Regression Analysis

Simple Linear Regression

Nicoleta Serban, Ph.D.
Professor
School of Industrial and Systems Engineering

Example 1: Testing the Theory of Purchasing Power Parity

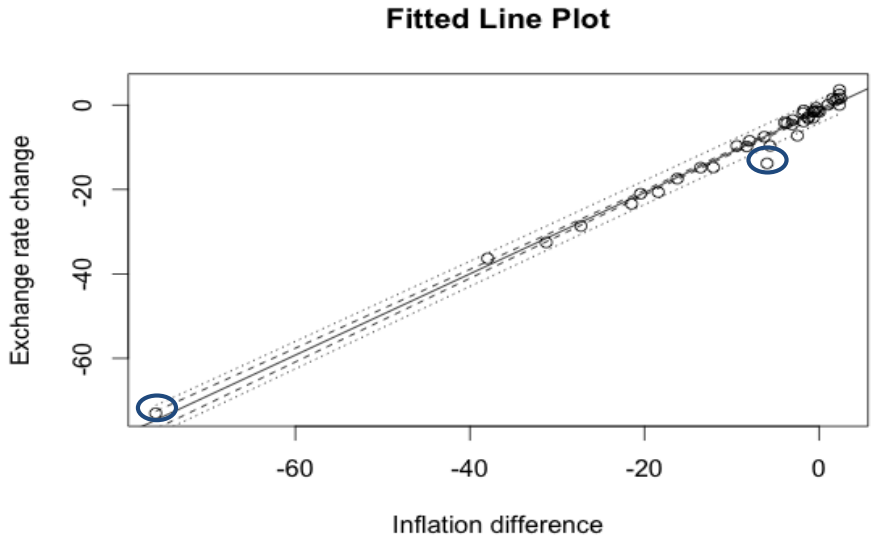
(Part 1)

About This Lesson

Confidence Bands in R

```
# Function for fitted line plot: See ppp-revised.R for this function
#regplot.confbands.fun = function(x, y, confidencelevel=.95, CImean=T, PI=T,
CIregline=F, legend=F){
    #### Modified from a function written by Sandra McBride, Duke University
….}
```

regplot.confbands.fun(Inflation.difference,Exchange.rate.change)



The fitted line plot shows several lines:

- The continuous line is the fitted regression line.
- The wider interrupted line band is the prediction confidence band.
- The narrower interrupted line band is the confidence band.
- The circles correspond to outliers.

Georgia Tech

Confidence and Prediction Intervals

```
# Confidence and prediction intervals for new observation
# Create new data point
    newppp = data.frame(Inflation.difference = c(-0.68))
# Specify whether a confidence or prediction interval
    predict(pppa,newppp,interval=c("confidence"))
        fit         lwr         upr
1 -2.173351   -2.756818   -1.589884
    predict(pppa,newppp,interval=c("prediction"))
        fit         lwr         upr
1 -2.173351   -5.554071    1.207369
## Why are the intervals different?
```

Interpretation of the two intervals:

- The 95% confidence limits of the average exchange rate change for all countries inflation difference equal to -0.68 are (-2.757,-1.590);

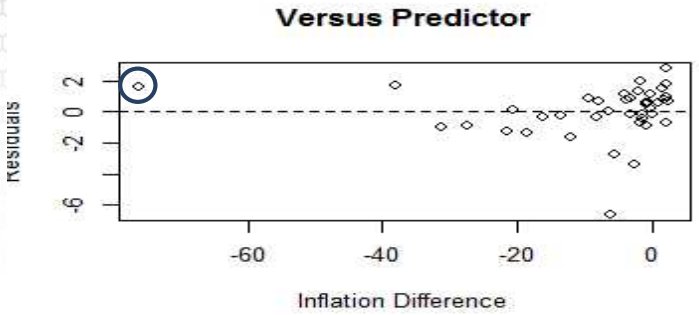
- The 95% confidence limits for the exchange rate change for one country with inflation difference equal to -0.68 are (-5.554,1.207).

