Project 2

Antonio Jurlina

2/17/2020

Time Series Econometrics

Dr. Wiesen

Project 2

1.) Let y_t be a univariate (scalar) stochastic process with unconditional moments $E(y_t) = \mu_t$, and $E(y_t - \mu_t)(y_{t-1} - \mu_{t-1})$, for $j = 0, 1, 2, ...$

- (a) What restrictions on these moments imply that y_t is stationary?

If $E(y_t)$ is constant, and if the autocovariance between a period t and a period $t - j$ depends only on j, (the difference between periods), then y_t is stationary.

- (b) What restrictions on these moments imply that y_t is white noise?

If $E(y_t)$ is 0, and if the autocovariance between two periods that are j apart is either 0 when $j \neq 0$ or σ^2 when $j = 0$, then y_t is white noise.

- (c) Suppose we restrict the autocovariances of the process to follow the sequence $\gamma_j = (-0.75)^j \gamma_0$ where $\gamma_0 = 2$. Compute γ_2 and interpret. Do the same for γ_3 and $\lim_{j\to\infty} \gamma_j$.

$$\gamma_2 = (-0.75)^2 * 2 = 1.125$$

$$\gamma_3 = (-0.75)^3 * 2 = -0.84375$$

$$\lim_{j\to\infty} \gamma_j =$$

$$= lim_{j\to\infty}(-0.75)^j * 2$$

$$= lim_{j\to\infty}(-1)^j (0.75)^j * 2$$

$$= 0$$

since

$$\lim_{j\to\infty} 0.75^j = 0$$

Autocovariances between period 0 and another period separated by an odd number of lags indicate ever decreasing negative values. This means that every value in a period $2n + 1$ (where $n = 0, 1, 2, ...$) lags away, varies inversely with the values in the first period. Furthermore, autocovariances between period 0 and another period separated by an even number of lags indicate ever decreasing positive values. This means that every value in a period $2n$ (where $n = 0, 1, 2, ...$) lags away, varies in the same direction with the values in the first period. Ultimately, as $j \to \infty$, these oscillations settle at 0.

2.) Suppose x_t is a univariate stochastic process over the period $t = 0, 1, 2, ..., 200$, which follows a linear deterministic trend with random error:

$$x_t = 100 + 10t + \varepsilon_t$$

where ε_t is mean-zero white noise with $E(\varepsilon_t^2) = 4$.

- (a) Compute an expression for the unconditional mean of x_t. Compute the mean in the last time period.

$$E(x_t) = E(100 + 10t + \varepsilon_t)$$
$$E(x_t) = E(100) + E(10t) + E(\varepsilon_t)$$
$$E(x_t) = 100 + 10E(t) + 0$$
$$E(x_t) = 100 + 10t$$
$$E(x_{200}) = 100 + 10(200) = 2100$$

- (b) Compute the variance of x_t.

$$var(x_t) = E(x_t^2) - [E(x_t)]^2$$
$$var(x_t) = E[(100 + 10t + \varepsilon_t)^2] - (100 + 10t)^2$$
$$var(x_t) = E(100^2 + 2000t + 100t^2 + 200\varepsilon_t + 20t\varepsilon_t + \varepsilon_t^2) - 100^2 - 2000t - 100t^2$$
$$var(x_t) = E(100^2) + E(2000t) + E(200\varepsilon_t) + E(100t^2) + E(200\varepsilon_t) + E(20t\varepsilon_t) + E(\varepsilon_t^2) - 100^2 - 2000t - 100t^2$$
$$var(x_t) = E(\varepsilon_t^2)$$
$$var(x_t) = 4$$

- (c) Compute the first autocorrelation coefficient of x_t (ρ_1).

$$\rho_1 = \frac{\gamma_1}{\gamma_0}$$
$$\rho_1 = \frac{E[(x_t - 100 - 10t)(x_{t+1} - 100 - 10(t+1))]}{4}$$
$$\rho_1 = \frac{E[(100 + 10t + \varepsilon_t - 100 - 10t)(100 + 10t + 10 + \varepsilon_{t+1} - 100 - 10t + 10)]}{4}$$
$$\rho_1 = \frac{E(\varepsilon_t * \varepsilon_{t+1})}{4}$$
$$\rho_1 = 0$$

- (d) Is x_t stationary? Why or why not? Is Δx_t stationary?

