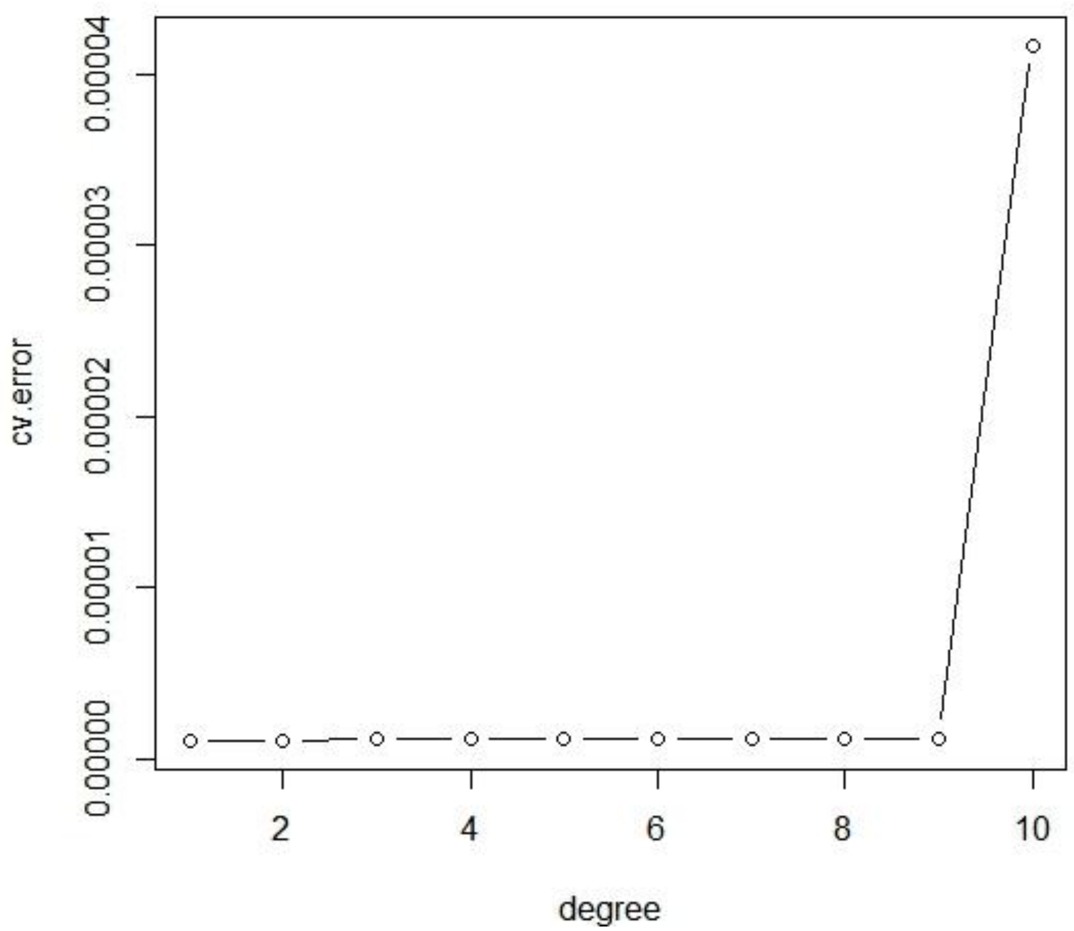
Following from our previous assignment we will now try to estimate the minimum MSE based on the degree of our linear model (For degree 2 to 10, with LOOCV method):

```
library(boot)
set.seed(1994)
cv.error=rep(0,10)
degree=1:10
for (d in degree){
    glm.fit=glm(test$"100regist2013"~poly(test$"100labour2013",d), data=test)
    cv.error[d] = cv.glm(test, glm.fit)$delta[1]
}
```

```
cv.error [1]

0.000001039363
0.000001067604
0.000001128878
0.000001138033
0.000001138054
0.000001161906
0.000001175736
0.000001190911
0.000001195136
0.000041671210
```

```
plot(degree, cv.error, type="b")
```



Also we will retrieve the summaries of Linear Models and as well as their plots (1st to 4th degree):

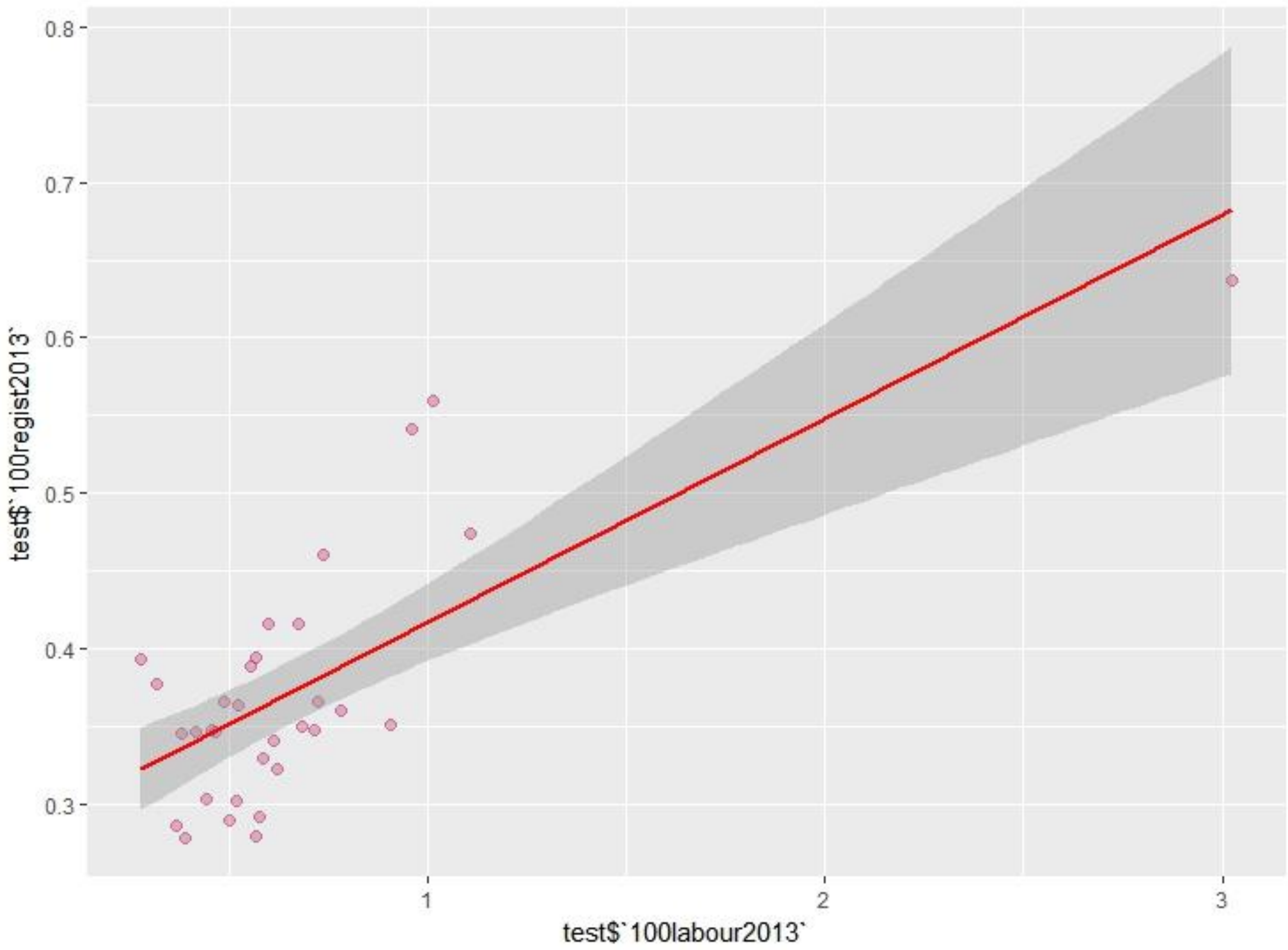
Linear (1st degree)

