Quantitative Analysis of Finance I

ECON90033

WEEK 3

STATIONARITY

DYNAMICS OF FINANCIAL TIME SERIES

THE AUTOCORELATION AND PARTIAL AUTOCORRELATION FUNCTIONS

UNIVARIATE TIME SERIES PROCESSES / MODELS

Reference:

HMPY: § 4.1-4.4

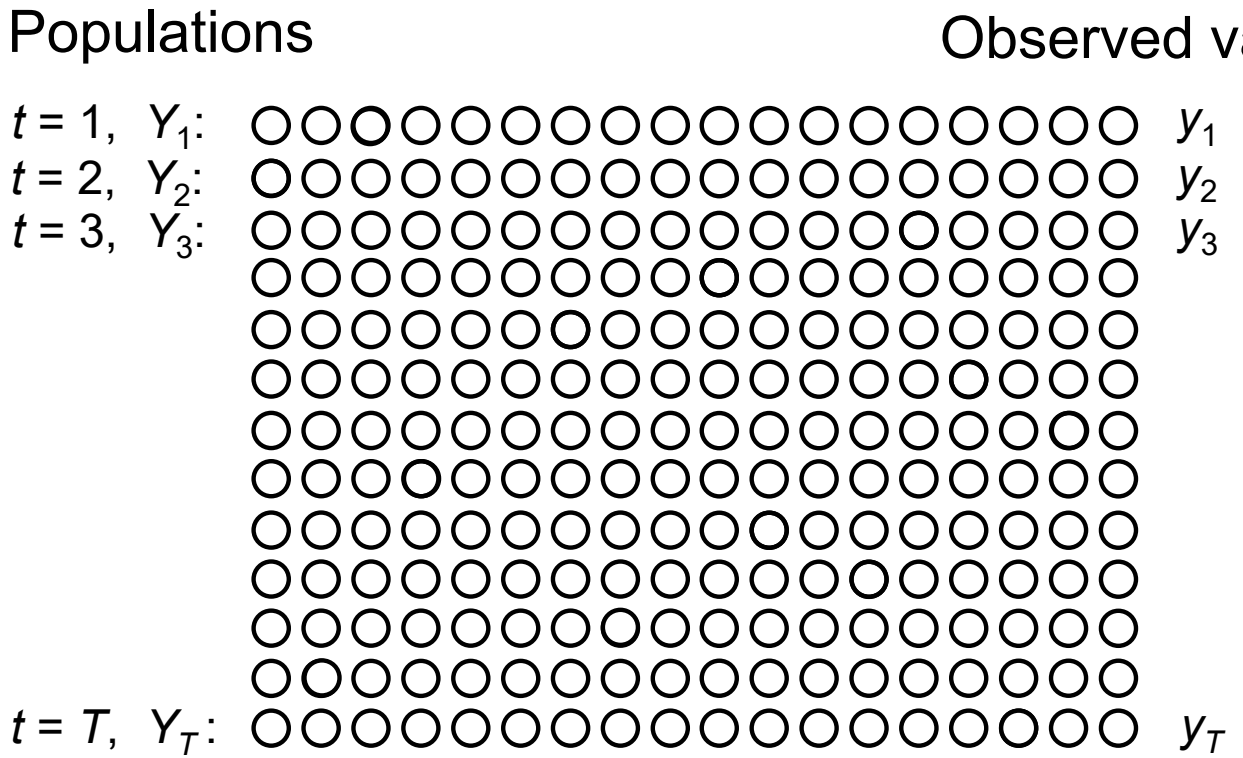
Dr László Kónya

January 2023

STATIONARY STOCHASTIC PROCESSES

- As we discussed in week 1, the modern approach to time series analysis assumes that an observed time series has been generated by a sequence of random variables {…, Y_1, Y_2, …, Y_t, …}, called stochastic process.

  ⟶ The observed time series is a particular realization (sample) of the underlying stochastic process (populations).



By treating the observed values as a random sample of size T, we tacitly assume that the sampled populations are not completely different, i.e., they meet some simplifying assumptions.

- In principle, a stochastic process is specified by its joint probability distribution, i.e., by the probabilities that any given time each y_t falls into some particular interval, and

  a stochastic process $\{y_t\}$ is said to be strongly or strictly stationary if its joint probability distribution at times $t_1, t_2, ..., t_m$ is the same as at times $t_1+k, t_2+k, ..., t_m+k$, where $m > 1$ and k are integers.

  ⟶ A strictly stationary process is invariant with respect to time and thus its properties are unaffected by a change of time origin.

  If $\{y_t\}$ has a normal joint probability distribution, or if it is linear in the sense that y_t is a linear combination of $\{y_{t-1}, y_{t-2}, \ldots\}$ and maybe of current and past values of some other processes, then its properties can be fully captured by the first and second (central) moments.

