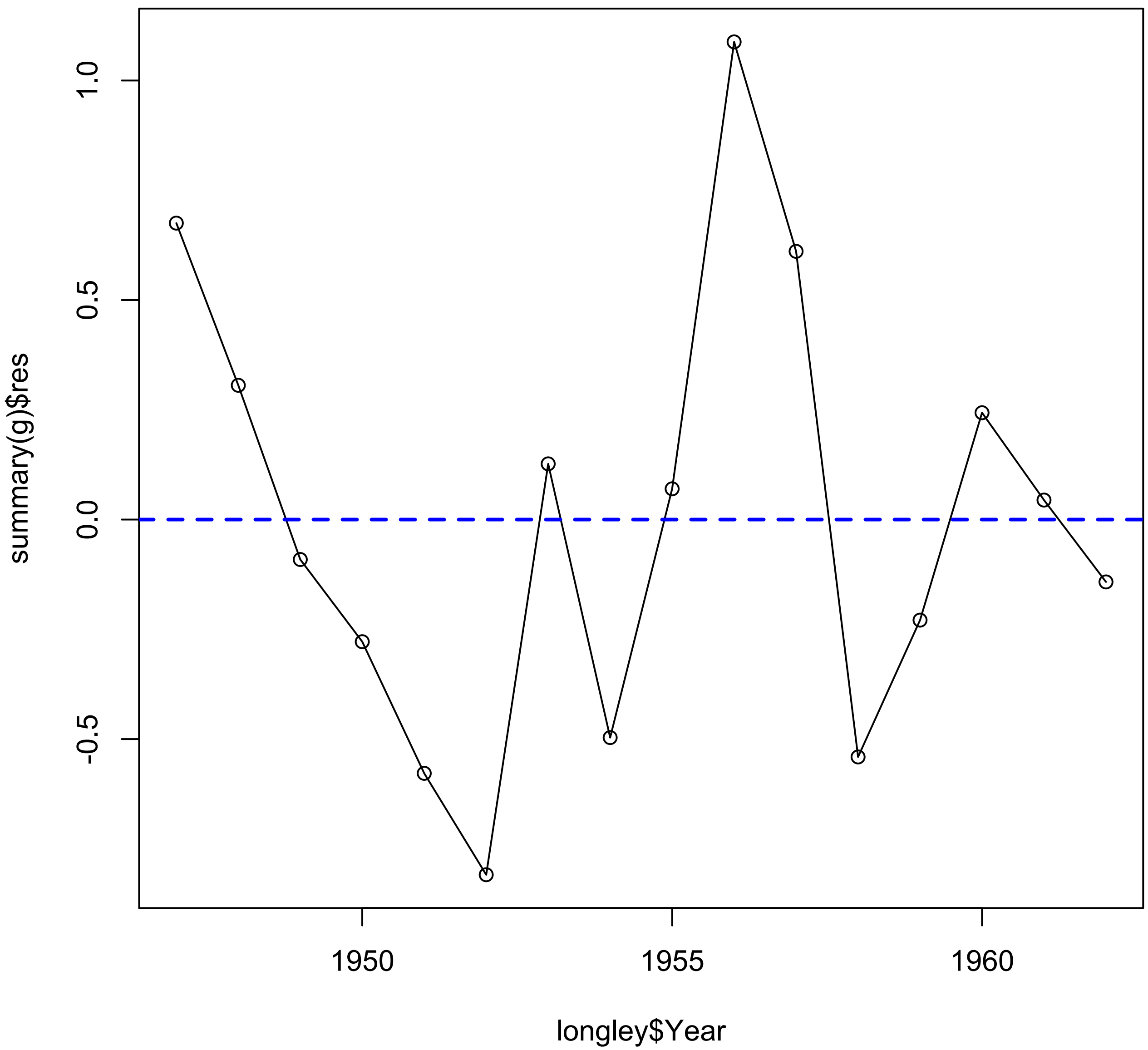
```
library(faraway)