```
Call:
lm(formula = test$`100regist2013` ~ poly(test$`100labour2013`,
    1))

Residuals:
      Min        1Q    Median        3Q       Max
-0.080301 -0.045076 -0.007448  0.035815  0.141547

Coefficients:
                            Estimate Std. Error t value
(Intercept)                 0.374262   0.009984  37.486
poly(test$`100labour2013`, 1) 0.344168   0.056478   6.094
                                        Pr(>|t|)
(Intercept)                 < 0.0000000000000002 ***
poly(test$`100labour2013`, 1)           0.00000107 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05648 on 30 degrees of freedom
Multiple R-squared:  0.5531, Adjusted R-squared:  0.5382
F-statistic: 37.13 on 1 and 30 DF,  p-value: 0.000001073
```



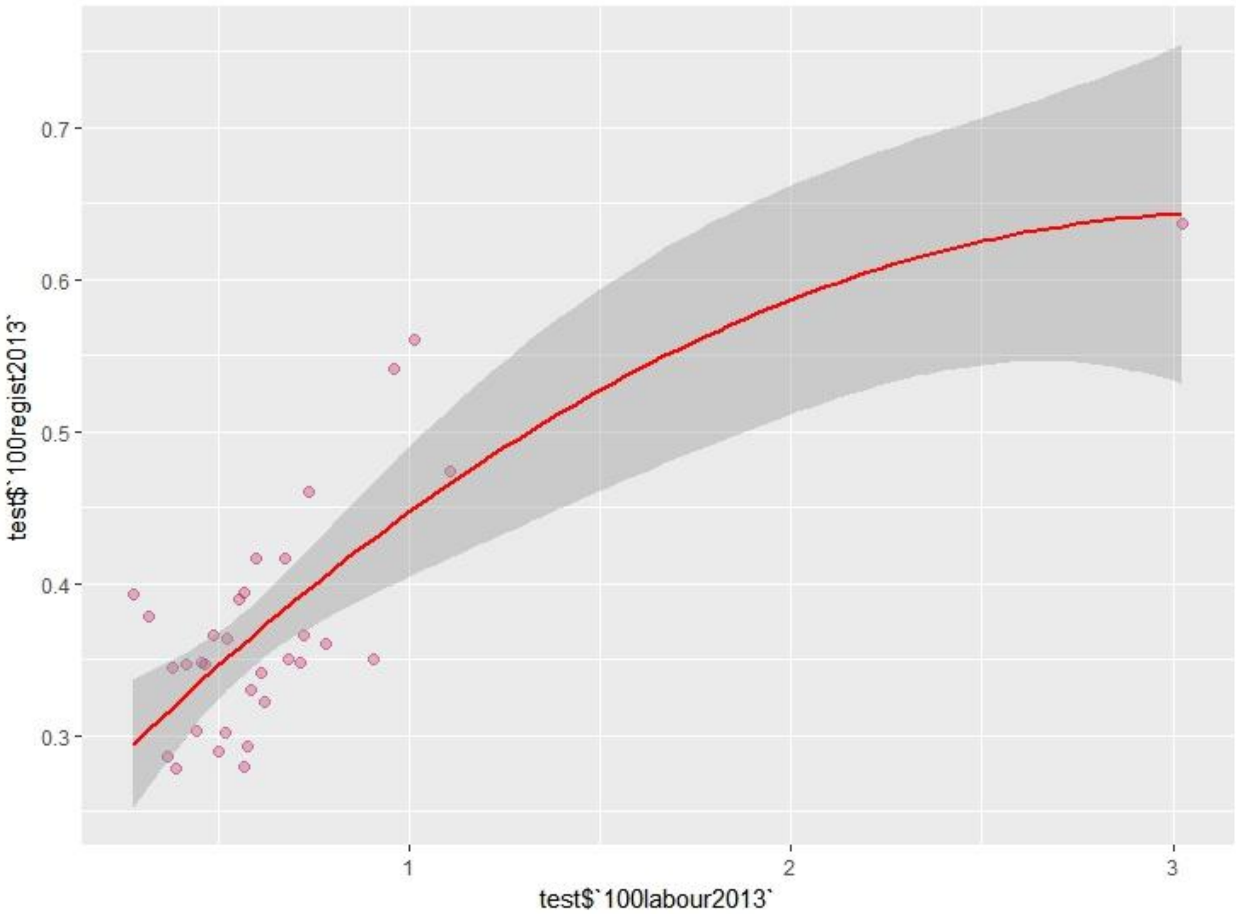
Quadratic (2nd degree)