  These moments are (over $t = 1, 2, \ldots, T$) the

  expected values (means): $E(y_1), E(y_2), ..., E(y_T)$

  and autocovariances: $$Cov(y_t, y_{t-k}) = E\left[(y_t - \mu_t)(y_{t-k} - \mu_{t-k})\right]$$ $$t = 1, 2, ..., T \;,\; k \in \mathbb{Z}$$

- In practice we usually focus on these central moments, even if we do not know for sure whether the joint probability distributions of the stochastic process are normal or linear.

  These moments are unknown population parameters and usually they are estimated from a single sample of T observations for $t = 1, 2, \ldots, T$.

  However, from $t = 1$ to $t = T$ the process has T number of means, T number of variances, and $T(T\text{-}1)/2$ number of different autocovariances ($k \neq 0$), so a sample of size T is not sufficient.

  ⟶ In order to reduce the number of unknown parameters some restrictions must be imposed on the process.

  A frequently made simplifying assumption that makes possible to estimate the moments from sufficiently long time averages instead of multiple observations of the same process over the same time period is weak stationarity.

  A stochastic process is said to be weakly or covariance stationary if it has constant and finite unconditional means, and autocovariances that do not depend on time. In brief,

$$E(y_t) = \mu < \infty \quad \text{and} \quad Cov(y_t, y_{t-k}) = E\left[(y_t - \mu)(y_{t-k} - \mu)\right] = \gamma_k < \infty$$

→ The autocovariances do not depend on *t*, but they might depend on the lag or displacement, *k*.

Moreover, since the variance is the zero order autocovariance

$$\sigma_y^2 = \gamma_0$$

and, since the covariance is a symmetrical measure, for weakly stationary stochastic processes

$$Cov(y_t, y_{t-k}) = Cov(y_{t-k}, y_t) = \gamma_k$$

The autocovariance has two disadvantages: it depends on the units of measurement, and it can assume any real number.

For these reasons, a more convenient measure is the dimensionless and bounded (-1, 1) autocorrelation coefficient, and the second requirement of weak stationarity is often given as

$$\rho_{y_t, y_{t-k}} = \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t) \times Var(y_{t-k})}} = \rho_{y_{t-k}, y_t} = \rho_{y_t, y_{t+k}} = \frac{\gamma_k}{\gamma_0} = \rho_k$$

Note:

a) It can be shown that for weakly stationary processes $\gamma_k \leq \gamma_0$ for any $k \neq 0$. Hence, if the variance is finite, then all other autocovariances are also finite.

b) Strict stationarity implies weak stationarity, granted that the first and second moments are finite, but the converse does not necessarily hold. However, in case of joint normality weak stationarity does imply strict stationarity.

c) From now on we use the term stationarity in the sense of weak stationarity.

d) A stationary process can be described with a model of fixed parameters.

- There are many different stationary processes.

  The simplest one $\{\varepsilon_t\}$, called white noise (*WN*), is a purely random series characterized by the following properties:

$$E(\varepsilon_t) = 0 \qquad Var(\varepsilon_t) = \sigma^2 \qquad \rho_k = 0 \quad (k > 0)$$

  These three properties of $\{\varepsilon_t\}$ are often briefly denoted as

$$\varepsilon_t \sim iid(0, \sigma^2)$$

  where *iid* stands for **i**ndependently and **i**dentically **d**istributed.

Note:

a) Due to the first property of white noise processes

$$Var(\varepsilon_t) = E(\varepsilon_t^2) \quad \text{and} \quad Cov(\varepsilon_t, \varepsilon_{t-k}) = E(\varepsilon_t \varepsilon_{t-k})$$

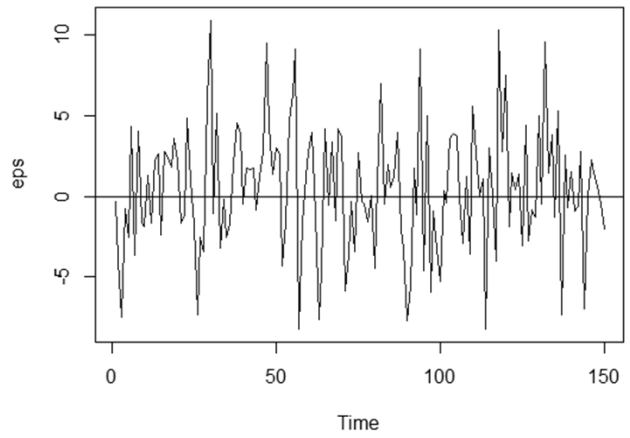
b) Serially uncorrelated series are not necessarily independent, unless they are normally distributed. Consequently, a normally distributed $\{\varepsilon_t\}$ white noise is not only serially uncorrelated but independent as well.

Ex 1:

a) Simulate and plot a normal white noise series for $t = 1, \ldots, 150$, assuming that $\varepsilon_t : N(\mu = 0, \sigma = 4)$.

```
eps = ts(rnorm(150, mean = 0, sd = 4))
ts.plot(eps, col = "blue")
abline(h = 0)
```

As expected, *eps* does not display any discernable pattern.