Since $E(x_t)$ depends on t, x_t is not stationary.

$$E(\Delta x_t) = E(x_t - x_{t-j}) = E(100 + 10t + \varepsilon_t - 100 - 10t + 10j - \varepsilon_{t-j}) = E(10j) = 10j$$
$$\gamma_j = E[(100 + 10t + \varepsilon_t - 100 - 10t)(100 + 10t - 10j + \varepsilon_{t-j} - 100 - 10t + 10j)]$$
$$\gamma_j = E(\varepsilon_t * \varepsilon_{t-j})$$

Since Δx_t has a constant expected value and the covariance between two lagged terms depends only on the difference between them, not on time, we can conclude that Δx_t is stationary.

3.) The "Hot Hand" in basketball refers to the hypothesis that players have an increased probability of making a shot following a made shot. A general way to interpret this hypothesis is that shooting success in the sport is a positively serially correlated stochastic process. (Here is a classic study.) Whether this hypothesis is valid is an empirical question that time series analysis can help us explore. Suppose we model a slightly altered version of the Hot Hand hypothesis with a first-order (stochastic) difference equation:

$$x_t = 16 + 0.2x_{t-1} + \varepsilon_t$$

where x_t measures a player's points-per-game in game t. (Note that in the context of this model t refers to game days, and does not have to represent equal time intervals like days or weeks.) ε_t is assumed to be a mean-zero white noise process. Thus, the above equation is an AR(1) process.

- (a) What is the player's "steady-state" points-per-game? (This can be interpreted as the player's unconditional scoring average.)

$$x_t = 16 + 0.2(16 + 0.2x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$

$$x_t = 16 + 0.2(16) + 0.2(0.2)(x_{t-2}) + 0.2\varepsilon_{t-1} + \varepsilon_t$$

$$x_t = 16 + 0.2(16) + 0.2(0.2)(16 + 0.2x_{t-3} + \varepsilon_{t-2}) + 0.2\varepsilon_{t-1} + \varepsilon_t$$

$$x_t = (0.2)^0 16 + (0.2)^1 16 + (0.2)^2 16 + (0.2)^3 x_{t-3} + (0.2)^0 \varepsilon_{t-0} + (0.2)^1 \varepsilon_{t-1} + (0.2)^2 \varepsilon_{t-2}$$

$$x_t = 16 \sum_{i=0}^{n} 0.2^i + 0.2^{n+1} x_{t-n-1} + \sum_{i=0}^{n} 0.2^i \varepsilon_{t-i}$$

$$x_t = \lim_{n\to\infty} [16 \sum_{i=0}^{n} 0.2^i + 0.2^{n+1} x_{t-n-1} + \sum_{i=0}^{n} 0.2^i \varepsilon_{t-i}]$$

$$x_t = 20 + 0 + \sum_{i=0}^{\infty} 0.2^i \varepsilon_{t-i}$$

$$E(x_t) = E(20) + E(\sum_{i=0}^{\infty} 0.2^i \varepsilon_{t-i})$$

$$E(x_t) = 20 + 0$$

$$E(x_t) = 20$$

- (b) In the first game of the season, the player scores 6 points above her average. How many points does the model predict she will score in her second game?

$$E(x_2 | x_1 = 26) =$$

$$= E(16 + 0.2(26) + \varepsilon_2 | x_1 = 26)$$

$$= 21.2 + E(\varepsilon_2 | x_1 = 26)$$

$$= 21.2 + 0$$

$$21.2$$

- (c) In her most recent game, the player's "error term" ε_t is 10 points. How does this fact alter her expected points-per-game in each of the following two games? That is, how does this "shock" to her shooting process affect her expected points-per-game at time $t+1$ and $t+2$?

