HW7

Tianyu Gao

2022-11-29

1

a)

Because the autocorrelation is high in the fitted model and the values close to each other in the model change fiercely (and also the intense seasonality in the model), I suspect that there is a stochastic trend in the Canadian inflation rate data. To test about this trend, we can use Dickey Fuller test to figure out whether there is a unit root trend in the model.

Suppose we want to test whether there data from 1 period before would determine the data at present(present t), equivalently

$$y_t \overset{?}{=} y_{t-1} + u_t$$

while for each t, u_t follows identical normal distributions independently. We take difference on y_t, denoting that $\Delta y_t = y_t - y_{t-1}$

In the fitted ARIMA model (pick $\mathrm{ARIMA}(1,0,1)$ to simplify the demonstration

$$y_t = \beta_1 y_{t-1} + \epsilon_{t-1} + u_t$$

we have $\Delta y_t = (\beta_1 - 1)y_{t-1} + \epsilon_{t-1} + u_t$

Denote $\varrho = \beta_1 - 1$, the test whether there is a unit root trend model then becomes the t test whether $\varrho = 0$

We have

$$\begin{aligned} H_0 &: \varrho = 0 \iff \beta_1 = 1 \iff \text{there is unit root} \\ H_A &: \varrho \neq 0 \iff \beta \neq 1 \iff \text{there is not unit root} \end{aligned}$$

For $\mathrm{ARIMA}(p,0,q)$

$$y_T = \sum_{t=1}^{p} \beta_t L^t y_T + \sum_{t=0}^{q} \alpha L^t u_T$$

The existence of unit root is equivalent to the equation $\sum_{t=1}^{p} \beta_t = 1$, similarly, we take difference on LHS, then on the RHS, we have

$$(\sum_{t=1}^{p} \beta_t - 1)y_{T-1} + \sum_{t=2}^{p} \beta_t L^t \Delta y_T$$

Then ADF test becomes to the test on the transformed regression whether the coefficient of Lag 1 term is equal to 0.

b)

```
-.10 / .05
```

```
## [1] -2
```

$$-\frac{.10}{.05} = -2 > -2.86$$

$$We fail to reject the t tests.

Then we fail to assume that there is no unit root in the model. When we consider how many lag terms to include, we should consider both the background of the problem (there is four quarters in a year), and also the plot of partial autocorrelation function to find out a benchmark model, with this benchmark model, we can use some model selection methods like step-wise selection to figure out the best model to use.

c)

For AR(1),

$$\Delta \widehat{\mathrm{InfC}_{2001:I}} = .002 - .31 * (-1.5) \approx .47$$

```
.002 - .31 *(-1.5)
```

```
## [1] 0.467
```

$$\widehat{\mathrm{InfC}_{2000:I}} = 1.3 + .47 \approx 1.8$$

$$\mathrm{FE} = \widehat{\mathrm{InfC}_{2000:1}} - \mathrm{InfC}_{2000:I} = 1.3 - 1.8 = -.5$$

For AR(4),

```
.021 + .46 * 1.5 + .39 * 1.4 - .25 * 3.5
```

```
## [1] 0.382
```

$$\Delta \widehat{\mathrm{InfC}_{2000:I}} = .021 - .46 * (-1.5) - .39 * (-1.4) - .25 * 3.5 = .382$$
$$\widehat{\mathrm{InfC}_{2000:I}} = 1.3 + .382 = 1.7$$
$$\mathrm{FE} = \widehat{\mathrm{InfC}_{2000:I}} - \mathrm{InfC}_{2000:I} = 2.1 - 1.7 = .4$$

For ADL(4, 1)

```
1.279 +.51 * 1.5 + .44 * 1.4 -.3 * 3.5 - .16 * 7
```

```
## [1] 0.49
```

$$\Delta \widehat{InfC_{2000:1}} = 1.279 + .51 * 1.5 + .44 * 1.4 - .3 * 3.5 - .16 * 7 = .49$$

$$\widehat{InfC_{2000:I}} = .49 + 1.3 \approx 1.8$$

$$FE = \widehat{InfC_{2000:I}} - InfC_{2000:I} = 2.1 - 1.8 = .3$$

d)

H_0 : The coeffiecient of URate is 0

H_A : The coefficient is not 0

$$|t_{act}| = |-\frac{1.6}{.07}| \approx 23 >> 1.96$$

Therefore we think there is Granger Causality Effect

```
-1.6/.07
```

```
## [1] -22.85714
```

Therefore we we suggest that the two variables are correlated.

