An overview of the data and some intuition wich possibly is way off

TMA4285 - Time Series

Christian Oppegård Moen

dd-mm-yyyy

Contents

Task

Model inflation rate along with three covariates of our choosing.

Exploratory Data Analysis (EDA)

Import data and format date.

```
pd <- read_csv("projectdataFormated.csv")
typeof(pd)
```

```
## [1] "list"
```

```
pd$Month = as.Date(paste0(sub("M", "-", pd$Month), "-01"), format = "%Y-%m-%d")
head(pd)
```

```
## # A tibble: 6 x 5
##   Month      Inflation Unemployed Bankrupt Consumption
##   <date>         <dbl>      <dbl>    <dbl>       <dbl>
## 1 2000-02-01       0.4      65736      331        -0.5
## 2 2000-03-01       0.1      62934      357        -2
## 3 2000-04-01       0.3      61231      269         3.6
## 4 2000-05-01       0.1      59553      331        -0.2
## 5 2000-06-01       0.5      60836      317        -2.4
## 6 2000-07-01      -0.4      68405      228         2.2
```

Inflation overview

Inflation over time:

```
pd$MonthInt = as.integer(format(pd$Month, "%m"))
pd$Season <- ifelse(month(pd$Month) %in% c(12, 1, 2), "Winter", ifelse(month(pd$Month) %in%
    c(3, 4, 5), "Spring", ifelse(month(pd$Month) %in% c(6, 7, 8), "Summer", "Autumn")))

# Season colored plot with 'loess' fitted curves ----
ggInflation = ggplot(pd, aes(x = Month, y = Inflation)) + geom_point(aes(col = Season)) +
    geom_line(linetype = 1, lwd = 0.5) + geom_smooth(aes(col = Season), se = F) +
    theme_minimal()
ggInflation
```

