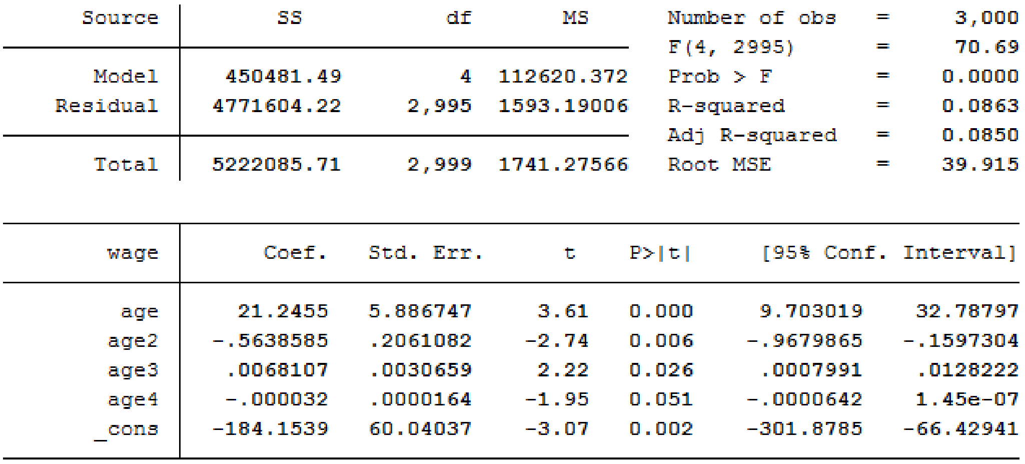
HW2:　Regression

——郑丽珊&宋悦溪

本次作业基于案例Wage Profile展开。首先，我们利用stata对案例Wage Profile部分内容进行复现；其次，以2017年上市公司为样本，利用董事会规模与企业研发投入的数据，分别使用Python、R、stata对数据回归。

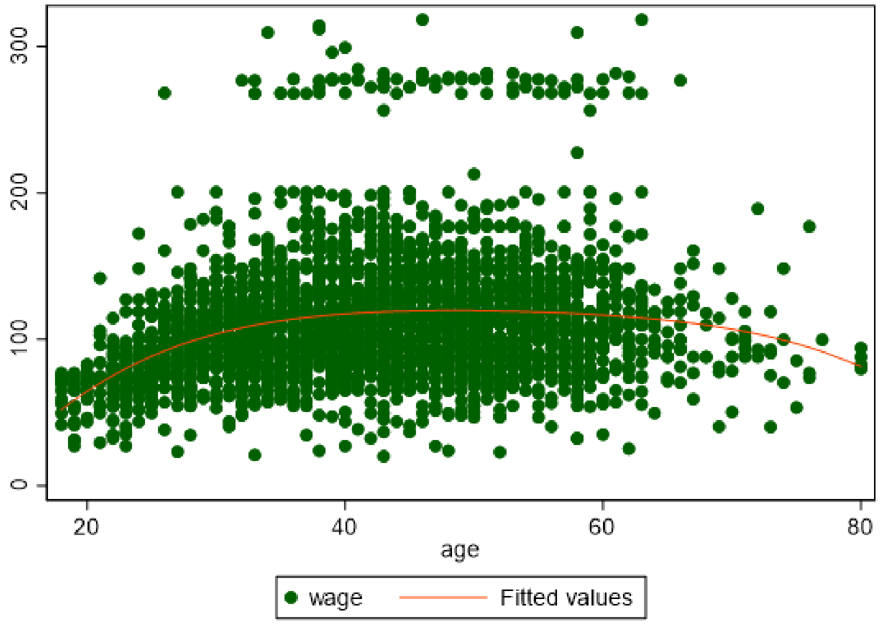
Part 1　Wage Profile 案例 stata 复现

```
gen age2=age^2
gen age3=age^3
gen age4=age^4
reg wage age age2 age3 age4
```

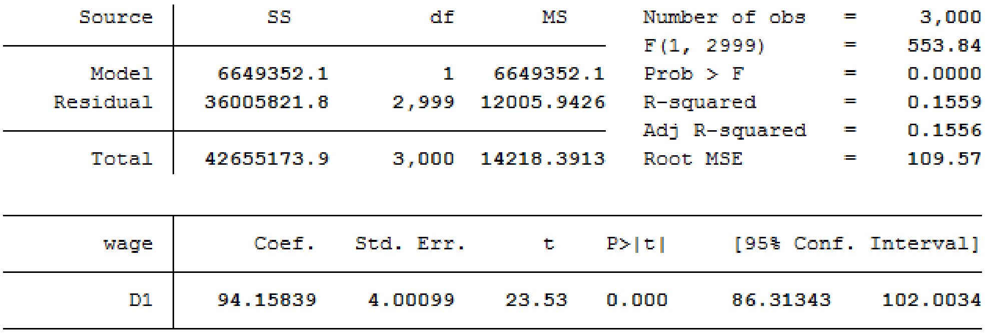


```
      Source |       SS           df       MS      Number of obs   =     3,000
-------------+----------------------------------   F(4, 2995)      =     70.69
       Model |   450481.49         4  112620.372   Prob > F        =    0.0000
    Residual |  4771604.22     2,995  1593.19006   R-squared       =    0.0863
-------------+----------------------------------   Adj R-squared   =    0.0850
       Total |  5222085.71     2,999  1741.27566   Root MSE        =    39.915

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |    21.2455   5.886747     3.61   0.000     9.703019    32.78797
        age2 |  -.5638585   .2061082    -2.74   0.006    -.9679865   -.1597304
        age3 |   .0068107   .0030659     2.22   0.026     .0007991    .0128222
        age4 |   -.000032   .0000164    -1.95   0.051    -.0000642    1.45e-07
       _cons |  -184.1539   60.04037    -3.07   0.002    -301.8785   -66.42941
------------------------------------------------------------------------------
```

```
predict fitted
scatter wage age || line fitted age,sort scheme(s1color)
```

