HE3021 Week 3 Tutorial 3 Attempt

Lui Yu Sen

2/3/2021

```r
knitr::opts_chunk$set(echo = TRUE)
library("tidyverse")
```

```
## Warning: package 'tidyverse' was built under R version 4.0.2

## -- Attaching packages -------------------------------------------------------------------------------

## v ggplot2 3.3.1     v purrr   0.3.4
## v tibble  3.0.1     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library("haven")
library("foreign")
library("ggplot2")
library("lmtest")
```

```
## Warning: package 'lmtest' was built under R version 4.0.3

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.2

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
library("sandwich")
```

```
## Warning: package 'sandwich' was built under R version 4.0.3
```

```r
library("forecast")
```

```
## Warning: package 'forecast' was built under R version 4.0.3

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
dat <- read_dta(file = filepath)
dat <- as.data.frame(dat)
dat$date <- as.Date(sprintf("%.2f.01", dat$date), format = "%Y.%m.%d")
```

1

(a)

(b)

```r
model <- tslm(rsp500~pcip+i3, ts(data = dat, frequency = 12, start = 1947))
summary(model)
```

```
##
## Call:
## tslm(formula = rsp500 ~ pcip + i3, data = ts(data = dat, frequency = 12,
##     start = 1947))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -157.871  -22.580    2.103   25.524  138.137
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 18.84306    3.27488   5.754 1.44e-08 ***
## pcip         0.03642    0.12940   0.281   0.7785
## i3          -1.36169    0.54072  -2.518   0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 40.13 on 554 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.01189,    Adjusted R-squared:  0.008325
## F-statistic: 3.334 on 2 and 554 DF,  p-value: 0.03637
```

An increase of 1 percentage point in the percentage of industrial production leads to an increase of 0.03642 percentage points in the monthly year-on-year SP500 returns on average, holding 3-month T-bill returns constant.
An increase of 1 percentage point in the 3-month T-bill returns leads to an decrease of 1.36169 percentage points in the monthly year-on-year SP500 returns on average, holding percentage of industrial production constant.

(c)

Only i3 is statistically significant.

(d)

$H_0 : \beta_1 = 1$, $H_a : \beta_1 \neq -1$, $\alpha = 0.05$, $df = 554$

```
critregion <- -1+c(-1,1)*qt(0.975, 554)
print(critregion)
```

```
## [1] -2.9642553  0.9642553
```

Since -1.36169 is in the critical region, then we have insufficient evidence to reject the hypothesis $\beta_1 = -1$.

(e)

Yes, it does imply that, since i3 is signficantly correlated with rsp500.

(f)

No, the model exhibits both non-stationarity and heteroscedascity. Values of rsp500 are autocorrelated with values from other time periods, and the squared residuals are significantly correlated with the explanatory variables.

```
heterotest <- cbind(model$model, na.trim(model$residuals))
colnames(heterotest)[4] <- "residualsquared"
heterotest <- mutate(heterotest, residualsquared = residualsquared**2)
heterotest <- tslm(residualsquared ~ pcip + i3, ts(heterotest, frequency = 12))
summary(heterotest)
```

```
##
## Call:
## tslm(formula = residualsquared ~ pcip + i3, data = ts(heterotest,
##     frequency = 12))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2746.4 -1387.1  -889.0   147.1 23207.6
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1236.13     240.18   5.147 3.69e-07 ***
## pcip          -19.35       9.49  -2.039   0.0419 *
## i3             86.47      39.66   2.180   0.0296 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2943 on 554 degrees of freedom
## Multiple R-squared:  0.01782,    Adjusted R-squared:  0.01427
## F-statistic: 5.026 on 2 and 554 DF,  p-value: 0.00687
```

For both of the explanatory variables, we have sufficient evidence to reject the null hypothesis that there is no correlation between the squared residuals and the explanatory variables.

```
box <- c()
for (n in 1:24){
    box_component <- Box.test(dat$rsp500, lag=n, type="Ljung-Box")$p.value
    box <- c(box, box_component)
    rm(box_component)
}
range(box)
```

```
## [1] 9.016716e-09 3.410245e-05
```

Testing for time lags of 1 to 24 months shows signficant autocorrelation between values across time periods, we have sufficient evidence to reject the null hypotheses that the values of each time period are independent of values from other time periods. All p-values are less than 0.05.
Thus, the conclusions are not reliable.