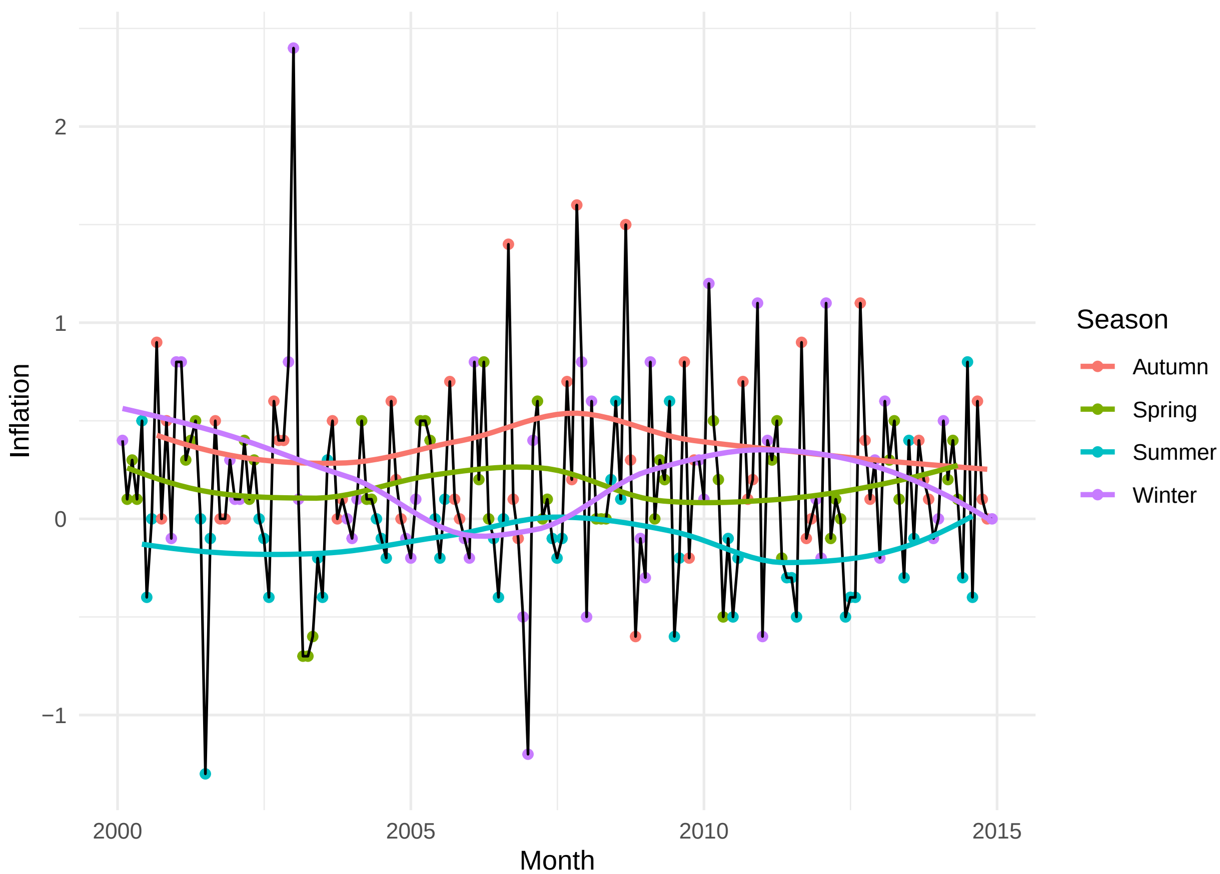


Figure 1: Inflation over time colored according to the four seasons. The colouerd lines are 'loess' fitted curves per season.

In Figure 1 the data seem normal, with timely changes in variation. There is not much evidence of any trends related to the four seasons. It seems like summer follow spring with a slight delay, suggesting that it could be beneficial to combine these two seasons.

```
mu <- mean(pd$Inflation)
sigma <- sd(pd$Inflation)
hist(pd$Inflation, breaks = 20, prob = TRUE, main = "Histogram with Normal Density Curve",
```

```r
    xlab = "Values", ylab = "Density")
curve(dnorm(x, mean = mu, sd = sigma), col = "cyan", lwd = 1, add = TRUE)
```

