***Team 4:***

*Adarsh Goyal*

*Anand Deshmukh*

*Jia Dai*

*Meena Kewlani*

*Shan Lin*

*Yash Ambegaokar*

10/4/2017

Team 4 – Assignment 6

Advanced Business Analytics – MGMT 67200

# About the dataset:

## Code:

Title 'Final Project Crude Oil Prices';

**data** Crude\_Prices;

input Year Month Prices;

time = \_n\_;

datalines;

1983 4 30.50

1983 5 30.16

1983 6 30.96

1983 7 31.59

1983 8 31.88

1983 9 31.24

1983 10 30.40

1983 11 29.86

1983 12 29.24

1984 1 29.65

1984 2 30.05

;

\*Note: First few datalines of the entire input data is shown here

\*Data file is submitted separately as an attachment

**Proc** **gplot** data=Crude\_Prices ;

symbol1 color=blue interpol=join v=dot ;

plot Prices\*time=**1** ;

**run** ;

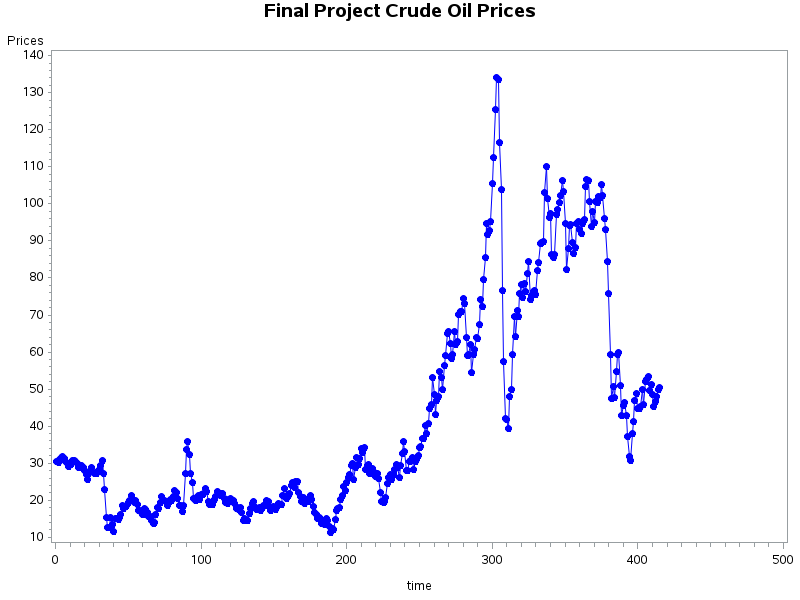


Figure: Crude Oil Monthly Prices vs time graph

Observation:

* We can see a change in variance
* We see a change in mean
* Hence, this model is not stationary and needs to be converted into a stationary model
* We take the log and difference to make the mean constant

/\*\*\*\*\*\*The model is non-stationary and differencing needs to be applied\*\*\*\*\*/

**data** Crude\_Prices\_diff;

set Crude\_Prices;

lnPrices=log(Prices);

diffPrices=dif(lnPrices);

**run**;

**Proc** **gplot** data=Crude\_Prices\_diff ;

symbol1 color=blue interpol=join v=dot ;

plot diffPrices\*time=**1** ;