THE AUTOCORRELATION AND PARTIAL AUTOCORRELATION FUNCTIONS

- The autocorrelation coefficients of a stochastic process (ρ_k) considered as a function of the time gap (k) are referred to as the autocorrelation function (*ACF*), and its plot is called correlogram.

  As an example, consider again a white noise. By definition, it is serially uncorrelated, so all of its autocovariances and autocorrelations are zero, except the ones for $k = 0$.

  In symbols:

$$\gamma_k = \begin{cases} \sigma^2 & if \quad k = 0 \\ 0 & if \quad k \geq 1 \end{cases} \longrightarrow ACF: \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} 1 & if \quad k = 0 \\ 0 & if \quad k \geq 1 \end{cases}$$ (Slide #5)

  In practice the data generating process behind some observed time series is typically unknown. In order to identify it, we estimate the *ACF* from the sample data and try to find a process with a theoretical *ACF* that closely matches the observed sample *ACF* (i.e., *SACF*).

- Certain stationary processes, however, cannot be characterized without ambiguity with their theoretical *ACF*s.

  ⟶ We need another tool, the partial autocorrelation function, *PACF*.

The idea behind the theoretical *PACF* can be best understood by considering the following three-variable population regression model:

$$Y = \alpha + \beta X + \gamma Z + \varepsilon$$

The simple (population) correlation coefficients, ρ_{yx}, ρ_{yz} and ρ_{xz}, measure the strength of linear association b/w *Y* and *X*, *Y* and *Z*, and *X* and *Z*, respectively, without controlling the third variable that might be related to both of them.

⟷ The partial (population) correlation coefficients measure the degree of linear association b/w *Y* and *X*, *Y* and *Z*, and *X* and *Z*, respectively, after having filtered out the potential influence of the third variable.

For example, controlling for the effect of *Z,* the partial correlation coefficient b/w *Y* and *X* is

$$\rho_{yx,z} = \frac{\rho_{yx} - \rho_{yz}\rho_{xz}}{\sqrt{(1-\rho_{yz}^2)(1-\rho_{xz}^2)}}$$

- The *k*-order partial autocorrelation between y_t and y_{t-k}, denoted as φ_{kk}, is the correlation b/w them after controlling for the effect of the intervening y_{t-1},..., y_{t-k+1} variables.

  ⟶ It is the regression coefficient of y_{t-k} in a population regression model of y_t on y_{t-1},..., y_{t-k+1}, y_{t-k} (assuming that $E(y) = 0$)

k = 1 $y_t = \varphi_{11} y_{t-1} + \varepsilon_t$ φ_{11} is the same as ρ_1, i.e., the correlation b/w y_t and y_{t-1};

k = 2 $y_t = \varphi_{21} y_{t-1} + \varphi_{22} y_{t-2} + \varepsilon_t$ φ_{22} is the partial correlation b/w y_t and y_{t-2} after controlling the effect of y_{t-1};

k = 3 $y_t = \varphi_{31} y_{t-1} + \varphi_{32} y_{t-2} + \varphi_{33} y_{t-3} + \varepsilon_t$ φ_{33} is the partial correlation b/w y_t and y_{t-3} after controlling the combined effect of y_{t-1} and y_{t-2} is.

etc.

⟶ The theoretical *PACF* is the sequence of these population partial autocorrelation coefficients φ_{11}, φ_{22}, φ_{33}, ..., and its plot is called partial correlogram.

For illustration, consider again a white noise. Due to the lack of serial correlation, the partial autocorrelations of a white noise are also zero for any nonzero displacement.

$$\longrightarrow \quad PACF: \quad \varphi_{kk} = \begin{cases} 1 & if \quad k = 0 \\ 0 & if \quad k \geq 1 \end{cases}$$

- In practice, the true mean, variance and autocorrelation coefficients of the actual data generating process are unknown, they must be estimated from a sample using the corresponding sample statistics.

  If the underlying stochastic process is *stationary* and *ergodic*, thus its sample moments converge to the corresponding parameters, the first and second population moments (μ, σ^2, ρ_k) can be estimated from a reasonably large sample of observations (y_1, y_2, ..., y_T) as follows:

$$\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$$

$$s_y^2 = \frac{1}{T}\sum_{t=1}^{T} (y_t - \bar{y})^2$$

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{Ts_y^2}$$

k-order sample autocorrelation coefficient

If $T \gg k$ (i.e., far greater), r_k ($k > 0$) can be obtained by estimating the following regression:

$$y_t = \alpha + \beta_k y_{t-k} + \varepsilon_t \longrightarrow r_k = \hat{\beta}_k \frac{s_{y_{t-k}}}{s_{y_t}} \approx \hat{\beta}_k$$

The collection of r_k ($k > 0$) is the sample autocorrelation function, *SACF*, while the sample partial autocorrelation function, *SPACF*, can be obtained by estimating the sequence of models shown on slide #10.

- If $\{y_t\}$ has been generated by a white noise process and T is moderately large, r_k and ϕ_{kk}-hat are approximately normally distributed.