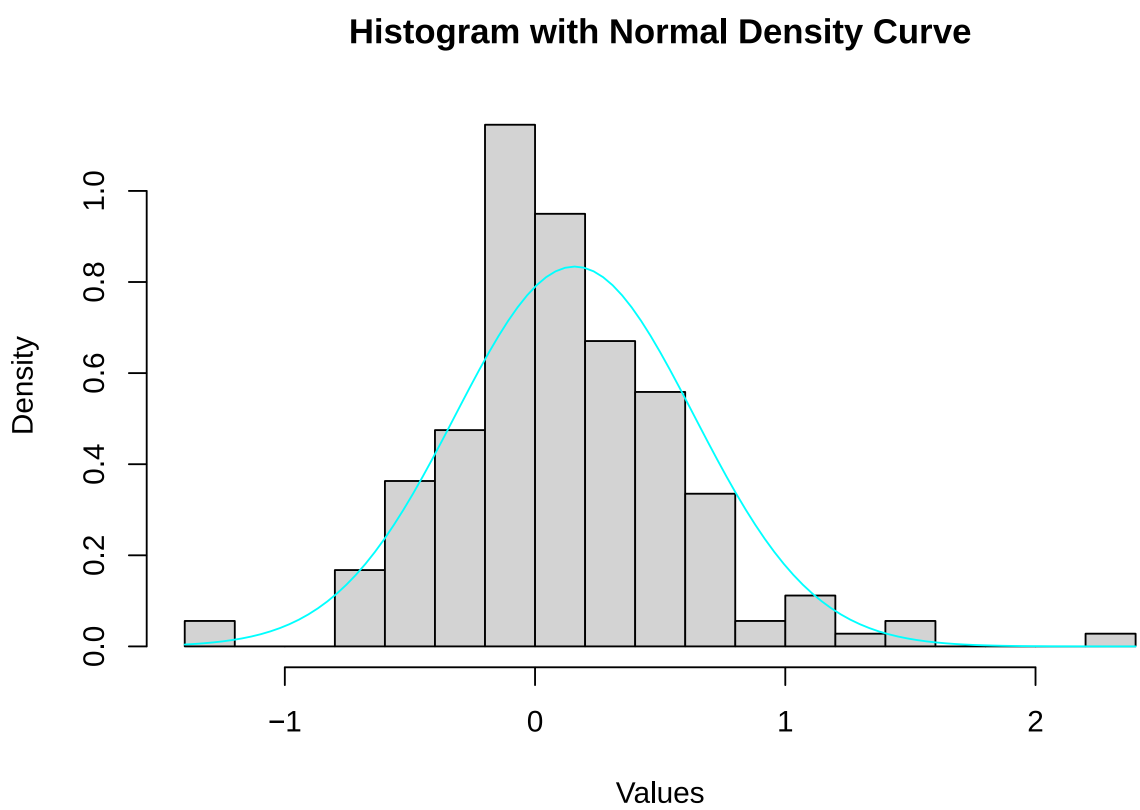


Figure 2: Histogram of inflation along with a normal curve of the sample mean and variance.

The histogram in Figure 2 also suggest normality. It also shows evidence of positive skewedness. The skewedness is significant per D'Agostino test. (Dunno much about this test. Guess we wont use it anyways.)

```r
agostino.test(pd$Inflation)
```

```
##
##  D'Agostino skewness test
##
## data:  pd$Inflation
## skew = 0.73703, z = 3.79240, p-value = 0.0001492
## alternative hypothesis: data have a skewness
```

```r
t = rep(1, 10)
for (i in 2:15) {
    t = c(t, rep(i, 12))
}
t = c(t, 16)
```

```r
pd$SeasonCount = paste0(pd$Season, t)

ggInflationSeason = ggplot(pd, aes(x = Month, y = Inflation, col = Season)) + geom_point() +
    stat_smooth(aes(fill = SeasonCount), method = "lm", se = F, show.legend = F,
        lwd = 0.5) + theme_minimal()
ggInflationSeason
```