2

a)

I will use a four-lag model, because AIC is preferred in model selection of AR models.

b)

$t_{act} = -2.186 > -2.86 \Rightarrow$ we fail to reject the null hypothesis that the model is stochastic$\Rightarrow$ There is a stochastic in the series and I will use difference of the data to train the model.

c)

H_0 : The model of unemployment rate is stochastic
$\iff \beta_1 + \beta_2 + \beta_3 + \beta_4 = 1 \text{ in } y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + u_t$

H_1 : The model of the unemployment rate is not stochastic
$\iff \beta_1 + \beta_2 + \beta_3 + \beta_4 \neq 1 \text{ in } y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} + \beta_4 y_{t-4} + u_t$

Because we have the statistic to be -4.791, the null hypothesis is rejected and we are sure that there is no stochastic process in the difference.

3.

a)

$$\text{Null ADL} : URCAN_t = \beta_0 + \beta_1 URCAN_{t-1} + \beta_2 URCAN_{t-2} + \beta_3 URCAN_{t-3} + \beta_4 URCAN_{t-4} + \alpha_1 URUSA_{t-1} + \alpha_2 URUSA_{t-2} + \alpha_3 URUSA_{t-3} + \alpha_4 URUSA_{t-4} + u_t$$

$$\text{Alternative ADL} : \beta_0 + \gamma_{t,0} + (\beta_1 + \gamma_{t,1})URCAN_{t-1} + (\beta_2 + \gamma_{t,2})URCAN_{t-2} + (\beta_3 + \gamma_{t,3})URCAN_{t-3} + (\beta_4 + \gamma_{t,4})URCAN_{t-4} + +(\alpha_1 + \eta_{t,1})URUSA_{t-1} + (\alpha_2 + \eta_{t,2})URUSA_{t-2} + (\alpha_3 + \eta_{t,3})URUSA_{t-3} + (\alpha_4 + \eta_{t,4})URUSA_{t-4} + u_t$$

Where in the alternative ADL

$$\begin{cases} \forall i \in [1,4], t > 1982 : II, \gamma_{t,i} = 1 \text{ otherwise} = 0 \\ \forall i[0,5], t > 1982 : II, \eta_{t,i} = 1 \text{ otherwise} = 0 \end{cases}$$

$$\because F_{9,130}(.95) = 1.95 < 1.96$$

Therefore we can reject the null hypothesis that the period before 1982 and after 1982 are identical.

b)

We can see that in the results the biggest F-statistic(QLR) > 3.11, so we can say that there is a break in the data points.

c)

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{.82}{\sqrt{24}} \approx .17$$

$$t_{act} = \frac{.003 - 0}{SE} \approx .018$$

```
.82/sqrt(24)
```

```
## [1] 0.1673818
```

```
.003 / .17
```

```
## [1] 0.01764706
```

Therefore we fail to reject the null hypothesis that the forecasting is useful. However, in the graph, the forecasted points do not move along with the actual points, and I suggest we could do some rescaling on the dataset to avoid the influence of non-normally distributed data, especially for some data pretty large.