  Namely:

$$r_k \sim N\left(0, \frac{1}{T}\right) \quad \text{and} \quad \hat{\varphi}_{kk} \sim N\left(0, \frac{1}{T}\right)$$

  Thus, the approximate 95% confidence intervals for ρ_k and ϕ_{kk} are

$$r_k \pm \frac{2}{\sqrt{T}} \quad \text{and} \quad \hat{\varphi}_{kk} \pm \frac{2}{\sqrt{T}}$$

  and at the 5% level $H_0 : \rho_k = 0$ ($\varphi_{kk} = 0$) can be rejected in favour of $H_A : \rho_k \neq 0$ ($\varphi_{kk} \neq 0$) if zero is outside the first (second) interval.

Note:

a) The previous decision rules are equivalent to

$$|r_k| > \frac{2}{\sqrt{T}} \quad \text{and} \quad |\hat{\varphi}_{kk}| > \frac{2}{\sqrt{T}}$$ respectively (Bartlett's test).

b) In practice, calculating the sample (partial) autocorrelation coefficients, the largest lag (*s*) should be relatively small compared to the sample size.

We are going to use the following rules of thumb: $s = min(10, T/5)$ for non-seasonal data and $s = min(2S, T/5)$ for seasonal data, where S is the number of seasons.

- From the approximate sampling distribution of r_k

$$\sqrt{T} r_k \sim N(0,1) \longrightarrow T r_k^2 \sim \chi_1^2$$

It can be also shown that the sample autocorrelation coefficients at various displacements are approximately independent of each other.

⟶ In addition to correlograms and confidence bands, one can also run joint hypothesis tests to find out whether the first *s* autocorrelation coefficients, as a group, are significant.

← $H_0: \rho_1 = \rho_2 = \ldots = \rho_s = 0$

versus H_A: at least one of $\rho_1 \ldots \rho_s$ is different from zero.

Box-Pierce (*BP*) test

$$Q_{BP} = T\sum_{k=1}^{s} r_k^2$$

Ljung-Box (*LB*) test

$$Q_{LB} = T(T+2)\sum_{k=1}^{s} \frac{r_k^2}{T-k}$$

If the data are generated by a white noise process, these test statistics are asymptotically chi-squared distributed with $df = s$.

→ Reject H_0 at the $\alpha \times 100\%$ significance level if $Q > \chi^2_{df,1-\alpha}$, or, if the *p*-value of the test is smaller than α.

Note:

a) Like in the *F*-test of overall significance, in the *BP* and *LB* tests H_0 is well specified, but H_A is not. This kind of tests are called portmanteau tests.

b) These tests also serve as a check to see if the residuals from a regression behave as a white noise. In that case $df = s - m$, where m is the number of parameters that were estimated.

c) In large samples, especially when $T \gg k$, $Q_{BP} \approx Q_{LB.}$

In relatively small samples, however, the *LB* test is more powerful than the *BP* test.

d) When performing these tests, it is important to select a reasonable *s* value.

If it is too small, the portmanteau tests might not be informative; if it is too large, the tests can be misleading.

We are going to use the rules of thumb mentioned on slide #13.

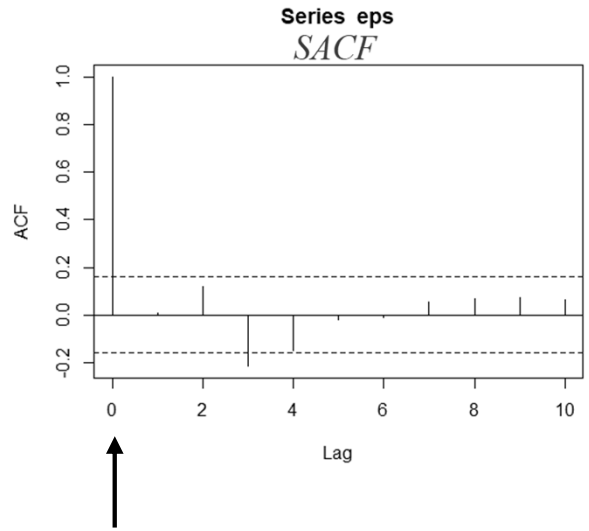
Ex 1: (cont.)

b) Calculate and plot *SACF* and *SPACF* of *eps*.

Since *min*(10, *T*/5) = *min*(10, 150/5) = 10, consider only the first 10 sample autocorrelation and partial autocorrelation coefficients.