# Longley's Economic Regression Data:
# A data frame with 7 economical variables, observed yearly from
# 1947 to 1962 (n=16).

# GNP.deflator: GNP implicit price deflator (1954=100)

# GNP: Gross National Product.

# Unemployed: number of unemployed.

# Armed.Forces: number of people in the armed forces.

# Population: ‘noninstitutionalized’ population ≥ 14 years of age.

# Year:

# Employed: number of people employed.
> g = lm(Employed ~ GNP + Population, data=longley)
> summary(g)
> plot(longley$Year, summary(g)$res, type='o')
> abline(h=0, lty=2, col="blue", lwd=2)
```



summary(g)$res
1.0
0.5
0.0
-0.5
1950
1955
1960
longley$Year

```
> library(lmtest)
> dwtest(g)

    Durbin-Watson test

data:  g
DW = 1.3015, p-value = 0.02245
alternative hypothesis: true autocorrelation is greater than 0

> library(nlme)
> g = gls(Employed ~ GNP + Population, correlation = corAR1(form= ~
Year), data=longley)
> summary(g)


Generalized least squares fit by REML
  Model: Employed ~ GNP + Population
  Data: longley
       AIC      BIC    logLik
  44.66377 47.48852 -17.33188

Correlation Structure: AR(1)
 Formula: ~Year
```

```
 Parameter estimate(s):
      Phi
0.6441692
Coefficients:
                Value Std.Error   t-value p-value
(Intercept) 101.85813 14.198932  7.173647  0.0000
GNP           0.07207  0.010606  6.795485  0.0000
Population   -0.54851  0.154130 -3.558778  0.0035

 Correlation:
           (Intr) GNP
GNP         0.943
Population -0.997 -0.966

Standardized residuals:
       Min         Q1        Med         Q3        Max