**run** ;

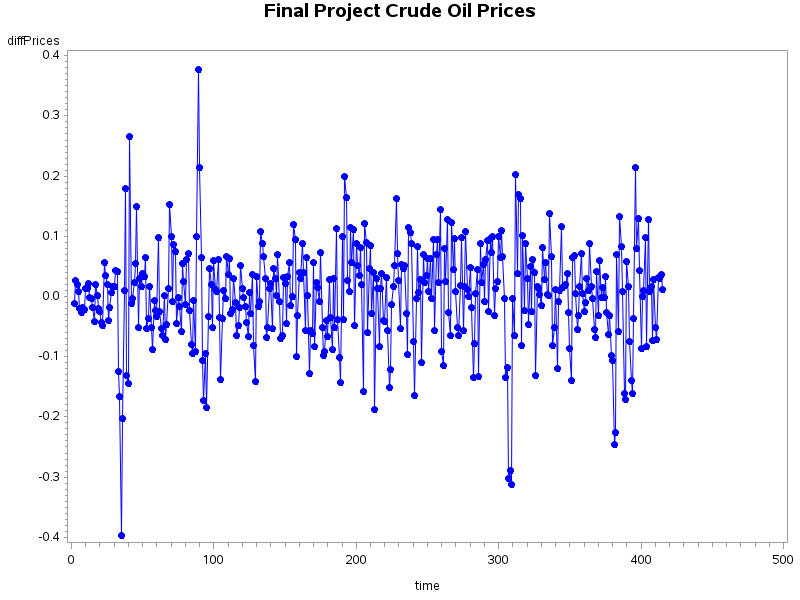


Figure: Model has now become stationary

## A] Identify

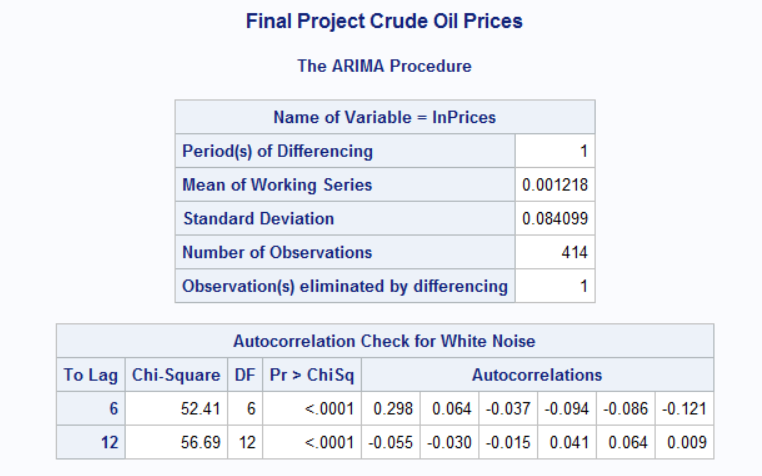
/\*\*\*\*\*\*\*\*\*\*\*\*\*arima modeling to check for ACF and PACF values\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\*\*\*\*We then predict our AR or MA model and check them using regression\*\*\*\*/

**Proc** **arima** data=Crude\_Prices\_diff;

identify var=lnPrices(**1**) nlag=**15** ;

**run**;



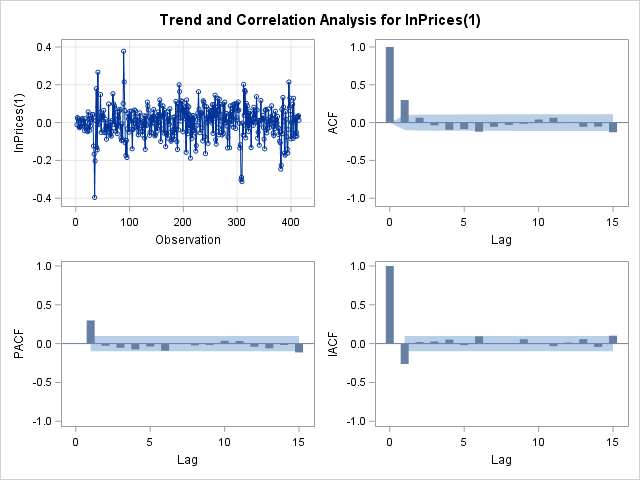


Figure: ARIMA Procedure and Seasonal-Trend Analysis for lnPrices(1) [differencing one]

Observation:

* No seasonal trend observed. Hence no need to take seasonal differencing.
* A tail-off at lag 1 in the ACF graph – indicates a AR model.
* A cut-off at lag 1 in the PACF graph – indicates AR(1) model
* Hence AR(1) models may be good fit tentatively and should be investigated further

To investigate the new models further we create lags and residuals

/\*\*\*\*Creating lags and residuals in the model\*\*\*\*/

**data** Crude\_Prices\_model;

set Crude\_Prices\_diff;

p1=lag(diffPrices);

p2=lag(p1);

p3=lag(p2);

p4=lag(p3);

p5=lag(p4);

**run**;

/\*\*\*\*This will give us residuals to test for adequacy\*\*\*\*/

**proc** **reg** data=Crude\_Prices\_model;

model diffPrices=p1 p2 p3 p4 p5;

output out=Crude\_Prices\_out r=residual;