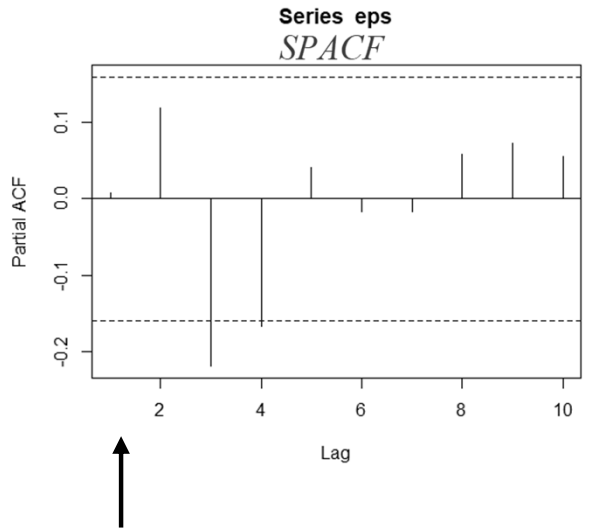
The sample autocorrelation and partial autocorrelation coefficients can be obtained by the *acf()* and *pacf()* *R* functions.

acf(eps, lag.max = 10, plot = TRUE)



$r_0 = 1$, by default.

pacf(eps, lag.max = 10, plot = TRUE)



$k = 1$ is the smallest lag.

On these correlograms the blue dashed lines are the boundaries of the approximate 95% confidence bands:

$$\pm\frac{2}{\sqrt{T}} = \pm\frac{2}{\sqrt{150}} = \pm 0.1632$$

Note: The *eps* series has been generated as a white noise, which is serially uncorrelated, yet r_3, ϕ_{33}-hat and ϕ_{44}-hat appear to be significant.

This is not quite unexpected because the probability of rejecting a true H_0 (i.e., Type I error) is 0.05 in any given test, but the familywise error rate, i.e., the probability of incorrectly rejecting H_0 in at least one of the 10 tests, is $1 - 0.95^{10} = 0.40$.

c) Perform the Box-Pierce and Ljung-Box tests on *eps* at the 5% significance level.

c) Perform the Box-Pierce and Ljung-Box tests on *eps* at the 5% significance level.

These tests can be performed by the *Box.test() R* function.

Box.test(eps, lag = 10, type = "Box-Pierce", fitdf = 0)

```
        Box-Pierce test

data:  eps
X-squared = 15.149, df = 10, p-value = 0.1267
```

Box.test(eps, lag = 10, type = "Ljung-Box", fitdf = 0)

```
        Box-Ljung test

data:  eps
X-squared = 15.782, df = 10, p-value = 0.106
```

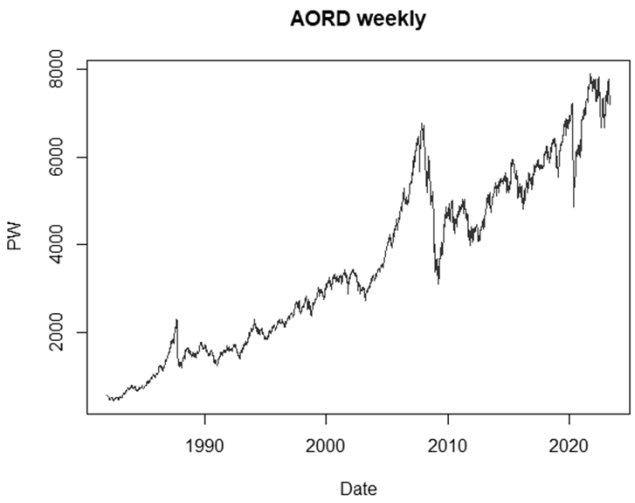
Both *p*-values are above 0.1, so neither test rejects the null hypothesis, not even at the 10% significance level.

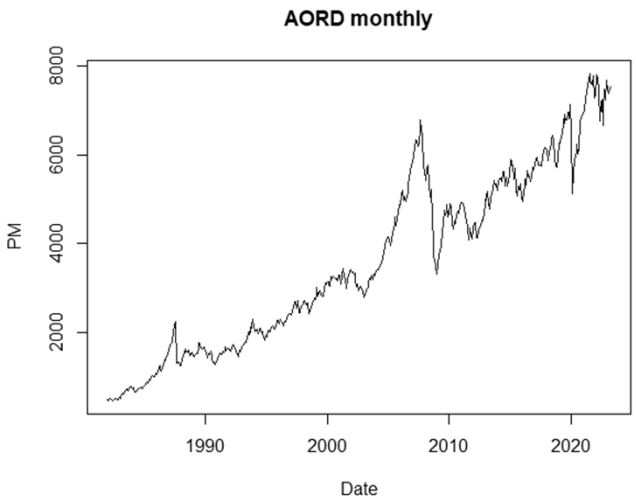
Hence, unlike the set of individual (Bartlett's) tests, these portmanteau tests maintain the null hypothesis of no autocorrelation or orders 1 to 12.