4.

a)

```
set.seed(7)
e = rnorm(n = 100, mean = 0, sd = 1)
```

```r
y = rep(0, times = 100)
e[0]
```

```
## numeric(0)
```

```r
y[1] = e[1]
```

```r
for (i in 2:100){
  y[i] = y[i-1] + e[i]
}
```

```r
set.seed(1206)
a = rnorm(n = 100, mean = 0, sd = 1)
x = rep(0, times = 100)
for (i in 2:100){
  x[i] = x[i-1] + a[i]
}
```

```r
c = summary(lm(y~x))
print(c)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.7235 -2.3555 -0.0214  3.1765  7.7671
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.1408     0.5047  18.112  < 2e-16 ***
## x             1.1243     0.1489   7.552 2.26e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.986 on 98 degrees of freedom
## Multiple R-squared:  0.3679, Adjusted R-squared:  0.3614
## F-statistic: 57.03 on 1 and 98 DF,  p-value: 2.263e-11
```

$$t_{act} = 7.552$$
$$R^2 = 0.3624$$

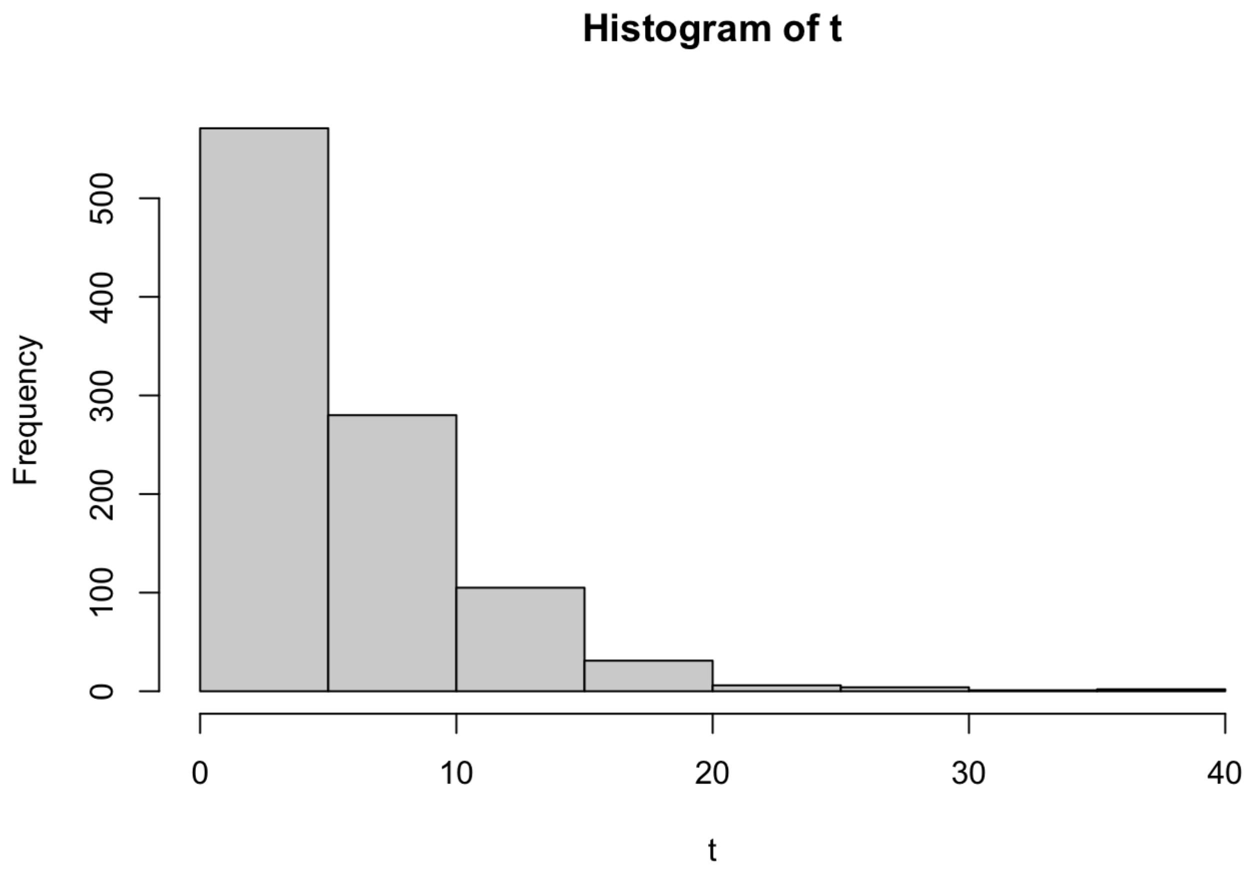
```r
sqrt(c$f)
```

```
##    value    numdf    dendf
## 7.551969 1.000000 9.899495
```