**run**;

/\*\*\*\*Creates lag for residuals\*\*\*\*/

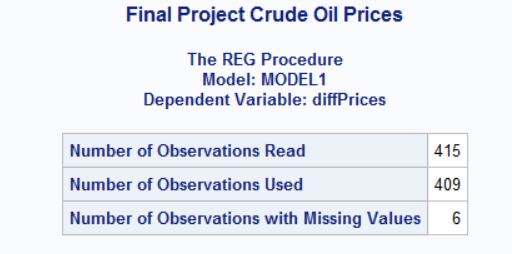
**data** Crude\_Prices\_out;

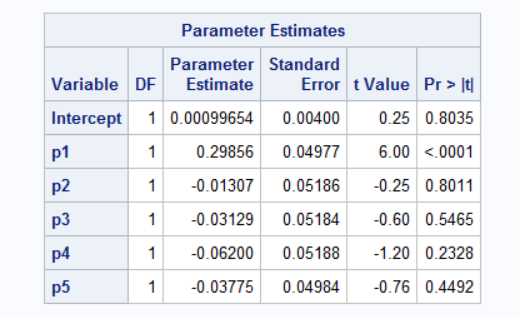
set Crude\_Prices\_out;

r1=lag(residual);

r2=lag(r1);

**run**;





We see only lag 1 (p1) in significant in t-test. We now check for regression and cp criterion

## B] Estimate

To investigate which of the models are most significant we use the Cp criterion. We also analyze p-values through regression results.

/\*\*\*We pick best two regression models after cp criterion for estimation for minimum SSE\*\*\*/

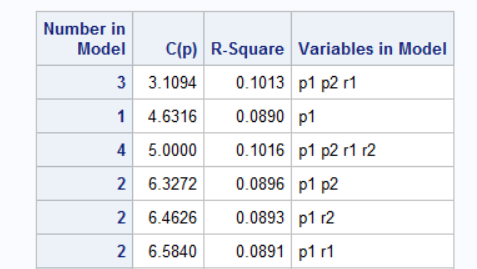
**proc** **reg** data=Crude\_Prices\_out;

model diffPrices = p1 p2 r1 r2/method=cp;

model diffPrices = p1;

model diffPrices = p1 p2 r1;

**run**;



…

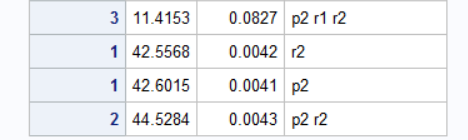


Figure: Cp criterion table: Three best regression models highlighted

## C] Adequacy

Checking ACF and PACF graphs for residuals

**Proc** **arima** data=Crude\_Prices\_out ;

identify var=residual nlag=**15** ;

**run**;

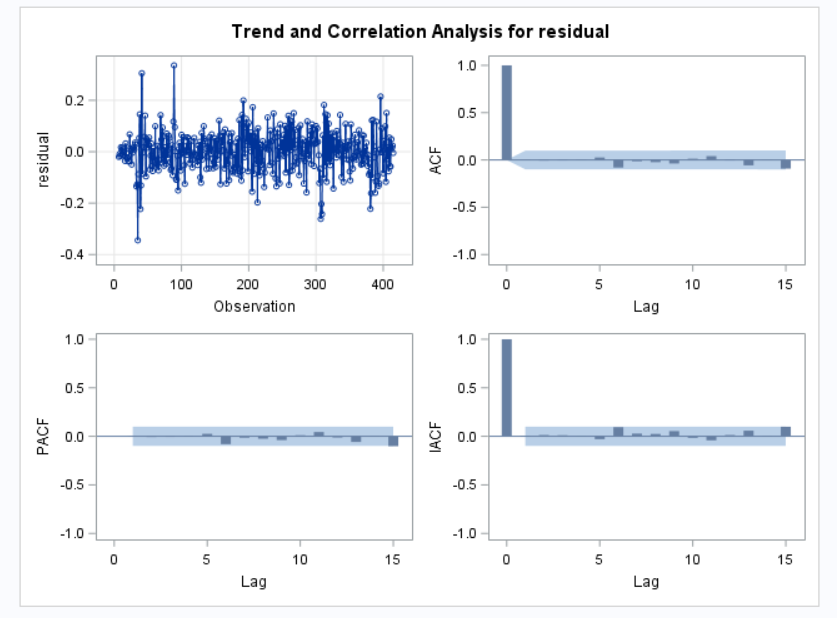


Figure: ARIMA Procedure for residuals

No significant ACF or PACF in residuals

## D] Evaluate

**proc** **arima** data=Crude\_Prices\_model;

identify var=lnPrices(**1**) ESACF;

estimate p = **1** plot;

estimate p = **2** q = **1** plot;