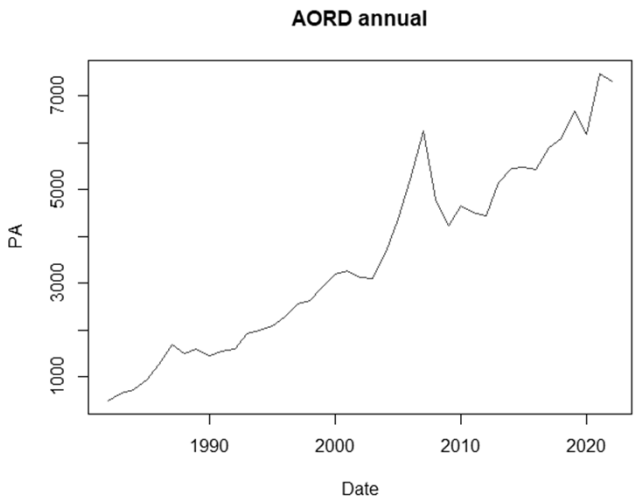
Ex 2:

In Ex 2 of week 1 we considered weekly ASX All Ordinaries Index (*PW*, from week 1 1982 to week 17 2023). Let's now see monthly *AORD* (*PM*, from January 1982 to April 2023) and annual *AORD (PA*, from 1982 to 2022) too.

a) Plot *PW*, *PM*, and *PA*. What can you observe?



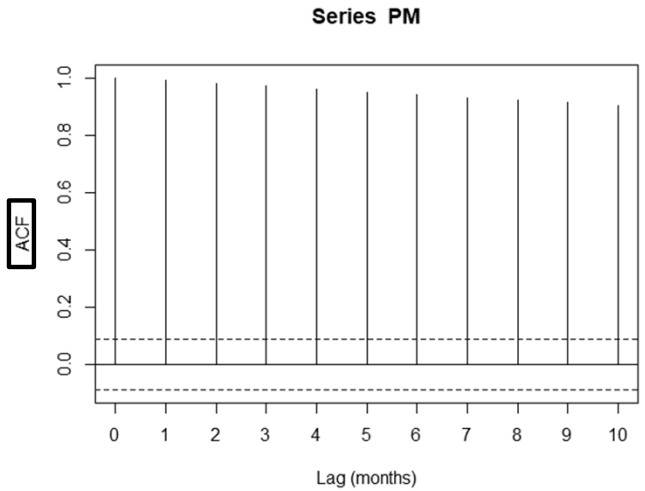




Irrespectively of the frequency, *AORD* has an upward trend, hence its mean changes in time. It is also visible that as the data is aggregated from weekly to monthly and then to annual, the curve becomes smoother and smoother.

b) Calculate and plot *SACF* and *SPACF* of *PM*.

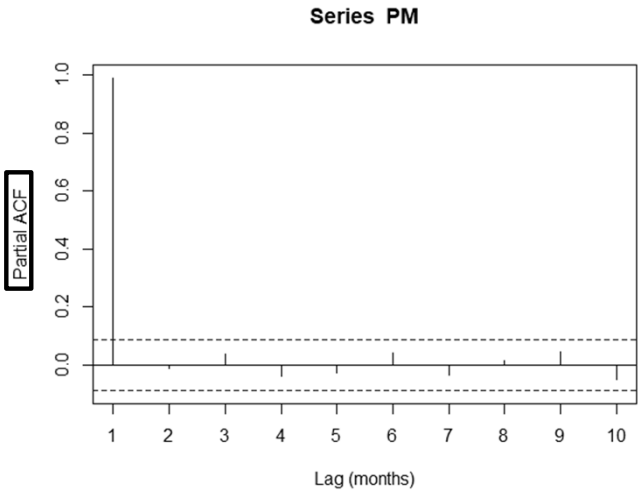
```
ka = floor(min(10, length(PM)/5))
acf(PM, lag.max = ka, xlab="Lag (months)",
     xaxt = "n", plot = TRUE)
axis(1, at = 0:ka/12, labels = 0:ka)
```



$r1$, r_2, …, r_{10} are all significant (at the 5% level) and appear to decay rather slowly.

s = 10 because $T/5$ = 496/5 > 10.

```
pacf(PM, lag.max = ka, xlab="Lag (months)",
     xaxt = "n", plot = TRUE)
axis(1, at = 0:ka/12, labels = 0:ka)
```



Only ϕ_{11}-hat is significant (at the 5% level).

c) Perform the Box-Pierce and Ljung-Box tests on *PM* at the 5% level.

```
Box.test(PM, lag = ka,
         type = "Box-Pierce", fitdf = 0)
```

```
        Box-Pierce test

data:  PM
X-squared = 4443.6, df = 10, p-value < 2.2e-16
```

```
Box.test(PM, lag = ka,
         type = "Ljung-Box", fitdf = 0)
```

```
        Box-Ljung test

data:  PM
X-squared = 4510.2, df = 10, p-value < 2.2e-16
```

Both tests reject the null hypothesis of no autocorrelation of orders 1 to 10 (*ka*) at any reasonable significance level.

⟶ The time series plot, the Bartlett's tests (i.e., the correlograms) and the *BP* and *LB* tests alike indicate that *PM* does not behave as a white noise.

UNIVARIATE TIME SERIES PROCESSES / MODELS

- Regression models based on cross-sectional data are always static in the sense that all observations belong to the same time period and hence the alleged relationship between the dependent and the independent variables is contemporaneous.

