IIA-3 Econometrics: Supervision 8

Emre Usenmez

Lent Term 2025

Topics Covered

Faculty Qs:

Supplementary Qs: stationary stochastic process; AR(1) process; random walk; correlogram; autocorrelation function (ACF); Box-Pierce Q statistic; Ljung-Box (LB) statistic; unit root test; Dickey-Fuller (DF) test; tau statistic; augmented Dickey-Fuller (ADF) test; overdifferencing; underdifferencing; difference-stationary process; trend-stationary process; detrended time series; autoregressive distributed lag (ARDL) model; spurious regression; cointegration; error correction mechanism (ECM); disequilibrium error; and error correction term

Related Reading:

Dougherty (2016), *Introduction to Econometrics*, 5th ed, OUP

> Chapter 11: Models Using Time Series Data

> Chapter 12: Autocorrelation

> Chapter 13: Introduction to Nonstationary Time Series

Wooldridge J M (2021) *Introductory Econometrics: A Modern Approach*, 7th ed,

> Chaoter 10: Basic Regression Analysis with Time Series Data

> Chapter 11: Further Issues in Using OLS with Time Series Data

> Chapter 12: Serial Correlation and Heteroskedasticity in Time Series Regressions

Gujarati, D N and Porter, D (2009) *Basic Econometrics*, 7th International ed, McGraw-Hill

> Chapter 21: Time Series Econometrics: Some Basic Concepts

Gujarati, D (2022) *Essentials of Econometrics*, 5th ed, Sage

> Chapter 21: Elements of Time-Series Econometrics

Stock, J H and Watson M W (2020) *Introduction to Econometrics*. 4th Global ed, Pearson

> Chapter 15: Introduction to Time Series Regression and Forecasting

FACULTY QUESTIONS

QUESTION A:

SUPPLEMENTARY QUESTIONS

QUESTION 1

(i) State the conditions under which a univariate stochastic process, z_t, is (covariance) stationary.

Answer: In the context of regression with time series data, the idea that historical relationships can be generalized to the future is formalized by the concept of stationarity. A time series is called *stationary* if its statistical properties remain constant over time. Formally, a stochastic process[1] z_t is *strictly* stationary if the joint distribution of $\{z_t : t = 1, 2, \dots\}$ is the same as the joint distribution of $\{z_{t+k} : t = 1, 2, \dots\} \ \ \forall \, k \in \mathbb{Z}^+$.

This means that the joint distribution of the first two terms in the sequence, (z_1, z_2), for example, must be the same as the joint distribution of (z_6, z_7), or of (z_t, z_{t+1}) for any $t \geq 1$. Note that this places no restrictions on how z_t and z_{t+1} are related to each other. What stationarity requires is that the nature of any correlation between adjacent terms is the same across all time periods.

On the other hand, a stochastic process z_t is *weakly stationary* if

- mean is constant over time: $\mathbb{E}(z_t) = \mu$;

- variance is constant over time: $Var(z_t) = \mathbb{E}(z_t - \mu)^2 = \sigma^2$; and

- covariance between the two time periods is constant over time and depends only on the distance / gap / lag between them: $Cov(z_t, z_{t+k}) = \mathbb{E}\big[(z_t - \mu)(z_{t+k} - \mu)\big] = \gamma_k$

Notice that μ, σ^2, and γ_k are finite-valued numbers that do not depend on time t, and covariance γ_k depends on k. Weakly stationary stochastic process is also referred to as *covariance stationary stochastic process, second order stationary stochastic process*, or *wide sense stationary stochastic process*. For the purposes of this class, and in most practical situations, this type of stationarity is usually sufficient.

A time series is, therefore, strictly stationary if all the moments of its probability distribution are invariant over time, and weak stationary if just the first two moments - mean and variance - are invariant over time. Yet, if the stationary process is normal, then the weakly stationary stochastic process is also strictly stationary, for the normal stochastic process is fully specified by its two moments, mean and variance.

(ii) Write out the moving average representation of the autoregressive process

$$z_t = \rho z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma^2)$$

where $NID(0, \sigma^2)$ denotes normally and independently distributed with mean zero and variance σ^2. Hence, show it is stationary provided that $|\rho| < 1$.

Answer: In order to derive the moving average representation of the given AR(1) process, begin by substituting z_{t-1}:

$$\begin{aligned} z_t &= \rho z_{t-1} + \varepsilon_t \\ &= \rho(\rho z_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \end{aligned}$$

---

[1]A stochastic process is a collection of random variables ordered in time. The word "stochastic" comes from the Greek word "stokhos" meaning "target" or "bull's eye".

$$= \rho^2 z_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

If we then continue with substituting z_{t-2}:

$$z_t = \rho^2(\rho z_{t-3} + \varepsilon_{t-2}) + \rho \varepsilon_{t-1} + \varepsilon_t$$
$$= \rho^3 z_{t-3} + \rho^2 \varepsilon_{t-2} + \rho \varepsilon_{t-1} + \varepsilon_t$$

If we do this type of substitution recursively t times, then we get the moving average representation:

$$z_t = \rho^t z_{t-t} + \sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i}.$$

Next, in order to show that it is stationary, the three criteria in part (i) need to be satisfied.

Criterion 1: $\mathbb{E}(z_t) = \mu$

Provided that $|\rho| < 1$, as $t \to \infty$, $\rho^t \to 0$, so asymptotically

$$z_t = \sum_{i=1}^{\infty} \rho^i \varepsilon_{t-i}.$$

Since $\varepsilon_t \sim NID(0, \sigma^2)$ and thus have 0 mean, $\mathbb{E}(z_t) = 0$. Alternatively, it can be constant if $\mathbb{E}(z_0) = 0$. Either approach satisfies this criterion.

Criterion 2: $Var(z_t) = \mathbb{E}(z_t - \mu)^2 = \sigma^2$

$$Var(z_t) = Var(\rho^t z_0) + Var\left(\sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i}\right)$$

$$= 0 + \sum_{i=0}^{t-1} Var(\rho^i \varepsilon_{t-i})$$

since $\rho^t z_0$ is constant its variance is 0, and since ε_t are independent, the variance of their sum equals the sum of their variances.

Furthermore, since

$$Var(\rho^i \varepsilon_{t-i}) = (\rho^i)^2 Var(\varepsilon_{t-i}) = \rho^{2i} \sigma_\varepsilon^2$$

the variance of the stochastic process z_t becomes:

$$Var(z_t) = \sum_{i=0}^{t-1} Var(\rho^i \varepsilon_{t-i})$$

$$= \sum_{i=0}^{t-1} \rho^{2i} \sigma_\varepsilon^2$$

$$= \sigma_\varepsilon^2 \sum_{i=0}^{t-1} \rho^{2i}$$

$$= \sigma_\varepsilon^2(1 + \rho^2 + \rho^4 + \rho^6 + \dots)$$

which is a geometric sum. That is,

$$Var(z_t) = \sigma_\varepsilon^2 \rho^0 + \sigma^2 \rho^2 + \sigma^2 \rho^4 + \sigma^2 \rho^6 + \dots + \sigma_\varepsilon \rho^{t-1} = \sigma_\varepsilon^2 \sum_{i=0}^{t-1} \rho^{2i}$$

$$\rho^2 Var(z_t) = \sigma_\varepsilon^2 \rho^2 + \sigma_\varepsilon^2 \rho^4 + \sigma_\varepsilon^2 \rho^6 + \dots + \sigma_\varepsilon \rho^{t+1}$$

$$Var(z_t) - \rho^2 Var(z_t) = \sigma_\varepsilon^2 \rho^0 - \sigma_\varepsilon^2 \rho^{t+1}$$
$$Var(z_t)(1 - \rho^2) = \sigma^2(1 - \rho^{t+1})$$
$$Var(z_t) = \sigma_\varepsilon^2 \left(\frac{1 - \rho^{t+1}}{1 - \rho^2} \right)$$
$$\lim_{t \to \infty} Var(z_t) = \lim_{t \to \infty} \frac{\sigma_\varepsilon^2(1 - \rho^{t+1})}{1 - \rho^2}$$
$$= \frac{\sigma_\varepsilon^2}{1 - \rho^2} - \frac{\sigma_\varepsilon^2}{1 - \rho^2} \lim_{t \to \infty} \rho^{t+1}$$
$$= \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

which can only converge if the stability condition $|\rho| < 1$ holds. Since that is the case in this question we therefore have a constant variance over time that is not dependent on t, thus the second criterion is also satisfied.

Criterion 3: $Cov(z_t, z_{t-k}) = \mathbb{E}\big[(z_t - \mu)(z_{t-k} - \mu)\big] = \gamma_k$

$$Cov(z_t, z_{t-k}) = \gamma_k = Cov\left(\sum_{i=0}^{t-1} \rho^i \varepsilon_{t-i} \ , \ \sum_{j=0}^{t-k-1} \rho^j \varepsilon_{t-k-j} \right)$$
$$= \sum_{i=0}^{t-1} \sum_{j=0}^{t-k-1} \rho^i \rho^j Cov(\varepsilon_{t-i} \ , \ \varepsilon_{t-k-1})$$

Since $\varepsilon_t \sim NID(0, \sigma^2)$, the error terms are independent which means their covariances are 0. The only time when we would have non-zero covariance is if they are the same error term where we have the variance, σ^2. This means,

$$Cov(\varepsilon_{t-i}, \varepsilon_{t-k-j}) = \begin{cases} \sigma^2 & \text{if they refer to the same error term} \\ 0 & \text{otherwise} \end{cases}$$

Therefore, most of the terms in the double summation are 0. We only need to sum over the cases where $t - i = t - k - j$ which only happens when $j = i - k$. Since j has to be non-negative, this also means $k \le i$. So the summation reduces to:

$$Cov(z_t, z_{t-k}) = \gamma_k = \sum_{i=0}^{t-1} \sum_{j=0}^{t-k-1} \rho^i \rho^j Cov(\varepsilon_{t-i} \ , \ \varepsilon_{t-k-1})$$
$$= \sum_{i=k}^{t-1} \rho^i \rho^{i-k} \sigma^2$$
$$= \sigma^2 \sum_{i=k}^{t-1} \rho^{2i-k}$$
$$= \sigma^2 \rho^{-k} \sum_{i=k}^{t-1} \rho^{2i}$$
$$= \sigma^2 \rho^k \sum_{i=k}^{t-1} \rho^{2(i-k)}$$

If we let $j = i - k$, then this becomes:

$$Cov(z_t, z_{t-k}) = \gamma_k = \sigma^2 \rho^k \sum_{j=0}^{t-1} \rho^{2j}$$

which is again a geometric sum. So we can tackle it accordingly:

$$\gamma_k = \sigma^2 \rho^k (\rho^0 + \rho^2 + \rho^4 + \cdots + \rho^{t-1})$$
$$\rho^2 \gamma_k = \sigma^2 \rho^k (\rho^2 + \rho^4 + \rho^6 + \cdots + \rho^{t+1})$$
$$\gamma_k - \rho^2 \gamma_k = \sigma^2 \rho^k (\rho^0 - \rho^{t+1})$$
$$\gamma_k = \sigma^2 \rho^k \left(\frac{1 - \rho^{t+1}}{1 - \rho^2} \right)$$
$$\lim_{t \to \infty} \gamma_k = \sigma^2 \rho^k \frac{1}{1 - \rho^2} - \sigma^2 \rho^k \frac{1}{1 - \rho^2} \lim_{t \to \infty} \rho^{t+1}$$
$$= \frac{\sigma^2 \rho^k}{1 - \rho^2} \quad \text{for } |\rho| < 1$$
$$= \rho^k Var(z_t).$$

Since the covariance is also not dependent on t and is a constant, the third criterion is also satisfied. Given that all three criteria are satisfied, this stochastic process is stationary.

(iii) What would the mean and variance of z_t be if $\rho = 1$?

Answer: Since ρ is a unity, the stochastic process would become

$$z_t = z_{t-1} + \varepsilon_t \quad , \quad t = 1, 2, 3, \ldots.$$

This process is called **random walk**.

The expected value of z_t can be obtained via recurring substitution as it was done in part (ii) above:

$$z_t = z_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$
$$= z_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$
$$\vdots$$
$$= z_{t-t} + \sum_{i=0}^{t-1} \varepsilon_{t-i}.$$

Therefore, the expected value of z_t is:

$$\mathbb{E}(z_t) = \mathbb{E}\left(z_0 + \sum_{i=0}^{t-1} \varepsilon_{t-i}\right)$$
$$= \mathbb{E}(z_0) + \mathbb{E}\left(\sum_{i=0}^{t-1} \varepsilon_{t-i}\right)$$
$$= \mathbb{E}(z_0) = z_0 \quad \forall \, t \geq 1$$

This means, the expected value of a random walk does not depend on t.

On the other hand, the variance of a random walk does change with t:

$$Var(z_t) = Var(z_0) + Var(\varepsilon_1) + \cdots + Var(\varepsilon_{t-2}) + Var(\varepsilon_{t-1}) + Var(\varepsilon_t)$$

If we assume that z_0 is nonrandom so that $Var(z_0) = 0$, then by the i.i.d. assumption for $\{\varepsilon_t\}$, the variance becomes:

$$Var(z_t) = t\sigma_\varepsilon^2$$

which means that the variance of a random walk increases as a linear function of time. This shows that the process cannot be stationary.

We can also look at the covariance between two time periods and check if the third criterion holds. For this, we can assume $z_0 = 0$ to simplify the algebra. This common assumption states that the process begins at 0, and $\mathbb{E}(z_0) = z_0 = 0$ for all t. With this assumption, we then have:

$$z_t = \sum_{i=0}^{t} \varepsilon_i \quad , \qquad z_{t-k} = \sum_{j=0}^{t-k} \varepsilon_j.$$

The autocovariance is then

$$Cov(z_t, z_{t-k}) = \gamma_k = Cov\Big(\sum_{i=0}^{t} \varepsilon_i \ , \ \sum_{j=0}^{t-k} \varepsilon_j \Big)$$

$$= \sum_{i=0}^{t} \sum_{j=0}^{t-k} Cov(\varepsilon_i, \varepsilon_j).$$

Since the error terms are independent

$$Cov(\varepsilon_i, \varepsilon_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Since j can only go up to $t - k$, and we have nonzero covariances when $i = j$, then $i \leq t - k$. This means, the effective range for the summation is from $i = 1$ to $i = t - k$:

$$\gamma_k = \sum_{i=1}^{t-k} \sigma^2 = (t - k)\sigma^2.$$

Therefore, the autocovariance function of a random walk with lag k is $\gamma_k = (t - k)\sigma^2$. This shows that the process is nonstationary and depends on t, thus not satisfying the third criterion either. As t increases, random walk process's statistical properties changes such that the autocovariance increases linearly for any fixed lag k. Similarly, for a fixed t, the autocovariance decreases linearly as the lag k increases.