-1.5924564 -0.5447822 -0.1055401  0.3639202  1.3281898

Residual standard error: 0.689207
Degrees of freedom: 16 total; 13 residual
> g1 = gls(Employed ~ GNP + Population, correlation = corARMA(p=1),
data=longley)
> summary(g1)  # should return the same result.
```

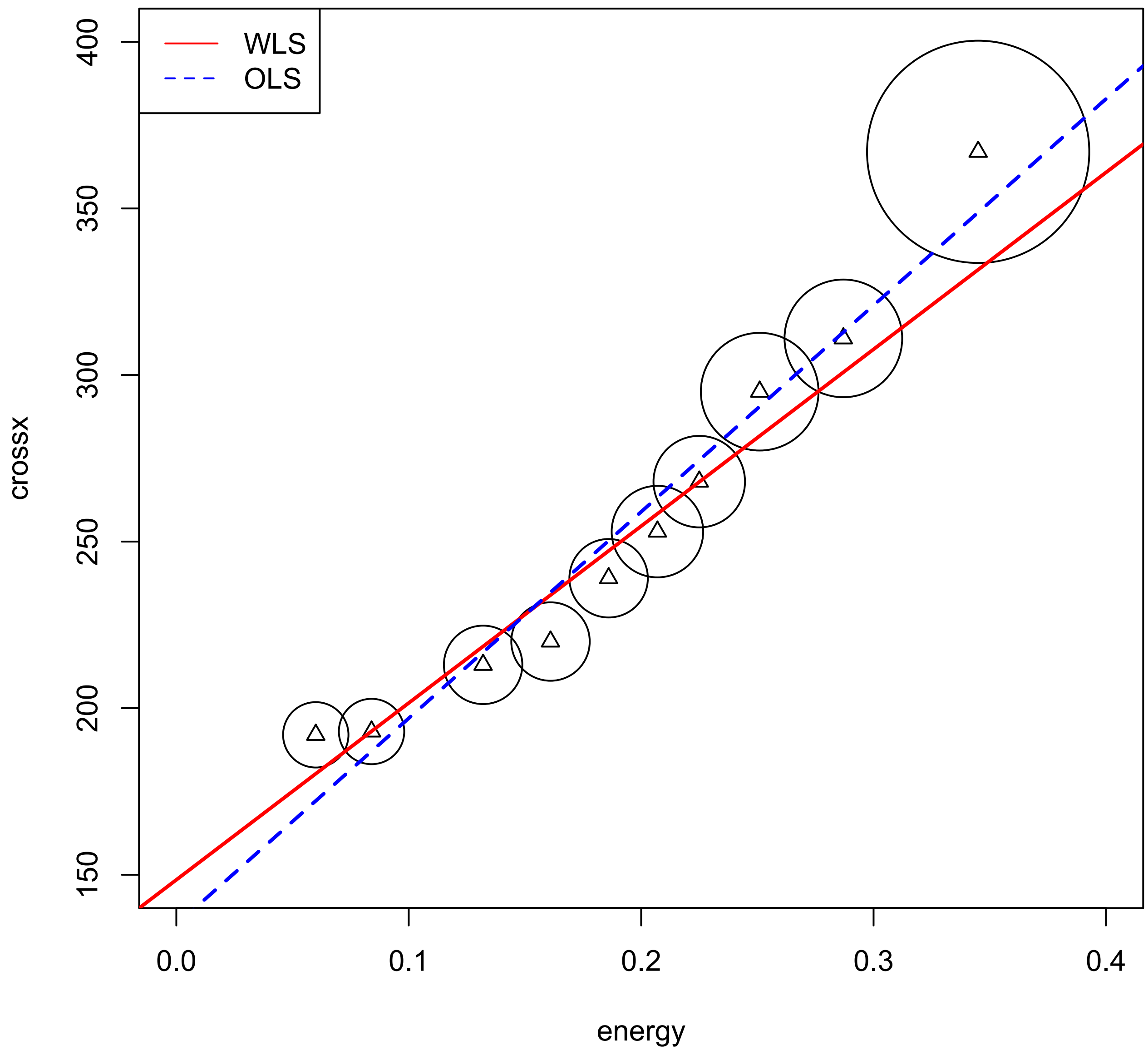
```
> g2 = gls(Employed ~ GNP + Population, correlation = corARMA(p=1),
data=longley, method="ML")
> summary(g2)
Generalized least squares fit by maximum likelihood
  Model: Employed ~ GNP + Population
  Data: longley
       AIC      BIC    logLik
  30.94792 34.81087 -10.47396

Correlation Structure: AR(1)
 Formula: ~1
 Parameter estimate(s):
      Phi
0.3651196

Coefficients:
               Value Std.Error   t-value p-value
(Intercept) 96.09369 13.967486  6.879813  0.0000
GNP          0.06822  0.010698  6.377462  0.0000
Population  -0.48716  0.153466 -3.174363  0.0073

......
```

```
> intervals(g)
Approximate 95% confidence intervals

 Coefficients:
                  lower         est.        upper
(Intercept) 71.18320460 101.85813305 132.5330615
GNP          0.04915865   0.07207088   0.0949831
Population  -0.88149053  -0.54851350  -0.2155365
attr(,"label")
[1] "Coefficients:"

 Correlation structure:
         lower      est.     upper
Phi -0.4432383 0.6441692 0.9645041
attr(,"label")
[1] "Correlation structure:"

 Residual standard error:
    lower      est.     upper
0.2477527 0.6892070 1.9172599

# DW-test shows the errors are significantly correlated,
# But the AR(1) coef rho is NOT significant. Contradiction?
```