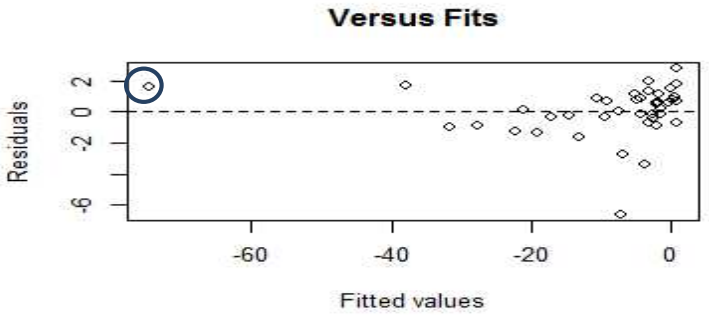
Georgia Tech

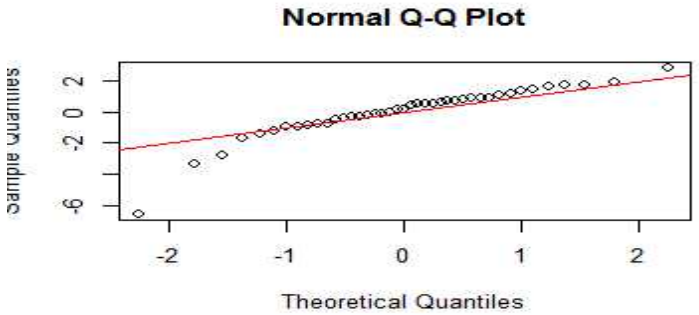
Residual Analysis in R

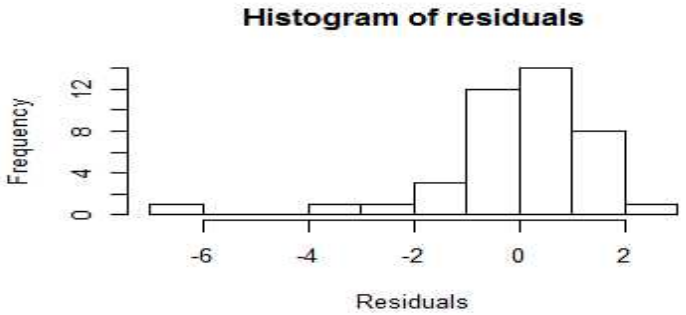
```
par(mfrow=c(2,2))
plot(Inflation.difference, residuals(pppa),xlab="Inflation
Difference",ylab="Residuals",main="Versus Predictor")
abline(h=0,lty=2)
plot(fitted(pppa),residuals(pppa),xlab="Fitted values",ylab="Residuals“, main="Versus Fits")
abline(h=0,lty=2)
qqnorm(residuals(pppa))
abline(0,1,lty=1,col="red")
hist(residuals(pppa),main="Histogram of residuals",xlab="Residuals")
```

Residual Analysis in R









Georgia Tech

Residual Analysis in R

***Leverage Points*:** The isolated point in residual plots is Brazil. Why is Brazil a leverage point?

- Brazil had a period of hyperinflation from 1980 to 1994, a time period during which prices went up by a factor of roughly 1 trillion.

Why do we care about leverage points?

- It can have a strong effect on the fitted regression, drawing the line away from the bulk of the points. It also can affect measures of fit like R-squared and t–statistics.

Influential Points in Regression Analysis

```
##### Repeat Analysis: Omit Brazil #########
## remove the data row corresponding to Brazil
    newppp = ppp[ppp$Country!="Brazil",]
    attach(newppp)
## Fit Linear Regression
    pppn = lm(Exchange.rate.change ~ Inflation.difference)
    summary(pppn)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.37222	0.30517	-4.497	6.31e-05
Inflation.difference	0.99152	0.02626	37.757	< 2e-16

Residual standard error: 1.62 on 38 degrees of freedom
Multiple R-squared: 0.974, Adjusted R-squared: 0.9734