– –

QUESTION 2

Use the Stata file BoatRace.dta, which contains information on the annual Oxford Cambridge boar race for the period 1946-2013, for each of the following:

(i) Obtain the time series plot of the variable `Winner` (Graphics/Time-series graphs/line plotes - (tsline)) and characterise the corresponding data generating process.
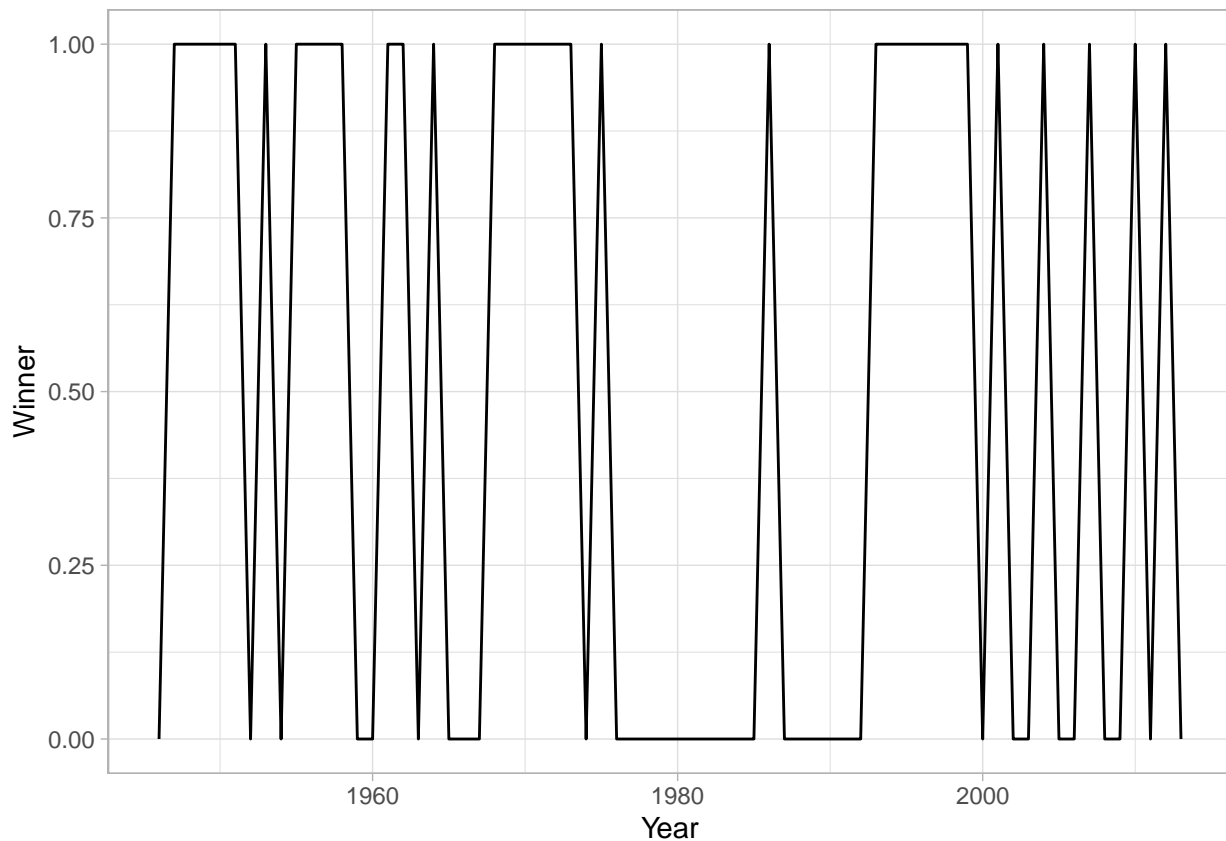
Answer:

In Stata:

```
quietly cd ..
use Data/BoatRace.dta
quietly tsset Year
tsline Winner
```

In R:

```r
boat_df <- read_dta("../Data/BoatRace.dta")
ggplot(boat_df,
       aes(x=Year, y=Winner)) +
  geom_line() +
  theme_light()
```


The plot switches from 0 to 1 and looks pretty random. It would be difficult to predict. There does not seem to be a general increase or decrease in `Winner` over time so it does not seem to be trending. Similarly, it does not look like there is any clear pattern in the data so it does not look like there is seasonality.

However, there may still be autocorrelation. It looks like there are periods where one team was more successful then the other, which may perhaps be due to the possibility that the team members or a good coach remain the same for more than 1 year while they are pursuing their studie. Therefore we cannot yet conclude that the data is weakly stationary.

(ii) Obtain the sample correlogram of `Winner` (Graphics/Time-Series graphs/Correlogram - (ac)). What does the correlogram tell us about serial correlation in `Winner`?

Answer:

In Stata:

```
quietly cd ..
use Data/BoatRace.dta
quietly tsset Year
ac Winner
```

In R:

```
ggAcf(boat_df$Winner) +
  ggtitle("Autocorrelations of `Winner`") +
  theme_light()
```


Autocorrelations of 'Winner'
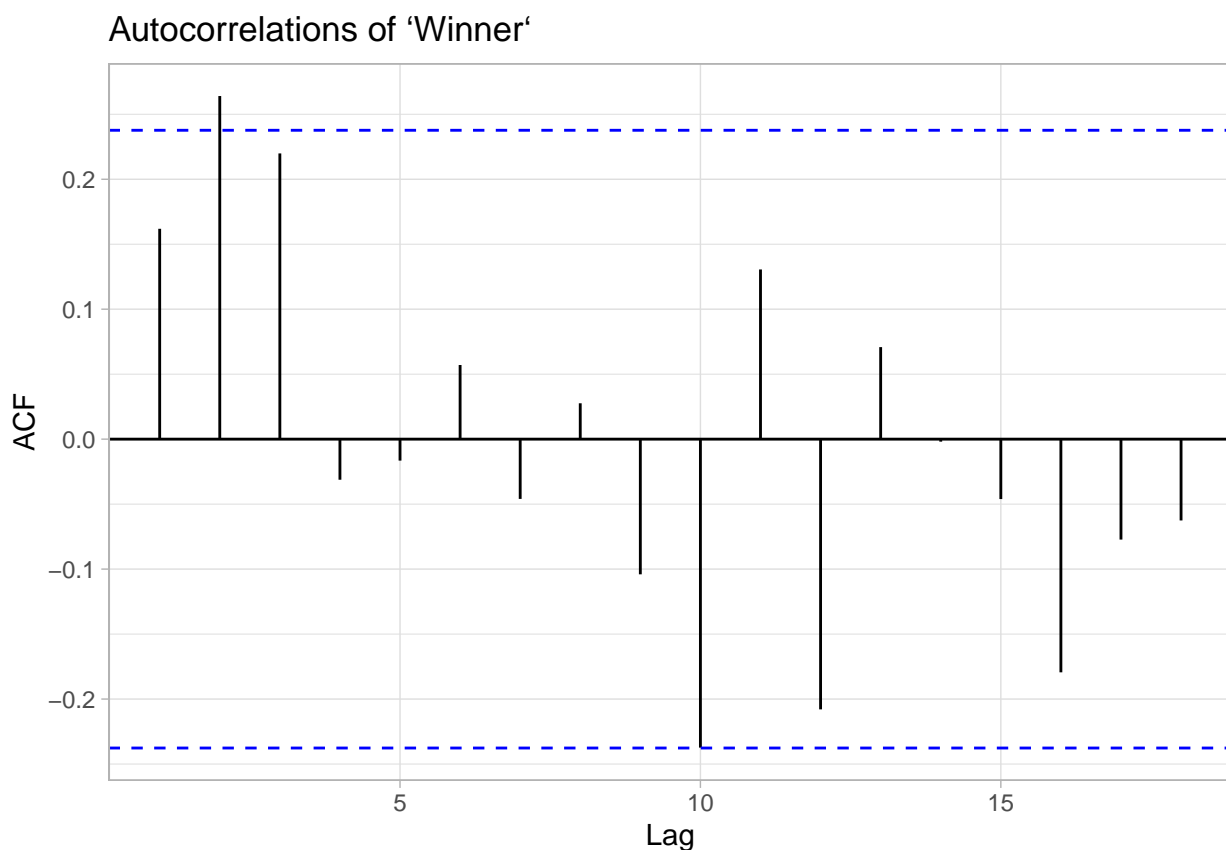
For stationary series, we would ultimately expect to see autocorrelations decaying to zero at higher lags. Here we see the ACF's not significantly different from 0, except for the second lag. So this is not giving us a clear picture with respect to stationarity.

Let's unpack what we just plotted. We are beginning to test for stationarity. This graphical test is called *autocorrelation function* (ACF).

Stationary Test 1: ACF

The ACF at lag k is defined as:

$$ACF_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{covariance at lag } k}{\text{variance}}.$$

Since both the numerator and the denominator are measured in the same units of measurement, ACF_k is a unitless number. It lies between +1 and -1 as any correlation coefficient.

We don't usually know the population but only a realization, i.e. sample, of a stochastic process, we can only compute the sample autocorrelation function, $\widehat{ACF}_k$:

$$\widehat{ACF}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\dfrac{\sum (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{n}}{\dfrac{\sum (Y_t - \bar{Y})^2}{n}}$$

The graph above is the sample correlogram that plots $\widehat{ACF}_k$ against k. The dashed lines gives the 95% confidence intervals that the autocorrelations are not different from 0. Bartlett (1946) has shown that for a time series that follows white noise, the standard error of the autocorrelation is approximately:[a]

$$se(\widehat{ACF}_k) \approx \frac{1}{\sqrt{n}}$$

The 95

$$CI_{95\%} = \pm \frac{1.96}{se(\widehat{ACF}_k)} = \pm \frac{1.96}{\sqrt{n}}$$

---

[a]Bartlett, M S (1946) "On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series", *Journal of the Royal Statistical Society*, Series B, 27:27-41.

In this question $n = 68$ so the 95% CI is ± 0.2377. The plot looks similar to a white noise where the autocorrelations seem to hover around 0 at various lags. At the same time there seems to be a positive correlation between W_t and W_{t-2} that is only just statistically significant. The third lag is also close to being significant. Perhaps this may be due to some combination of the boat staying the same for over 2 or 3 years, which is the length of stay for the undergraduates.

We can also use the **Q statistic**.

Stationarity Test 2: Box-Pierce Q statistic[a]

[a]Box G E P, and Pierce, D A (1970) "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models" *Journal of the American Statistical Association*, 65:1509-1526.

Box-Pierce Q statistic[a] which instead of testing the statistical significance of any individual autocorrelation coefficient tests the joint hypothesis that all the $\widehat{ACF}_k$ up to certain lags are simultaneously equal to 0:

$$Q = n \sum_{k=1}^{m} \widehat{ACF}_k^2$$

where n is the sample size and m is the lag length. In large samples, it is *approximately* distributed as χ^2 distribution with m degrees of freedom. If the computed Q is greater than the critical Q value from the χ^2 distribution at the specified significance level, then the null hypothesis is rejected, meaning that at least one or more of the correlation coefficients are nonzero.

---

[a]

A variant of the Box-Pierce Q statistic is the **Ljung-Box (LB) statistic**[2]

> ### Stationarity Test 3: Ljung-Box (LB) statistic
>
> LB statistic is defined as:
> $$LB = n(n+2) \sum_{k=1}^{m} \left(\frac{\widehat{ACF}_k^2}{n-k} \right) \sim \chi_m^2.$$
>
> Although in large samples both Q and LB statistics follow the chi-square distribution with m degrees of freedom, the LB statistic has been found to have more statistically powerful small-sample properties than the Q statistic.

Lets see what LB test would reveal for our boatrace. We can check for

```
quietly cd ..
use Data/BoatRace.dta
ac Winner
wntestq Winner, lags(4)
wntestq Winner, lags(5)
corrgram Winner
```

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    10.5031
 Prob > chi2(4)            =     0.0328
```

```
Portmanteau test for white noise
---------------------------------------
 Portmanteau (Q) statistic =    10.5237
 Prob > chi2(5)            =     0.0617
```

| | | | | | -1   0   1 | -1   0   1 |
LAG	AC	PAC	Q	Prob>Q	[Autocorrelation]	[Partial autocor]
1	0.1619	0.1640	1.862	0.1724	|-	|-
2	0.2640	0.2471	6.8902	0.0319	|--	|-
3	0.2199	0.1611	10.431	0.0152	|-	|-
4	-0.0312	-0.1649	10.503	0.0328	|	-|
5	-0.0165	-0.1088	10.524	0.0617	|	|
6	0.0571	0.0934	10.774	0.0956	|	|
7	-0.0460	0.0090	10.939	0.1413	|	|
8	0.0276	0.0340	10.999	0.2017	|	|
9	-0.1040	-0.1934	11.872	0.2206	|	-|
10	-0.2374	-0.2832	16.496	0.0863	-|	--|
11	0.1306	0.4187	17.92	0.0835	|-	|---
12	-0.2079	-0.2294	21.595	0.0423	-|	-|
13	0.0708	0.1971	22.029	0.0549	|	|-
14	-0.0019	-0.2162	22.03	0.0780	|	-|
15	-0.0461	0.0044	22.22	0.1022	|	|
16	-0.1794	-0.3340	25.167	0.0669	-|	--|
17	-0.0773	0.1831	25.725	0.0796	|	|-
18	-0.0626	-0.0176	26.097	0.0976	|	|
19	-0.1370	-0.2397	27.922	0.0849	-|	-|

[2]Ljung, G M, and Box, G E P (1978) "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, 66:66-72.