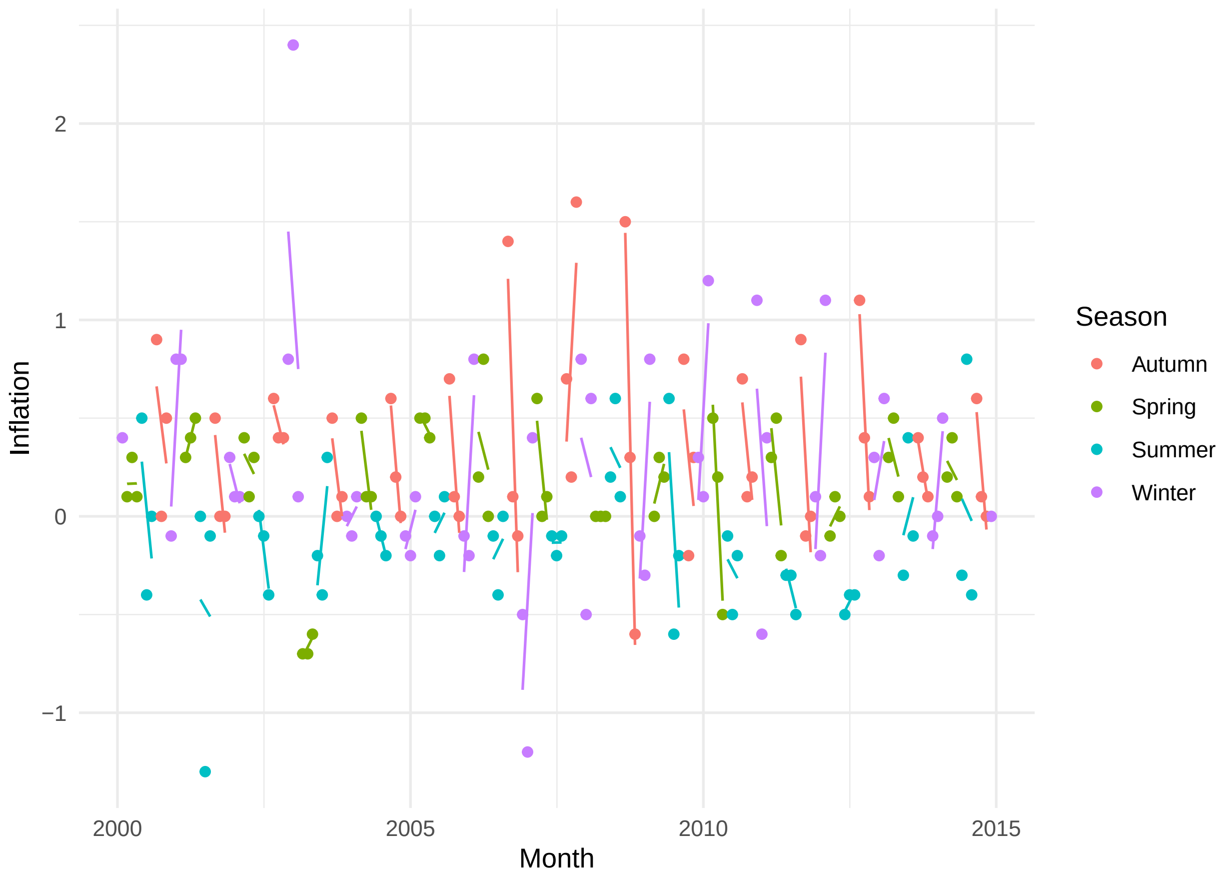


Figure 3: Linear fit to each season.

It is somewhat difficult to see any trends per season in Figure 3, so we compute some values below and show the slopes in a Boxplot.

```r
# Find season slopes ----
seasonSlopes = coef(lmList(Inflation ~ Month | SeasonCount, data = pd))["Month"]
seasons = unique(pd$Season)

seasonTrends = data.frame(row.names = c("mean", "var"))
seasonSlopes2 = list()
for (s in seasons) {
    st = na.omit(seasonSlopes$Month[startsWith(rownames(seasonSlopes), s)])
    seasonSlopes2[[s]] = st
    seasonTrends[[s]] = c(mean(st), var(st))
}
seasonTrends
```

```
##              Winter        Spring        Summer        Autumn
## mean 5.645161e-03 -2.712290e-03 -1.618297e-03 -0.0100979338
## var  9.795485e-05  2.787469e-05  2.549012e-05  0.0001140914
```

```r
boxplot(seasonSlopes2, notch = T)
abline(h = 0, lty = 2)
# abline(h=seasonTrends['mean', 'Winter'])
abline(h = CI(seasonSlopes2$Autumn, ci = 0.95), lty = 2, col = "gray")
```