```
# The experiment was designed to test certain theories about the
# nature of the strong interaction. The cross-section(crossx)
# variable is believed to be linearly related to the inverse of the
# energy(energy - has already been inverted). At each level of the
# momentum, a very large numbe of observations were taken so that
# it was possible to accurately estimate the standard deviation
# of the response(sd).

> strongx
   momentum energy crossx sd
1         4  0.345    367 17
2         6  0.287    311  9
3         8  0.251    295  9
4        10  0.225    268  7
5        12  0.207    253  7
6        15  0.186    239  6
7        20  0.161    220  6
8        30  0.132    213  6
9        75  0.084    193  5
10      150  0.060    192  5
```

```
> g=lm(crossx ~ energy, strongx, weights=1/sd^2)
> summary(g)
> gu=lm(crossx ~ energy, strongx);
> summary(gu)

# Lack-of-fit Test
> 1 - pchisq(summary(g)$sig^2*8, 8)
[1] 0.005004345


# How we estimate sigma-square?
> cbind(summary(gu)$sig^2, sum(gu$res^2)/8)
 [1,] 161.1616 161.1616

# For WLS, these two do not agree.
> summary(g)$sig^2
 [1] 2.744081
> sum(g$res^2)/8
[1] 248.7354
> sum(g$res^2/strongx$sd^2)/8
[1] 2.744081
```

```
> plot(crossx ~ energy, data=strongx, cex=sd, xlim=c(0, 0.4),
ylim=c(150, 400));

> points(crossx ~ energy, data=strongx, pch=2)

> abline(g, col="red", lty=1, lwd=2);
> abline(gu, col="blue", lty=2, lwd=2);

> legend("topleft", col=c("red", "blue"), lty=c(1,2),
legend=c("WLS", "OLS"));
```



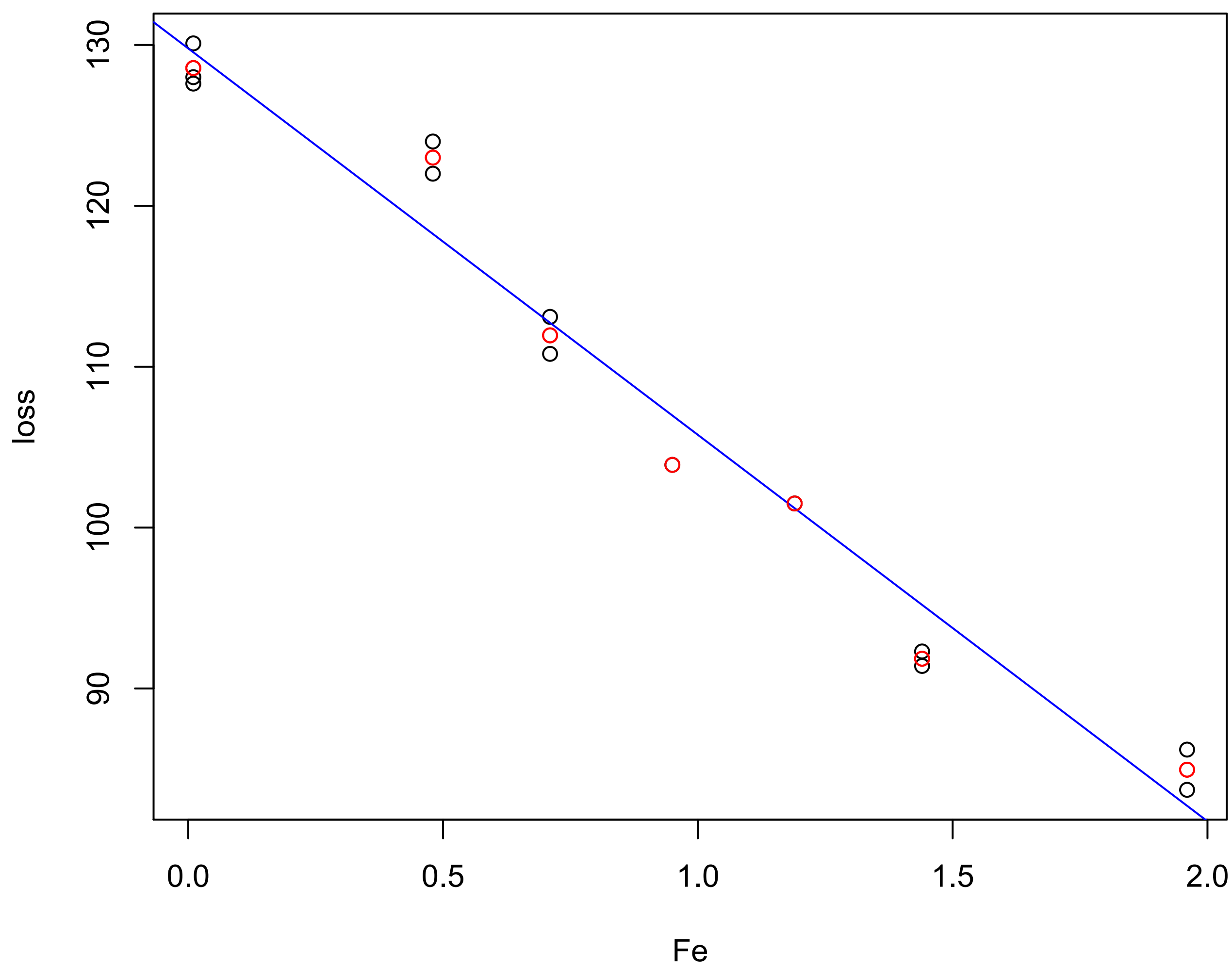
WLS
OLS
crossx
150
200
250
300
350
400
0.0
0.1
0.2
0.3
0.4
energy

```
# Then try polynomials.
> g2 =lm(crossx ~ energy + I(energy^2) , strongx, weights=1/sd^2)
> anova(g, g2)

Model 1: crossx ~ energy
Model 2: crossx ~ energy + I(energy^2)
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1      8 21.9526