```
20       -0.0635   -0.0381    28.321   0.1020          |                    |
21        0.0101    0.3063    28.331   0.1310          |                    |--
22        0.1140    0.1719    29.676   0.1265          |                    |-
23        0.0110    0.0740    29.689   0.1585          |                    |
24        0.2023    0.0717    34.116   0.0826          |-                   |
25        0.0992   -0.1290    35.206   0.0845          |                  -|
26        0.1140   -0.2462    36.678   0.0799          |                  -|
27       -0.1068   -0.4262    38.003   0.0778          |               ---|
28       -0.0047    0.0150    38.006   0.0983          |                    |
29       -0.0186   -0.1144    38.048   0.1213          |                    |
30       -0.1216   -0.2116     39.9    0.1068          |                  -|
31       -0.0766   -0.6229    40.655   0.1149          |              ----|
32       -0.0905   -1.2845    41.736   0.1164          |          --------|
```

This looks pretty indecisive. We cannot reject the null that all the autocorrelation coefficients are jointly equal to 0 at 5% significance when we use 5 lags, but we can reject when we use 4 lags.

There is not a direct equivalent command in R but we can replicate the table above in R by building our own function:

```r
r_corrgram_extended <- function(time_series, lag.max = 20) {
  n <- length(time_series)

  # Calculate ACF and PACF
  acf_result <- acf(time_series, lag.max = lag.max, plot = FALSE)
  pacf_result <- pacf(time_series, lag.max = lag.max, plot = FALSE)

  # Calculate standard errors
  se_acf <- rep(1/sqrt(n), lag.max)

  # Initialize vectors for Q statistics and p-values
  q_stats <- numeric(lag.max)
  p_values <- numeric(lag.max)

  # Calculate Q statistics for each lag
  for (i in 1:lag.max) {
    box_test <- Box.test(time_series, lag = i, type = "Ljung-Box")
    q_stats[i] <- box_test$statistic
    p_values[i] <- box_test$p.value
  }

  # Create results data frame
  results <- data.frame(
    Lag = 1:lag.max,
    AC = round(acf_result$acf[-1],4),  # Remove lag 0
    se_AC = round(se_acf,4),
    PAC = round(pacf_result$acf,4),
    Q = round(q_stats,4),
    Prob_Q = round(p_values,4)
  )

  return(results)
}
```

```
r_corrgram_extended(boat_df$Winner)
```

Another test we can use to check for stationarity is the **unit root test**.

Stationarity Test 4: Unit Root test

Consider the stochastic process

$$Y_t = \rho Y_{t-1} + u_t \quad , \quad -1 \leq \rho \leq 1$$

where u_t is a white noise error term. If $\rho = 1$, i.e when we have a unit root, then the stochastic process above is a random walk model without drift, which is a nonstationary stochastic process. The idea of the test is then to regress Y_t on its one lagged period and check if the estimated coefficient is equal to 1 or not. If it is, then Y_t is nonstationary.

However, we cannot estimate this equation by OLS and use the usual t test to test if the coefficient is equal to 1 because the test itself is severely biased in the case of a unit root. To tackle this, we subtract Y_{t-1} from both sides to obtain:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + u_t$$
$$\Delta Y_t = (\rho - 1)Y_{t-1} + u_t$$
$$\Delta Y_t = \delta Y_{t-1} + u_t.$$

It is this equation that is then estimated. The null hypothesis is that $\delta = 0$ with the alternative being $\delta < 0$ since $|\rho| < 1$ for stationarity to hold.[a]

The only thing left to consider now for the unit root test is which test to use to see if the coefficient of Y_{t-1} is 0 or negative in this first differenced equation. We cannot use t test because the t value of $\hat{\delta}$ does not follow the t distribution even in large samples; i.e. it does not have an asymptotic normal distribution.

---
[a] If $\delta = 0$ then $\Delta Y_t = (Y_t - Y_{t-1}) = u_t$. Since u_t is a white noise term it is stationary which means the first differences of a random walk time series are stationary.

The alternative approach is to use the **Dickey-Fuller (DF) test**, which is also called the **tau statistic**, or **tau test**.[3]

Stationarity Test 5: Dickey-Fuller (DF) test

The procedure for implementing the DF test involves several decisions. First, recall that a random walk process can have a drift, or it may have both deterministic and stochastic trends. To allow for various possibilities, the DF test is estimated in three different forms; i.e. under three different null hypotheses:

- Y_t is a random walk: $\Delta Y_t = \delta Y_{t-1} + u_t$
- Y_t is a random walk with a drift: $\Delta Y_t = \beta_0 + \delta Y_{t-1} + u_t$
- Y_t is a random walk with a drift around a deterministic trend: $\Delta Y_t = \beta_0 + \beta_1 \, t + \delta Y_{t-1} + u_t$
   where t is the time or trend variable.

In each case the hypotheses are:
$$\mathbb{H}_0 : \delta = 0 \quad , \quad \mathbb{H}_1 : \delta < 0.$$

If the null is rejected it means that either:

1. Y_t is stationary with 0 mean in the case of random walk stochastic process,
2. Y_t is stationary with nonzero mean in the case of random walk with a drift stochastic process,
3. Y_t we can test for $\delta < 0$ (i.e. no stochastic trend) and $\beta_0 \neq 0$ (i.e. existence of a deterministic

---
[3] Dicket, D A, and Fuller, W A (1979) "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, Vol 74, p427-431; see also Fuller, W A (1976) "Introduction to Statistical Time Series", John Wiley & Sons, NY.

> trend) simultaneously using the F test but using the critical values tabulated by Dickey and Fuller.
>
> It is <u>important</u> to note that the critical values of the tau test to test the hypothesis that $\delta = 0$ are *different* for each of the preceding three specifications of the DF test.
>
> The steps for this test are as follows:
>
> 1. Estimate using OLS one of the three random walk stochastic processes (with/without drift or with drift around a deterministic trend)
> 2. Divide the estimated coefficient of Y_{t-1} in each case by its standard error to compute the tau (τ) statistic
> 3. Look up the DF tables. If $|\tau|$ statistic exceeds the $|\tau|$ critical values then we reject the null hypothesis, in which case the time series is stationary. Otherwise, we fail to reject the null, in which case the time series is nonstationary.
>
>   In most cases the tau value will be negative. Therefore, alternatively, we can say that if the computed negative tau value is smaller than (i.e. more negative than) the critical tau value then we reject the null hypothesis, in which case the time series is stationary.

In this question, we can run the test as follows:

```
quietly cd ..
use Data/BoatRace.dta
dfuller Winner
dfuller Winner, drift
dfuller Winner, trend
```

```
Dickey-Fuller test for unit root              Number of obs  = 67
Variable: Winner                              Number of lags =  0

H0: Random walk without drift, d = 0

                                     Dickey-Fuller
                    Test        -------- critical value ---------
                 statistic        1%            5%            10%
-----------------------------------------------------------------
  Z(t)            -6.833        -3.556        -2.916        -2.593
-----------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000.
```

```
Dickey-Fuller test for unit root              Number of obs  = 67
Variable: Winner                              Number of lags =  0

H0: Random walk with drift, d = 0

                                     t-distribution
                    Test        -------- critical value ---------
                 statistic        1%            5%            10%
-----------------------------------------------------------------
  Z(t)            -6.833        -2.385        -1.669        -1.295
-----------------------------------------------------------------
p-value for Z(t) = 0.0000
```

```
Dickey-Fuller test for unit root              Number of obs  = 67
Variable: Winner                              Number of lags =  0
```

```
H0: Random walk with or without drift

                                Dickey-Fuller
                  Test        -------- critical value ---------
                statistic         1%            5%            10%
        -------------------------------------------------------------
         Z(t)         -7.074       -4.113        -3.483        -3.170
        -------------------------------------------------------------

MacKinnon approximate p-value for Z(t) = 0.0000.
```

This tells us that in any of the specifications the stochastic process is stationary.

In conducting the DF test it was assumed that the error term u_t were uncorrelated. In case they are correlated, Dickey and Fuller have developed the **augmented Dickey-Fuller (ADF) test**.

Stationarity Test 6: Augmented Dicky-Fuller (ADF) test

The ADF test is conducted by augmenting the preceding three equations by adding the lagged values of the dependent variable ΔY_t. So, for example, suppose we are using random walk with a drift around a deterministic trend equation. The ADF test then consists of estimating the following regression:

$$\Delta Y_t = \beta_0 + \beta_1\ t + \delta\ Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + \varepsilon_t$$

where ε_t is a pure white noise error term and where $\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2}), \Delta Y_{t-2} = (Y_{t-2} - Y_{t-3})$ etc. **Note:** The number of lagged difference terms to include is often determined empirically, the idea being to include enough terms so that the error term is serially uncorrelated, which would allow us to obtain unbiased estimate of δ.

Lets try with 1 lag on the random walk without a drift or trend equation:

```
quietly cd ..
use Data/BoatRace.dta
dfuller Winner, lags(1) regress
```

```
Augmented Dickey-Fuller test for unit root

Variable: Winner                          Number of obs   = 66
                                          Number of lags =   1

H0: Random walk without drift, d = 0

                                Dickey-Fuller
                  Test        -------- critical value ---------
                statistic         1%            5%            10%
        -------------------------------------------------------------
         Z(t)         -3.945       -3.558        -2.917        -2.594
        -------------------------------------------------------------

MacKinnon approximate p-value for Z(t) = 0.0017.

Regression table
-------------------------------------------------------------------------
    D.Winner | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+-----------------------------------------------------------
      Winner |
```

```
      L1. |  -.6159696    .1561211    -3.95   0.000    -.9279526   -.3039865
      LD. |  -.2471483    .1220797    -2.02   0.047    -.4911051   -.0031915
          |
    _cons |   .2965779    .0975263     3.04   0.003     .1016871    .4914688
------------------------------------------------------------------------------
```

Based on $Z(t)$ output, we can reject the null hypothesis at all common significance levels. The estimated $\delta = -0.616$ implies that $\rho = (1 - 0.616) = 0.384$. On the other hand, if we try this with 5 lags, we cannot reject the null hypothesis even at 10% significance level.

In R we can obtain the same results via the following:

```r
# Basic Dickey-Fuller test (equivalent to Stata's dfuller y)
df_test <- ur.df(boat_df$Winner, type = "none", lags = 0)
summary(df_test)

# With trend (equivalent to Stata's dfuller y, trend)
df_trend_test <- ur.df(boat_df$Winner, type = "trend", lags = 0)
summary(df_trend_test)

# With drift/constant (equivalent to Stata's dfuller y, drift)
df_drift_test <- ur.df(boat_df$Winner, type = "drift", lags = 0)
summary(df_drift_test)

# Augmented Dickey-Fuller with lags (equivalent to Stata's dfuller y, lags(1))
adf_test <- ur.df(boat_df$Winner, type = "none", lags = 1)
summary(adf_test)

# With automatic lag selection (similar to Stata's dfuller y, lags(AIC))
max_lags <- floor(12 * (length(boat_df$Winner)/100)^(1/4))  # Rule of thumb
adf_auto <- ur.df(boat_df$Winner, type = "none", selectlags = "AIC", lags = max_lags)
summary(adf_auto)
```

(iii) Make a time series plot of the variable `WinTime` and briefly describe its behavior.
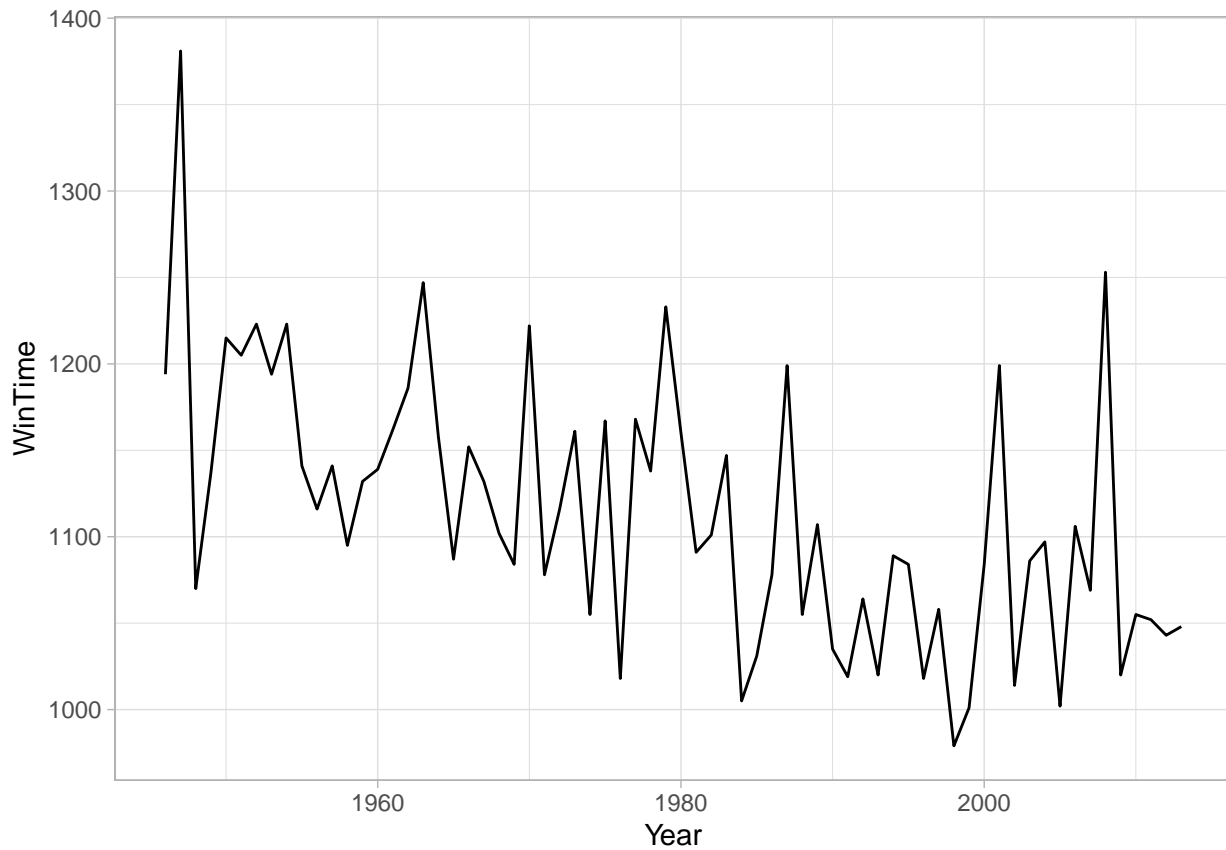
Answer:

In STATA:

```stata
quietly cd ..
use Data/BoatRace.dta
quietly tsset Year
tsline WinTime
```
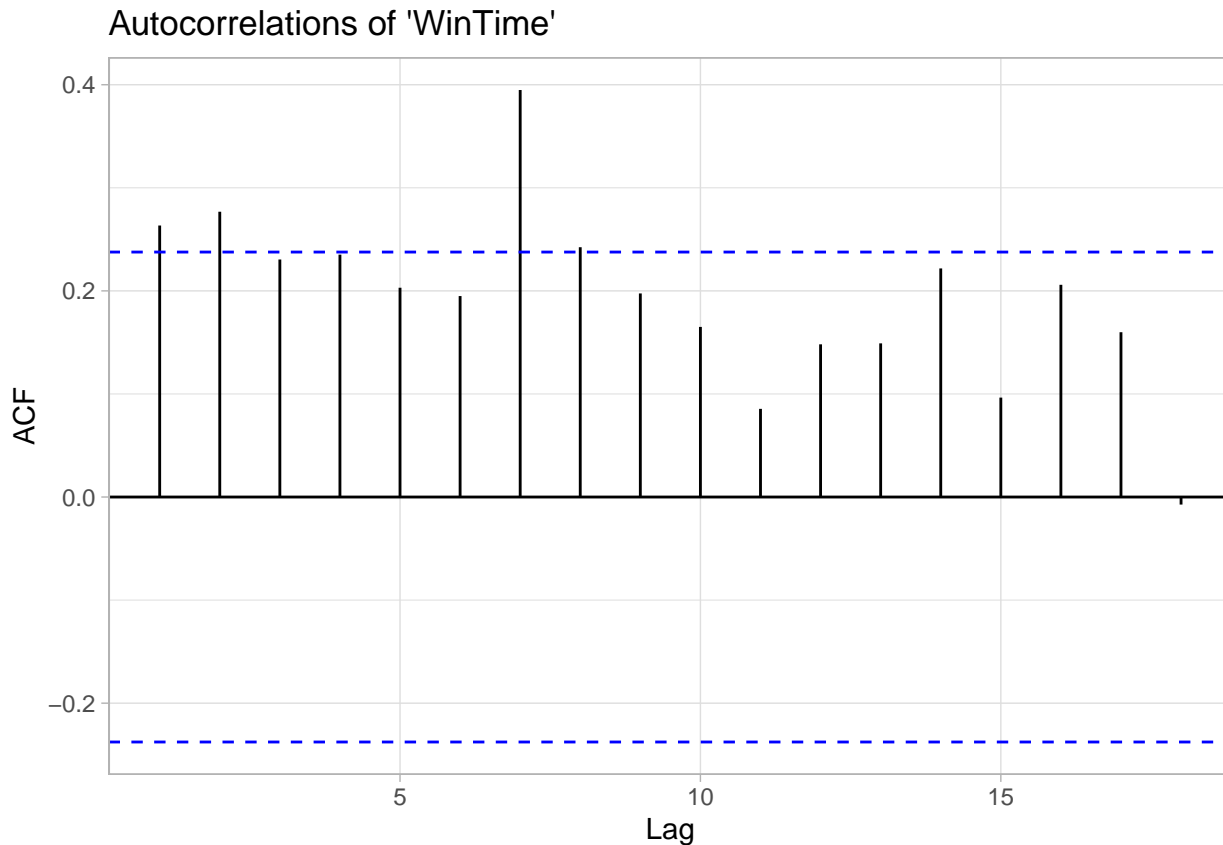
In R:

```r
ggplot(boat_df,
       aes(x=Year, y=WinTime)) +
  geom_line() +
  theme_light()
```

It looks like there is a trend of decline in time, possibly bottoming out later in the series. We can check for stationarity looking at the autocorrelations:

```
ggAcf(boat_df$WinTime) +
  ggtitle("Autocorrelations of 'WinTime'") +
  theme_light()
```

Autocorrelations of 'WinTime'

We can run a DF test using random walk with trend since there seems to be a downward trend.