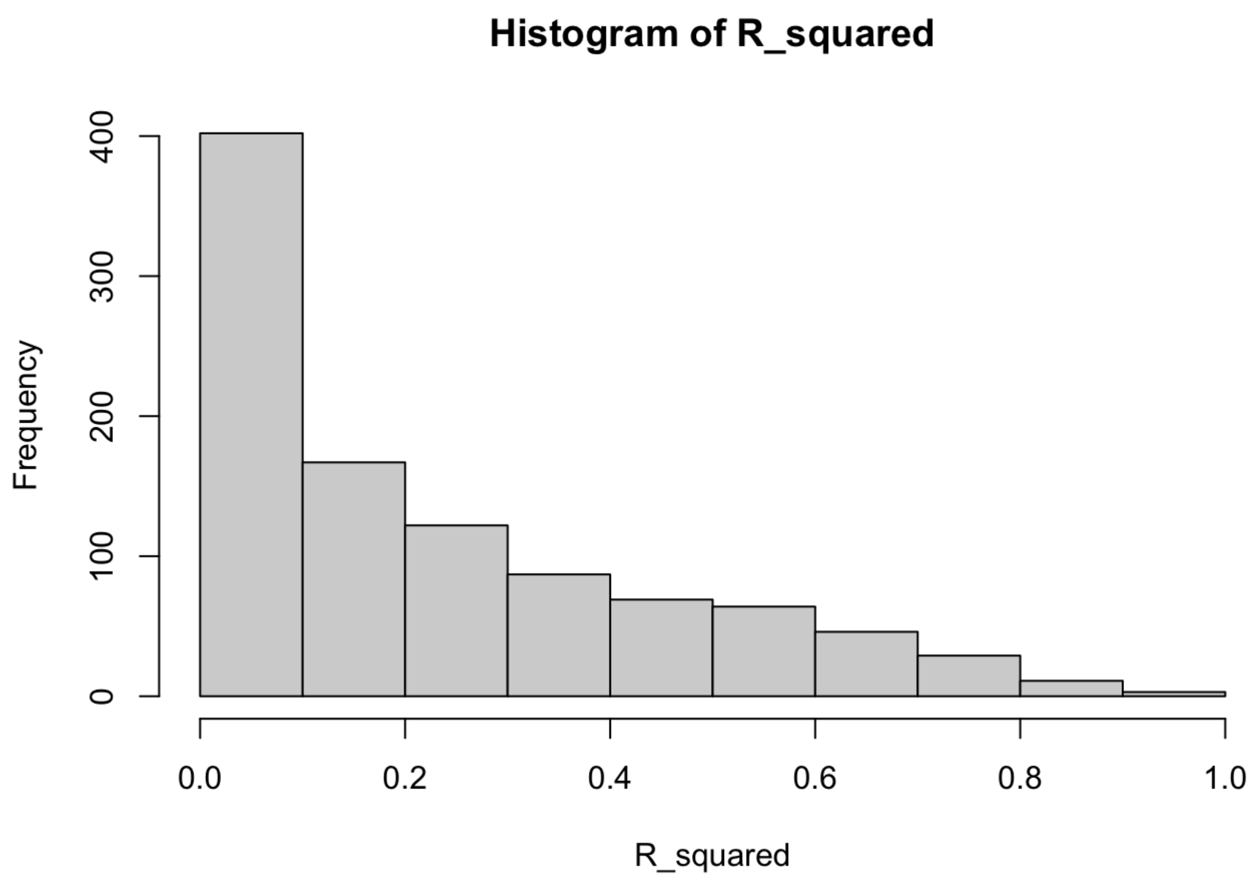
b)

```
walk = function(new){
  chain <<- append(chain, chain[length(chain)] + new)
}
t = c()
R_squared = c()
for (i in 1:1000){
 set.seed(i)
 e = rnorm(n = 100)
 chain = c(0)
 lapply(e, walk)
 x = chain[2:101]
 set.seed(15*i + 33)
 a = rnorm(n = 100)
 chain = c(0)
 lapply(a, walk)
 y = chain[2:101]
 model = summary(lm(y~x))
 t = append(t, sqrt(model$fstatistic[1]))
 R_squared = append(R_squared, model$r.squared)
}
```

```
hist(t)
```



```
hist(R_squared)
```



```
quantile(R_squared, .95)
```

```
##       95%
## 0.6795352
```

```
quantile(R_squared, 0.5)
```

```
##       50%
## 0.1451138
```

```
quantile(R_squared, 0.05)
```

```
##          5%
## 0.001872425
```