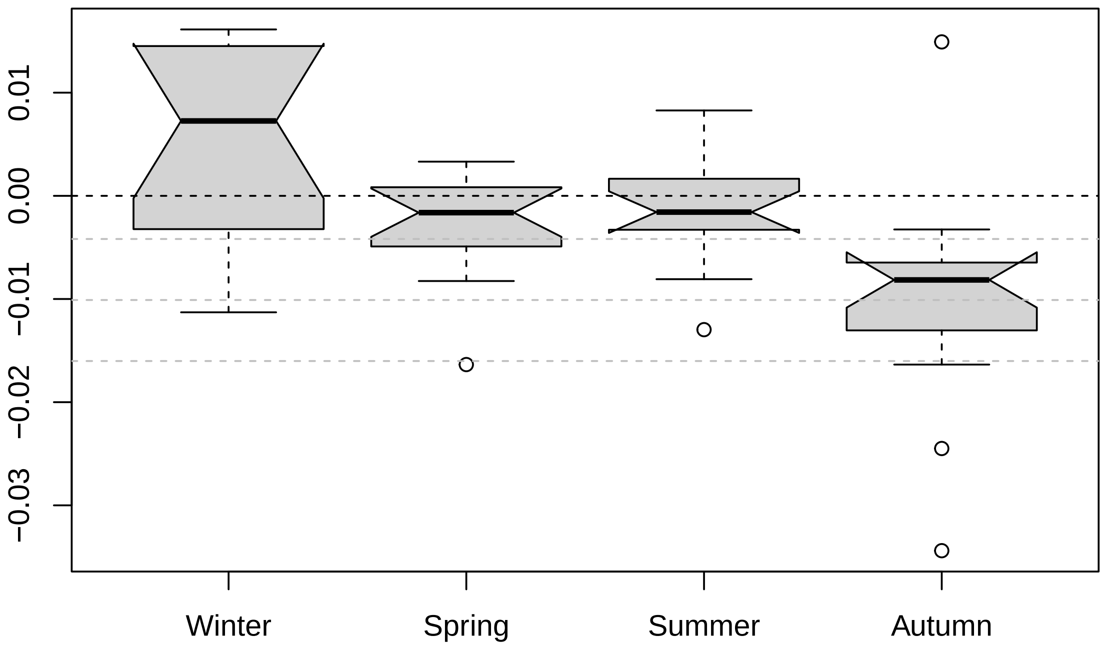


Figure 4: Boxplot of the slopes of inflation per season. Confidence interval and mean for Autumn is shown by dashed grey lines.

Figure 4 suggest that Autumn is significantly negative. That is, there is a significant decrease in inflation during the Autumn. Notches say that it is significantly different from the other seasons. Why negative inflation in the autumn tho? Saving money for the holidays?

Inflation Time Series

Let's look at acf, pacf and so on.

```r
set.seed(420)
par(mfrow = c(2, 2))
normalData = rnorm(dim(pd)[1])
```

```r
acf(pd$Inflation, lag.max = 25)
pacf(pd$Inflation, lag.max = 25)
acf(normalData, lag.max = 25)
pacf(normalData, lag.max = 25)
```

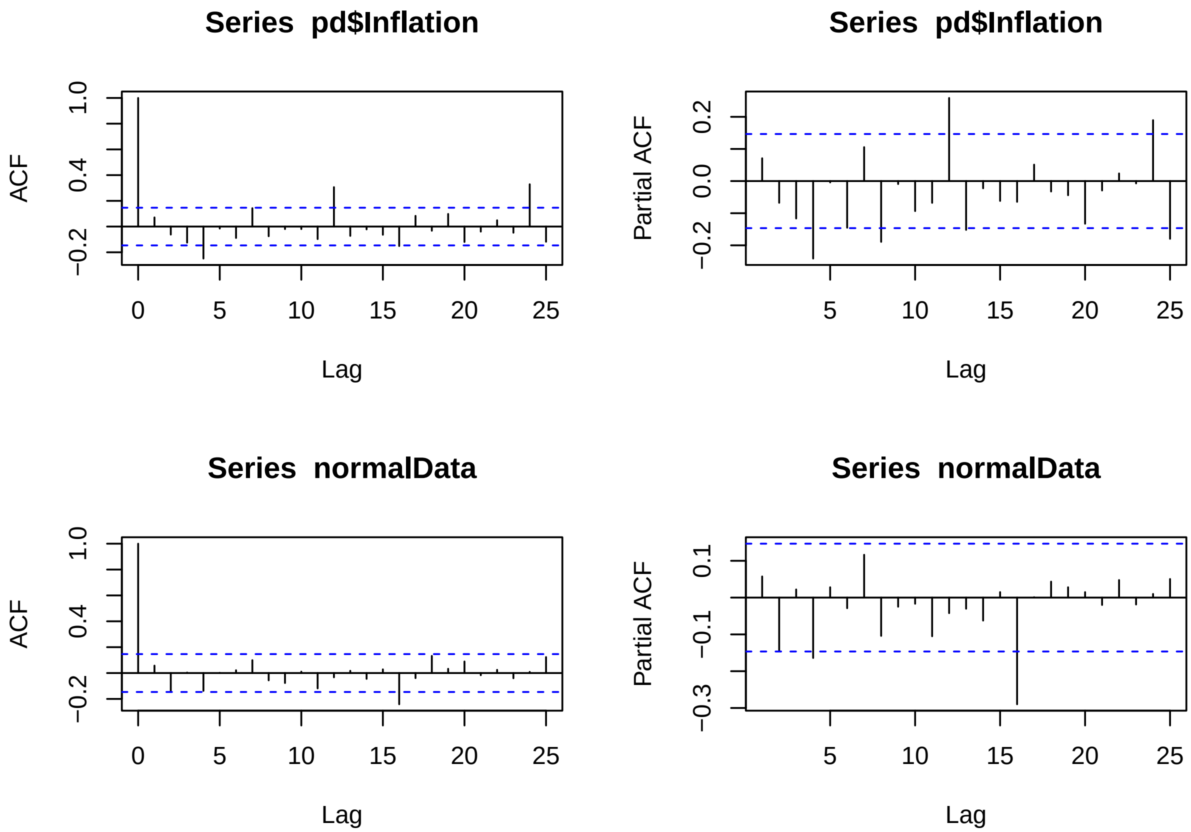


Figure 5: cap

ACF show somewhat cut off immediately, but we notice the significant 4, 12 and 24 lag.