```
quietly cd ..
use Data/BoatRace.dta
dfuller WinTime, trend regress
```

```
Dickey-Fuller test for unit root              Number of obs  = 67
Variable: WinTime                             Number of lags =  0

H0: Random walk with or without drift

                                     Dickey-Fuller
                      Test      -------- critical value ---------
                   statistic        1%           5%          10%
------------------------------------------------------------------
 Z(t)                 -8.601       -4.113       -3.483       -3.170
------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000.

Regression table
-------------------------------------------------------------------------------
   D.WinTime | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+-----------------------------------------------------------------
     WinTime |
         L1. |  -1.072409    .1246814    -8.60   0.000    -1.321489    -.8233293
             |
      _trend |  -2.372236    .4925508    -4.82   0.000     -3.35622    -1.388253
       _cons |   1274.021    149.3008     8.53   0.000     975.7581     1572.283
```

From the $Z(t)$ results we can reject the null hypothesis at all common significant levels and conclude that $\delta < 0$ and thus the stochastic process is stationary. $\delta = -1.07$ implies that $\rho = 1 - 1.07 = -0.07$.

(iv) Note that `DWinTime` in your data set is obtained by taking the first differences of `WinTime`. Suppose that the data generating process for `WinTime` is given by the following:

$$WinTime_t = \gamma + \delta\, Year_t + v_t$$

where v_t are i.i.d. shocks. Derive the population autocorrelations ρ_j, $j = 1, 2, 3, \ldots$ for `DWinTime`.

Answer:

Since `DWinTime` is the first differences of `WinTime` it can be expressed as:

$$
\begin{aligned}
DWinTime_t &= \gamma + \delta\, Year_t + v_t - (\gamma + \delta\, Year_{t-1} + v_{t-1}) \\
&= \delta(\Delta\, Year_t) + \Delta v_t \\
&= \delta + \Delta v_t \quad \text{since change in Year is always 1}
\end{aligned}
$$

We can use this expression to derive the population autocorrelations:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Cov(DWinTime_t\ ,\ DWinTime_{t-k})}{\sqrt{Var(DWinTime_t)\ ,\ Var(DWinTime_{t-k})}} = \frac{Cov(DWinTime_t\ ,\ DWinTime_{t-k})}{Var(DWinTime_t)}$$

To do so, first calculate the autocovariance function at lag k. Since v_t are i.i.d. shocks, we can assume them to be white noise process with mean 0 and variance σ_v^2. This also means $Cov(v_t, v_s) = 0$ for $t \neq 0$.

To obtain γ_k first calculate the mean of `DWinTime`:

$$
\begin{aligned}
\mathbb{E}(DWinTime_t) &= \mathbb{E}(\delta + \Delta v_t) \\
&= \delta + \mathbb{E}(v_t - v_{t-1}) \\
&= \delta
\end{aligned}
$$

We can next obtain the variance of `DWinTime`, which is the same as autocovariance at lag 0:

$$
\begin{aligned}
\gamma_0 &= Var(DWinTime) \\
&= Var(\delta + \Delta v_t) \\
&= Var(\Delta v_t) \\
&= Var(v_t - v_{t-1}) \\
&= Var(v_t) + Var(-v_{t-1}) \quad \text{since } v_t \text{ and } v_{t-1} \text{ are independent} \\
&= \sigma_v^2 + \sigma_v^2 \\
&= 2\sigma_v^2
\end{aligned}
$$

Thus, `DWinTime` $\sim (\delta, 2\sigma_v^2)$.

For lag 1, the autocovariance is:

$$
\begin{aligned}
\gamma_1 &= Cov(DWinTime_t\ ,\ DWinTime_{t+1}) \\
&= Cov(\delta + \Delta v_t\ ,\ \delta + \Delta v_{t+1}) \\
&= Cov(v_t - v_{t-1}\ ,\ v_{t+1} - v_t)
\end{aligned}
$$

$$\begin{aligned}
&= Cov(v_t \ , \ v_{t+1}) - Cov(v_t \ , \ v_t) - Cov(v_{t-1} \ , \ v_{t+1}) + Cov(v_{t-1} \ , \ v_t) \\
&= 0 - \sigma_v^2 - 0 + 0 \\
&= -\sigma_v^2.
\end{aligned}$$

And for lag k where $k > 1$, the autocovariance is:

$$\begin{aligned}
\gamma_k &= Cov(DWinTime_t \ , \ DWinTime_{t+k}) \\
&= Cov(v_t - v_{t-1} \ , \ v_{t+k} - v_{t+k-1}) \\
&= Cov(v_t \ , \ v_{t+k}) - Cov(v_t \ , \ v_{t+k-1}) - Cov(v_{t-1} \ , \ v_{t+k}) + Cov(C_{t-1, v_{t+k-1}}) \\
&= 0 - 0 - 0 + 0 \\
&= 0
\end{aligned}$$

Using these, now we can calculate the autocorrelation function $\rho_k = \gamma_k/\gamma_0$:

$$\rho_0 = \frac{\gamma_0}{\gamma_0} = 1$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\sigma_v^2}{2\sigma_v^2} = -0.5$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{0}{2\sigma_v^2} = 0 \quad \text{for } k > 1$$

Therefore, the autocorrelation function for the first-differenced series $DWinTime_t$ is:

$$\rho_k \begin{cases} 1 & \text{if } k = 0 \\ -0.5 & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

This specific pattern of first lag autocorrelation of -0.5 and 0 thereafter is a signature characteristic of overdifferencing a time series. **Overdifferencing** refers to the the fact that if a trend series is a trend-stationary process but we treat it as a difference-stationary process. The reverse of this is called **underdifferencing**.

- *difference-stationary process*: This refers to the fact that if a time series has a unit root, the first differences of such time series are stationary.

- *trend-stationary process*: The most straightforward way of making a time series with trendline stationary is to regress it on time and the residuals from this regression will be stationary.

  That is, when we run the following regression

  $$Y_t = \beta_0 + \beta_1 \ t + u_t$$

  the linearly **detrended time series**, $\hat{u}_t$

  $$\hat{u}_t = (Y_t - \hat{\beta}_0 - \hat{\beta}_1 \ t)$$

  will be stationary.

  On the other hand, the trend line may be nonlinear. For example, $Y_t = \beta_0 + \beta_1 \ t + \beta_2 \ t^2 + u_t$ is a quadratic series. The residuals from this will now be quadratically detrended time series.

Since we overdifferenced here, lets see if we can get trend-stationary process:

```
quietly cd ..
use Data/BoatRace.dta
reg WinTime t
```

```
      Source |       SS           df       MS        Number of obs   =        68
-------------+----------------------------------      F(1, 66)        =     31.94
       Model |  128954.652          1  128954.652     Prob > F        =    0.0000
    Residual |  266506.569         66  4037.97832     R-squared       =    0.3261
-------------+----------------------------------      Adj R-squared   =    0.3159
       Total |  395461.221         67  5902.40628     Root MSE        =    63.545


--------------------------------------------------------------------------------
     WinTime | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+------------------------------------------------------------------
           t |  -2.21867    .3926053    -5.65   0.000    -3.002532   -1.434808
       _cons |  1190.382    15.58351    76.39   0.000     1159.269    1221.496
--------------------------------------------------------------------------------
```
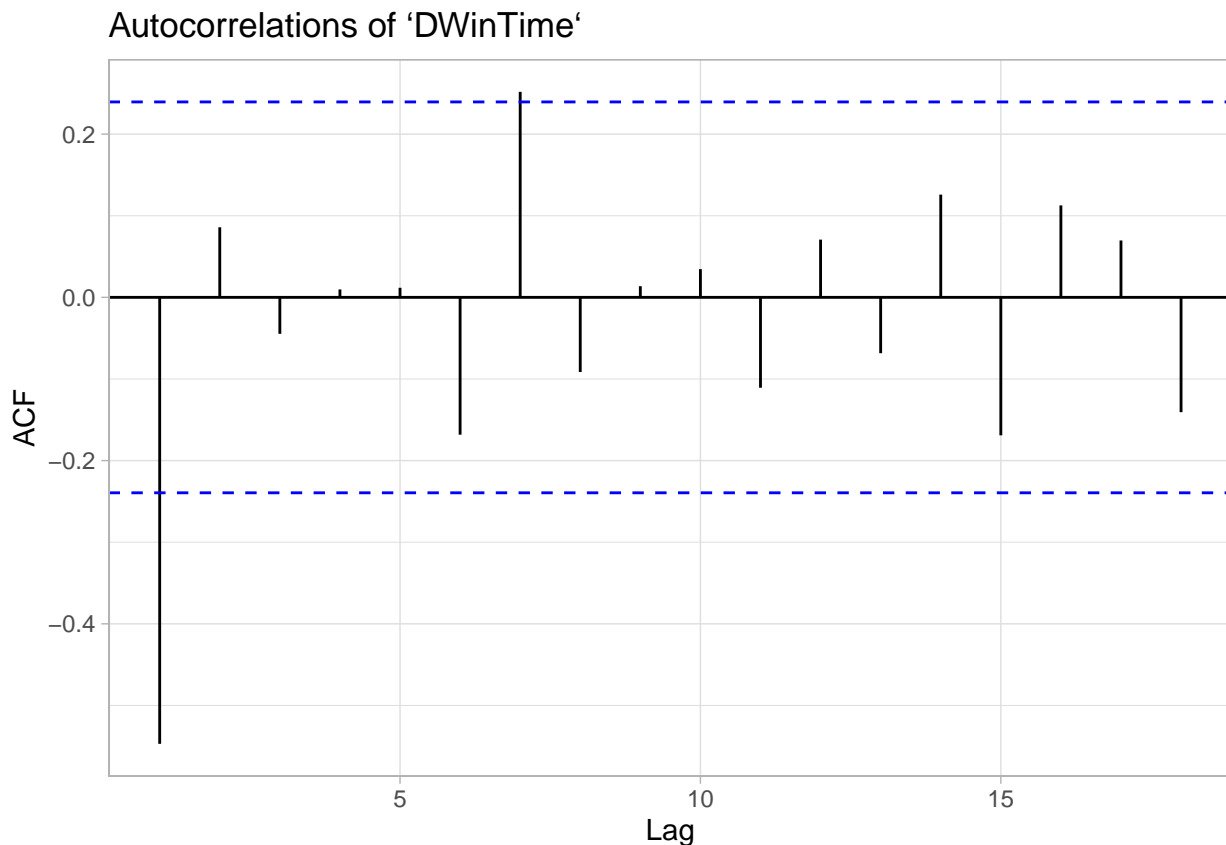
and if we plot the residuals we will see that they are stationary.

(v) Generate the sample correlogram of `DWinTime`. Are your results expected?

Answer:

```
ggAcf(boat_df$DWinTime) +
  ggtitle("Autocorrelations of `DWinTime`") +
  theme_light()
```


Autocorrelations of 'DWinTime'

This confirms that $\rho_1 \approx -0.5$ which we would expect for $k = 1$.

— —

QUESTION 3

(i) Using data from the file Timeseries1.dta, estimate the following equations.

$$C_t = \rho + \psi Y_t + \varepsilon_t \tag{1}$$
$$C_t = \alpha + \beta Y_t + \gamma Y_{t-1} + \lambda C_{t-1} + \varepsilon_t \tag{2}$$

where C_t is the log of household consumption expenditure and Y_t is the log of household disposable income at time t. Comment on your results.

Answer:

The first model provides an expression for the long run or equilibrium relationship between income and consumption

$$Consumption_t = \theta \ Income_t^{\psi}$$

where $\rho = \ln \theta$. Here, ψ is the long-run elasticity of `Consumption` with respect to `Income`. When `Consumption` takes its equilibrium value with respect to `Income`, this equation holds. However, economic systems are rarely in equilibrium.

What we usually observe instead is a short-run disequilibrium relationship involving lagged values of both variables. In the second equation only first lags are included but in practice second or higher order lags can appear. This second equation implies that `Consumption` takes time to adjust fully to variations in `Income`.

We can start regressing the first, static model and then the second model. The first model is static because it is modeling a contemporaneous relationship between C and Y. Usually, a static model is postulated when a change in Y at time t is believed to have an immediate effect on C. That is, $\Delta C_t = \psi Y_t$ when $\Delta \varepsilon_t = 0$. Static regression models are also used when we are interested in knowing the tradeoff between C and Y. Though, here we would not expect a trade-off between consumption and income.

Before we regress, lets see what the relationship between C and Y looks like.