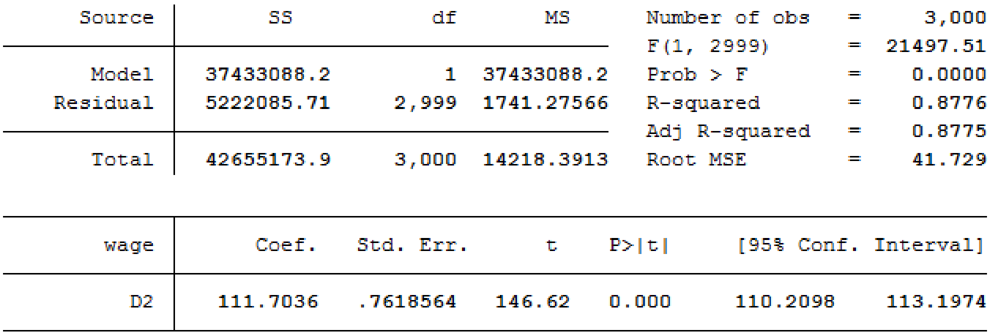


gen D1=(age<=33.5)
gen D2=(33.5<age<=49)
gen D3=(49<age<=64.5)
gen D4=(age>64.5)
reg wage D1,nocons



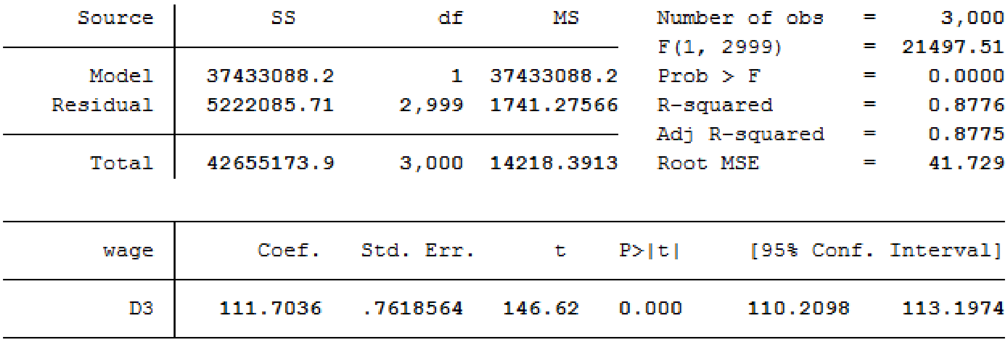
```
      Source |       SS           df       MS      Number of obs   =     3,000
-------------+----------------------------------   F(1, 2999)      =    553.84
       Model |   6649352.1         1   6649352.1   Prob > F        =    0.0000
    Residual |  36005821.8     2,999  12005.9426   R-squared       =    0.1559
-------------+----------------------------------   Adj R-squared   =    0.1556
       Total |  42655173.9     3,000  14218.3913   Root MSE        =    109.57

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          D1 |   94.15839    4.00099    23.53   0.000     86.31343    102.0034
------------------------------------------------------------------------------
```

reg wage D2,nocons



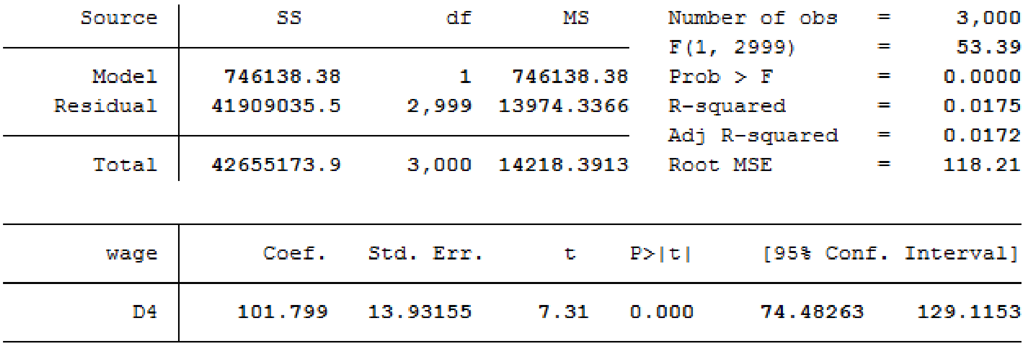
```
      Source |       SS           df       MS      Number of obs   =     3,000
-------------+----------------------------------   F(1, 2999)      =  21497.51
       Model |  37433088.2         1  37433088.2   Prob > F        =    0.0000
    Residual |  5222085.71     2,999  1741.27566   R-squared       =    0.8776
-------------+----------------------------------   Adj R-squared   =    0.8775
       Total |  42655173.9     3,000  14218.3913   Root MSE        =    41.729

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          D2 |   111.7036    .7618564   146.62   0.000     110.2098    113.1974
------------------------------------------------------------------------------
```