PACF show that the 4, 8, 12, 13 and 24 lag is significant.

Covariate/Model selection

Not much to go on but the normality shown in figure 2. Suppose the inflation can be model as a regression with autocorrelated errors. Namely,

$$Y_t = \sum_{i=1}^{r} \beta_i z_{ti} + x_t, \tag{1}$$

where x_t has some covariance structure $\gamma_x(s,t)$, e.g., ARMA, z_{ti} is (unknown?) covariate i at time t and β_i are cofficients. In OLS we assume that the noise is Gaussian, that is, $\gamma_x(s,t) = \sigma_x^2$ for $s = t$, else zero. That is, the temporal x_t captures the stationary process, but not the variance variation we discussed earlier(?). More on this in Shumway & Stoffer ch. 3.8 p.143.

Assuming that the covariates are known, we can standardize the variance by multiplying by the root of the covariance matrix. Then we get the standardized (not centralized) equation

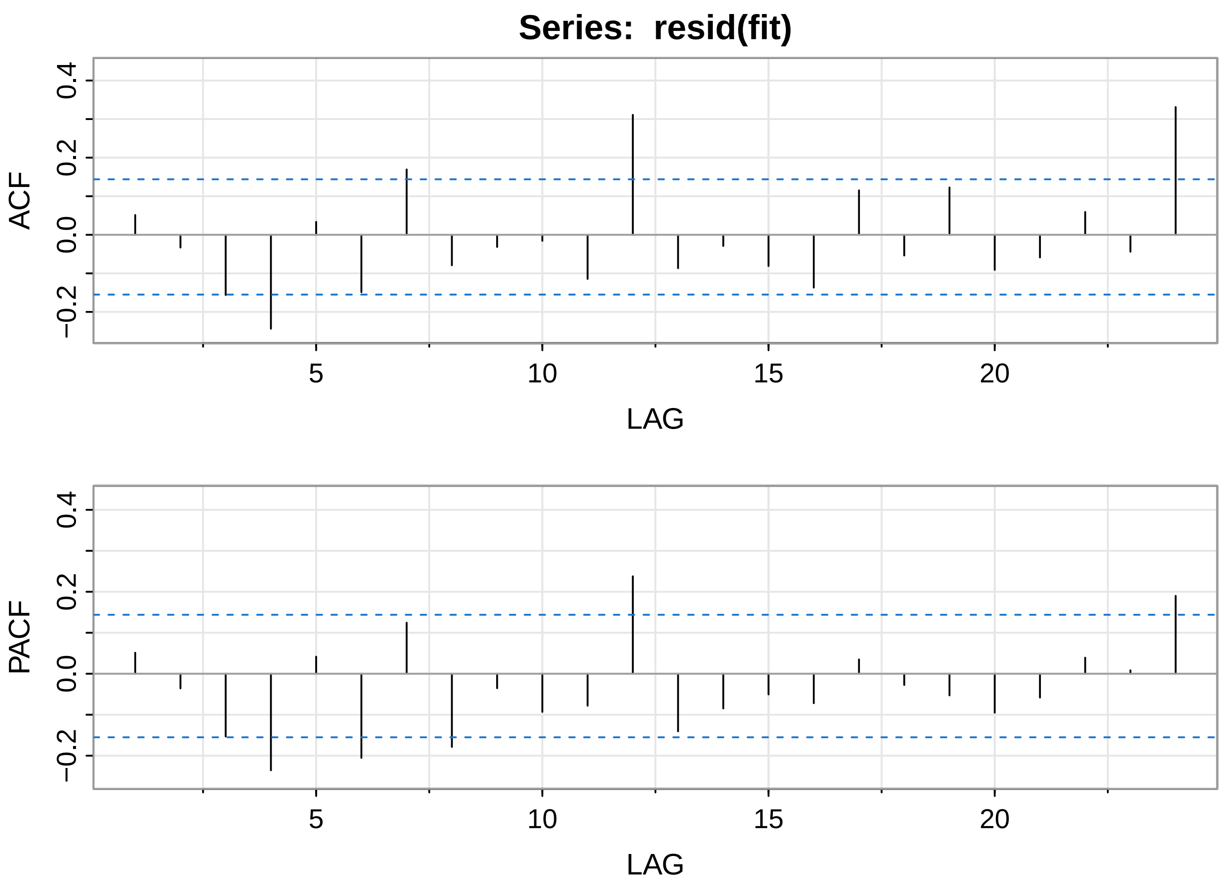
$$\boldsymbol{y}^* = \mathbf{Z}^*\boldsymbol{\beta} + \boldsymbol{\delta}, \tag{2}$$