```
ts_df <- read_dta("../Data/Timeseries1.dta")
ggplot(ts_df,
       aes(x = Year)) +
  geom_line(aes(y = C), linetype = "dashed") +
  geom_line(aes(y = Y)) +
  theme_light() +
  labs(y = "log Consumption (- -) & log Income")
```
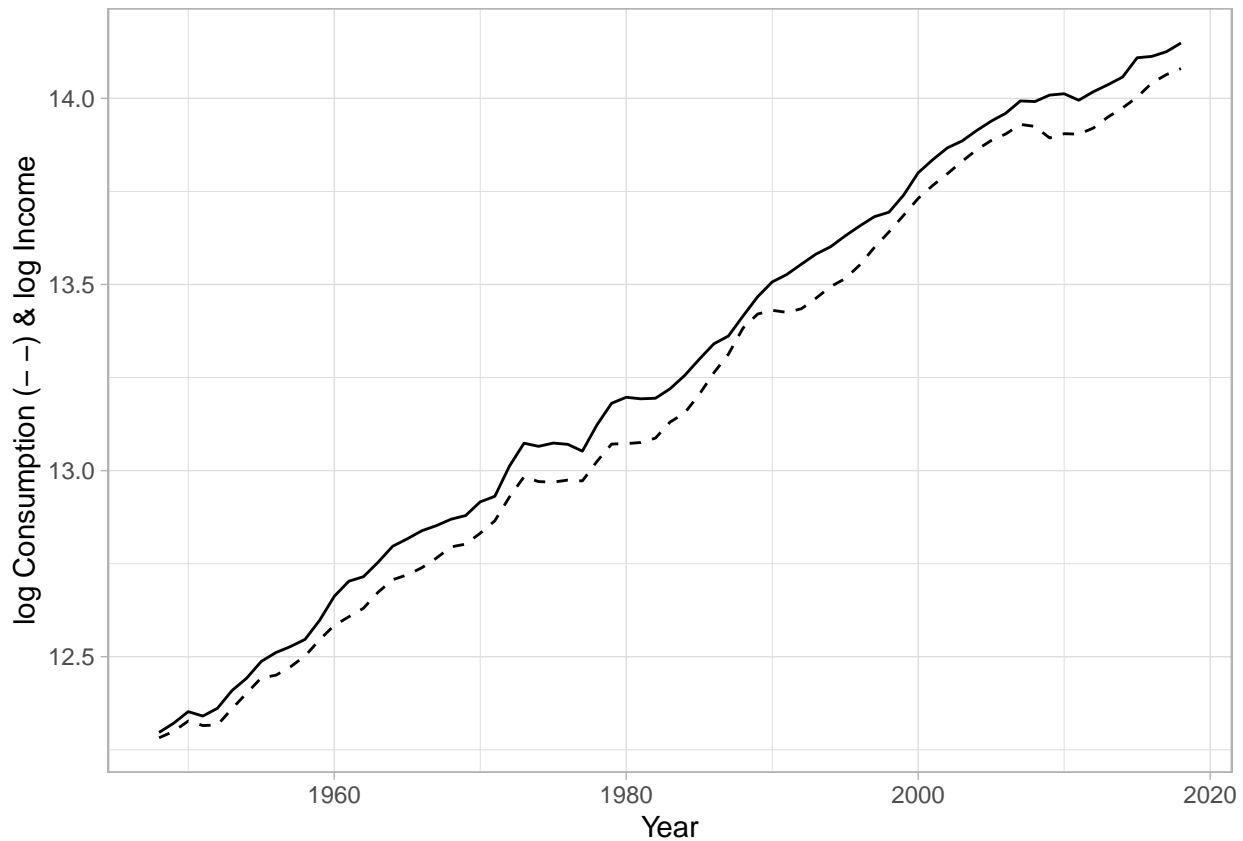
They both show a clear upwards trend and a very close relationship with each other. We can estimate that relationship using the first, static model:

```
lm1_SQ3i <- lm(C ~ Y, data = ts_df)
```

and in Stata

```
quietly cd ..
use Data/Timeseries1.dta
regress C Y
```

```
      Source |       SS           df       MS      Number of obs   =        71
-------------+----------------------------------   F(1, 69)        =  34857.35
       Model |  22.133814         1   22.133814    Prob > F        =    0.0000
    Residual |  .043813803        69   .000634983   R-squared       =    0.9980
-------------+----------------------------------   Adj R-squared   =    0.9980
       Total |  22.1776278        70   .316823255   Root MSE        =     .0252
```

```
------------------------------------------------------------------------------
           C | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           Y |    .985157   .0052766    186.70   0.000     .9746304    .9956836
       _cons |   .1201545   .0701098      1.71   0.091    -.0197107    .2600197
------------------------------------------------------------------------------
```

The equation does not suggest a tradeoff between Y and C since $\hat{\psi} > 0$, which we would not have expected anyway.

Now notice that R^2 is extremely high at 99.8% and so is the t-statistic for the coefficient of Y. This may be because they are strongly correlated or due to potential presence of **spurious regression**.

> **Spurious Regression**
>
> The nonsensical correlations were first discovered by Yule $(1926)^a$ who showed that spurious correlations could persist in nonstationary time series even if the sample is very large. In a simulation study, Granger and Newbold $(1974)^b$ examined spurious relationships between two independent but non-stationary variables X and Y - both were generated by the random walk processes - and estimated the relationship $\hat{Y}_t = \hat{\alpha} + \hat{\beta}X_t$. They showed that a potential indicator of a spurious correlation is a low Durbin-Watson statistic combined with an acceptable R^2. That is, a good rule of thumb to suspect that the estimated regression is spurious is to check if $R^2 > d$.
>
> Recall from Supervision 4 that the d statistic is the ratio of the sum of squared differences between successive residuals to the RSS:
>
> $$DW = d = \frac{\sum_{i=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{i=1}^{n}\hat{u}_t{}^2}$$
>
> *Note*: the number of observations in the numerator of the d statistic is one less than the denominator, i.e. $n-1$, because one observation is lost in taking successive differences.
>
> ---
> ^{a}Yule, G U (1926) "Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of Time Series", *Journal of the Royal Statistical Society, 89:1-64*
> ^{b}Granger C W J, and Newbold, P (1974) "Spurious Regressions in Econometrics" *Journal of Econometrics" 2:111-120*

Let's obtain the d statistic for this regression. For this, in R we use `durbinWatsonTest()` function from the `car` package, or `dwtest` from the `lmtest` package:

```r
car::durbinWatsonTest(lm1_SQ3i)
# or
dwtest(lm1_SQ3i)
```

The same calculation in STATA can be done via the `estat dwatson` command

```stata
quietly cd ..
use Data/Timeseries1.dta
quietly regress C Y
quietly tsset Year
estat dwatson
```

```
Durbin-Watson d-statistic(  2,    71) =   .3458889
```

Which gives us a d statistic of 0.3459 which is much lower than the R^2. So this could indicate a possibility of spurious correlation which, if so, would not mean that there may not be a relationship between C and Y but that the R^2 and t statistic from this regression would be misleading and that the t statistics would not be distributed as Student's t distribution, meaning it can't be used for testing hypotheses about the parameters.

This can arise because the model we used assumes that the stochastic process is **integrated of order 0**, denoted as $I(0)$ but the individual variables are I(1). We can check the correlograms to see if there are autocorrelations:

```r
plot1 <- ggAcf(ts_df$C) +
  ggtitle("Autocorrelations of log Consumption") +
  theme_light()
plot2 <- ggAcf(ts_df$Y) +
  ggtitle("Autocorrelations of log Income") +
  theme_light()
grid.arrange(plot1, plot2, ncol=2)
```
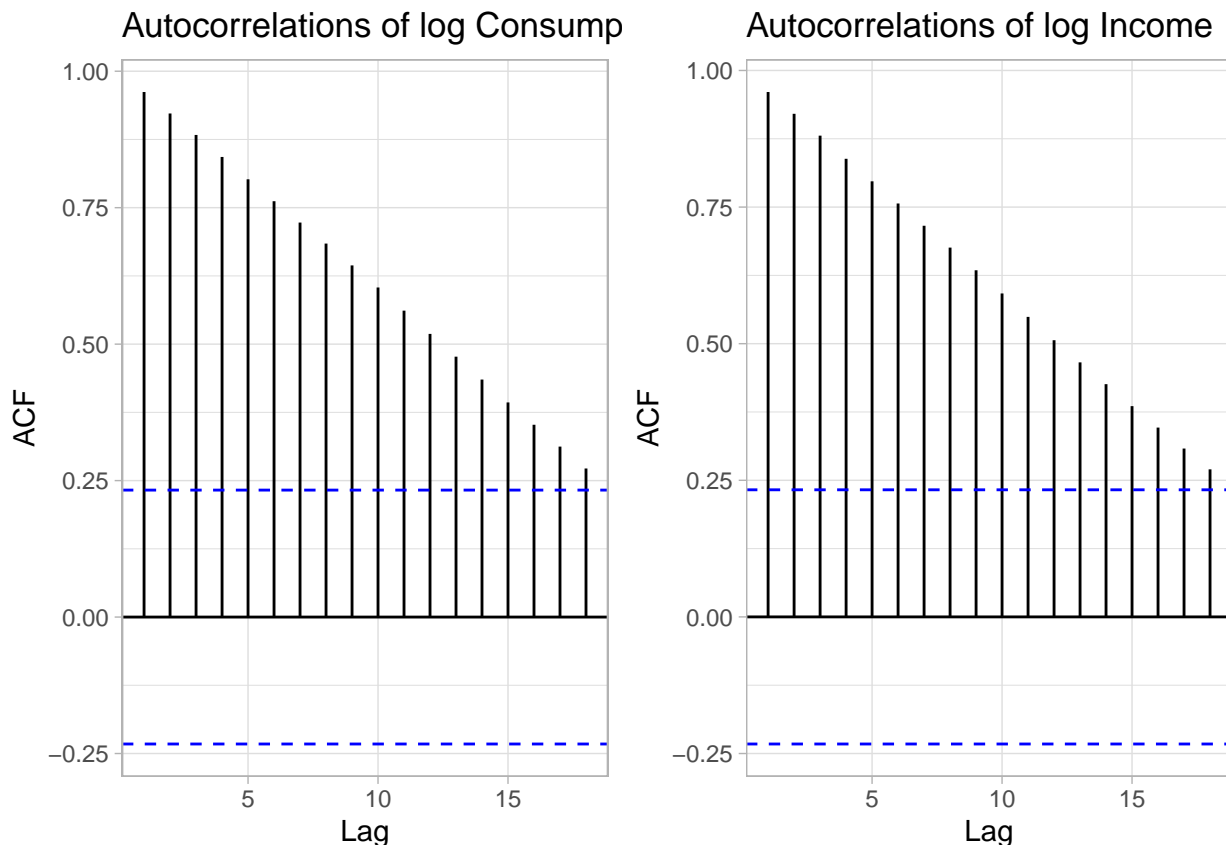
which shows that we have strong autocorrelations in both variables that go back a number of lags. They are both decaying to 0, albeit slowly, ao there may or may not be stationarity.

On the other hand this can be a **cointegrating regression**. That is, the time series individually contain stochastic trend, i.e. $I(1)$ but the two series *share the same common trend* so that the regression of one on the other will not be necessarily spurious. Economically speaking, two variables will be **cointegrated** if they have a long-term, or equilibrium, relationship between them. This explanation is more likely in this case. We can nevertheless check for cointegration.[4] For this, we will perform **Engle-Granger (EG)** and **augmented Engle-Granger (AEG)** tests.[5]

The first step in EG test is to perform a unit root test on the residuals of the regression:

```
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly regress C Y
predict e, residuals
dfuller e, lags(0)
dfuller e, lags(1)
```

```
Dickey-Fuller test for unit root          Number of obs  = 70
Variable: e                               Number of lags =  0

H0: Random walk without drift, d = 0


                        Dickey-Fuller
```

---

[4]For a good introduction to cointegration testing see Dickey D A, Jansen, D W, and Thornton D (1991) "A Primer on Cointegration with an Application to Money and Income" *Economic Review*, Federal Reserve Bank of St Louis, March-April:59.
[5]Engle, R F and Granger C W (1987) "Co-integration and Error Correction: Representation, Estimation, and Testing" *Econometrica*, 55:251-276.

```
          Test        -------- critical value ---------
        statistic         1%            5%           10%
--------------------------------------------------------------
 Z(t)        -2.904      -3.552        -2.914       -2.592
--------------------------------------------------------------
```

MacKinnon approximate p-value for Z(t) = 0.0449.

Augmented Dickey-Fuller test for unit root

```
Variable: e                              Number of obs  = 69
                                         Number of lags =  1
```

H0: Random walk without drift, d = 0

```
                              Dickey-Fuller
          Test        -------- critical value ---------
        statistic         1%            5%           10%
--------------------------------------------------------------
 Z(t)        -3.574      -3.553        -2.915       -2.592
--------------------------------------------------------------
```

MacKinnon approximate p-value for Z(t) = 0.0063.

We cannot reject the null of unit root at 5% but we can at 10% so it appears that residuals from the regression are almost but not quite stationary at 5% level. It would be difficult to accept this conclusion, for economic theory suggests that there should be a stable relationship between these two variables. Interestingly, the residuals are stationary when we run the augmented DF (ADF) test with one lag. This suggests that the inclusion of lagged term helps to stabilize the residuals, indicating a potential relationship between the time series variables that is better captured with at least one lag. This implies that the model with the lagged term is more appropriate for testing cointegration as it accounts for the dynamic relationship between the variables more accurately.

Lets now look at the second equation where we have the lagged terms. This model is called **autoregressive distributed lag (ARDL) model**.

```stata
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
regress C Y L.Y L.C
predict e, residuals
dfuller e, lags(0)
```

```
      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(3, 66)        =  39478.85
       Model |   21.31537         3  7.10512335    Prob > F        =    0.0000
    Residual |  .011878213        66  .000179973   R-squared       =    0.9994
-------------+----------------------------------   Adj R-squared   =    0.9994
       Total |  21.3272483        69  .309090554   Root MSE        =    .01342


--------------------------------------------------------------------------------
         C | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+------------------------------------------------------------------
         Y |
        --. |   .7491507   .0838085     8.94   0.000     .5818217    .9164798
        L1. |  -.6091466   .0903504    -6.74   0.000    -.7895371   -.4287561
            |
```

```
         C |
       L1. |    .859517    .0672937    12.77   0.000     .7251608    .9938732
           |
     _cons |   .0013938    .0392322     0.04   0.972    -.0769358    .0797235
-------------------------------------------------------------------------------
```

(1 missing value generated)

```
Dickey-Fuller test for unit root              Number of obs  =  69
Variable: e                                   Number of lags =   0

H0: Random walk without drift, d = 0

                               Dickey-Fuller
                 Test      -------- critical value ---------
               statistic      1%          5%          10%
-----------------------------------------------------------------
  Z(t)          -5.705      -3.553      -2.915      -2.592
-----------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000.
```

In this second equation we see that these residuals are now stationary. Even if we use ADF with several lags, the residuals are still stationary.

This result supports the idea that these two are cointegrated but `Consumption` takes time to adjust fully to variations in `Income`.

(ii) Estimate the following and interpret your results (and explain the significance of your results for the interpretation of equations (1) and (2) in part (i) on page 22):

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t \tag{3}$$

$$\Delta C_t = \beta_0 + \beta_1 C_{t-1} + \varepsilon_t \tag{4}$$

$$\Delta C_t = \gamma_0 + \gamma_1 C_{t-1} + \gamma_2 \Delta C_{t-1} + \varepsilon_t \tag{5}$$

Verify your results using the appropriate test in Stata and explain why equation 5 is an improvement on 4 in this particular case (esp. see Thomas, Chapter 13 and 14).

Answer: The effect of nonstationarity can often be removed by first differencing. This is because even though the log of income and log of consumption exhibit upwards trend, this would not normally be the case for the annual change output. Since C and Y are logarithms of 'Consumption' and 'Income', the first differences are the proportional rate of growth:

$$\ln(Y_t) - \ln(Y_{t-1}) = \ln\left(\frac{Y_t}{Y_{t-1}}\right) \approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

The proportionate or percentage growth rate is far more likely to remain roughly constant over time than is the absolute growth rate $Y_t - Y_{t-1}$. It follows that the equation in first differences, especially in logarithmic form, is likely to involve stationary variables only.

As such, we are effecetively running a unit root (DF/ADF) test with these equations. Recall from Question 2(ii) Stationary Tests 4-6, we want to check if the lagged variable is different from 0. If it is not, then it is stationary.