**run**;

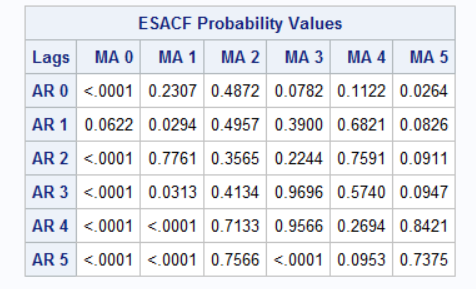


Figure: ESACF method

From ESACF method also we get model1: AR(2) and MA(1) and model2: AR(1) best. We select them tentatively and check for AIC & SBC model.

Option 1: Variables in Model –

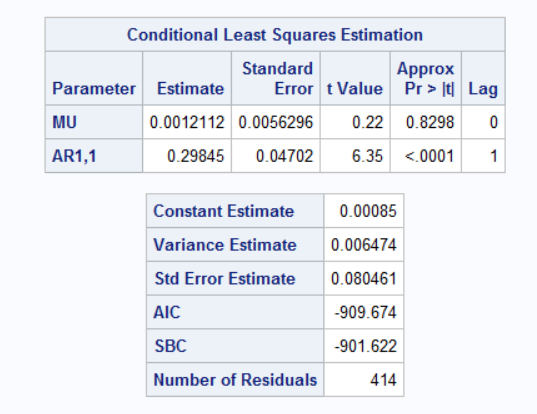
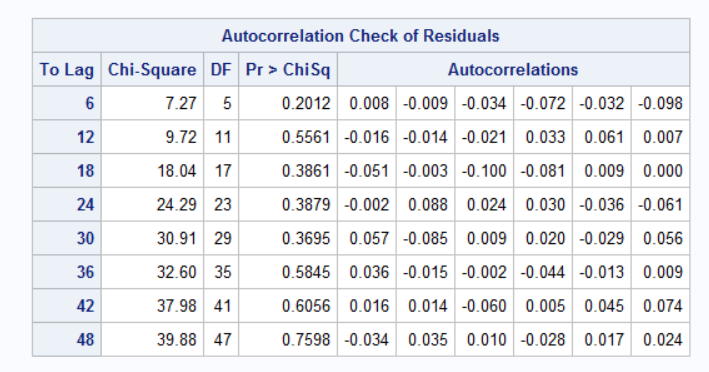


Figure: model characteristics



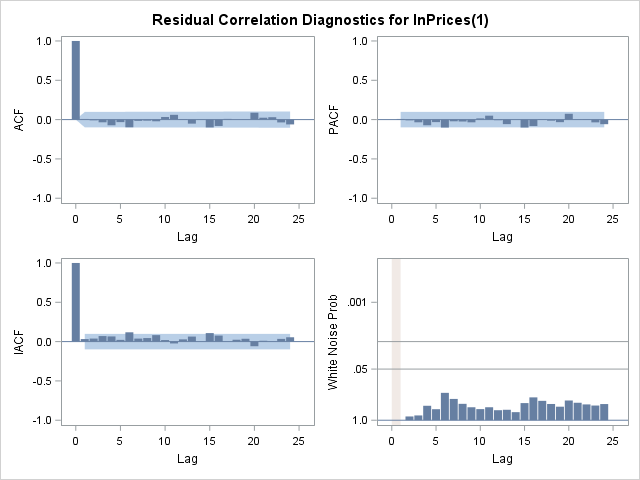


Figure: model residual diagnostics: Residuals are insignificant at 95% confidence interval

Model AR(1) is adequate

Equation for Model AR(1)