```
Call:
lm(formula = test$`100regist2013` ~ poly(test$`100labour2013`,
   2))

Residuals:
     Min       1Q   Median       3Q      Max
-0.08059 -0.04251  0.00083  0.03162  0.11089

Coefficients:
                              Estimate Std. Error t value           Pr(>|t|)
(Intercept)                   0.374262   0.009663  38.730 < 0.0000000000000002 ***
poly(test$`100labour2013`, 2)1  0.344168   0.054664   6.296        0.000000709 ***
poly(test$`100labour2013`, 2)2 -0.095063   0.054664  -1.739             0.0926 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05466 on 29 degrees of freedom
Multiple R-squared:  0.5953, Adjusted R-squared:  0.5674
F-statistic: 21.33 on 2 and 29 DF,  p-value: 0.000002008
```



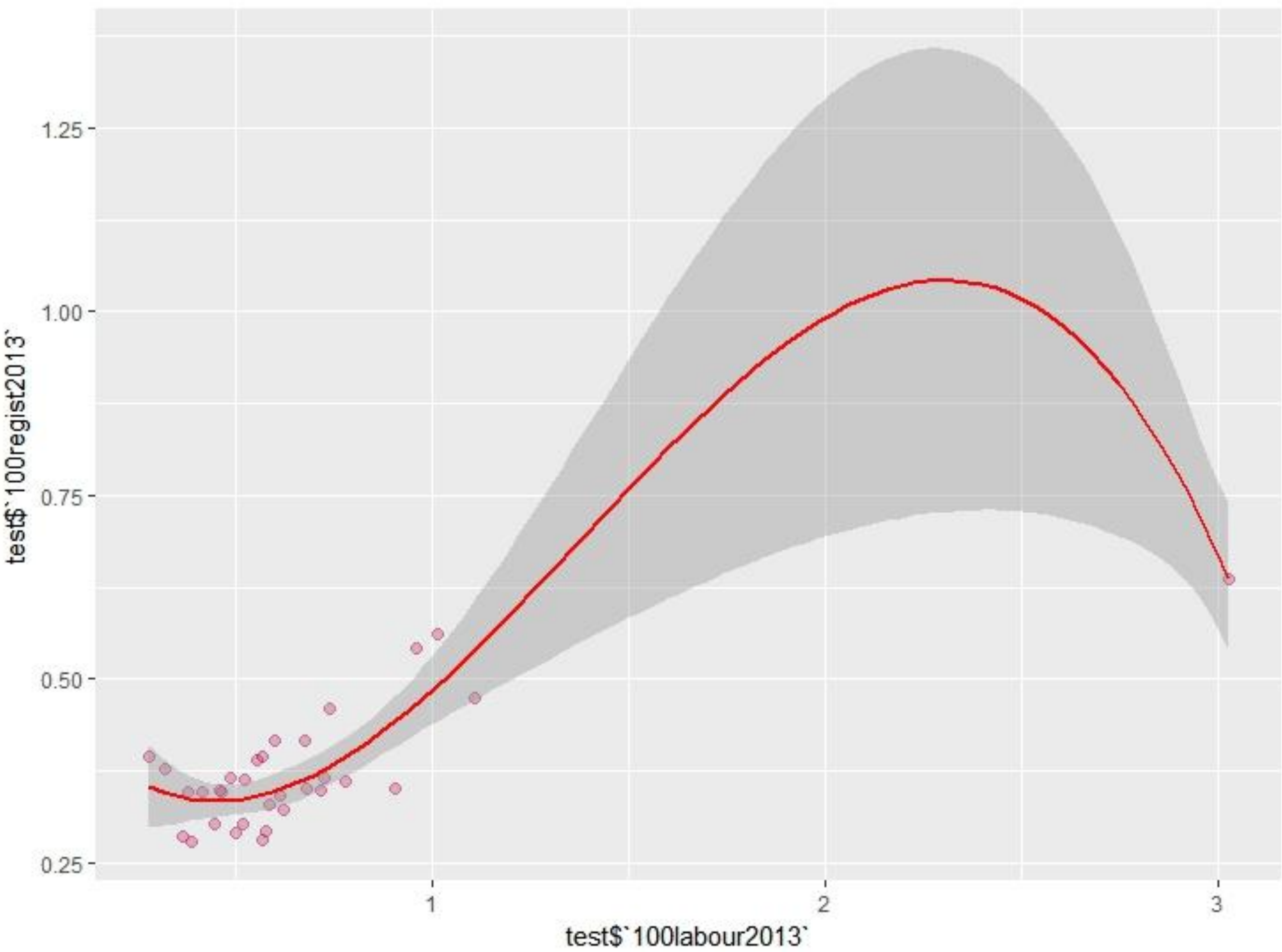
Qubic (3rd degree)

```