```
## Test whether the slope is equal to 1 (PPP
theory)
    tvalue = (0.9915−1)/ 0.02626
    pvalue = 2*(1-pt(abs(tvalue),38))
```

$\hat{\beta}_0$ = -1.372, se($\hat{\beta}_0$) = 0.305
Statistical significance for β_0:
t-value= -4.497, p-value ≈ 0

$\hat{\beta}_1$ = 0.9915, se($\hat{\beta}_1$) = 0.02626
Test the null hypothesis $\beta_1 = 1$:
p-value = 0.748

We are seeing violations of PPP with respect to intercept only.

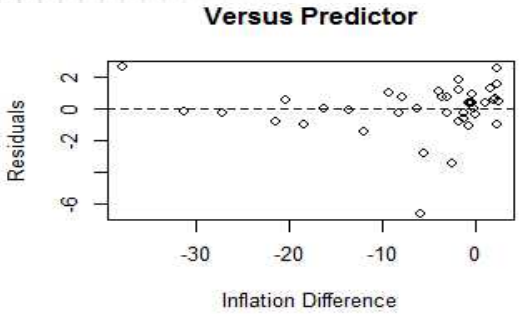
Georgia Tech

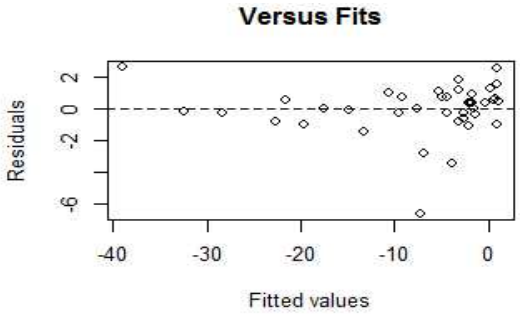
Residual Analysis: Model without Brazil

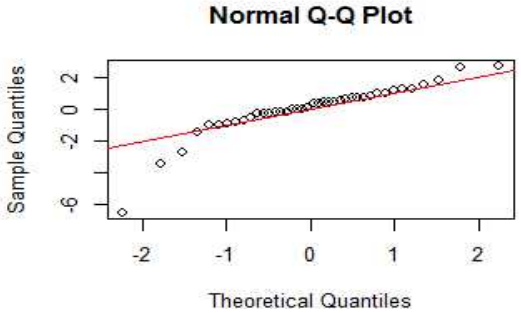
```
par(mfrow=c(2,2))
plot(Inflation.difference, residuals(pppn),xlab="Inflation Difference",ylab="Residuals",
main="Versus Predictor")
abline(h=0,lty=2)
plot(fitted(pppn),residuals(pppn),xlab="Fitted values",ylab="Residuals",main="Versus Fits")
abline(h=0,lty=2)
qqnorm(residuals(pppn))
abline(0,1,lty=1,col="red")
hist(residuals(pppn),main="Histogram of residuals",xlab="Residuals")
```

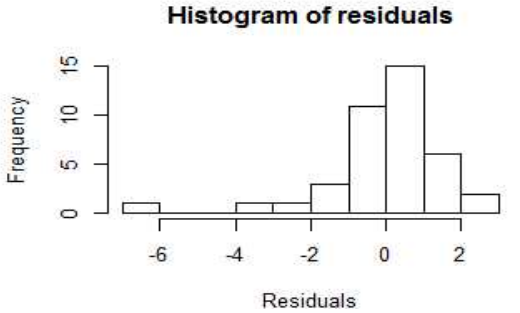
Georgia Tech

Residual Analysis: Model without Brazil









Assumptions:
Linearity: No pattern in the residuals with respect to the predicting variable.
Constant Variance: The variance is higher for higher fitted values. Does not hold.
Uncorrelated Errors: No grouping of the residuals
Normality: Except for the presence of an outlier, it is reasonably symmetric.

Outliers *(observations for which the residual value is away from the range)***:**
The isolated point in the residual plots is Indonesia. Would omitting Indonesia change anything? The strength of the relationship would increase, but so the rejection of PPP.

Testing the Theory of Purchasing Power Parity

Findings:

- Support is decidedly mixed
- Developed countries:
    - Changes in inflation difference do seem to be balanced by exchange rate changes
    - One outlier: Greece
- Developing countries:
    - The case for PPP is considerably weaker;
    - Brazil and Indonesia
- PPP is not robust to unusual economic or political conditions

Summary