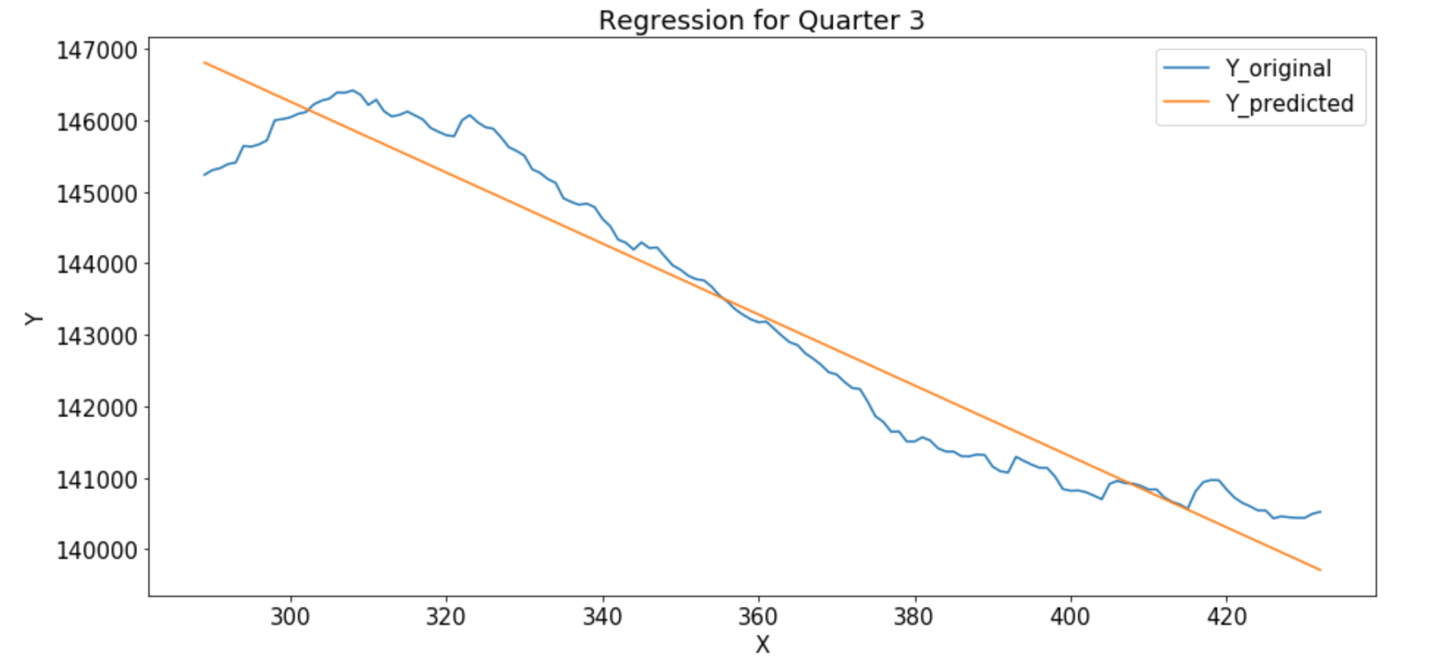
Seasonal has quite a lot of variation from the true value and fits the data worst among all the 3.