```r
#Generate lagged variables and differences
ts_df <- ts_df %>%
  mutate(
    Y_lag = lag(Y, 1, default=NA),
    C_lag = lag(C, 1, default=NA),
    DY = Y - Y_lag,
    DC = C- C_lag,
    DC_lag = lag(DC, 1, default = NA)
        )
# Run the three regressions
lm1_SQ3ii <- lm(DY ~ Y_lag, data = ts_df)
lm2_SQ3ii <- lm(DC ~ C_lag, data = ts_df)
lm3_SQ3ii <- lm(DC ~ C_lag + DC_lag, data = ts_df)

# Store models in a list with descriptive names
models <- list(
  "Eq (3)" = lm1_SQ3ii,
  "Eq (4)" = lm2_SQ3ii,
  "Eq (5)" = lm3_SQ3ii
)

# Create a coefficient map for nice variable names
coef_map <- c(
  "Y_lag" = "Y_lag",
  "C_lag" = "C_lag",
  "DC_lag" = "DC_lag",
  "(Intercept)" = "Constant"
)

# Generate the table with modelsummary
modelsummary(models,
             coef_map = coef_map,
             stars = TRUE,
             title = "SQ3(ii) Regression Results",
             gof_map = c("nobs", "r.squared", "adj.r.squared", "F"),
             fmt = 5,  # 4 decimal places
             output = "default")  # "default" for console output
```

Recall from Supervision 4, we can use Breusch-Godfrey (Lagrange Multiplier) Test to check for serial correlation

```stata
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly generate DY = Y - Y[_n-1]
quietly generate DC = C - L.C

quietly regress DY L.Y
bgodfrey

quietly regress DC L.C
bgodfrey

quietly regress DC L.C L.DC
```

Table 1: SQ3(ii) Regression Results

	Eq (3)	Eq (4)	Eq (5)
Y_lag	−0.005 09		
	(0.004 32)		
C_lag		−0.002 05	−0.002 30
		(0.004 25)	(0.003 92)
DC_lag			0.450 54***
			(0.109 63)
Constant	0.093 94	0.052 69	0.044 52
	(0.057 33)	(0.056 13)	(0.051 91)
Num.Obs.	70	70	69
R2	0.020	0.003	0.208
R2 Adj.	0.006	−0.011	0.184
F	1.388	0.232	8.648

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

bgodfrey

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2                df               Prob > chi2
-------------+-------------------------------------------------------------
        1    |         3.741                1                   0.0531
---------------------------------------------------------------------------
               H0: no serial correlation
```

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2                df               Prob > chi2
-------------+-------------------------------------------------------------
        1    |        14.172                1                   0.0002
---------------------------------------------------------------------------
               H0: no serial correlation
```

```
Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2                df               Prob > chi2
-------------+-------------------------------------------------------------
        1    |         3.215                1                   0.0730
---------------------------------------------------------------------------
               H0: no serial correlation
```

We see from the regression results that the first lags are not different from 0. Thus it appears that the first differencing is stationary. This also means the initial regression of equation (1) produces inconsistent

estimates.

When we check for serial correlation, it appears that only equation (4) is serially correlated. We can visually check for the autocorrelation using correlogram:

```r
ggAcf(ts_df$DC) +
  ggtitle("Autocorrelations of log Consumption") +
  theme_light()
```


Autocorrelations of log Consumption
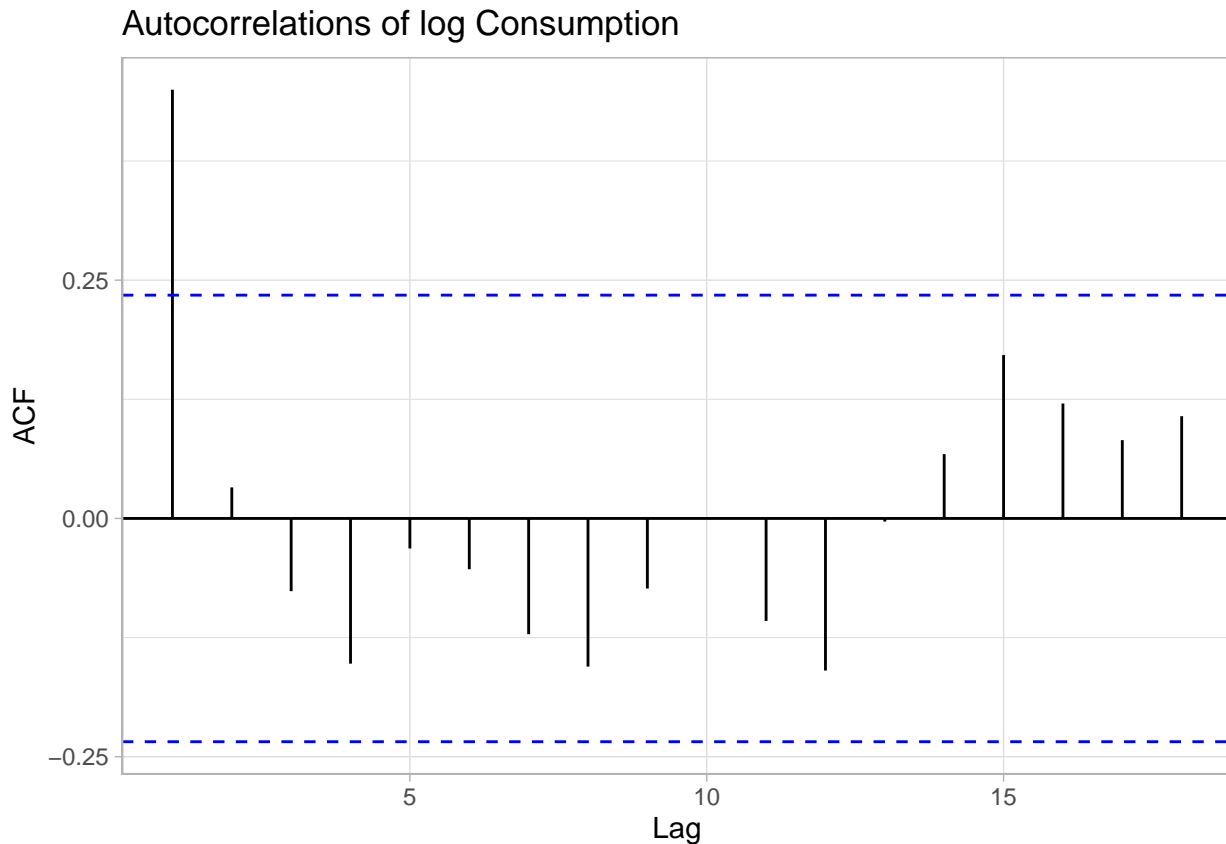
There seems to be a statistically significant correlation with its first lag. This is why equation (5) is an improvement on (4). It takes into account the first lag of ΔC_t as well.

We can also look at if after differencing these variables still exhibit upward trends:

```r
ggplot(ts_df,
       aes(x = Year)) +
  geom_line(aes(y = DC), linetype = "dashed", na.rm=TRUE) +
  geom_line(aes(y = DY), na.rm=TRUE) +
  theme_light() +
  labs(y = "D log Consumption (- -) & D log Income")
```
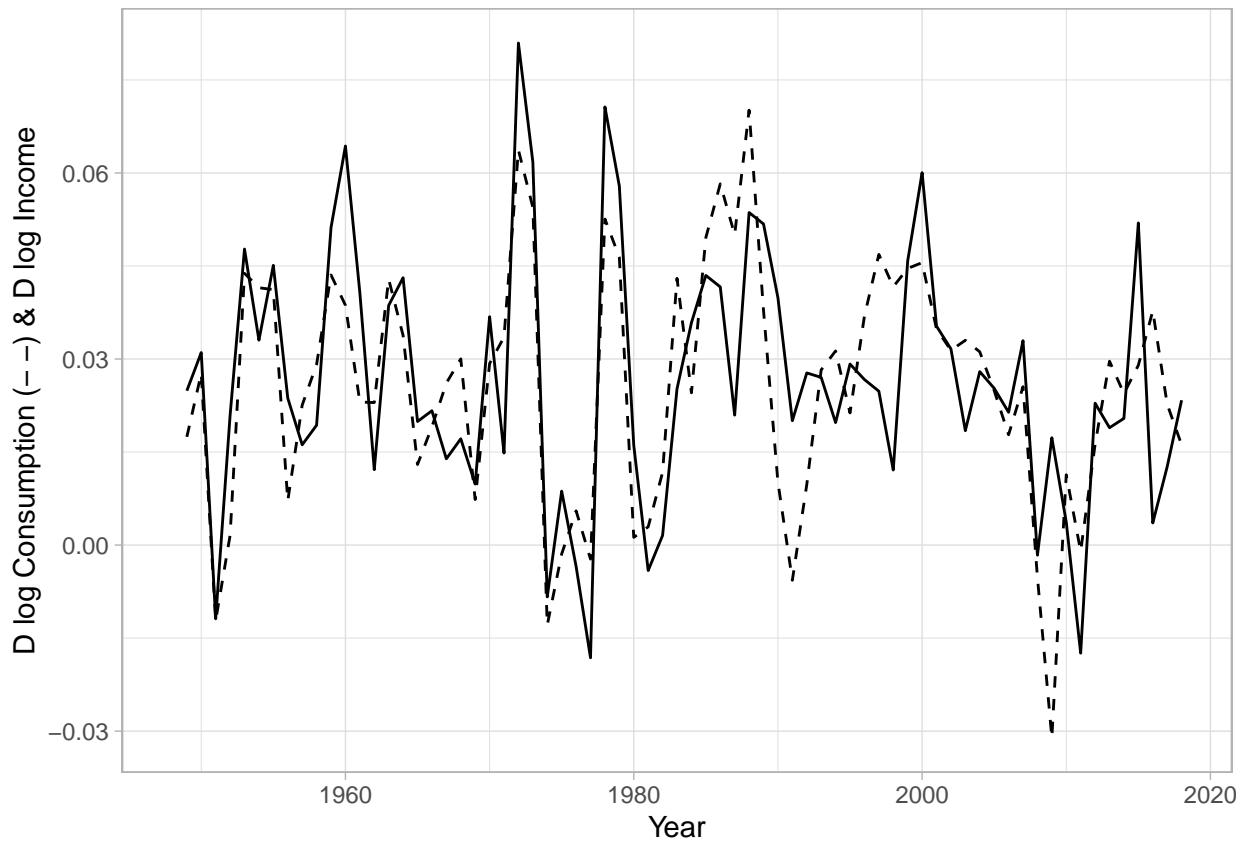
The graphs shows that we no longer have any discernable trend.

We can finally check if the differenced values are stationary using ADF test. If so, then C and Y are individually $I(1)$, supporting the cointegration argument.

```
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly generate DY = Y - L.Y
quietly generate DC = C - L.C

dfuller DY, lag(0)
dfuller DC, lag(0)
```

```
Dickey-Fuller test for unit root                Number of obs  = 69
Variable: DY                                    Number of lags =  0

H0: Random walk without drift, d = 0

                                   Dickey-Fuller
                   Test     -------- critical value ---------
                 statistic         1%          5%         10%
-------------------------------------------------------------
 Z(t)              -6.385     -3.553      -2.915      -2.592
-------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000.
```

```
Dickey-Fuller test for unit root              Number of obs  = 69
Variable: DC                                  Number of lags =  0

H0: Random walk without drift, d = 0

                              Dickey-Fuller
              Test      -------- critical value ---------
            statistic        1%           5%          10%
-------------------------------------------------------------
  Z(t)          -5.030      -3.553       -2.915       -2.592
-------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000.
```

which confirms that both C and Y are $I(1)$ as ΔC_t and ΔY_t are individually stationary.

(iii) Now estimate the following equation:

$$\Delta C_t = \alpha + \beta\ \Delta Y_t + (1-\lambda)(Y_{t-1} - C_{t-1}) + \varepsilon_t \tag{6}$$

Answer:

```
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly generate DY = Y - L.Y
quietly generate DC = C - L.C
quietly generate ECM = Y - C

regress DC DY L.ECM
```

```
      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(2, 67)        =     40.93
       Model |  .014518875         2  .007259437   Prob > F        =    0.0000
    Residual |  .011882695        67  .000177354   R-squared       =    0.5499
-------------+----------------------------------   Adj R-squared   =    0.5365
       Total |  .026401569        69  .000382631   Root MSE        =    .01332


------------------------------------------------------------------------------
          DC | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          DY |   .7496825   .0831291     9.02   0.000     .5837562    .9156089
             |
         ECM |
         L1. |   .1372176   .0635656     2.16   0.034      .01034    .2640951
             |
       _cons |   -.004719   .0061754    -0.76   0.447    -.0170451    .0076072
------------------------------------------------------------------------------
```

(iv) Show that equation (6) is a restricted version of equation (2) on page 22 where the long-run elasticity is unity, and test this restriction using your results from (i) and (iii).

Answer: The fact that equation (6) is a restricted version of equation (2) is demonstrated in the next part below.

To test whether the long-run elasticity is unity:

