Spurious Regression

Zahid Asghar

Spurious Regression

$y_t = a_0 + a_1 z_t + e_t$

where y_t is the dependent variable, z_t is the independent variable, a_0 and a_1 are the parameters, and e_t is the error term. Assumptions of the classical model:

- both the $\{y_t\}$ and $\{z_t\}$ sequences be stationary
- the errors have a zero mean and a finite variance.

In the presence of nonstationary variables, there might be what Granger and Newbold (1974) call a spurious regression.

> A spurious regression has a high R^2 and t-statistics that appear to be significant, but the results are without any economic meaning. The regression output "looks good" because the least-squares estimates are not consistent and the customary tests of statistical inference do not hold.

CASE 1: Both {yt} and {zt} are stationary. The classical regression model is appropriate.

CASE 2: The {yt} and {zt} sequences are integrated of different orders. Regression equations using such variables are meaningless.

CASE 3: The nonstationary {yt} and {zt} sequences are integrated of the same order and the residual sequence contains a stochastic trend. This is the case in which the regression is spurious. In this case, it is often recommended that the regression equation be estimated in first differences.

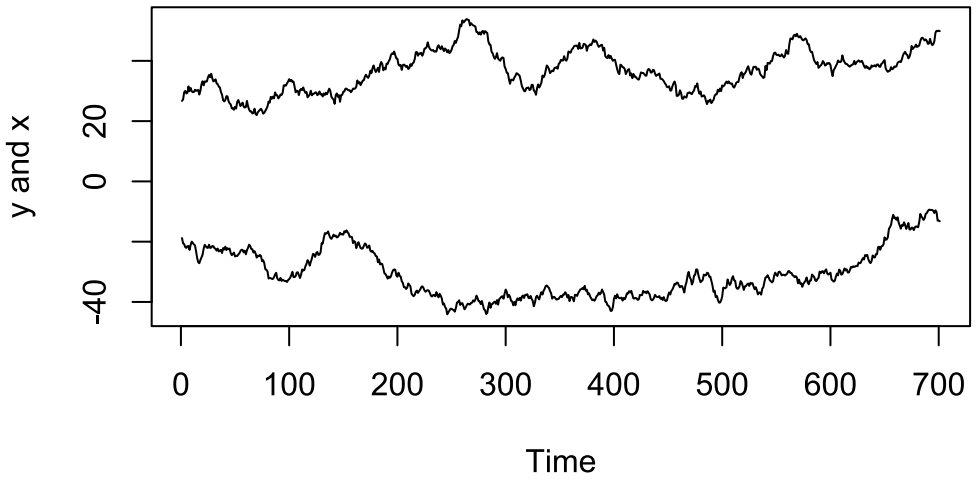
CASE 4: The nonstationary {yt} and {zt} sequences are integrated of the same order and the residual sequence is stationary. In this circumstance, {yt} and {zt} are cointegrated.

Load libraries and generate data

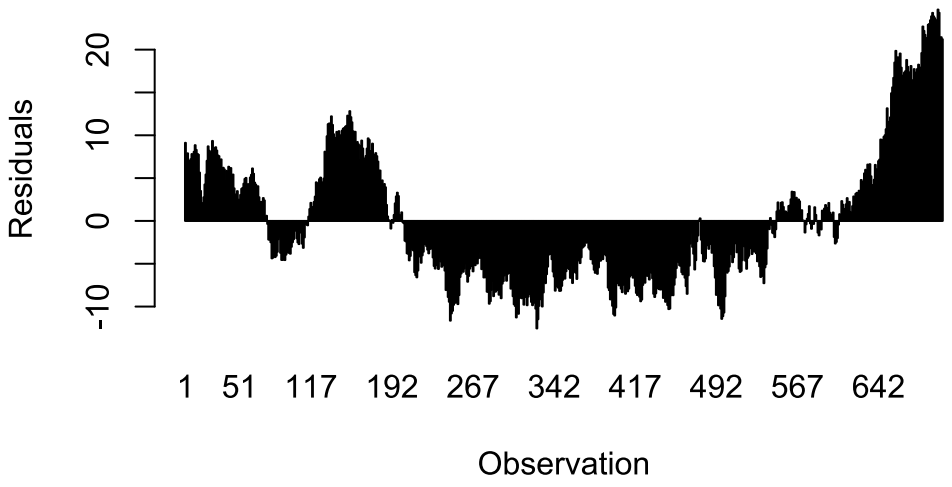
```
library(tidyverse) # An umbrella package that installs the tidyverse packages

library(readxl) # To read excel files
```

```r
library(broom) # To tidy up the regression output

library(moderndive) # To get regression points

library(ggfortify) # To plot regression diagnostics

library(ggResidpanel) # To plot regression diagnostics
```

```r
T <- 1000
set.seed(1357)
y <- ts(rep(0,T))
vy <- ts(rnorm(T))
for (t in 2:T){
  y[t] <- y[t-1]+vy[t]
}

set.seed(4365)
x <- ts(rep(0,T))
vx <- ts(rnorm(T))
for (t in 2:T){
  x[t] <- x[t-1]+vx[t]
}
y <- ts(y[300:1000])
x <- ts(x[300:1000])
ts.plot(y,x, ylab="y and x")
```