reg wage D3,nocons



```
      Source |       SS           df       MS      Number of obs   =     3,000
-------------+----------------------------------   F(1, 2999)      =  21497.51
       Model |  37433088.2         1  37433088.2   Prob > F        =    0.0000
    Residual |  5222085.71     2,999  1741.27566   R-squared       =    0.8776
-------------+----------------------------------   Adj R-squared   =    0.8775
       Total |  42655173.9     3,000  14218.3913   Root MSE        =    41.729

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          D3 |   111.7036    .7618564   146.62   0.000     110.2098    113.1974
------------------------------------------------------------------------------
```

reg wage D4,nocons



```
      Source |       SS           df       MS      Number of obs   =     3,000
-------------+----------------------------------   F(1, 2999)      =     53.39
       Model |   746138.38         1   746138.38   Prob > F        =    0.0000
    Residual |  41909035.5     2,999  13974.3366   R-squared       =    0.0175
-------------+----------------------------------   Adj R-squared   =    0.0172
       Total |  42655173.9     3,000  14218.3913   Root MSE        =    118.21

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          D4 |    101.799    13.93155     7.31   0.000     74.48263    129.1153
------------------------------------------------------------------------------
```

Part 2　董事会规模与企业创新

——基于 Wage Profile 案例的应用

在现代企业治理结构中，董事会拥有企业的最高决策权，是公司治理结构的核心，对企业发展有着举足轻重的作用。当今，创新已成为推动经济增长的重要力量，推动企业创新发展已成为主流。信息理论认为，董事会规模的扩大，董事经验、技能和社会网络关系有助于企业获取更多的研发信息及机会；代理理论也指出，董事长规模的扩大能更加科学与准确地评价管理层的经营行为，抑制管理层机会主义，缓解管理层与董事会成员的信息不对称程度，从而有利于企业研发战略的实施和技术创新活动的开展。那么，企业研发投入与董事长规模是否一定关系呢？

基于以上背景，我们利用2017年上市公司的数据，分析董事会规模与企业创新的关系。数据来源于CSMAR数据库。其中，采用董事会人数衡量董事会规模，单位：个人；采用研发投入（R&D）衡量企业创新，单位：百万元。（如需要源数据，可联系作者）