```
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly generate DY = Y - L.Y
quietly generate DC = C - L.C
quietly generate ECM = Y - C

quietly regress C Y L.Y L.C
test Y + L.Y + L.C = 1
```

```
 ( 1)   Y + L.Y + L.C = 1


       F(  1,    66) =     0.02
            Prob > F =    0.8751
```

Thus we cannot reject the null hypothesis that the that the long-run elasticity is unity.

(v) Carefully explain the reasoning behind the formulation given in equation (6), especially discussing the use of the error correction term $(Y_{t-1} - C_{t-1})$.

Answer: In order to understand the reasoning behind equation (6) we need to discuss **error correction mechanism**, or ECM. This was first used by Sargan (1964).[6] Suppose the long-run or equilibrium relationship between consumption and income is given by equation (1) on page (22) where ψ is long-run consumption elasticity. Recall in part (i) we said that economic systems are rarely in equilibrium. When C takes on a different value to its equilibrium value with respect to Y, then we have *disequilibrium error*:

$$\text{disequilibrium error} = u_t = C_t - \rho - \psi Y_t$$

which measures the extent of disequilibrium between the two variables. Thus it is 0 when C and Y are in equilibrium. By taking into account first-order lags, equation (2) incorporates the idea that C is not always at its equilibrium value relative to Y and that it takes time to adjust full to variations in Y.

This equation (2) however involves levels of variables that may not be stationary. To address this we can first-difference it:

$$C_t - C_{t-1} = \alpha + \beta Y_t + \gamma Y_{t-1} + \lambda C_{t-1} + \varepsilon_t - C_{t-1}$$
$$= \alpha + \beta Y_t + \gamma Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$

The next step involves a slight trick whereby we add and subtract βY_{t-1} from the right hand side so that we can also have an expression for ΔY_t:

$$\Delta C_t = \alpha + \beta Y_t - \beta Y_{t-1} + \beta Y_{t-1} + \gamma Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$

---

[6]Sargan J D (1964) "Wages and Prices in the United Kingdom: A Study in Econometric Methodology" in K F Wallis and D F Hendry (eds.), *Quantitative Economics and Econometric Analysis*, Basil Blackwell, Oxford, UK.

$$= \alpha + \beta \ \Delta Y_t + \beta Y_{t-1} + \gamma Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$
$$= \alpha + \beta \ \Delta Y_t + (\beta + \gamma)Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$

Notice that for this to match equation (6) on page 32, we need to have $\beta + \gamma = 1 - \lambda$ because if that is the case then we have:

$$\Delta C_t = \alpha + \beta \ \Delta Y_t + (1 - \lambda)Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$
$$= \alpha + \beta \ \Delta Y_t + (1 - \lambda)Y_{t-1} - (1 - \lambda)C_{t-1} + \varepsilon_t$$
$$= \alpha + \beta \ \Delta Y_t + (1 - \lambda)(Y_{t-1} - C_{t-1}) + \varepsilon_t.$$

This restricted version therefore means that $\beta + \gamma + \lambda = 1$. This equation (6) way of expressing equation (2) can be interpreted in an interesting new way. Equation (2) is the short-run disequilibrium relationship. Accordingly, the term in parenthesis in equation (6), $(Y_{t-1} - C_{t-1})$ is the disequilibrium error from period $t - 1$, or the error correction term. Therefore, this equation states that the current change in C depends both on the change in Y as well as on the extent of disequilibrium error, or the equilibrium error correction term. If this correction term is 0, then the model is in equilibrium. If it is nonzero, then the model is out of equilibrium.

Suppose $\Delta Y_t = 0$ and the error correction term, $(Y_{t-1} - C_{t-1})$, is negative. This means C_{t-1} is above its equilibrium value of $\alpha + \beta Y_{t-1}$. Since $(1 - \lambda)$ is expected to be positive, the term $(1 - \lambda)(Y_{t-1} - C_{t-1})$ is negative and therefore ΔC_t will be negative to restore the equilibrium. That is, if C_t is above its equilibrium value, it will start falling in the next period to correct the disequilibrium error. This is why this mechanism is called Error Correcting Mechanism (ECM).

Similarly, if $(Y_{t-1} - C_{t-1})$ is positive while $\Delta Y_t = 0$ - i.e. C_t is below its equilibrium value - then $(1 - \lambda)(Y_{t-1} - C_{t-1})$ will also be positive, which will cause ΔC_t to be positive, leading C_t to rise in period t.

Thus, the absolute value of $(1 - \lambda)$ determines how quickly the equilibrium is restored. From the regression results in part (iii) above we see that C adjusts to Y with a lag: only about 13.7% of the discrepancy between long-term and short-term C is corrected within one time period. From the same regression result we see that the short-run consumption elasticity is about 0.75. From the regression of the first equation we also know that the long-run consumption elasticity is about 0.99.

(vi) Plot the error correction term used to estimate equation (6) on page 32 and the residuals from your estimate of equation (1) on page 22 over time. What is the relationship between the two series? Perform unit root tests on both series and compare the results. Discuss the significance of these results - especially how these results might be used in situations where the long-run elasticity is not unity. (You might also estimate equation (6) again but this time using errors from equation (1) lagged by one year, instead of the error correction term).

Answer: To plot the error correction term and the residuals from the first equation in R we need to first create new variables in our data frame for the residuals and the error correction term.
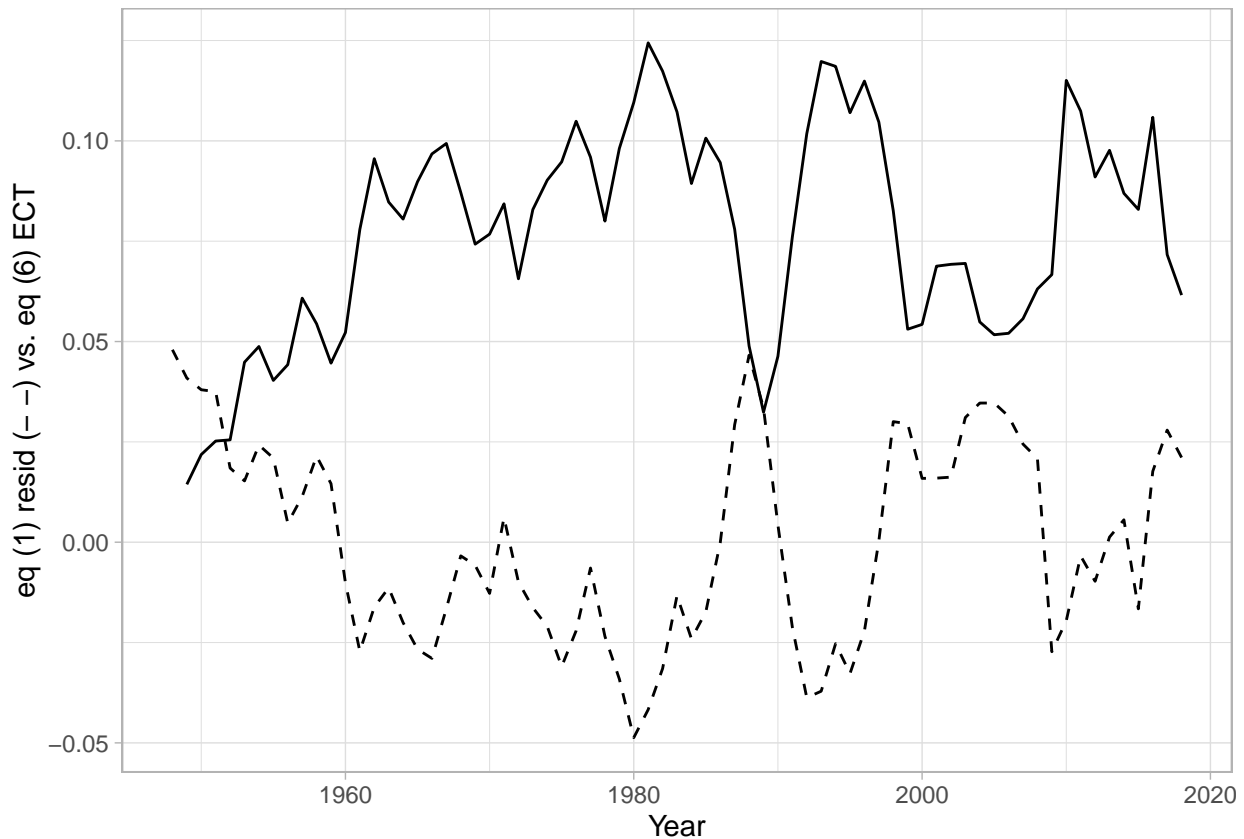
in STATA:

```
quietly cd ..
use Data/BoatRace.dta
quietly tsset Year
reg C Y
predict res, residuals
gen ECT = L.Y - L.C
tsline ECT res
```

and in R:

```r
ts_df <- ts_df %>%
  mutate(
    ECT = Y_lag - C_lag,
    eq1resid = residuals(lm1_SQ3i)
  )


ggplot(ts_df,
       aes(x = Year)) +
  geom_line(aes(y = eq1resid), linetype = "dashed", na.rm=TRUE) +
  geom_line(aes(y = ECT), na.rm=TRUE) +
  theme_light() +
  labs(y = "eq (1) resid (- -) vs. eq (6) ECT")
```


The ECM from equation (6) and the residuals from equation (1) are strongly negatively correlated, though lagged by 1 period. To see why this is the case, recall that the residuals of equation (1) is expressed as:

$$\varepsilon_t = C_t - \rho - \psi Y_t$$
$$= -(\psi Y_t - C_t) - \rho$$

where, ψ in this model is the long un elasticity. If it is assumed to be unitary, then the residuals are:

$$\varepsilon_t = -(Y_t - C_t) - \rho$$

which means

$$\varepsilon_{t-1} = -(Y_{t-1} - C_{t-1}) - \rho$$
$$= -ECM - \rho$$

$$\varepsilon_{t-1} + \rho = -(Y_{t-1} - C_{t-1})$$

This means if we add ρ to the residuals and then multiply by -1 and we should get ECM with 1 lag. The intercept in the first model was not statistically different from 0. Lets plot this:

```r
ts_df <- ts_df %>%
  mutate(
    neg_lag_eq1resid = -(lag(residuals(lm1_SQ3i),1))
  )

ggplot(ts_df,
       aes(x = Year)) +
  geom_line(aes(y = neg_lag_eq1resid), linetype = "dashed", na.rm=TRUE) +
  geom_line(aes(y = ECT), na.rm=TRUE) +
  theme_light() +
  labs(y = "eq (1) -resid (- -) vs. eq (6) ECT")
```
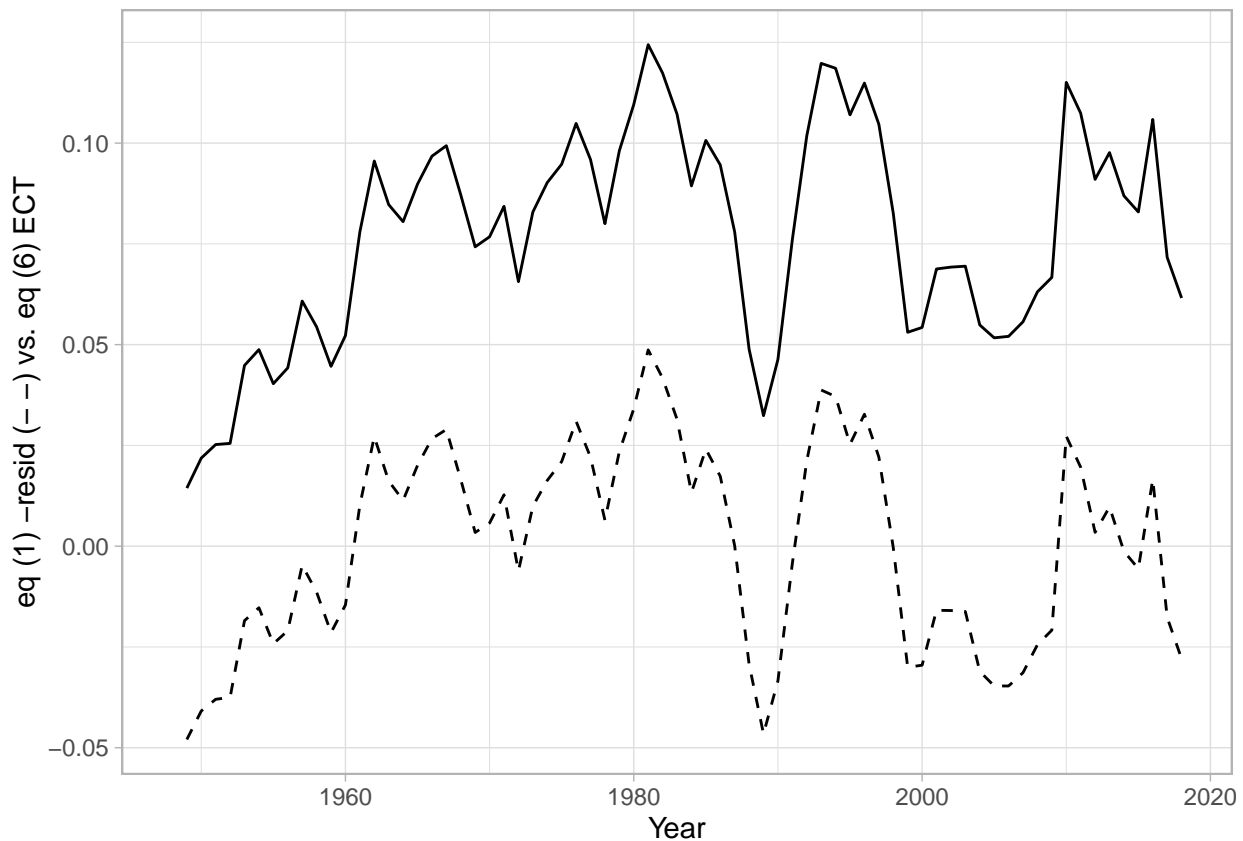

So multiplying the lagged residuals by -1 closely tracks the ECM though not identically.

Next, lets run unit root tests on both series.

