DATA621 LMR Ex 8.2

Chun Yip

2022/4/16

R Markdown

Using the divusa data, fit a regression model with divorce as the response and unemployed, femlab, marriage, birth and military as predictors. (a) Make two graphical checks for correlated errors. What do you conclude? (b) Allow for serial correlation with an AR(1) model for the errors. (Hint: Use maximum likelihood to estimate the parameters in the GLS fit by gls(. . . , method="ML", . . .)). What is the estimated correlation and is it significant? Does the GLS model change which variables are found to be significant? (c) Speculate why there might be correlation in the errors

```
data(divusa, package="faraway")
head(divusa)
```

```
##   year divorce unemployed femlab marriage birth military
## 1 1920     8.0        5.2  22.70     92.0 117.9   3.2247
## 2 1921     7.2       11.7  22.79     83.0 119.8   3.5614
## 3 1922     6.6        6.7  22.88     79.7 111.2   2.4553
## 4 1923     7.1        2.4  22.97     85.2 110.5   2.2065
## 5 1924     7.2        5.0  23.06     80.3 110.9   2.2889
## 6 1925     7.2        3.2  23.15     79.2 106.6   2.1735
```

```
summary(divusa)
```

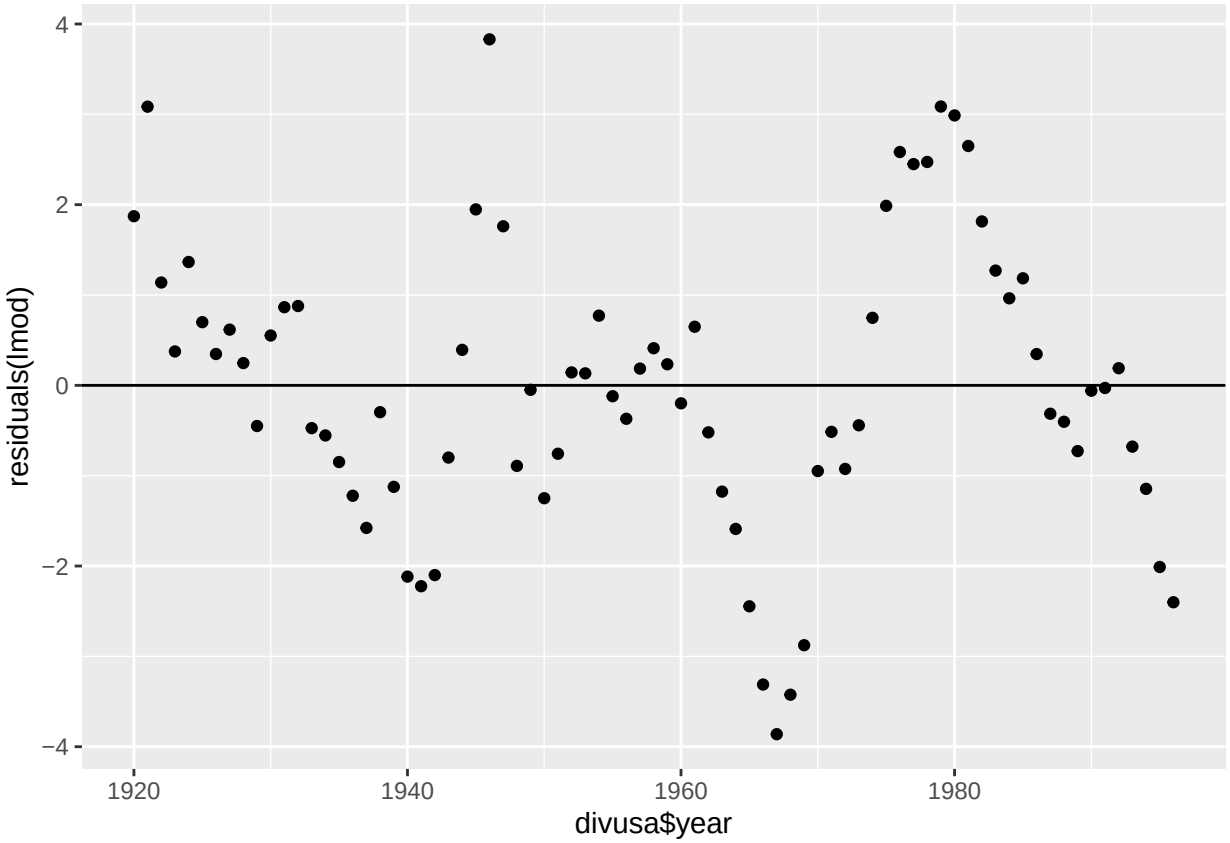
```
##       year         divorce        unemployed         femlab
##  Min.   :1920   Min.   : 6.10   Min.   : 1.200   Min.   :22.70
##  1st Qu.:1939   1st Qu.: 8.70   1st Qu.: 4.200   1st Qu.:27.47
##  Median :1958   Median :10.60   Median : 5.600   Median :37.10
##  Mean   :1958   Mean   :13.27   Mean   : 7.173   Mean   :38.58
##  3rd Qu.:1977   3rd Qu.:20.30   3rd Qu.: 7.500   3rd Qu.:47.80
##  Max.   :1996   Max.   :22.80   Max.   :24.900   Max.   :59.30
##     marriage          birth           military
##  Min.   : 49.70   Min.   : 65.30   Min.   : 1.940
##  1st Qu.: 61.90   1st Qu.: 68.90   1st Qu.: 3.469
##  Median : 74.10   Median : 85.90   Median : 9.102
##  Mean   : 72.97   Mean   : 88.89   Mean   :12.365
##  3rd Qu.: 80.00   3rd Qu.:107.30   3rd Qu.:14.266
##  Max.   :118.10   Max.   :122.90   Max.   :86.641
```

