Econometrics – Problem Set #8

Junwoo Yang

June 14, 2021

Exercise 15.2

The Index of Industrial Production (IP_t) is a monthly time series that measures the quantity of industrial commodities produced in a given month. This problem uses data on this index for the United States. All regressions are estimated over the sample period 1986:M1–2017:M12 (that is, January 1986 through December 2017). Let $Y_t = 1200 \times \ln(IP_t/IP_{t-1})$.

(a) A forecaster states that Y_t shows the monthly percentage change in IP, measured in percentage points per annum. Is this correct? Why?

(b) Suppose she estimates the following AR(4) model for Y_t:

$$\hat{Y}_t = 0.749 + 0.071Y_{t-1} + 0.170Y_{t-2} + 0.216Y_{t-3} + 0.167Y_{t-4}.$$

Use this AR(4) to forecast the value of Y_t in January 2018, using the following values of IP for July 2017 through December 2017:

	2017:M7	2017:M8	2017:M9	2017:M10	2017:M11	2017:M12
IP	105.01	104.56	104.82	106.58	106.86	107.30

(c) Worried about potential seasonal fluctuations in production, she adds Y_{t-12} to the autoregression. The estimated coefficient on Y_{t-12} is -0.061, with a standard error of 0.043. Is this coefficient statistically significant?

Answer. (a) The Maclaurin series of $\ln(1+x)$ is as follows:

$$\ln(1+x) = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^n}{n} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \quad \text{for } |x| < 1.$$

We can simplify $\ln(1+x)$ by considering only the first term in its Maclaurin series, that is, first order approximation $\ln(1+x) \approx x$. Then, we get

$$Y_t = 1200 \ln\left(\frac{IP_t}{IP_{t-1}}\right) = 1200 \ln\left(1 + \frac{IP_t - IP_{t-1}}{IP_{t-1}}\right) \approx 1200\left(\frac{IP_t - IP_{t-1}}{IP_{t-1}}\right)$$

which is exactly the monthly percentage change in IP that converted to annual percentage change. Thus, the statement is correct. Here, we assumed $\left|\frac{IP_t - IP_{t-1}}{IP_{t-1}}\right| < 1$.

(b) First we calculate Y_t as following:

	2017:M7	2017:M8	2017:M9	2017:M10	2017:M11	2017:M12
IP	105.01	104.56	104.82	106.58	106.86	107.30
Y		-5.153417	2.980229	19.98154	3.148428	4.9309

The forecasted value of Y_t in January 2018, $\hat{Y}_{2018:M1|2017:M12}$, is

$$0.749 + 0.071 \times 4.9309 + 0.170 \times 3.148428 + 0.216 \times 19.98154 + 0.167 \times 2.980229 = 6.448038.$$

(c) The t-statistics on Y_{t-12} is $\frac{-0.061}{0.043} = -1.418605 > -t_{0.025}$. It is not included in rejection region, so the coefficient of Y_{t-12} is not statistically significant at $\alpha = 0.05$.

Exercise 15.7

Suppose Y_t follows the stationary AR(1) model $Y_t = 2.5 + 0.7Y_{t-1} + u_t$, where u_t is i.i.d. with $E(u_t) = 0$ and $\text{Var}(u_t) = 9$.

(a) Compute the mean and variance of Y_t.

(b) Compute the first two autocovariances of Y_t.

(c) Compute the first two autocorrelations of Y_t.

(d) Suppose $Y_T = 102.3$. Compute $Y_{T+1|T} = E(Y_{T+1}|Y_T, Y_{T-1}, \cdots)$.

Answer. To show that Y_t is stationary, we need to prove following proposition.

Proposition. In AR(1) model of the form

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t, \tag{1}$$

the $\{Y_t\}$ process is stationary if $|\beta_1| < 1$ and $u_t \overset{\text{i.i.d.}}{\sim} (0, \sigma_u^2)$.