2

(a)

An increase of 1 year increases GDP by β_1 on average, holding other factors constant.
An increase of 1 percentage point of interest rate increases GDP of the same time period by β_2 on average, holding other factors constant. An increase of 1 percentage point of interest rate increases GDP of the next time period by β_3 on average, holding other factors constant, a lagged effect of 1 period.

(b)

Family income over generations in low income countries. As family income increases, they should have more resources to allow their children to consume better education, which should allow these children to secure better jobs and allow their own children to consume even better education. The regression model might look like this:

$$inc\tilde{o}me_t = \beta_0 + \beta_1 educationyears + u_i \ where \ income_t = \alpha_0 + \alpha_1 income_{t-18} + inc\tilde{o}me_t$$

(c)

This would mean oscillating values over time, as an increase in the previous period causes a decrease in the current one, and vice versa.
An example might be the Taylor rule for central banks. As inflation increases, so does the interest rate, which then leads to a decrease in inflation and thus the central bank reduces the interest rate. The regression model might look like this:

$$intere\tilde{s}trate_t = \beta_0 + \beta_1 inflationrate + u_i \ where \ interestrate_t = \alpha_0 + \alpha_1 interestrate_{t-1} + intere\tilde{s}trate_t$$

(d)

No, strict exogeneity will not hold. Domestic consumption should be correlated with the previous time period's unemployment, as it takes time for newly unemployed individuals to adjust their consumption habits. It is also correlated with GDP, since $Y = C + I + G + x - M$, where C is domestic consumption and %Y% is national income, which can be accounted using GDP.

(e)

No, we cannot assume that time series observations are independently distributed, since the order matters and the realisations are not randomly distributed, and each observation is at least dependent on its preceding and succeeding observation.

3

(a)

$$E[\begin{pmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u_T} \end{pmatrix} | \begin{pmatrix} \mathbf{x_{11}} \\ \vdots \\ \mathbf{x_{T1}} \end{pmatrix}, \begin{pmatrix} \mathbf{x_{12}} \\ \vdots \\ \mathbf{x_{T2}} \end{pmatrix}, \begin{pmatrix} \mathbf{x_{13}} \\ \vdots \\ \mathbf{x_{T3}} \end{pmatrix}] = 0$$

This means that the average error must be 0 for all time periods, independent of x-values for any time period 1 to T.

$$E(\mathbf{x_{j,k}^T} * \mathbf{u_i}) = \mathbf{E}[\begin{pmatrix} 0 & \cdots & x_{j,k} & \cdots & 0 \end{pmatrix}_{\mathbf{1*T}} \begin{pmatrix} 0 \\ \vdots \\ u_i \\ \vdots \\ 0 \end{pmatrix}_{\mathbf{T*1}}] = \mathbf{0}\ \forall \mathbf{i}, \mathbf{j} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \cdots, \mathbf{T} \textbf{ and for } \mathbf{k} = \mathbf{1}, \mathbf{2}, \mathbf{3}$$

Decomposing the vectors of x-variables of all time periods into basis vectors, this means that for each independent variable x_k, all periods' values must be orthogonal with the error term from all periods, not just when $i = j$. Since $E(\mathbf{u_i}) = \mathbf{0}$, then this means that $E(\mathbf{x_{j,k}^T} * \mathbf{u_i}) - \mathbf{E}(\mathbf{x_{j,k}^T})\mathbf{E}(\mathbf{u_i}) = \mathbf{E}(\mathbf{x_{j,k}^T} * \mathbf{u_i}) - \mathbf{0} = \mathbf{0}$.

(b)

$$E(\mathbf{x_{j,k}^T} * \mathbf{u_i}) = \mathbf{E}[\begin{pmatrix} 0 & \cdots & x_{j,k} & \cdots & 0 \end{pmatrix}_{\mathbf{1*T}} \begin{pmatrix} 0 \\ \vdots \\ u_j \\ \vdots \\ 0 \end{pmatrix}_{\mathbf{T*1}}] = \mathbf{0}\ \forall \mathbf{j} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \cdots, \mathbf{T} \textbf{ and for } \mathbf{k} = \mathbf{1}, \mathbf{2}, \mathbf{3}$$

The subscript being j for the vector of error terms means that this only needs to be true for the same time period j, for each independent variable x_k.