Call:
lm(formula = test$`100regist2013` ~ poly(test$`100labour2013`,
   3))

Residuals:
      Min        1Q    Median        3Q       Max
-0.090959 -0.033512 -0.004153  0.034715  0.079960

Coefficients:
                             Estimate Std. Error t value           Pr(>|t|)
(Intercept)                  0.37426    0.00865  43.265 < 0.0000000000000002 ***
poly(test$`100labour2013`, 3)1  0.34417    0.04893   7.033        0.000000119 ***
poly(test$`100labour2013`, 3)2 -0.09506    0.04893  -1.943            0.06217 .
poly(test$`100labour2013`, 3)3 -0.14003    0.04893  -2.862            0.00789 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04893 on 28 degrees of freedom
Multiple R-squared:  0.6869, Adjusted R-squared:  0.6534
F-statistic: 20.48 on 3 and 28 DF,  p-value: 0.0000003172
```



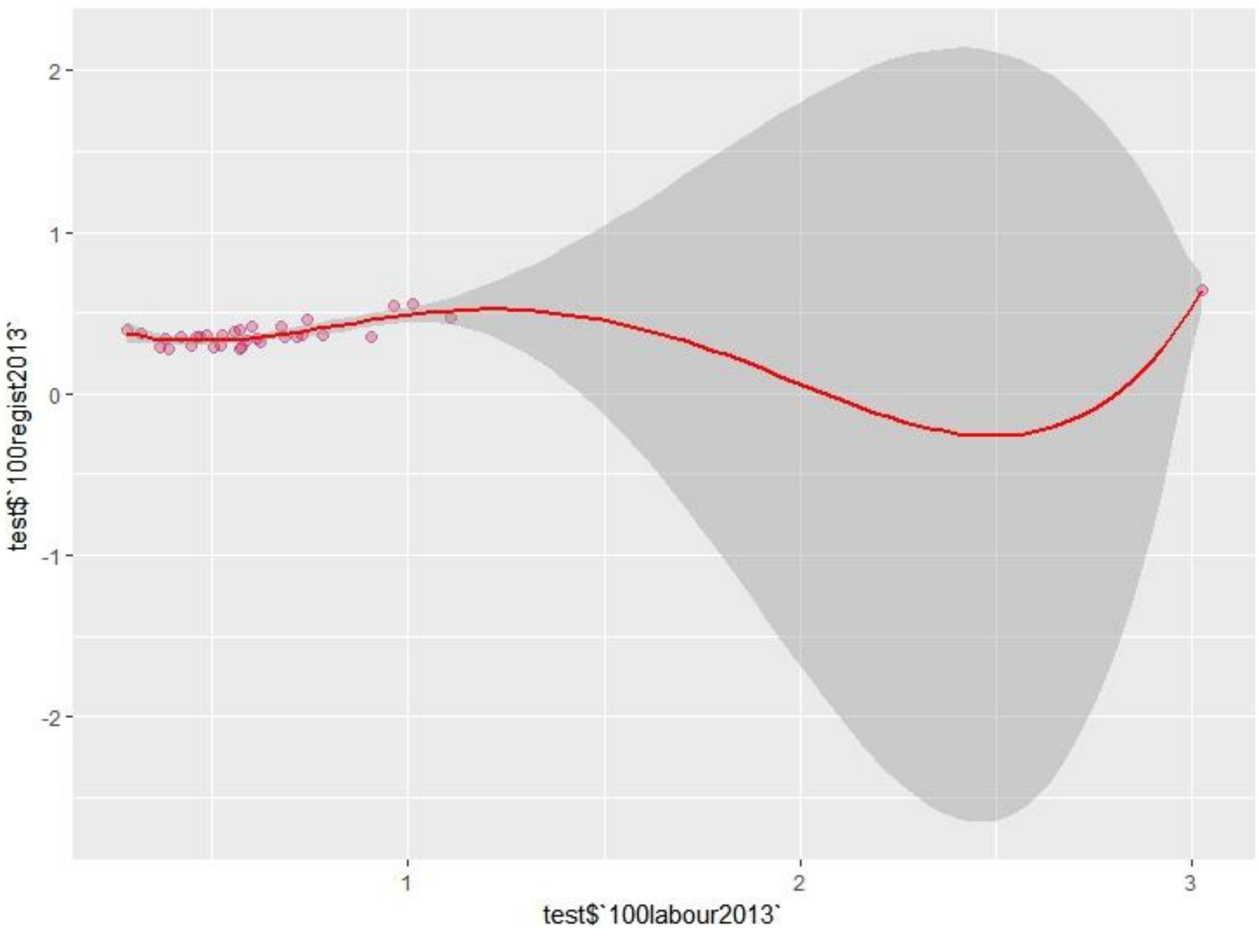
Quartic (4^{th} degree)

```
Call:
lm(formula = test$`100regist2013` ~ poly(test$`100labour2013`,
    4))

Residuals:
      Min        1Q    Median        3Q       Max
-0.105093 -0.035615 -0.004116  0.034964  0.071204

Coefficients:
                               Estimate Std. Error t value            Pr(>|t|)
 (Intercept)                   0.374262   0.008615  43.445 < 0.0000000000000002 ***
poly(test$`100labour2013`, 4)1  0.344168   0.048732   7.063         0.000000136 ***
poly(test$`100labour2013`, 4)2 -0.095063   0.048732  -1.951             0.06154 .
poly(test$`100labour2013`, 4)3 -0.140028   0.048732  -2.873             0.00782 **
poly(test$`100labour2013`, 4)4  0.054125   0.048732   1.111             0.27650
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04873 on 27 degrees of freedom
Multiple R-squared:  0.7006, Adjusted R-squared:  0.6562
F-statistic: 15.79 on 4 and 27 DF,  p-value: 0.0000008896
```



Findings

From the estimations of MSEs for different polynomial degrees, we can understand that our scenario can be best described with the use of a linear model (1^{st} degree). Because of insufficient number of observations, but also because of an extreme value (that one that has as labour ratio ~3.0), any higher polynomial is not more accurate and seems that cannot offer any safer prediction.
It is recommended to have more observations in order to eliminate the problems that we have encountered because of the extreme value.
All in all, the MSE seems to be almost the same until the 9^{th} polynomial degree.