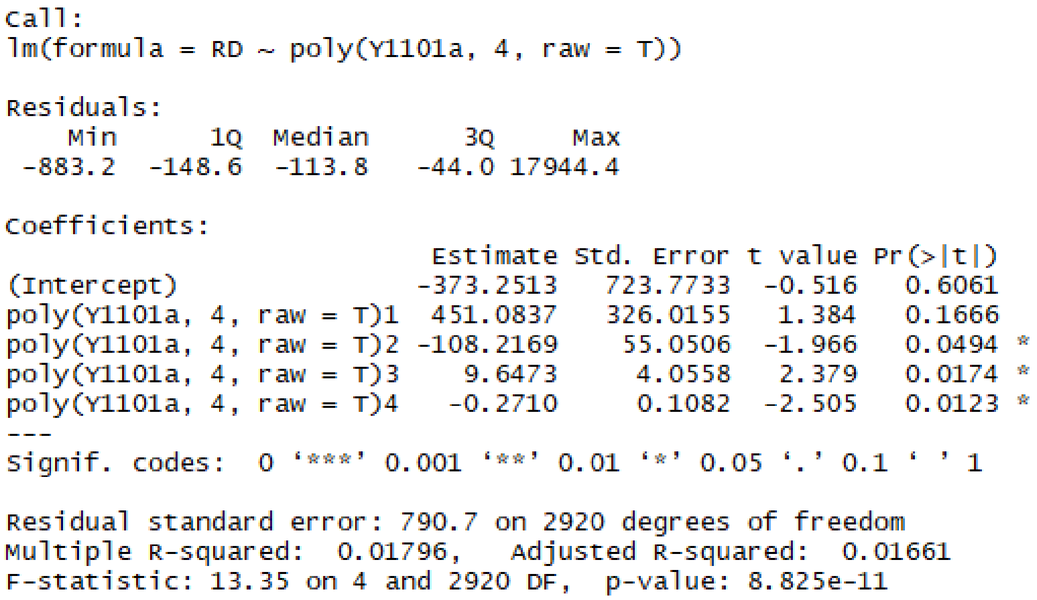
Part 2-A　基于 Python 分析

详细代码及结果见附件：cg.html。

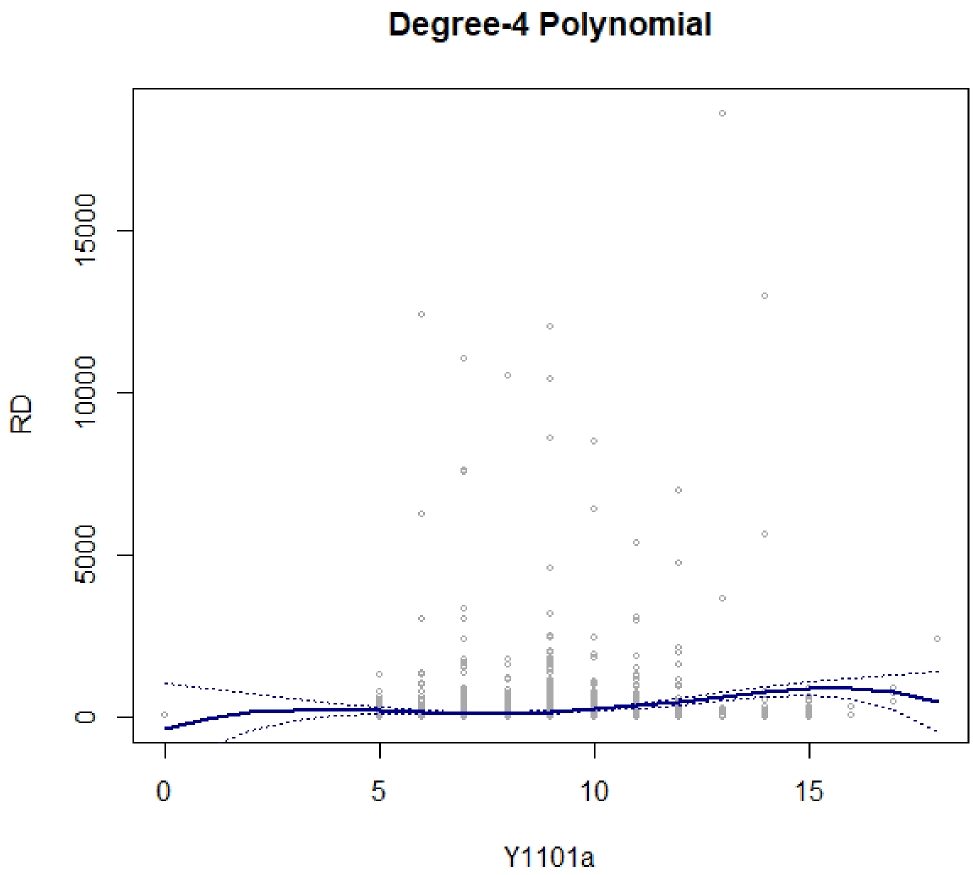
Part 2-B　基于 R 分析

详细代码见附件：cg.R，结果如下。

1. Polynomial

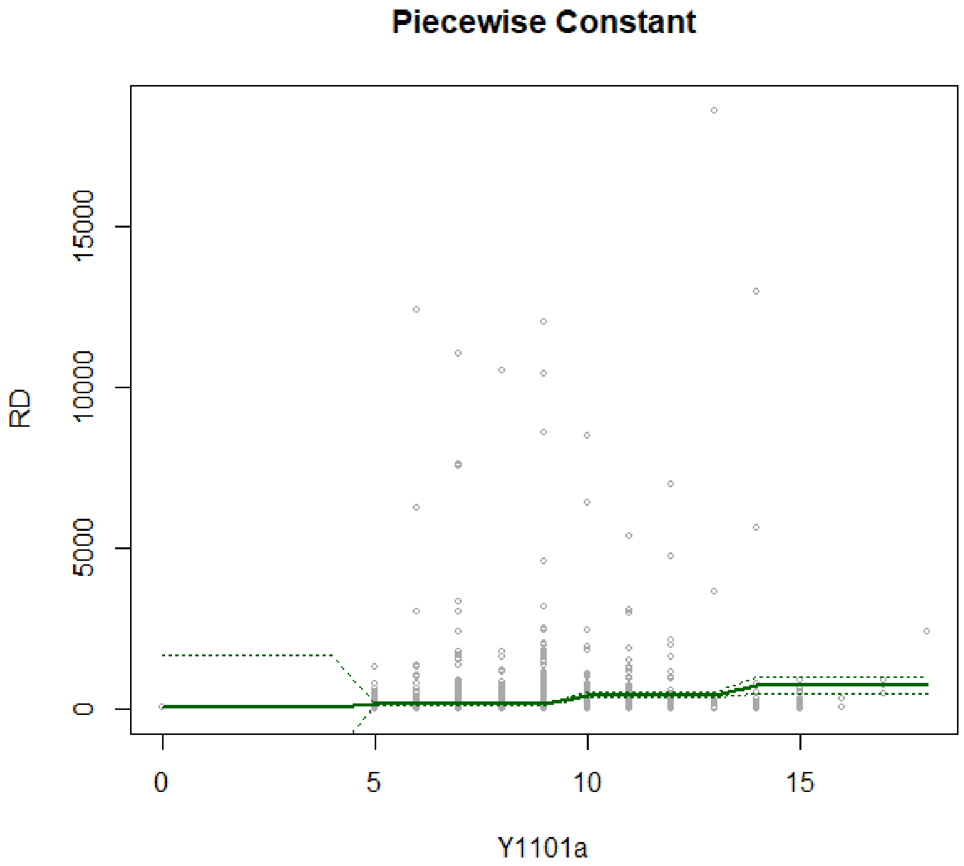


```
Call:
lm(formula = RD ~ poly(Y1101a, 4, raw = T))

Residuals:
    Min      1Q  Median      3Q     Max 
 -883.2  -148.6  -113.8   -44.0 17944.4 

Coefficients:
                            Estimate Std. Error t value Pr(>|t|)  
(Intercept)                -373.2513   723.7733  -0.516   0.6061  
poly(Y1101a, 4, raw = T)1   451.0837   326.0155   1.384   0.1666  
poly(Y1101a, 4, raw = T)2  -108.2169    55.0506  -1.966   0.0494 *
poly(Y1101a, 4, raw = T)3     9.6473     4.0558   2.379   0.0174 *
poly(Y1101a, 4, raw = T)4    -0.2710     0.1082  -2.505   0.0123 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 790.7 on 2920 degrees of freedom
Multiple R-squared:  0.01796,	Adjusted R-squared:  0.01661 
F-statistic: 13.35 on 4 and 2920 DF,  p-value: 8.825e-11
```