$$x_t = 16 + 0.2x_{t-1} + 10$$

$$x_{t+1} = 16 + 0.2x_t + \varepsilon_{t+1}$$

$$x_{t+1} = 16 + 0.2(16 + 0.2x_{t-1} + 10) + \varepsilon_{t+1}$$

$$x_{t+1} = (0.2)^0 16 + (0.2)^1 16 + (0.2)^2 16 + (0.2)^3 x_{t-3} + (0.2)^0 \varepsilon_{t+1} + (0.2)^1 10 + (0.2)^2 \varepsilon_{t-2}$$

and since error terms have expected value 0, we end up with

$$E(x_{t+1}|\varepsilon_t = 10) = 20 + 0.2(10) = 22.$$

Similarly, for x_{t+2}, we end up with

$$E(x_{t+2}|\varepsilon_t = 10) = 20 + 0.2^2(10) = 20.4.$$

Table 1: Natural log of industrial production

mean	variance	mean (first diff)	variance (first diff)
3.900904	0.3740432	0.0023825	8.94e-05

4.) Download the monthly time series for industrial production (*INDPRO.xlsx*) from Blackboard. The time sample in the file covers January 1947 to December 2018.

- (a) For the *natural log* of the industrial production and the first difference of the log of industrial production, compute the sample mean, the sample variance, and the first 12 autocorrelations. Report these results in a table.

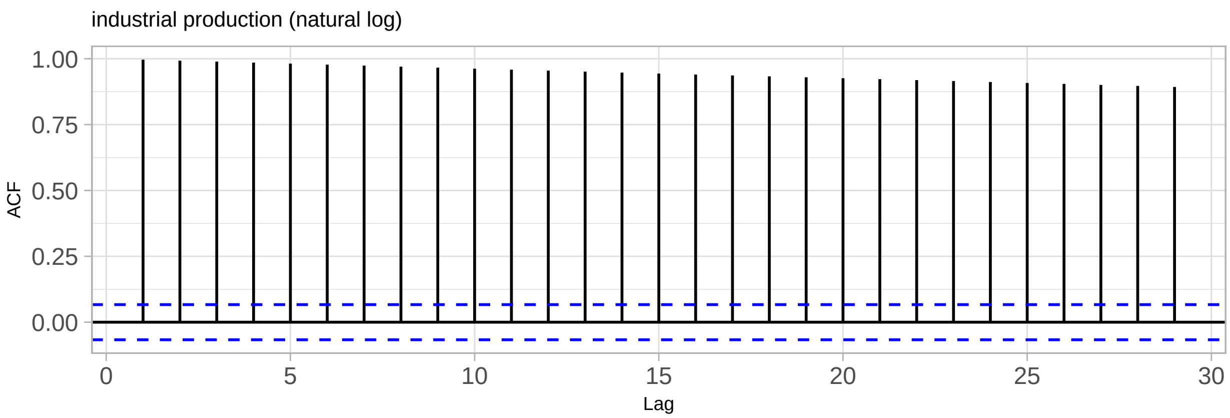
```r
table1 <- tibble(`mean` = mean(log_industrial_production),
                 `variance` = var(log_industrial_production))
table2 <- tibble(`mean (first diff)` = mean(log_diff_industrial_production,
                                            na.rm = TRUE),
                 `variance (first diff)` = var(log_diff_industrial_production,
                                               na.rm = TRUE))

table3 <- tibble(lag = c(1:12),
          `ln(industrial production)` = acf1(log_industrial_production,
                                             plot = FALSE,
                                             max.lag = 12),
          `ln(industrial production first diff)` = acf1(log_diff_industrial_production,
                                                        plot = FALSE,
                                                        max.lag = 12))
```