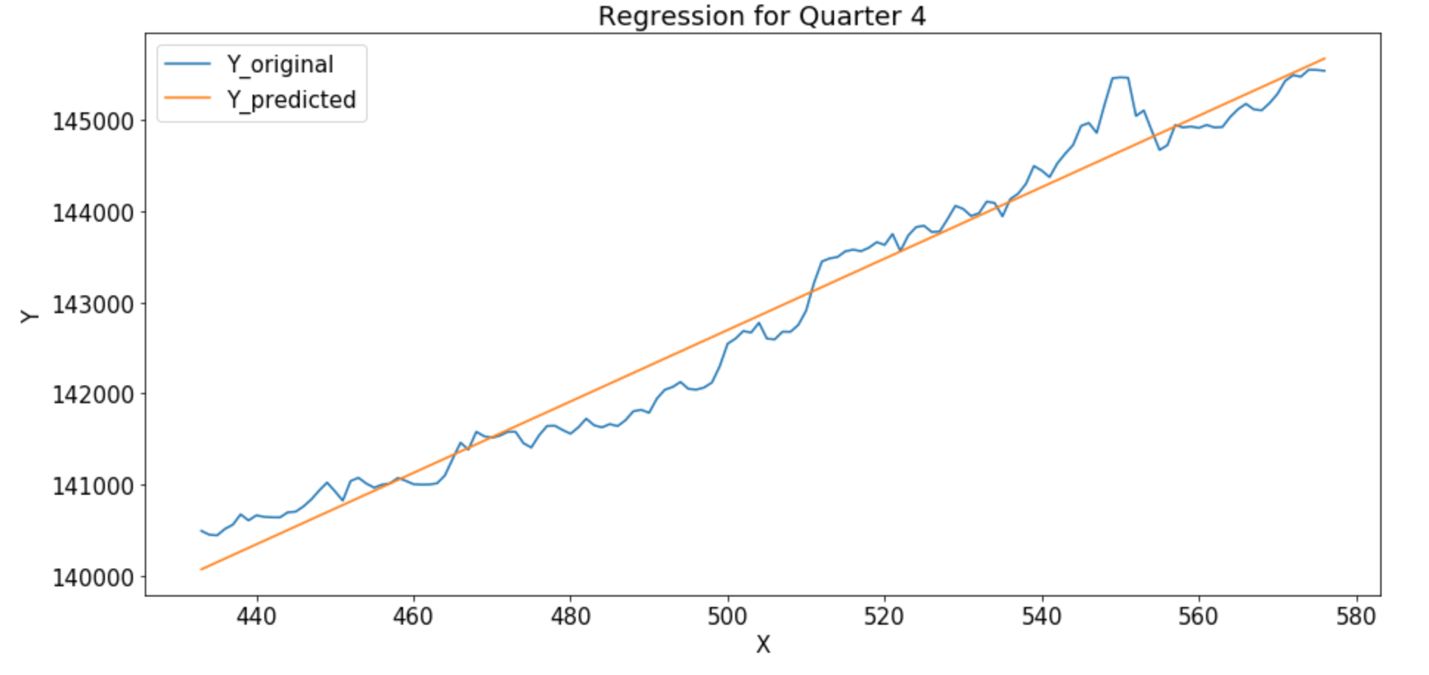
**Question – 6**

**Part (a)**

****

****





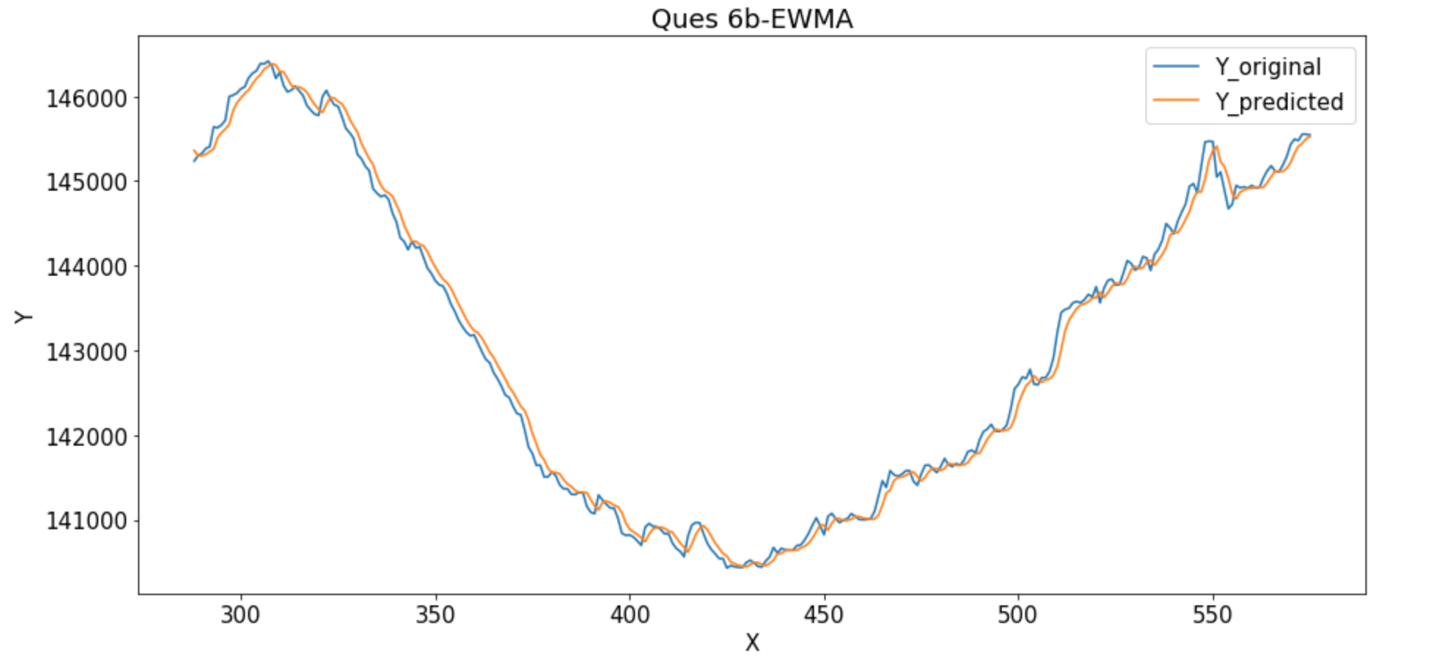
SSE for Quarter 1 28176831.901159

SSE for Quarter 2 10656145.204724