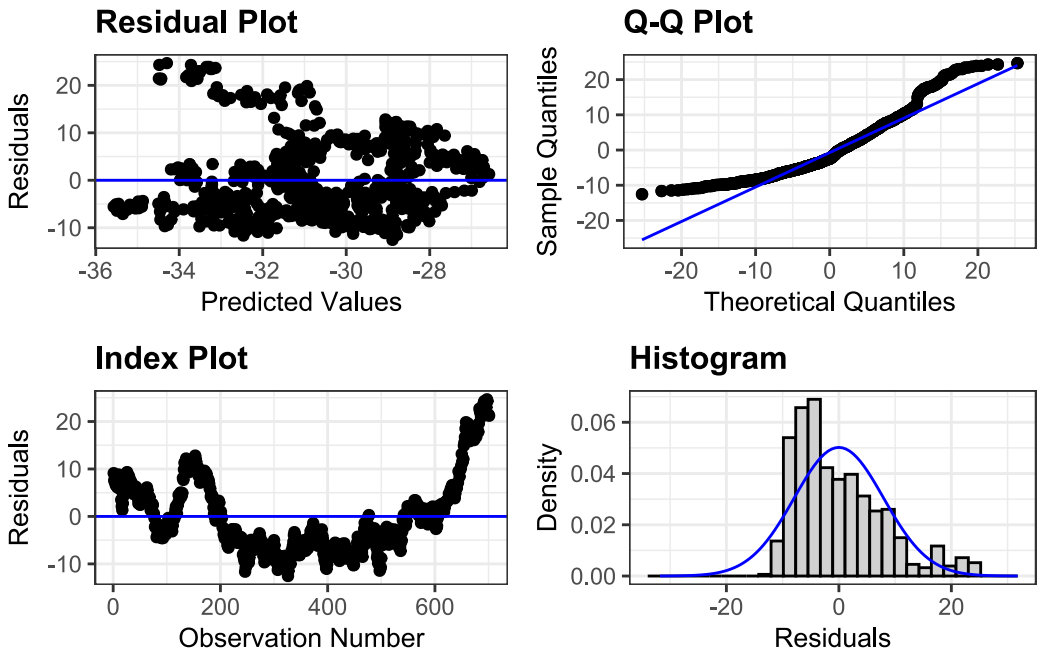


```r
# barplot of residuals
resid <- residuals(lm(y~x))
barplot(resid, ylab="Residuals", xlab="Observation", main="Residuals of y~x")
```



Residuals panel

```r
resid_panel(lm(y~x))
```


Residual Plot
Residuals
20
10
0
-10
-36
-34
-32
-30
-28
Predicted Values
Q-Q Plot
Sample Quantiles
20
10
0
-10
-20
-20
-10
0
10
20
Theoretical Quantiles
Index Plot
Residuals
20
10
0
-10
0
200
400
600
Observation Number
Histogram
Density
0.06
0.04
0.02
0.00
-20
0
20
Residuals