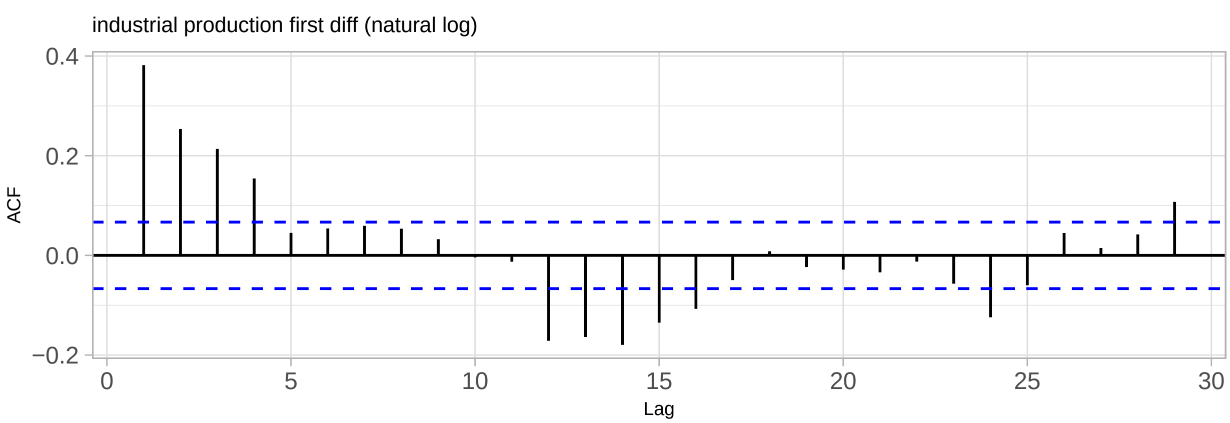
Table 2: Autocorrelation

lag	ln(industrial production)	ln(industrial production first diff)
1	0.9964364	0.3817068
2	0.9927981	0.2536130
3	0.9891164	0.2136951
4	0.9853429	0.1543489
5	0.9815361	0.0451260
6	0.9777238	0.0540477
7	0.9738711	0.0592840
8	0.9700274	0.0535316
9	0.9661658	0.0325128
10	0.9623471	-0.0039447
11	0.9585790	-0.0127197
12	0.9548251	-0.1713938

- (b) Do either of these series appear to be white noise? Why or why not?

```
plot_grid(ggAcf(log_industrial_production) +
            theme_light() +
            ggtitle("industrial production (natural log)") +
            theme(title = element_text(size = 7)),
          ggAcf(log_diff_industrial_production) +
            theme_light() +
            ggtitle("industrial production first diff (natural log)") +
            theme(title = element_text(size = 7)),
          ncol = 1, nrow = 2)
```





Neither of these series appear to be white noise given that they both show statistically significant autocorrelation.

- (c) Over this sample period, what is the average *annual* growth rate of industrial production? (Again use the log growth rate here.)

```
reg_c <- lm(log_industrial_production ~ year(date), project_2_data) %>%
  tidy() %>%
  mutate(term = c("intercept", "year"))
```

term	estimate	std.error	statistic	p.value
intercept	-53.1094349	0.4165566	-127.4963	0
year	0.0287568	0.0002101	136.8685	0

The average *annual* growth rate is 0.0288.

- (d) Estimate a simple linear-trend model for the log of industrial production. Report the constant and slope coefficient estimates from this model, along with estimated standard errors, p-values, and interpret those coefficients.

```
reg_d <- lm(log_industrial_production ~ seq_along(log_industrial_production)) %>%
  tidy() %>%
  mutate(term = c("intercept", "slope"))
```