Proof. (1) can be rewritten as

$$Y_t - \alpha = \beta_1(Y_{t-1} - \alpha) + u_t$$

where $\alpha = \frac{\beta_0}{1-\beta_1}$. Let $Z_t = Y_t - \alpha$. Then, we obtain

$$\begin{aligned} Z_t &= \beta_1 Z_{t-1} + u_t \\ &= \beta_1(\beta_1 Z_{t-2} + u_{t-1}) + u_t \\ &= \beta_1(\beta_1(\beta_1 Z_{t-3} + u_{t-2}) + u_{t-1}) + u_t \\ &= \cdots \\ &= u_t + \beta_1 u_{t-1} + \beta_1^2 u_{t-2} + \cdots . \end{aligned} \tag{2}$$

From (2), AR(1) model is expressed as a linear function of u_{t-i} for $i \geq 0$. Thus,

$$E[Z_t] = E[Y_t - \alpha] = E[Y_t] - \alpha = 0$$

and $E[Y_t] = \alpha$. Since $\{u_t\}$ are mutually independent, $\text{Cov}(u_t, u_{t-j}) = E[u_t u_{t-j}] = 0$. Then, we get

$$\begin{aligned} \text{Cov}(Y_{t-1}, u_t) &= E[(Y_{t-1} - \alpha)u_t] \\ &= E[(u_{t-1} + \beta_1 u_{t-2} + \beta_1^2 u_{t-3} + \cdots)u_t] \\ &= E[u_t u_{t-1} + \beta_1 u_t u_{t-2} + \beta_1^2 u_t u_{t-3} + \cdots] \\ &= 0, \end{aligned}$$

and

$$\begin{aligned} \text{Var}(Y_t) &= E[(Y_t - \alpha)^2] \\ &= E[(\beta_1(Y_{t-1} - \alpha) + u_t)^2] \\ &= \beta_1^2 \text{Var}(Y_{t-1}) + \sigma_u^2 \\ &= \beta_1^2(\beta_1^2 \text{Var}(Y_{t-2}) + \sigma_u^2) + \sigma_u^2 \\ &= \beta_1^2(\beta_1^2(\beta_1^2 \text{Var}(Y_{t-3}) + \sigma_u^2) + \sigma_u^2) + \sigma_u^2 \\ &= \cdots \\ &= \sigma_u^2 + \sigma_u^2\beta_1^2 + \sigma_u^2\beta_1^4 + \cdots \\ &= \frac{\sigma_u^2}{1 - \beta_1^2}. \end{aligned}$$

The j-th autocovariance is

$$\begin{aligned}
\text{Cov}(Y_t, Y_{t-j}) &= E[(Y_t - \alpha)(Y_{t-j} - \alpha)] \\
&= E[(u_t + \beta_1 u_{t-1} + \beta_1^2 u_{t-2} + \cdots)(u_{t-j} + \beta_1 u_{t-j-1} + \beta_1^2 u_{t-j-2} + \cdots)] \\
&= E[\beta_1^k u_{t-j} u_{t-j} + \beta_1^{j+1} u_{t-j-1} \beta_1 u_{t-j-1} + \cdots] \\
&= \beta_1^j \sigma_u^2 + \beta_1^{j+2} \sigma_u^2 + \beta_1^{j+4} \sigma_u^2 + \cdots \\
&= \sigma_u^2 (\beta_1^j + \beta_1^{j+2} + \beta_1^{j+4} + \cdots) \\
&= \frac{\sigma_u^2 \beta_1^j}{1 - \beta_1^2}.
\end{aligned}$$

$E[Y_t]$, $\text{Var}(Y_t)$ and $\text{Cov}(Y_t, Y_{t-j})$ are all time invariant. Thus, Y_t is stationary. □

(a) Because of stationarity of Y_t,

$$E[Y_t] = \frac{\beta_0}{1 - \beta_1} = 8.33333, \quad \text{Var}(Y_t) = \frac{\sigma_u^2}{1 - \beta_1^2} = 17.64706.$$

(b) The first two autocovariances of Y_t is

$$\text{Cov}(Y_t, Y_{t-1}) = \frac{\sigma_u^2 \beta_1}{1 - \beta_1^2} = 12.35294, \quad \text{Cov}(Y_t, Y_{t-2}) = \frac{\sigma_u^2 \beta_1^2}{1 - \beta_1^2} = 8.647059.$$

(c) The first two autocorrelations of Y_t is

$$\text{Corr}(Y_t, Y_{t-1}) = \frac{\text{Cov}(Y_t, Y_{t-1})}{\text{Var}(Y_t)} = \beta_1 = 0.7, \quad \text{Corr}(Y_t, Y_{t-2}) = \frac{\text{Cov}(Y_t, Y_{t-2})}{\text{Var}(Y_t)} = \beta_1^2 = 0.49.$$

(d) The forecasted value is

$$Y_{T+1|T} = 2.5 + 0.7 Y_T = 2.5 + 0.7 \times 102.3 = 74.11.$$

Exercise 15.12

Consider the stationary AR(1) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$, where u_t is i.i.d. with mean 0 and variance σ_u^2. The model is estimated using data from time periods $t = 1$ through $t = T$, yielding the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$. You are interested in forecasting the value of Y at time $T + 1$, that is, Y_{T+1}. Denote the forecast by $\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$.