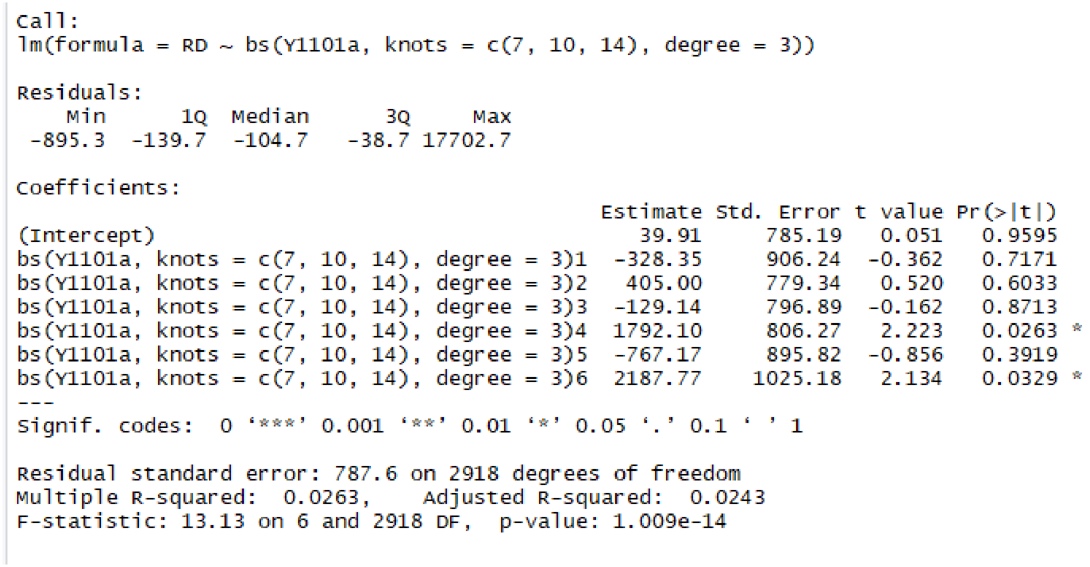


2. Piecewise Constant

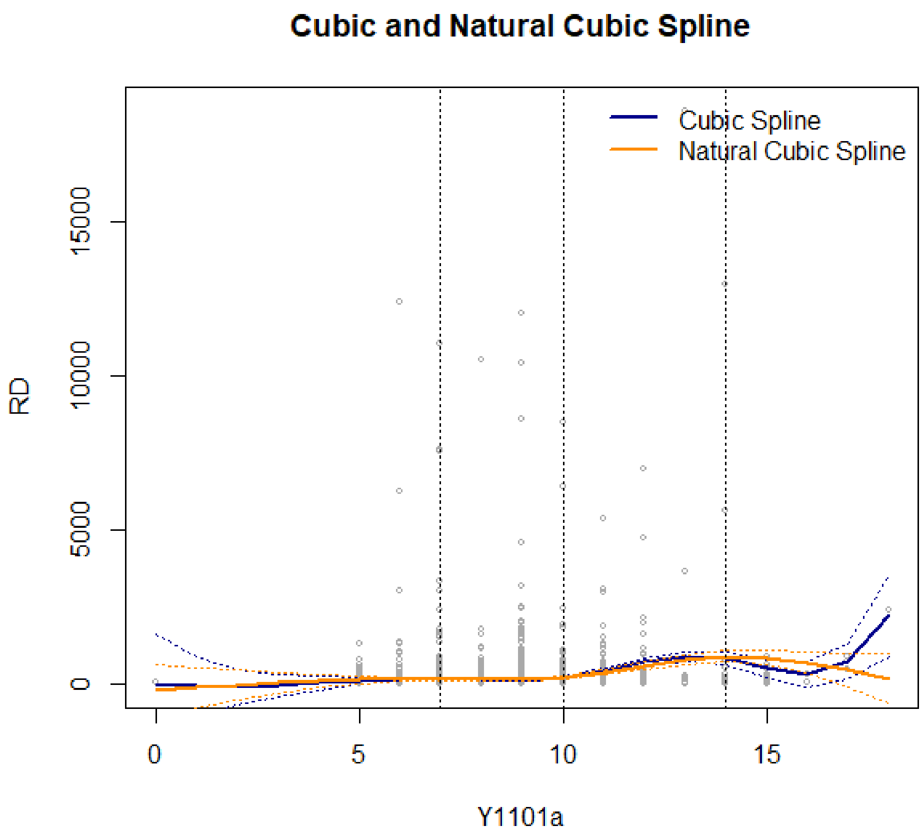
```
Call:
lm(formula = RD ~ 0 + cut(Y1101a, 4))

Residuals:
    Min      1Q  Median      3Q     Max 
 -736.5  -140.5  -113.5   -43.5 18154.9 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
cut(Y1101a, 4)(-0.018,4.5]    80.00     790.54   0.101    0.919    
cut(Y1101a, 4)(4.5,9]        158.55      15.51  10.224  < 2e-16 ***
cut(Y1101a, 4)(9,13.5]       446.10      46.91   9.510  < 2e-16 ***
cut(Y1101a, 4)(13.5,18]      736.51     123.46   5.966 2.73e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 790.5 on 2921 degrees of freedom
Multiple R-squared:  0.07316,	Adjusted R-squared:  0.07189 
F-statistic: 57.64 on 4 and 2921 DF,  p-value: < 2.2e-16
```