where $\boldsymbol{y}^* = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{y}$, $\mathbf{Z}^* = \boldsymbol{\Gamma}^{-1/2}\mathbf{Z}$, blablabla... Since we are in OLS territory we want to write it in matrix notation to make coefficient estimation neat and familiar by the hat-matrix. Let's try by following r code on page 144. What sarima tho, perhaps $(1, 0, 0) \times (P =?, D =?, Q =?)_{12}$

```
summary(fit <- lm(Inflation ~ Unemployed + Bankrupt + Consumption, data = pd))  # dayum, inline variabl
```

```
##
## Call:
## lm(formula = Inflation ~ Unemployed + Bankrupt + Consumption,
##     data = pd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.26459 -0.26469 -0.07747  0.22755  2.30889
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.191e-01  1.944e-01   1.127   0.2614
## Unemployed  -3.905e-06  2.688e-06  -1.453   0.1481
## Bankrupt     6.633e-04  4.888e-04   1.357   0.1765
## Consumption -5.795e-02  2.346e-02  -2.470   0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4694 on 175 degrees of freedom
## Multiple R-squared:  0.0533, Adjusted R-squared:  0.03707
## F-statistic: 3.284 on 3 and 175 DF,  p-value: 0.02216
```

```
acf2(resid(fit), 24)
```