```stata
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly regress C Y
predict e, residuals
dfuller e, lags(1)

quietly generate ECT = Y - C
```

```
dfuller ECT, lags(1)
```

```
Augmented Dickey-Fuller test for unit root

Variable: e                                    Number of obs  = 69
                                               Number of lags =  1

H0: Random walk without drift, d = 0

                                 Dickey-Fuller
                   Test      -------- critical value ---------
                 statistic       1%            5%           10%
-----------------------------------------------------------------
  Z(t)            -3.574       -3.553        -2.915        -2.592
-----------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0063.
```

```
Augmented Dickey-Fuller test for unit root

Variable: ECT                                  Number of obs  = 69
                                               Number of lags =  1

H0: Random walk without drift, d = 0

                                 Dickey-Fuller
                   Test      -------- critical value ---------
                 statistic       1%            5%           10%
-----------------------------------------------------------------
  Z(t)            -3.646       -3.553        -2.915        -2.592
-----------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0049.
```

Thus we reject the null that there is a unit root in the residuals of first model and/or in the ECM even at 1% level. Both are stationary with 1 lags.

In order to understand the relationship between these two and how it can be used in situations where the long-run elasticity is not unity, lets reproduce the two equations here and develop an understanding of their components.

1. Long run relations: $C_t = \rho + \psi Y_t + \varepsilon_t$

2. Error correction mechanism: $\Delta C_t = \alpha + \beta \ \Delta Y_t + (1 - \lambda)(Y_{t-1} - C_{t-1}) + \varepsilon_t$

where

- the residuals from the long-run relationship are: $\varepsilon_t = C_t - \rho - \psi Y_t$

- the error correction term in the ECM is: $(Y_{t-1} - C_{t-1})$

The key insight is that these terms are related but not identical. To see the relationship, substitute long run equation for C_t in the error correction term:

$$
\begin{aligned}
(Y_{t-1} - C_{t-1}) &= Y_{t-1} - \rho - \psi Y_{t-1} - \varepsilon_{t-1} \\
&= (1 - \psi)Y_{t-1} - \rho - \varepsilon_{t-1}
\end{aligned}
$$

Now we can check for the behaviors with respect to the long-run elasticity:

- If unity elasticity, $\psi = 1$, and no constant term, $\rho = 0$, then $Y_t = C_t - \varepsilon_t$, and in the case of no errors: $Y_t = C_t$.

  Under these conditions, the error correction term would be the negative of the lagged residuals $-\varepsilon_{t-1}$ since $(Y_t - C_t) = (1 - 1)Y_{t-1} - 0 - \varepsilon_{t-1}$.

- When $\psi = 1$ and $\rho \neq 0$ the error correction term becomes $Y_{t-1} = C_{t-1} = -\rho - \varepsilon_t$. This introduces a constant shift to the error correction process.

  If $\rho > 0$, there is a consistent negative bias in the error correction term

  If $\rho < 0$, therei s a consistent positive bias in the error correction term

- When $\psi \neq 0$ and $\rho = 0$ the error correction term becomes $Y_{t-1} = C_{t-1} = (1 - \psi)Y_{t-1} - \varepsilon_t$.

  Here $(1 - \psi)Y_t$ means the error correction term now depends on the absolute level of income, not just on deviations from the equilibrium, i.e. disequilibrium error.

  If $\psi < 1$ then $(1-\psi)$ is positive and higher income levels contribute positively to the error correction term

  If $\psi > 1$ then $(1 - \psi)$ is negative and higher income levels contribute negatively to the error correction term

  The further ψ is from 1, the larger the influence of income level on the error correction process

- If $\psi \neq 0$ and $\rho \neq 0$ then the error term contains all three components: $Y_{t-1} = C_{t-1} = (1-\psi)Y_{t-1}-\rho-\varepsilon_t$. This creates a complex adjustment process that depends on:

  the income level, scaled by how far ψ is from 1

  A constant shift based on $-\rho$

  the disequilibrium error

This complexity has some important implications:

1. *misspecification concerns*: Using $(Y_t - C_t)$ as the error correction term when $\psi \neq 0$ or $\rho \neq 0$ could lead to biased estimates of adjustment speeds and inccorect inferences about the dynamics.

2. *parameter interpretation*: The coefficient $(1 - \lambda)$ on this complex error correction term does not purely represent the adjustment speed anymore, as it is applied to multiple components. To see this, substitute in the error correction term in the ECM:

$$\Delta C_t = \alpha + \beta \ \Delta Y_t + (1 - \lambda)(1 - \psi)(Y_{t-1} - \rho - \varepsilon_{t-1}) + \varepsilon_t$$

which shows that $(1 - \lambda)$ multiplies three different terms: the income level effect Y_{t-1}, the constant term ρ, and the disequilibrium error ε_{t-1}. This means, $(1 - \lambda)$ simultaneously controls (i) how quickly deviations are corrected, (ii) how much the income level itself affects consumption changes, (iii) how much the constant term affects consumption changes. Because it is scaling other effects, we can no longer say that $(1 - \lambda)$ purely represents adjustment speed. Finally, this ECM implicitly forces the adjusment speed for all components to be identical, which may not reflect the true dynamics.

If we want to incorporate non-unity elasticity into the ECM, we can also modify the error correction term and use the errors from equation (1) lagged by one year, $\varepsilon_{t-1} = C_{t-1} - \rho - \psi Y_{t-1}$:

$$\Delta C_t = \alpha + \beta \ \Delta Y_t + (1 - \lambda)(C_{t-1} - \rho - \psi Y_{t-1}) + \varepsilon_t$$

This way, we have:

- $(1 - \lambda)$ represents the speed of adjustment

- $(C_{t-1} - \rho - \psi Y_{t-1})$ is the deviation from the long run equilibrium, i.e. disequilibrium error

- ψ is the non-unity elasticity parameter.

This formulation directly connect the ECM to the long-run relationship while allowing for non-unity elasticity. It ensures that consumption converges to $C_t = \rho + \psi Y_t$ in the long run while capturing short run dynamics through $\beta \Delta Y_t$.

To derive this expression, begin with the autoregressive distributed lag (ARDL) model of equation (2) on page (2) and subtract C_{t-1} from both sides:

$$C_t = \alpha + \beta Y_t + \gamma Y_{t-1} + \lambda C_{t-1} + \varepsilon_t$$
$$C_t - C_{t-1} = \alpha + \beta Y_t + \gamma Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$

if we then add and subtract βY_{t-1} we have:

$$\Delta C_t = \alpha + \beta(Y_t - Y_{t-1} + Y_{t-1}) + \gamma Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$
$$= \alpha + \beta \ \Delta Y_{t-1} + (\beta + \gamma)Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$

Now notice that mathematically, for a stable system $|\lambda| < 1$ is required. This is because it is the coefficient of the lagged dependent variable. So, for simplification, consider $C_t = \lambda C_{t-1} +$ other terms. As we repeatedly substitute we get $C_t = \lambda^n C)t - n+$ accumulated terms. As n increases, the term $\lambda^n C_{t-n}$ represents how shocks from the distant past (C_{t-n}) affect the present. If $|\lambda| < 1$ then past shocks eventually die out and the system converges to a stable equilibrium. That is, consumption eventually returns to an equilibrium relationship with income after any shock. If $|\lambda| = 1$ then past shocks persist indefinitely and we have a unit root process whereby the system has no equilibrium point. That is, shocks accumulate and the series doesn't revert to any mean. This would mean, consumption never fully adjusts to shocks and temporary disruptions have permanent effects. If $|\lambda| > 1$ then past shocks get amplified over time and small deviations grow explosively resulting in a system that diverges without a bound. This means consumption moves further and further away from any equilibrium relationship with income, which is economically implausible.

For stability, therefore, $|\lambda| < 1$ is required which means $(\lambda - 1) < 0$. Set $-\eta = (\lambda - 1)$ so that we have a positive value for η for stability. η represents the speed of adjustment - how quickly consumption moves back towards equilibrium after a shock. So $\delta > 0$ ensures that deviations from equilibrium are corrected rather than amplified.

Similarly, we can link the income coefficients $(\beta + \gamma)$ to the long run elasticity parameter. Consider what happens in long run equilibrium where $C_t = C_{t-1} = C$ and $Y_t = Y_{t-1} = Y$. The ARDL model, i.e. equation (2), becomes $C = \alpha + \beta Y + \gamma Y + \lambda C$. Solving for C, we have $C(1 - \lambda) = \alpha + (\beta + \gamma)Y$ or $C = \frac{\alpha}{1-\lambda} + \frac{\beta+\gamma}{1-\lambda}Y$. For this to match the long run relationship $C = \rho + \psi Y$, we need $\frac{\beta+\gamma}{1-\lambda} = \psi$. Since we defined $\eta = -(\lambda - 1)$ this expression becomes $\frac{\beta+\gamma}{\eta} = \psi$ or $(\beta + \gamma) = \eta\psi$. This ensures that the dynamic model converges exactly to our specified long-run relationship with the correct elasticity parameter.

We also need the constant term $\frac{\alpha}{1-\lambda}$ to be consistent with the long run relationship. This means, $\frac{\alpha}{\eta} = \rho$ or $\alpha = \eta\rho$.

These three restrictions ensure that:

- the system is stable and returns to equilibrium after shocks,

- the long run relationship between income and consumption has the correct elasticity, ψ,

- the long run relationship has the correct intercept, ρ.

We then apply these restrictions to the ARDL model:

$$\Delta C_t = \alpha + \beta \ \Delta Y_{t-1} + (\beta + \gamma)Y_{t-1} + (\lambda - 1)C_{t-1} + \varepsilon_t$$
$$= \eta\rho + \beta \ \Delta Y_{t-1} + \eta\psi \ Y_{t-1} - \eta \ C_{t-1} + \varepsilon_t$$
$$= \eta\rho + \beta \ \Delta Y_{t-1} - \eta(C_{t-1} - \psi \ Y_{t-1}) + \varepsilon_t$$
$$= \beta \ \Delta Y_{t-1} - \eta(C_{t-1} - \rho - \psi Y_{t-1}) + \varepsilon_t$$

We can estimate this in two ways:

1. We can either estimate the model

$$\Delta C_t = \alpha + \beta \ \Delta Y_t + \eta \ u_{t-1} + \varepsilon_t$$

where $u_{t-1} = C_{t-1} - \rho - \psi Y_{t-1}$. We know from the regression results of equation (1) that ρ is not statistically different from 0 and $\psi = 0.985157$ with 95% CI = [0.9746304, 0.9956836]. Therefore, $u_{t-1} = C_{t-1} - 0.985157 \ Y_{t-1}$.

2. Alternatively, we can estimate the following model

$$\Delta C_t = \alpha + \beta \ \Delta Y_t + \eta \ C_{t-1} - \eta \ \rho - \eta \ \psi \ Y_{t-1} + \varepsilon_t$$

Since ρ is statistically not different from 0 and if we denote $\zeta = -\eta\psi$. Then this expression becomes:

$$\Delta C_t = \alpha + \beta \ \Delta Y_t + \eta \ C_{t-1} + \zeta \ Y_{t-1} + \varepsilon_t$$

In R:

```r
ts_df <- ts_df %>%
  mutate(
    eq1resid_lag = lag(eq1resid, 1, default=NA)
  )
lm1_SQ3vi <- lm(DC ~ DY + eq1resid_lag, data = ts_df)
lm2_SQ3vi <- lm(DC ~ DY + C_lag + Y_lag, data = ts_df)

summary(lm1_SQ3vi)
summary(lm2_SQ3vi)

# obtain psi
-lm2_SQ3vi$coefficients[4]/lm2_SQ3vi$coefficients[3]
# check if psi is different than 1
car::linearHypothesis(lm2_SQ3vi, "Y_lag + C_lag = 0")
```

and in Stata:

```stata
quietly cd ..
use Data/Timeseries1.dta
quietly tsset Year
quietly regress C Y
predict u, residuals

quietly generate DY = Y - L.Y
quietly generate DC = C - L.C

regress DC DY L.u
regress DC DY L.C L.Y
/* obtain psi */
display -_b[L.Y]/_b[L.C]
/*check if psi is different than 1 */
lincom _b[L.Y] + _b[L.C]
```

```
      Source |       SS           df       MS      Number of obs   =        70
-------------+----------------------------------   F(2, 67)        =     40.61
       Model |  .014467673         2  .007233837   Prob > F        =    0.0000
    Residual |  .011933896        67  .000178118   R-squared       =    0.5480
-------------+----------------------------------   Adj R-squared   =    0.5345
       Total |  .026401569        69  .000382631   Root MSE        =    .01335
```

```
--------------------------------------------------------------------------
        DC | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------+--------------------------------------------------------------
        DY |   .7425324    .0825309    9.00   0.000      .5778      .9072648
           |
         u |
       L1. |  -.1396295    .0669286   -2.09   0.041    -.2732195   -.0060395
           |
     _cons |   .0059936    .0027089    2.21   0.030     .0005865    .0114007
--------------------------------------------------------------------------


     Source |       SS          df       MS        Number of obs   =       70
------------+----------------------------------     F(3, 66)        =    26.90
      Model |  .014523357       3   .004841119      Prob > F        =   0.0000
   Residual |  .011878213      66   .000179973      R-squared       =   0.5501
------------+----------------------------------     Adj R-squared   =   0.5296
      Total |  .026401569      69   .000382631      Root MSE        =   .01342


--------------------------------------------------------------------------
        DC | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------+--------------------------------------------------------------
        DY |   .7491507    .0838085    8.94   0.000     .5818217    .9164798
           |
         C |
       L1. |   -.140483    .0672937   -2.09   0.041    -.2748392   -.0061268
           |
         Y |
       L1. |   .1400042    .0664235    2.11   0.039     .0073854    .2726229
           |
     _cons |   .0013938    .0392322    0.04   0.972    -.0769358    .0797235
--------------------------------------------------------------------------
```

.99659163

(1) L.C + L.Y = 0

```
--------------------------------------------------------------------------
        DC | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-----------+--------------------------------------------------------------
       (1) |  -.0004788    .0030342   -0.16   0.875    -.0065368    .0055791
--------------------------------------------------------------------------
```

which gives us:

- $\alpha \sim 0$. Thus the short run autonomous change in consumption is around 0 to 0.006 (the latter is just significantly different from 0 at 5%). This represents the trend or drift component in consumption growth and measures the constant change in consumption that occurs regardless of income changes. Since it is either 0 or tiny little bit positive, it suggests that consumption tends to remain unchanged or grows tiny little bit over time even when income is unchanged which could reflect changing preferences, technology, or other structural factors.

- $\beta \sim 0.74$. Thus, the short run marginal propensity to consume, i.e. the immediate impact of income changes on consumption is around 0.74. In other words it reflects that not its entirety but only part of

income changes are consumed immediately.

- $-\eta = (1 - \lambda) = -0.014$ shows the speed of adjustment to equilibrium. In the ECM of equation (6) this would be positive, and in the ECM of the alternative, more conventional, specification here it would be negative. Thus, this result tells us that about 1.4% of disequilibrium is corrected each period.

  The half life of any deviation can be calculated as $\ln(0.5)/\ln(1 + \eta) = -49.8563$

- $\psi = -\zeta/\eta = -.1400042/-.140483 = 0.99659163$ which is the long run income elasticity of consumption that measures the proportional relationship between income and consumption in the long run. When we test if it is different than 1, we cannot reject that null. Thus a 1% increase in income leads to about 1% increase in consumption in the long run, meaning a constant savings rate since consumption increases proportionally with income in the long run.

- $\rho = 0$. This measures the long run autonomous consumption. It represents the baseline level of consumption that exists even when income is zero. It could reflect subsistence consumption, consumption from wealth, or social safety nets, and thus usually positive since some minimal consumption is necessary for survival. In this question it is 0. In logs, it affects the proportion of income consumed at different income levels.

Usually we would expect long-run MPC (i.e. ψ) > short-run MPC (i.e. β) as households adjust consumption patterns over time. Full adjustment time is approximately $3/(1-\lambda)$ periods. Also note that temporary income change mainly affects consumption by β and permanent income change eventually affects consumption by ψ.

These parameters are central to testing economic theories such as the "Permanent Income Hypothesis", "Absolute Income Hypothesis", and "Relative Income Hypothesis" each of which makes different predictions about these coefficients' values.