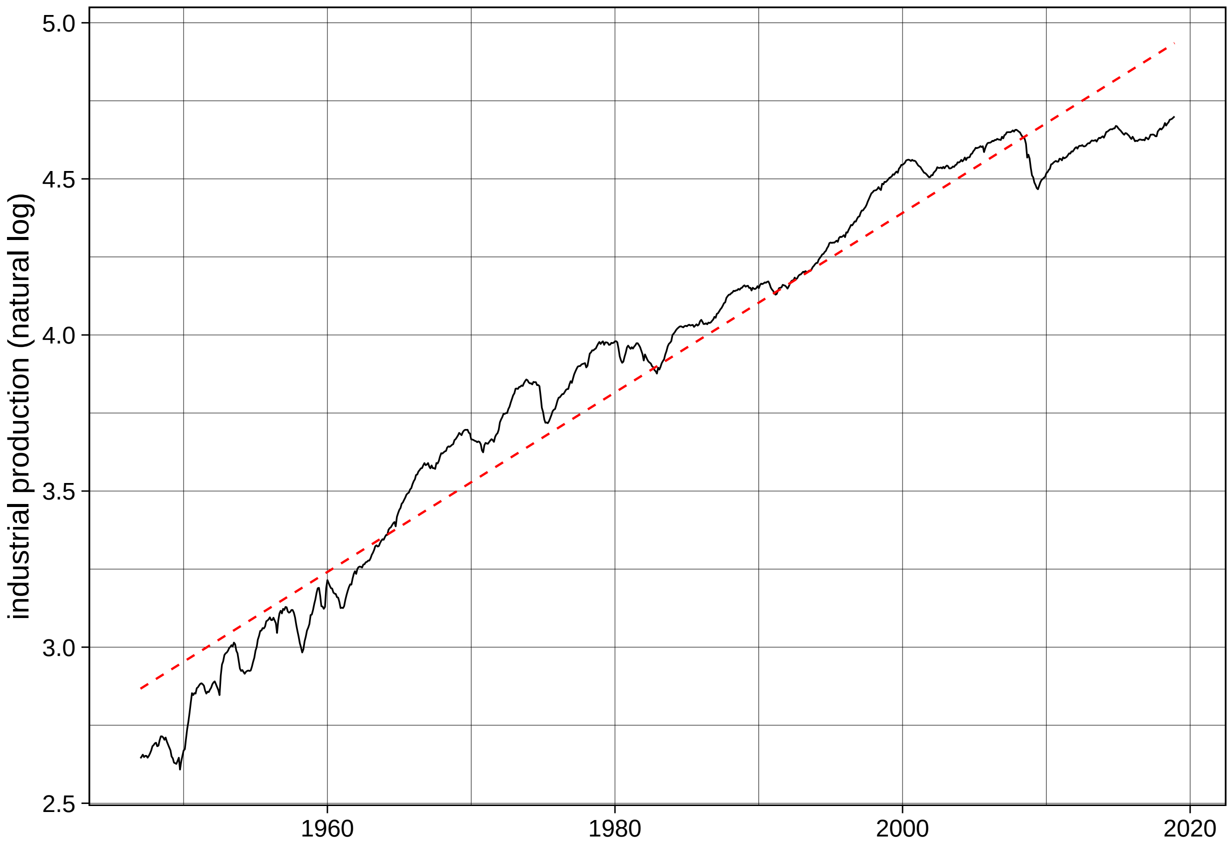
term	estimate	std.error	statistic	p.value
intercept	2.8644617	0.0087226	328.3972	0
slope	0.0023964	0.0000175	137.1656	0

The expected value for the natural log of the industrial production is 2.8645, at the very start of the series. Slope of 0.0024 shows an average expected growth in the industrial production of 0.24 percent monthly.

- (e) Plot on the same graph the log of industrial production and its fitted linear time trend.

```
ggplot(project_2_data, aes(date, log_industrial_production)) +
  geom_line(size = 0.3) +
  stat_smooth(method = "lm", se = FALSE, size = 0.4, linetype = "dashed", color = "red") +
  theme_linedraw() +
  ylab("industrial production (natural log)") +
  theme(
    axis.title.x = element_blank()
  )
```

```
## `geom_smooth()` using formula 'y ~ x'
```



More at https://github.com/antoniojurlina/time_series_econometrics