3. Cubic Spline and Natural Cubic Spline



```
Call:
lm(formula = RD ~ bs(Y1101a, knots = c(7, 10, 14), degree = 3))

Residuals:
    Min      1Q  Median      3Q     Max 
 -895.3  -139.7  -104.7   -38.7 17702.7 

Coefficients:
                                             Estimate Std. Error t value Pr(>|t|)  
(Intercept)                                     39.91     785.19   0.051   0.9595  
bs(Y1101a, knots = c(7, 10, 14), degree = 3)1 -328.35     906.24  -0.362   0.7171  
bs(Y1101a, knots = c(7, 10, 14), degree = 3)2  405.00     779.34   0.520   0.6033  
bs(Y1101a, knots = c(7, 10, 14), degree = 3)3 -129.14     796.89  -0.162   0.8713  
bs(Y1101a, knots = c(7, 10, 14), degree = 3)4 1792.10     806.27   2.223   0.0263 *
bs(Y1101a, knots = c(7, 10, 14), degree = 3)5 -767.17     895.82  -0.856   0.3919  
bs(Y1101a, knots = c(7, 10, 14), degree = 3)6 2187.77    1025.18   2.134   0.0329 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 787.6 on 2918 degrees of freedom
Multiple R-squared:  0.0263,	Adjusted R-squared:  0.0243 
F-statistic: 13.13 on 6 and 2918 DF,  p-value: 1.009e-14
```