  Time series regressions are also static when the current value of the dependent variable is modelled exclusively with the current value(s) of the independent variable(s). Static models, however, are often too restrictive and unrealistic because real-life relationships tend to be dynamic, i.e., the current value of the dependent variable is related to its own past and/or to lagged value(s) of other variable(s).

  This week we briefly review dynamic models for a single stationary process. These models are dynamic either through the own lags of the dependent variable, or through the lags of the error terms, or both.

  We are going to consider three dynamic univariate time series models, popularized by Box, G. and Jenkins, G. (1970): the moving average (*MA*), autoregressive (*AR*), and mixed autoregressive moving average (*ARMA*) models.

MOVING AVERAGE (*MA*) PROCESSES / MODELS

- A moving average model order q, $MA(q)$, assumes that the dependent variable is fully determined by current and lagged white noise innovations (shocks), i.e.

$$y_t = \alpha + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}$$

(θ: theta)

$\theta_0 \equiv 1$, $\varepsilon_t \sim iid(0, \sigma^2)$ ⟶ ε_t is a white noise, so it is stationary.

The unconditional mean and variance of y_t are

$$E(y_t) = \alpha + E\left[\sum_{i=0}^{q} \theta_i \varepsilon_{t-i}\right] = \alpha + \sum_{i=0}^{q} \theta_i \underbrace{E(\varepsilon_{t-i})}_{0} = \alpha$$

$$Var(y_t) = Var\left[\sum_{i=0}^{q} \theta_i \varepsilon_{t-i}\right] = \sum_{i=0}^{q} \theta_i^2 \underbrace{Var(\varepsilon_{t-i})}_{\sigma^2} = \sigma^2 \sum_{i=0}^{q} \theta_i^2 = \gamma_0$$

... and it can be also shown that the autocovariances are

$$Cov\,(y_t, y_{t-k}) = \gamma_k = \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{k+i} \text{ if } k \le q \text{ and } 0 \text{ if } k > q$$

These unconditional first and second moments do not depend on *t*, so all finite-order *MA*(*q*) processes are stationary.

From the autocovariances the autocorrelation coefficients are

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{i=0}^{q-k} \theta_i \theta_{k+i}}{\sum_{i=0}^{q} \theta_i^2} \text{ if } k \le q \text{ and } 0 \text{ if } k > q$$

⟶ *ACF* is nonzero for $k \le q$ and cuts to zero for $k > q$.

⟶ The memory of an *MA*(*q*) process is *q* periods long.

In practice, this property can be used to determine the order of the *MA* model. Namely, if the *SACF* of some observed time series has a distinct cut off point at lag *q* (i.e., r_k is significant for $k \le q$ but insignificant for $k > q$), then it might be described with an *MA*(*q*) model.

AUTOREGRESSIVE (*AR*) PROCESSES / MODELS

- An autoregressive model of order p, $AR(p)$, is specified as

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t$$

$$\varepsilon_t \sim iid(0, \sigma^2)$$

(φ: phi)

⟶ ε_t affects y_t, and it might also affect y_{t+1}, y_{t+2}, etc., but it has no effect on y_{t-1}, y_{t-2}, etc.

- Consider the special case of $AR(1)$, i.e., $p = 1$:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t \quad , \quad \varphi_1 \neq 0$$

⟶ $$E(y_t) = \varphi_0 + \varphi_1 E(y_{t-1})$$ … and assuming that y_t is stationary,

$$E(y) = \varphi_0 + \varphi_1 E(y) \longrightarrow E(y) = \frac{\varphi_0}{1 - \varphi_1} \quad , \quad \varphi_1 \neq 1$$

$$\longrightarrow \quad Var(y_t) = Var(\varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t)$$

$$= \varphi_1^2 Var(y_{t-1}) + Var(\varepsilon_t) + 2\varphi_1 \underbrace{Cov(y_{t-1}, \varepsilon_t)}_{0}$$

… and assuming again that y_t is stationary,

$$Var(y) = \varphi_1^2 Var(y) + \sigma^2$$

$$\longrightarrow \quad Var(y) = \gamma_0 = \frac{\sigma^2}{1-\varphi_1^2} \;,\; 1-\varphi_1^2 > 1 \quad \longrightarrow \quad |\varphi_1| < 1$$

Stationarity condition of *AR*(1)

Given that this condition holds, it can be shown that the autocovariances and autocorrelations are

$$Cov(y_t, y_{t-k}) = \gamma_k = \varphi_1^k \frac{\sigma^2}{1-\varphi_1^2} \quad \longrightarrow \quad \rho_k = \frac{\gamma_k}{\gamma_0} = \varphi_1^k$$

$\longrightarrow$ The *ACF* of a stationary *AR*(1) process does not have a cut-off point, so the memory of this process is infinitely long.

Its *PACF*, however, does have a cut-off point at lag 1 because y_t depends directly only on y_{t-1}.

This property can be generalized and in practice it can be used to determine the order of the *AR* model. Namely, if the *SPAC*F of some observed time series has a distinct cut off point at lag p (i.e., φ_{kk}-hat is significant for $k \leq p$ but insignificant for $k > p$), then it might be described with an *AR*(p) model.