Make two graphical checks for correlated errors

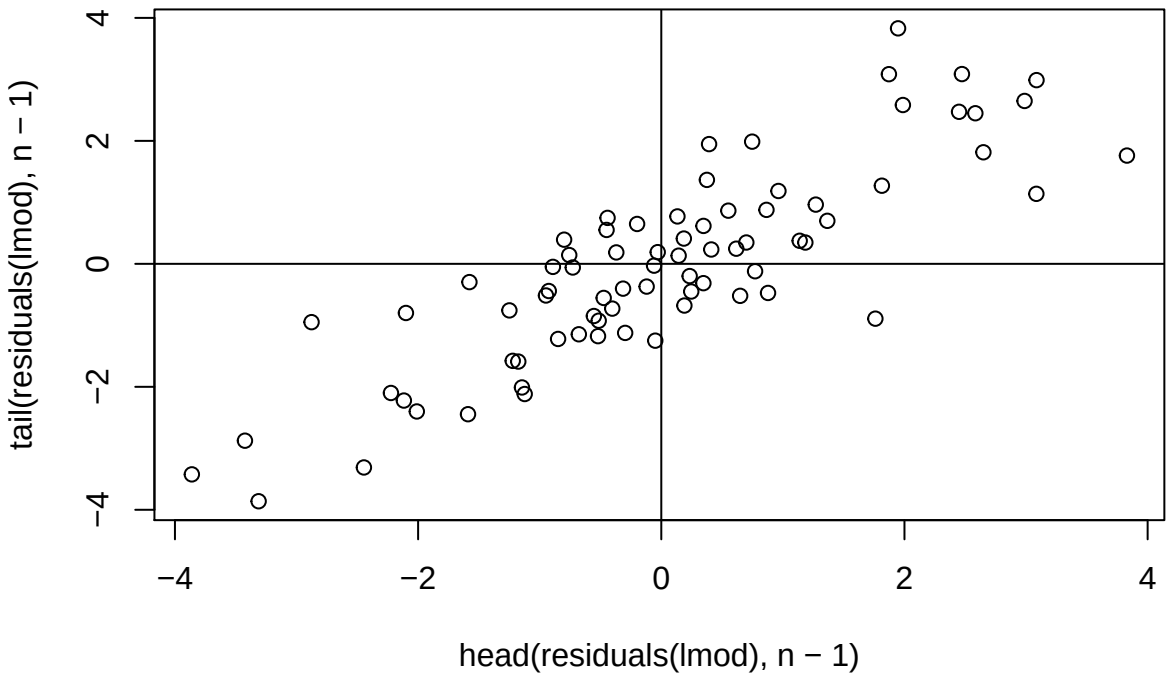
```r
lmod <- lm(divorce ~ unemployed+femlab+marriage+birth+military, divusa)
summary(lmod)
```

```
##
## Call:
## lm(formula = divorce ~ unemployed + femlab + marriage + birth +
##     military, data = divusa)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8611 -0.8916 -0.0496  0.8650  3.8300
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.48784    3.39378   0.733   0.4659
## unemployed  -0.11125    0.05592  -1.989   0.0505 .
## femlab       0.38365    0.03059  12.543  < 2e-16 ***
## marriage     0.11867    0.02441   4.861 6.77e-06 ***
## birth       -0.12996    0.01560  -8.333 4.03e-12 ***
## military    -0.02673    0.01425  -1.876   0.0647 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.65 on 71 degrees of freedom
## Multiple R-squared:  0.9208, Adjusted R-squared:  0.9152
## F-statistic: 165.1 on 5 and 71 DF,  p-value: < 2.2e-16
```

```r
ggplot(lmod, aes(x=divusa$year, y=residuals(lmod)), ylab='Residuals')+
  geom_point()+
  geom_hline(yintercept=0)
```




```
n<-length(residuals(lmod))
plot(tail(residuals(lmod),n-1) ~ head(residuals(lmod),n-1))
abline(h=0, v=0)
```




```
cor(residuals(lmod)[-1],residuals(lmod)[-length(residuals(lmod))])
```

```
## [1] 0.8469792
```

There are positive correlation serial correlation shown in both plots. The correlation is 0.85 between successive

```
glsmod<-gls(divorce ~ unemployed+femlab+marriage+birth+military, divusa, correlation=corAR1(form=~year))
summary(glsmod)