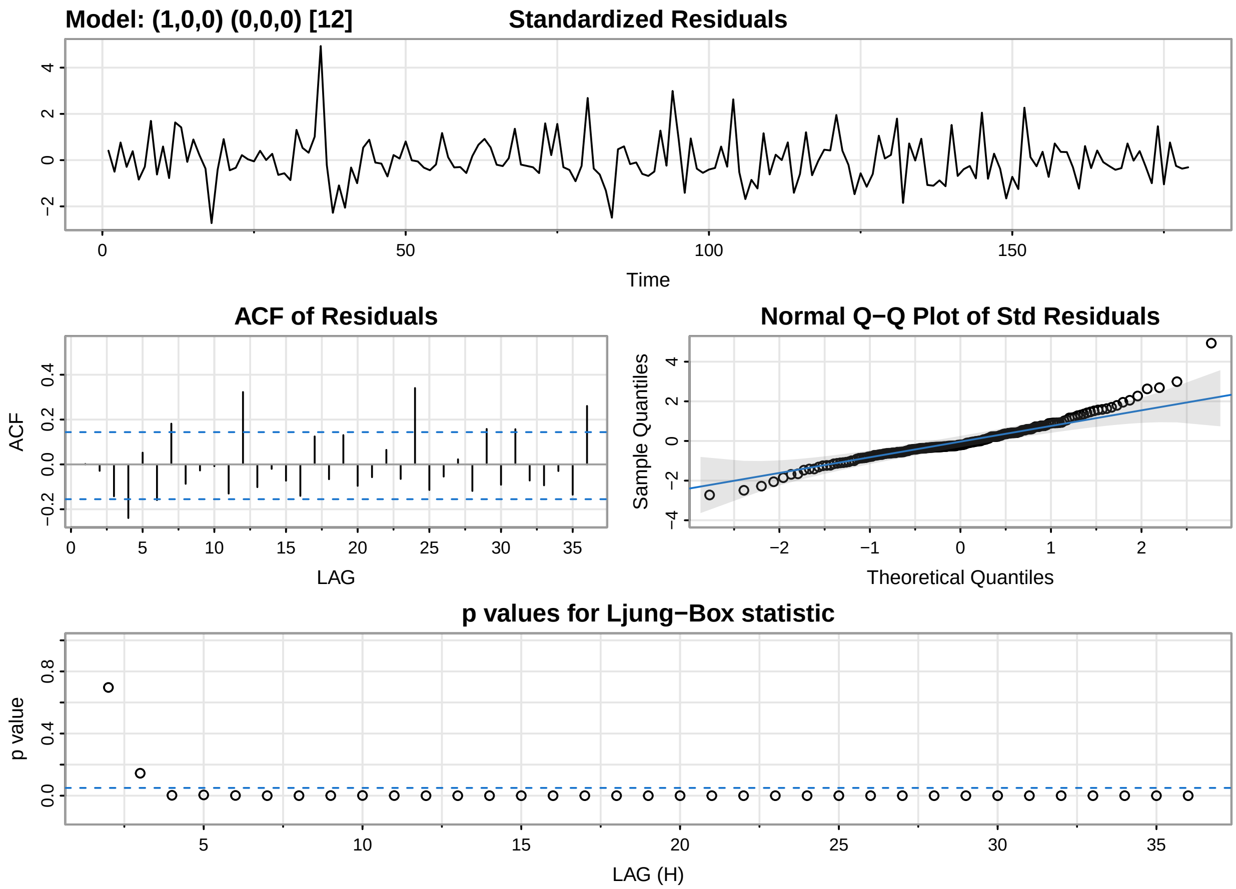



```
##      [,1]  [,2]  [,3]  [,4] [,5]  [,6] [,7]  [,8]  [,9] [,10] [,11] [,12] [,13]
## ACF  0.05 -0.03 -0.16 -0.24 0.03 -0.15 0.17 -0.08 -0.03 -0.02 -0.11  0.31 -0.09
## PACF 0.05 -0.04 -0.15 -0.24 0.04 -0.21 0.12 -0.18 -0.04 -0.09 -0.08  0.24 -0.14
##      [,14] [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23] [,24]
## ACF  -0.03 -0.08 -0.14  0.12 -0.05  0.12 -0.09 -0.06  0.06 -0.04  0.33
## PACF -0.08 -0.05 -0.07  0.03 -0.03 -0.05 -0.10 -0.06  0.04  0.01  0.19
```

```
# sarima: Seasonal components: P, D, Q, S = SAR order, seasonal diff., SMA
# order
sarima(pd$Inflation, 1, 0, 0, S = 12, xreg = cbind(Unemp = pd$Unemployed, Bankr = pd$Bankrupt,
    Cons = pd$Consumption))
```

```
## initial  value -0.765189
## iter   2 value -0.766512
## iter   3 value -0.766536
## iter   4 value -0.766541
## iter   5 value -0.766541
## iter   5 value -0.766541
## iter   5 value -0.766541
## final  value -0.766541
## converged
## initial  value -0.768858
## iter   2 value -0.768861
## iter   3 value -0.768862
```

```
## iter    3 value -0.768862
## iter    3 value -0.768862
## final  value -0.768862
## converged
```




```
## $fit
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xreg, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##     REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1  intercept  Unemp  Bankr     Cons
##       0.0520     0.2281  0e+00  6e-04  -0.0560
## s.e.  0.1122     0.2020  1e-04  5e-04   0.0236
##
## sigma^2 estimated as 0.2149:  log likelihood = -116.36,  aic = 244.73
##
## $degrees_of_freedom
## [1] 174
##
## $ttable
##             Estimate     SE t.value p.value
```

```
## ar1          0.0520 0.1122  0.4640  0.6432
## intercept   0.2281 0.2020  1.1288  0.2605
## Unemp        0.0000 0.0001 -0.0354  0.9718
## Bankr        0.0006 0.0005  1.2788  0.2027
## Cons       -0.0560 0.0236 -2.3711  0.0188
##
## $AIC
## [1] 1.367193
##
## $AICc
## [1] 1.36913
##
## $BIC
## [1] 1.474033
```