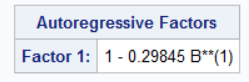


Figure: model

Option 2: Variables in Model –

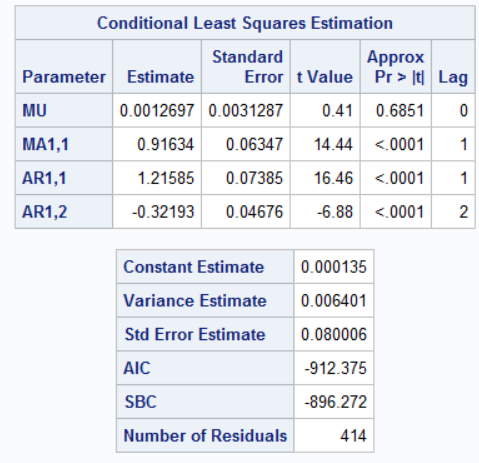
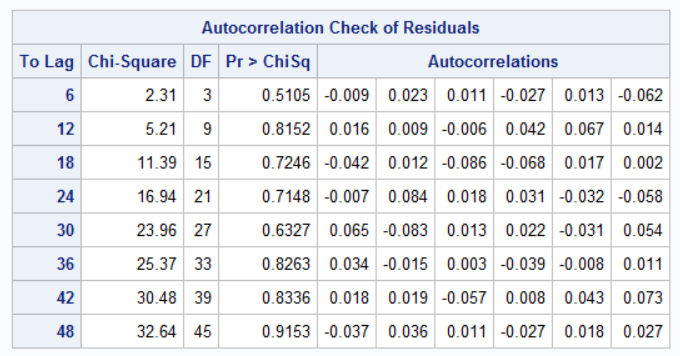


Figure: model characteristics



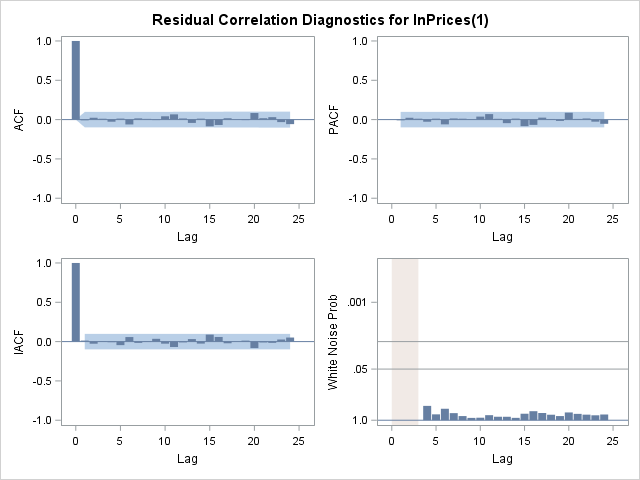


Figure: model residual diagnostics: Residuals are insignificant at 95% confidence interval

Model AR(2) & MA(1) is adequate

Equation for Model AR(2) and MA(1)

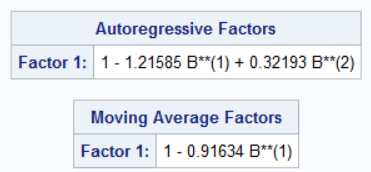


Figure: model

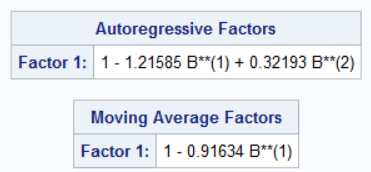
## Comparison of two models

|  |  |  |  |
| --- | --- | --- | --- |
| Option Number | Variables | AIC | SBC |
| 1 |  | -912.375 | -896.272 |
| 2 |  | -909.674 | -901.622 |

The most negative AIC factor is for Option 1: Variables used -

Additionally, the Cp criterion also suggested that the AR(2) & MA(1) model is the most appropriate model.

Hence the AR(2) & MA(1) model is selected:



Equation:

## E] Forecast

**proc** **arima** data=Crude\_Prices\_model;

identify var=lnPrices(**1**) nlag = **15**;

**run**;

estimate p=**2** q =**1** plot;

**run**;

forecast out=future\_prices lead=**6** interval= month;

**run**;

**data** final;

set future\_prices;

price = exp(lnPrices);

forecast = exp(forecast + std\*std/**2** );

l95 = exp( l95 );

u95 = exp( u95 );

time=\_n\_;

**run**;

symbol1 color=blue interpol= none v = star;

symbol2 color=red interpol = join v = dot;

symbol3 color=green interpol=join v = circle;