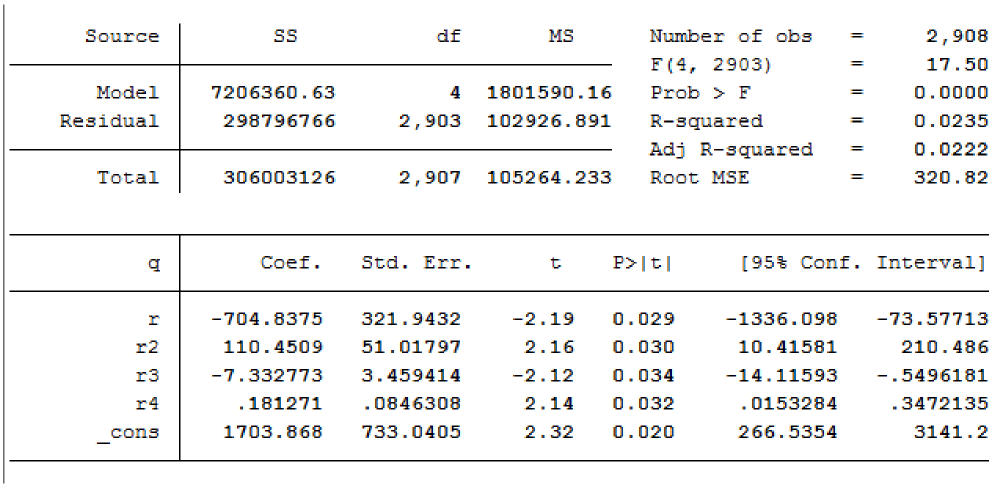
Part 2-C　基于 stata 分析

gen r2=r^2

gen r3=r^3

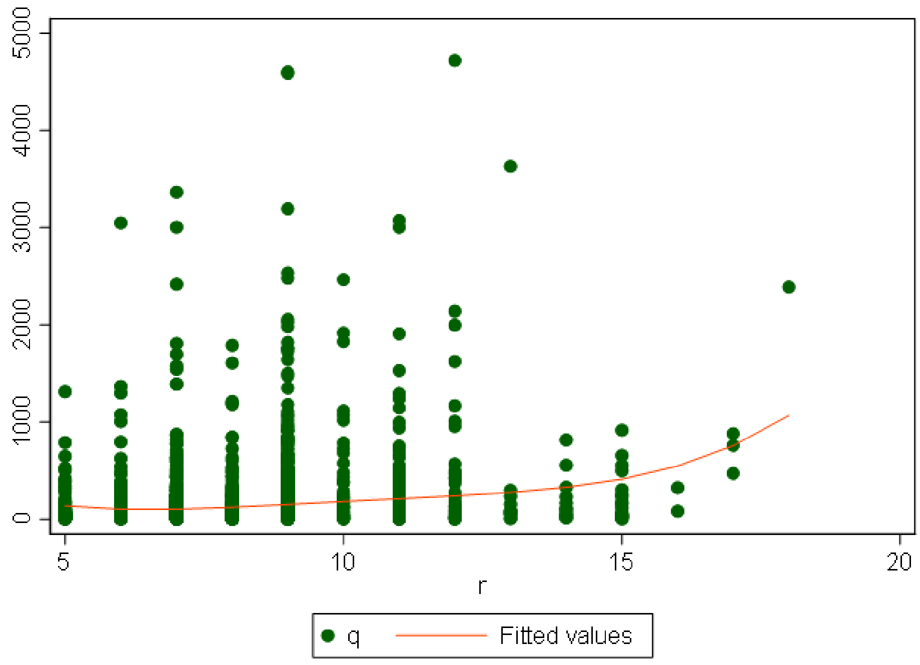
gen r4=r^4

reg q r r2 r3 r4



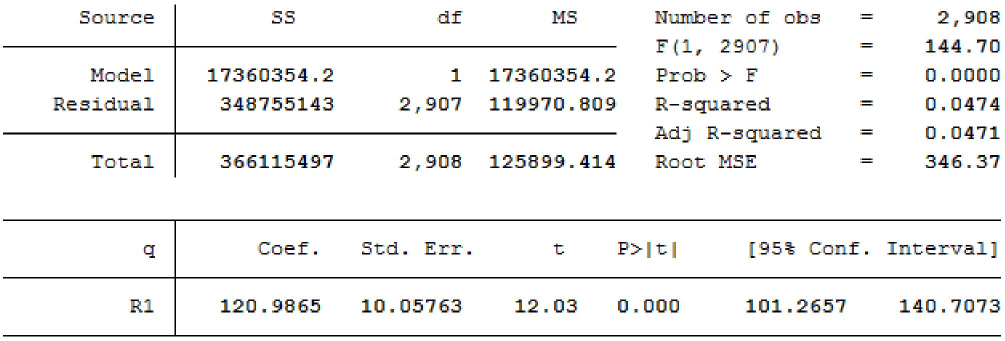
```
      Source |       SS           df       MS      Number of obs   =     2,908
-------------+----------------------------------   F(4, 2903)      =     17.50
       Model |  7206360.63         4  1801590.16   Prob > F        =    0.0000
    Residual |   298796766     2,903  102926.891   R-squared       =    0.0235
-------------+----------------------------------   Adj R-squared   =    0.0222
       Total |   306003126     2,907  105264.233   Root MSE        =    320.82

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           r |  -704.8375   321.9432    -2.19   0.029    -1336.098   -73.57713
          r2 |   110.4509   51.01797     2.16   0.030     10.41581     210.486
          r3 |  -7.332773   3.459414    -2.12   0.034    -14.11593   -.5496181
          r4 |    .181271   .0846308     2.14   0.032     .0153284    .3472135
       _cons |   1703.868   733.0405     2.32   0.020     266.5354      3141.2
------------------------------------------------------------------------------
```

predict fitted
scatter q r || line fitted r,sort scheme(s1color)



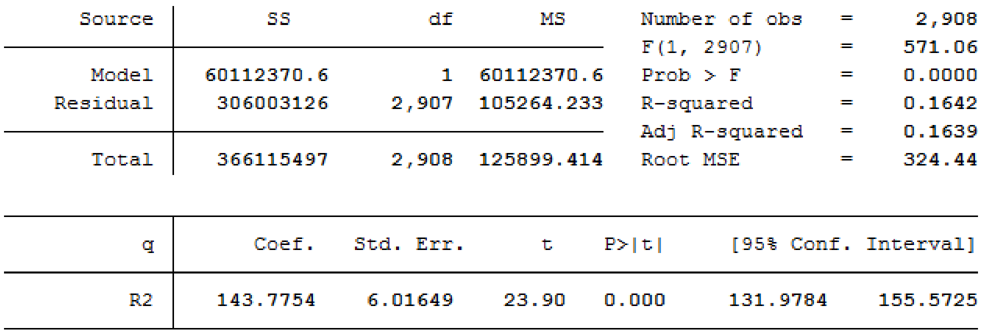
gen R1=(r<=8.25)
gen R2=(8.25<r<=11.5)
gen R3=(11.5<r<=14.75)
gen R4=(r>14.75)