2      7  3.2255  1    18.727 40.641 0.0003761 ***

> g3 =lm(crossx ~ energy + I(energy^2) + I(energy^3), strongx,
weights=1/sd^2)

> anova(g2, g3)

Model 1: crossx ~ energy + I(energy^2)
Model 2: crossx ~ energy + I(energy^2) + I(energy^3)

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      7 3.2255
2      6 2.0678  1    1.1577 3.3593 0.1165
```

```
#################################################################
# Data consist of thirteen specimens of 90/10 Cu-Ni alloys with
# varying iron content in percent. The specimens were submerged
# in sea water for 60 days and the weight loss due to corrosion
# was recorded in units of milligrams per square decimeter per
# day.
# Fe           Iron content in percent
# loss         Weight loss in mg per square decimeter per day
> corrosion[order(corrosion$Fe),]
     Fe  loss
1  0.01 127.6
6  0.01 130.1
11 0.01 128.0
2  0.48 124.0
7  0.48 122.0
3  0.71 110.8
9  0.71 113.1
4  0.95 103.9
5  1.19 101.5
8  1.44  92.3
12 1.44  91.4
10 1.96  83.7
13 1.96  86.2
```

```
> g = lm(loss ~ Fe, data=corrosion);
> plot(loss~Fe, data=corrosion); abline(coef(g));

> ga=lm(loss ~ factor(Fe), data=corrosion);
> cbind(corrosion, ga$fitted)[order(corrosion$Fe),]

     Fe  loss ga$fitted
1   0.01 127.6  128.5667
6   0.01 130.1  128.5667
11  0.01 128.0  128.5667
2   0.48 124.0  123.0000
7   0.48 122.0  123.0000
3   0.71 110.8  111.9500
9   0.71 113.1  111.9500
4   0.95 103.9  103.9000
5   1.19 101.5  101.5000
8   1.44  92.3   91.8500
12  1.44  91.4   91.8500
10  1.96  83.7   84.9500
13  1.96  86.2   84.9500
```

```
> anova(g, ga)
Analysis of Variance Table

Model 1: loss ~ Fe
Model 2: loss ~ factor(Fe)
  Res.Df     RSS Df Sum of Sq      F   Pr(>F)
1     11 102.850
2      6  11.782  5    91.069 9.2756 0.008623 **

> 1-pf(9.2756,5,6)
[1] 0.008622884
```



loss
130
120
110
100
90
0.0
0.5
1.0
1.5
2.0
Fe