```

```
## Generalized least squares fit by maximum likelihood
##   Model: divorce ~ unemployed + femlab + marriage + birth + military
##   Data: divusa
##        AIC      BIC    logLik
##   179.9523 198.7027 -81.97613
##
## Correlation Structure: AR(1)
##  Formula: ~year
##  Parameter estimate(s):
##       Phi
## 0.9715486
##
## Coefficients:
##                 Value Std.Error   t-value p-value
```

```
## (Intercept) -7.059682  5.547193 -1.272658  0.2073
## unemployed   0.107643  0.045915  2.344395  0.0219
## femlab       0.312085  0.095151  3.279878  0.0016
## marriage     0.164326  0.022897  7.176766  0.0000
## birth       -0.049909  0.022012 -2.267345  0.0264
## military     0.017946  0.014271  1.257544  0.2127
##
##  Correlation:
##             (Intr) unmply femlab marrig birth
## unemployed -0.420
## femlab     -0.802  0.240
## marriage   -0.516  0.607  0.307
## birth      -0.379  0.041  0.066 -0.094
## military   -0.036  0.436 -0.311  0.530  0.128
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.4509327 -0.9760939 -0.6164694  1.1375377  2.1593261
##
## Residual standard error: 2.907665
## Degrees of freedom: 77 total; 71 residual
```

The residual standard error is larger than the one of linear regression. The autocorrelation is significant and estimated phi is 0.97. The GLS doesn't really change any significant predictors. If a couple is in military and unemployed in previous year, it's quite possible to get divorce since the couple will have financial problem and separate during the military service.