reg q R1,nocons



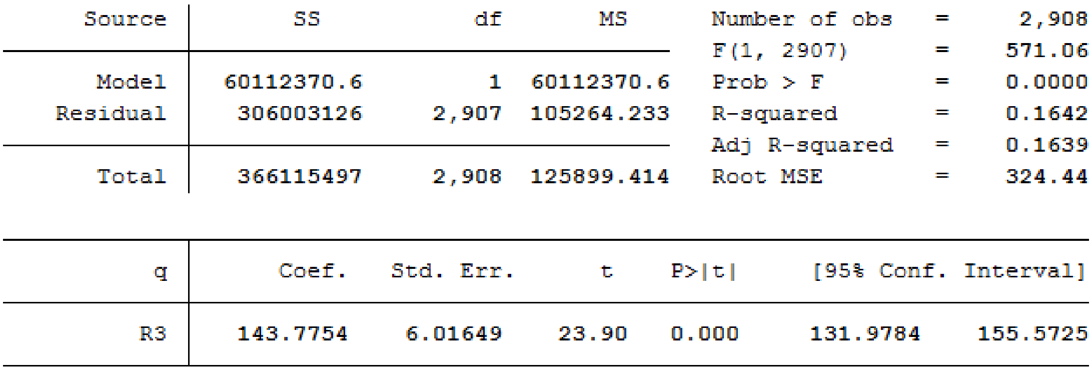
```
      Source |       SS           df       MS      Number of obs   =     2,908
-------------+----------------------------------   F(1, 2907)      =    144.70
       Model |  17360354.2         1  17360354.2   Prob > F        =    0.0000
    Residual |   348755143     2,907  119970.809   R-squared       =    0.0474
-------------+----------------------------------   Adj R-squared   =    0.0471
       Total |   366115497     2,908  125899.414   Root MSE        =    346.37

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          R1 |   120.9865   10.05763    12.03   0.000     101.2657    140.7073
------------------------------------------------------------------------------
```

reg q R2,nocons



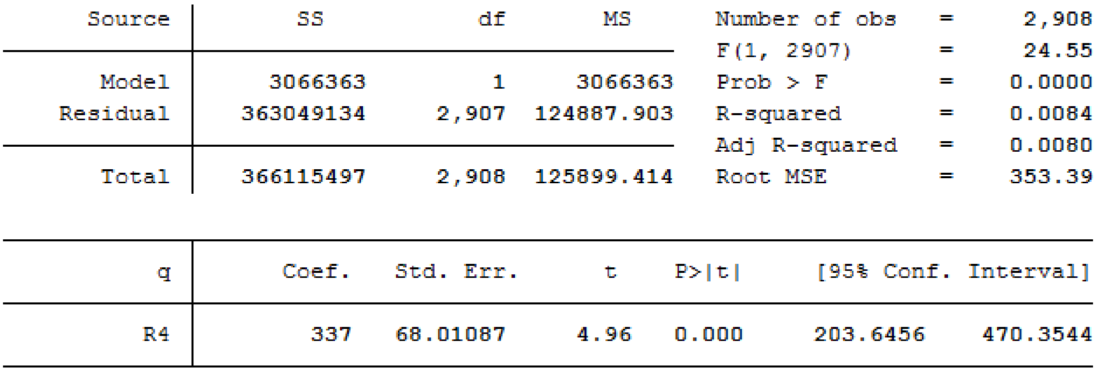
```
      Source |       SS           df       MS      Number of obs   =     2,908
-------------+----------------------------------   F(1, 2907)      =    571.06
       Model |  60112370.6         1  60112370.6   Prob > F        =    0.0000
    Residual |   306003126     2,907  105264.233   R-squared       =    0.1642
-------------+----------------------------------   Adj R-squared   =    0.1639
       Total |   366115497     2,908  125899.414   Root MSE        =    324.44

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          R2 |   143.7754    6.01649    23.90   0.000     131.9784    155.5725
------------------------------------------------------------------------------
```

reg q R3,nocons



```
      Source |       SS           df       MS      Number of obs   =     2,908
-------------+----------------------------------   F(1, 2907)      =    571.06
       Model |  60112370.6         1  60112370.6   Prob > F        =    0.0000
    Residual |   306003126     2,907  105264.233   R-squared       =    0.1642
-------------+----------------------------------   Adj R-squared   =    0.1639
       Total |   366115497     2,908  125899.414   Root MSE        =    324.44

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          R3 |   143.7754    6.01649    23.90   0.000     131.9784    155.5725
------------------------------------------------------------------------------
```

reg q R4,nocons



```
      Source |       SS           df       MS      Number of obs   =     2,908
-------------+----------------------------------   F(1, 2907)      =     24.55
       Model |     3066363         1     3066363   Prob > F        =    0.0000
    Residual |   363049134     2,907  124887.903   R-squared       =    0.0084
-------------+----------------------------------   Adj R-squared   =    0.0080
       Total |   366115497     2,908  125899.414   Root MSE        =    353.39

------------------------------------------------------------------------------
           q |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          R4 |        337   68.01087     4.96   0.000     203.6456    470.3544
------------------------------------------------------------------------------
```

上述一系列研究表明，代表企业创新的企业研发投入与董事会规模成正相关关系，即董事会规模越大，研发投入费用越多，企业创新程度越高，这也与信息理论和代理理论的预测结论不谋而合。如今科技创新已成为国际竞争中成败的主导因素，科技竞争力决定着一个国家或地区在未来世界竞争格局中的位置。那么对于企业来说，技术创新的重要作用不言而喻，它是每个企业赖以生存并持续发展的重要依仗。

该实证研究可以为促进中国企业技术创新和研发活动提供相关建议。企业可以在保持合理董事会规模的前提下，适度扩大董事会的规模，这样不仅可能获得更多的创新资源，同时还能减少信息不对称以助于企业创新活动的开展，进而达到持续推动企业技术创新的效果。

参考文献

[1]陈强. (2014). *高级计量经济学及Stata应用. 第2版.* 高等教育出版社.

[2]刘小元, & 李永壮. (2012). 董事会、资源约束与创新环境影响下的创业企业研发强度——来自创业板企业的证据. *软科学, 26*(6).