SSE for Quarter 3 43197449.811269

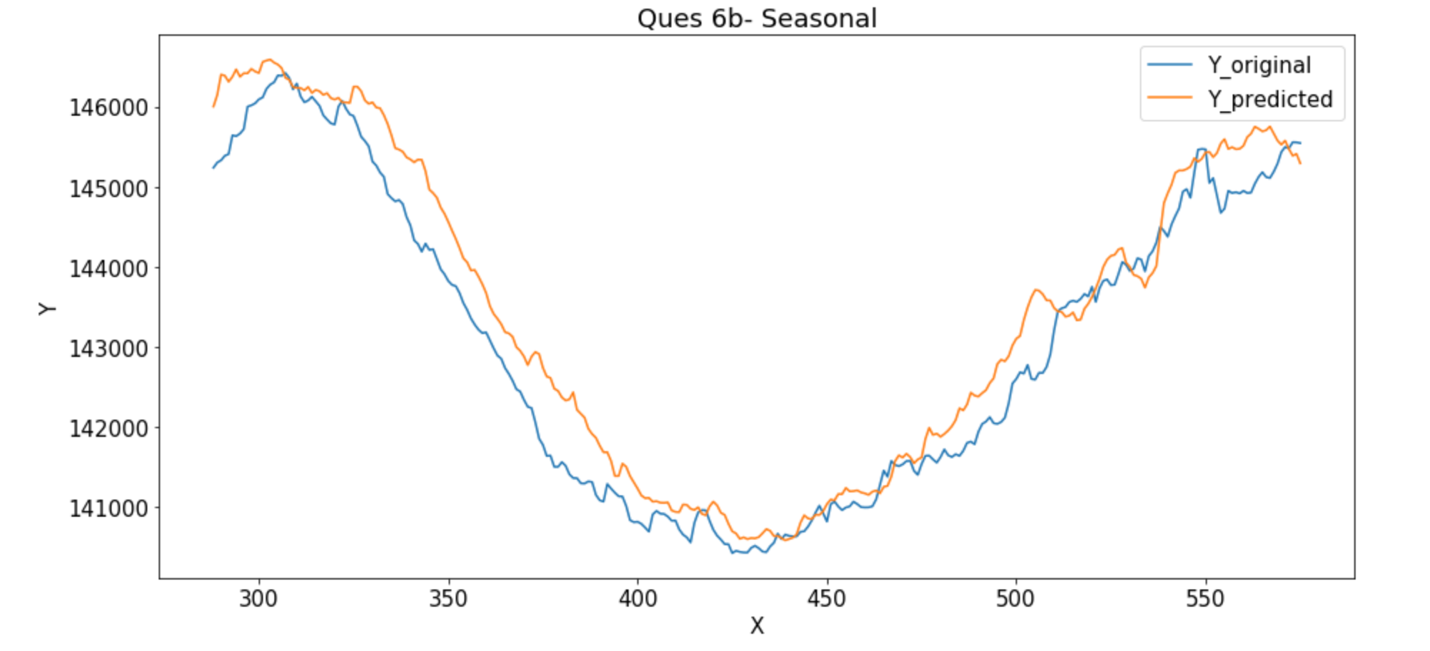
SSE for Quarter 4 10948958.734072

**Part (b)**

****

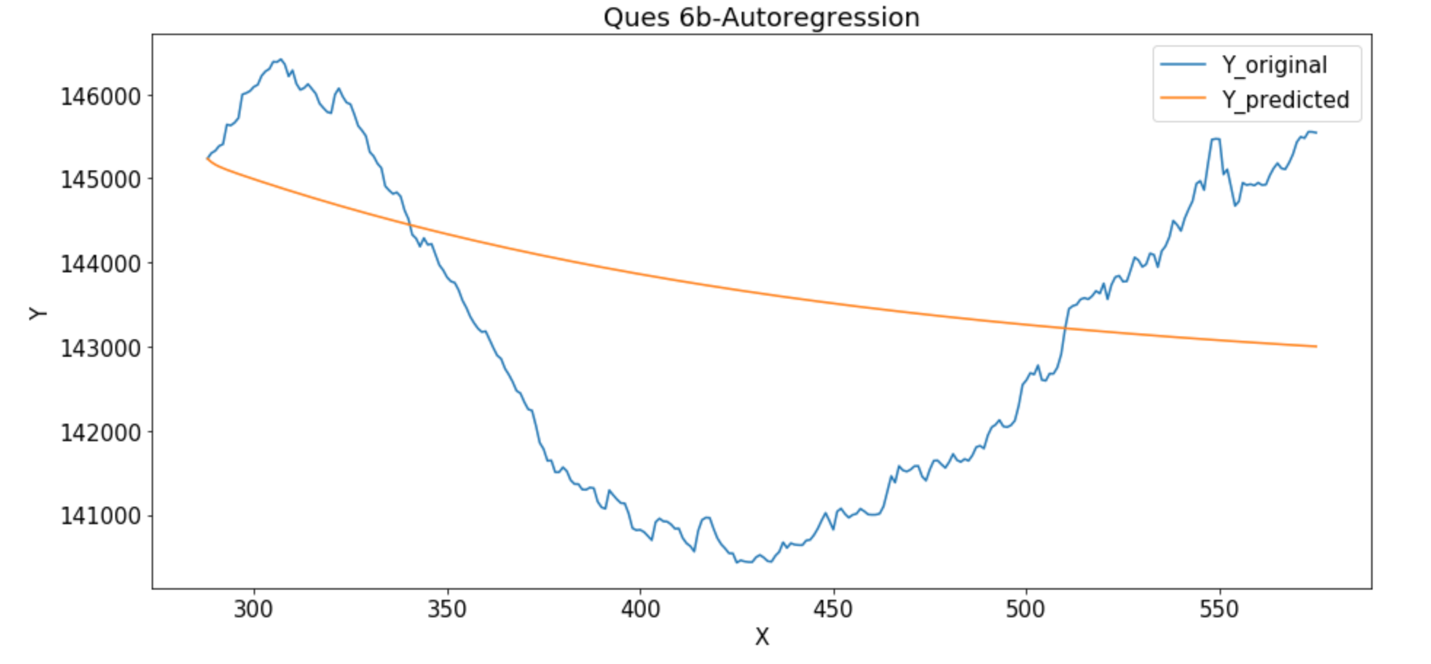
Average Error is 112.57530569648468

Average Error percent is 0.07852972594855746

****

Average Error is 403.86805555555554

Average Error percent is 0.2819586242684633

****

Average Error is 1696.9368242402622

Average Error percent is 1.1919321066834467