**proc** **gplot** data = final;

plot price\*time =**1** forecast\*time =**2**/overlay;

**run**;

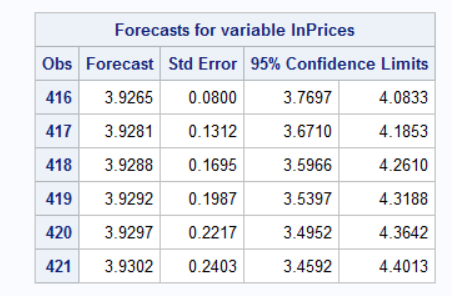
**proc** **gplot** data = final;

where time >**400**;

plot price\*time =**1** forecast\*time =**2** l95\*time=**3** u95\*time=**3**/overlay;

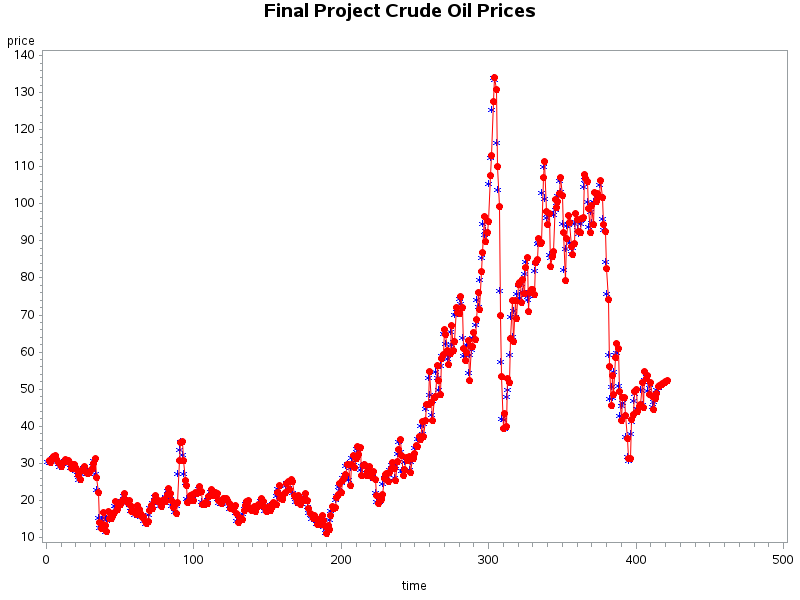
**run**;

**quit**;



Forecast values using model

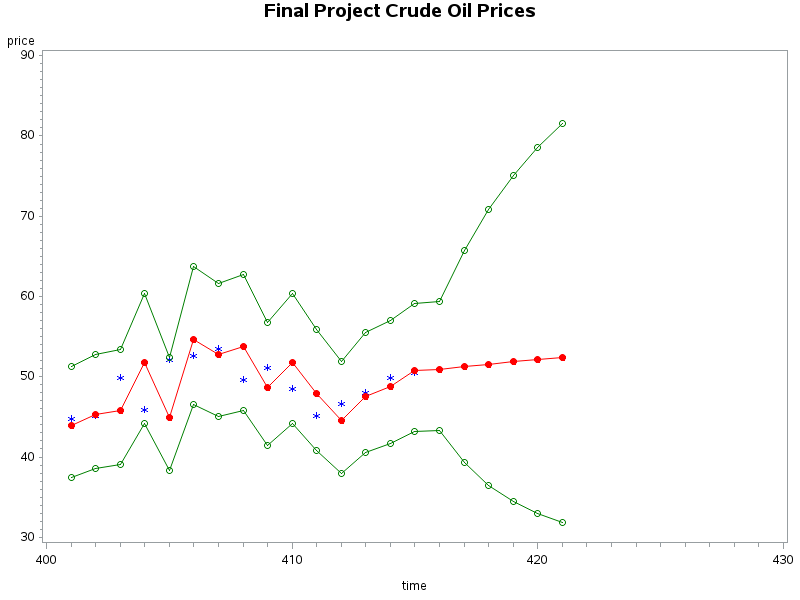
We don’t see much variations in forecasted oil prices for the next six months. That suggests oil prices are likely to be stable for next six months.



FORECAST

HISTORY

Historical and Forecast values using model over the entire dataset



HISTORY

FORECAST

Historical and Forecast values using model after 400th observation with 95% confidence limits interval