Both the t-statistic and the R^2 are exponentially distributed.

For R^2 .95 quantile is .67, .5 quantile is .15, .05 quantile is .002

```
quantile(t, .95)
```

```
##      95%
## 14.41547
```

```
quantile(t, .5)
```

```
##      50%
## 4.078618
```

```
quantile(t, .05)
```

```
##        5%
## 0.4287558
```

For t-statistics, .95 quantile is 14.42, .5 quantile is 4.08, .05 quantile is.43

```
sum(abs(t) > 1.96)/ length(t)
```

```
## [1] 0.732
```

Of all 1000 simulations, I rejected the null hypothesis for 732 times, the fraction is .732.

c)

```
walk = function(new){
  chain <<- append(chain, chain[length(chain)] + new)
}
t = c()
R_squared = c()
for (i in 1:1000){
 set.seed(i)
 e = rnorm(n = 50)
 chain = c(0)
 lapply(e, walk)
 x = chain[2:51]
 set.seed(15*i + 33)
 a = rnorm(n = 50)
 chain = c(0)
 lapply(a, walk)
 y = chain[2:51]
 model = summary(lm(y~x))
 t = append(t, sqrt(model$fstatistic[1]))
 R_squared = append(R_squared, model$r.squared)
}
```

```
sum(abs(t) > 1.96)/length(t)
```

```
## [1] 0.662
```

```
walk = function(new){
  chain <<- append(chain, chain[length(chain)] + new)
}
t = c()
R_squared = c()
for (i in 1:1000){
 set.seed(i)
 e = rnorm(n = 200)
 chain = c(0)
 lapply(e, walk)
 x = chain[2:201]
 set.seed(15*i + 33)
 a = rnorm(n = 200)
 chain = c(0)
 lapply(a, walk)
 y = chain[2:201]
 model = summary(lm(y~x))
 t = append(t, sqrt(model$fstatistic[1]))
 R_squared = append(R_squared, model$r.squared)
}
```

```
sum(t > 1.96) / length(t)
```

```
## [1] 0.822
```

```
t = c()
R_squared = c()
for (i in 1:1000){
 set.seed(i)
 e = rnorm(n = 300)
 chain = c(0)
 lapply(e, walk)
 x = chain[2:301]
 set.seed(15*i + 33)
 a = rnorm(n = 300)
 chain = c(0)
 lapply(a, walk)
 y = chain[2:301]
 model = summary(lm(y~x))
 t = append(t, sqrt(model$fstatistic[1]))
 R_squared = append(R_squared, model$r.squared)
}
```

```
sum(t > 1.96) / length(t)
```

```
## [1] 0.861
```

```
walk = function(new){
  chain <<- append(chain, chain[length(chain)] + new)
}
t = c()
R_squared = c()
for (i in 1:1000){
 set.seed(i)
 e = rnorm(n = 10000)
 chain = c(0)
 lapply(e, walk)
 x = chain[2:10001]
 set.seed(15*i + 33)
 a = rnorm(n = 10000)
 chain = c(0)
 lapply(a, walk)
 y = chain[2:10001]
 model = summary(lm(y~x))
 t = append(t, sqrt(model$fstatistic[1]))
 R_squared = append(R_squared, model$r.squared)
}
```

```
sum(t > 1.96)/length(t)
```

```
## [1] 0.978
```

By induction, the values seems to converge at 1.