(a) Show that the forecast error is $Y_{T+1} - \hat{Y}_{T+1|T} = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) Y_T]$.

(b) Show that u_{T+1} is independent of Y_T.

(c) Show that u_{T+1} is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

(d) Show that $\text{Var}(Y_{T+1|T} - \hat{Y}_{T+1|T}) = \sigma_u^2 + \text{Var}[(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) Y_T]$.

Answer. (a) $Y_{T+1} - \hat{Y}_{T+1|T} = (\beta_0 + \beta_1 Y_T + u_{T+1}) - (\hat{\beta}_0 + \hat{\beta}_1 Y_T) = u_{T+1} - [(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1) Y_T]$.

(b) By (2), Y_T is a function of u_{T-i} for $i \leq 0$, and u_t are mutually independent. Thus, Y_T and u_{T+1} are independent.

(c) The OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of $Y_1, \cdots, Y_T$ which in turn are functions of u_{T-i} for $i \geq 0$. Similarly, u_{T+1} is independent of $\hat{\beta}_0$ and $\hat{\beta}_1$.

(d) This follows from (a)–(c) because the two terms are independent, and therefore have a zero covariance.

Empirical Exercise 15.2 (a)

Repeat the calculations reported in Table 15.2 using regressions estimated over the 1932:M1–2002:M12 sample period.

Answer.

```
library(readxl)
stock <- read_excel("Stock_Returns_1931_2002/Stock_Returns_1931_2002.xlsx")
stock <- data.frame(stock)

library(quantmod)
stock$ExReturn.L1 <- Lag(stock$ExReturn, 1)
stock$ExReturn.L2 <- Lag(stock$ExReturn, 2)
stock$ExReturn.L3 <- Lag(stock$ExReturn, 3)
stock$ExReturn.L4 <- Lag(stock$ExReturn, 4)
stock <- subset(stock, time!=1931)

ar1 <- lm(ExReturn ~ ExReturn.L1, data = stock)
ar2 <- lm(ExReturn ~ ExReturn.L1 + ExReturn.L2, data = stock)
ar3 <- lm(ExReturn ~ ExReturn.L1 + ExReturn.L2 + ExReturn.L3, data = stock)
ar4 <- lm(ExReturn ~ ExReturn.L1 + ExReturn.L2 + ExReturn.L3 + ExReturn.L4, data = stock)

library(sandwich)
rob_se <- list(sqrt(diag(sandwich(ar1))),sqrt(diag(sandwich(ar2))),sqrt(diag(sandwich(ar4))))

library(stargazer)
stargazer(ar1, ar2, ar4, se = rob_se, digits = 3, header = F, type = "text",
          omit.stat = "rsq", column.labels=c("AR(1)","AR(2)","AR(4)"), out = "Results.txt")

##
## ==========================================================================================
##                                              Dependent variable:
##                        -------------------------------------------------------------------
##                                                   ExReturn
##                                AR(1)                 AR(2)                 AR(4)
##                                 (1)                   (2)                   (3)
## ------------------------------------------------------------------------------------------
## ExReturn.L1                    0.098                 0.102*                0.099*
##                               (0.061)               (0.061)               (0.058)
##
## ExReturn.L2                                          -0.040                -0.029
##                                                     (0.057)               (0.054)
##
## ExReturn.L3                                                                -0.098*
##                                                                           (0.054)
##
## ExReturn.L4                                                                 0.006
##                                                                           (0.046)
##
## Constant                      0.524***              0.543***              0.590***
##                               (0.181)               (0.186)               (0.199)
##
## ------------------------------------------------------------------------------------------
## Observations                    852                   852                   852
## Adjusted R2                    0.009                 0.009                 0.016
## Residual Std. Error       5.135 (df = 850)      5.134 (df = 849)      5.115 (df = 847)
## F Statistic          8.359*** (df = 1; 850) 4.864*** (df = 2; 849) 4.497*** (df = 4; 847)
## ==========================================================================================
## Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```