- Expand the *AR*(1) model as follows:

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t$$

$$= \varphi_0 + \varphi_1(\varphi_0 + \varphi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varphi_0 + \varphi_1\varphi_0 + \varphi_1^2 y_{t-2} + \varepsilon_t + \varphi_1\varepsilon_{t-1}$$

$$= \ldots = \varphi_0(1 + \varphi_1 + \varphi_1^2 + \ldots + \varphi_1^k) + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \ldots + \varphi_1^k\varepsilon_{t-k} + \varphi_1^{k+1} y_{t-k+1}$$

$$\frac{1-\varphi_1^{k+1}}{1-\varphi_1}, \quad \varphi_1 \neq 1$$

MA(∞) representation

⟶ If $|\varphi_1| < 1$ and $k \to \infty$,

$$y_t = \frac{\varphi_0}{1-\varphi_1} + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \varphi_1^2\varepsilon_{t-2} + \ldots$$

In general, a finite order *AR*(*p*) process is stationary if its characteristic roots are inside the unit circle, and in this case it has an equivalent *MA*(∞) representation.

- Similarly, the *MA*(1) model can be expanded as follows:

$$y_t = \alpha + \varepsilon_t + \theta_1 \varepsilon_{t-1} = \alpha + \varepsilon_t + \theta_1 (y_{t-1} - \alpha - \theta_1 \varepsilon_{t-2})$$

$$= \alpha(1-\theta_1) + \varepsilon_t + \theta_1 y_{t-1} - \theta_1^2 \varepsilon_{t-2} = \ldots$$

$$= \alpha(1-\theta_1-\theta_1^2-\ldots) + \sum_{i=1}^{\infty} (-1)^{i+1} \theta_1^i y_{t-i} + \varepsilon_t \quad \longleftarrow \quad AR(\infty)$$

$$\frac{\alpha}{1-\theta_1}, \ |\theta_1| < 1 \quad \longrightarrow$$

If $|\theta_1| < 1$, θ_1^i vanishes as *i* approaches infinity, and *AR*(∞) is stationary.

$\longrightarrow$ If $|\theta_1| < 1$, the *MA*(1) model has an equivalent stationary *AR*(∞) representation, and it is said to be invertible.

In general, every *MA*(*q*) model that has an equivalent *AR*(∞) representation is called invertible.

MIXED AUTOREGRESSIVE MOVING AVERAGE (*ARMA*) PROCESSES / MODELS

- The combination of an *AR*(*p*) model and a *MA*(*q*) model is called a mixed autoregressive moving average model, *ARMA*(*p*,*q*).

$$\longrightarrow \quad y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \quad , \quad \varepsilon_t \sim iid(0, \sigma^2)$$

ARMA(*p,q)* is stationary and has a *MA*(∞) representation if the *AR* component is stationary, and is invertible and has an *AR*(∞) representation if the *MA* component is invertible.

For example, in the special case of *ARMA*(1,1), i.e., $p = q = 1$,

$$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad , \quad \varphi_1 \neq 0 \, , \theta_1 \neq 0$$

stationarity requires $|\varphi_1| < 1$ and invertibility requires $|\theta_1| < 1$.

(Ex 2)

In part (c) we developed the sample correlograms for *PM* and observed that its *SACF* exhibited a rather slowly, maybe exponentially, decaying pattern, while its *SPACF* had an abrupt drop after lag 1. Based on these observations, an *AR*(1) model seems to be a good option. Nevertheless, for the sake of illustration, we are going to consider other specifications, as well.

d) Fit an *AR*(1) model augmented with a deterministic linear trend to *PM.*

We use this trend variable to account for the upward trend in the *PM* data series.

```
tm = ts(1:length(PM), frequency = 12,
    start = c(1982, 1))
```

An *AR*(1) model, or *AR*(*p*) models in general, can be estimated by OLS, similarly to standard multiple linear regression models.

Nevertheless, it is better to rely on the *R Arima()* function of the *forecast* package that was specifically designed for *ARMA* models, of which *AR*(1) is just a special case. If there is some extra regressor in the model, like a deterministic trend, this function fits a regression model with *ARMA* errors, i.e., puts all the dynamics in the error term.

In the current example this means that instead of

$$y_t = \varphi_0 + \lambda t + \varphi_1 y_{t-1} + \varepsilon_t \quad , \quad \varepsilon_t \sim iid(0, \sigma^2)$$

Arima() estimates

$$y_t = \tilde{\varphi}_0 + \tilde{\lambda} t + \eta_t$$
$$\eta_t = \varphi_1 \eta_{t-1} + \varepsilon_t$$

where ε_t is the same white noise error as above, but η_t is an *AR*(1) error, so it is not a white noise.

Note: These two specifications are equivalent and produce the same estimates of y_t.

ar1ma0_PM = Arima(PM, c(1,0,0), xreg = tm)
summary(ar1ma0_PM)

```
Series: PM
Regression with ARIMA(1,0,0) errors

Coefficients:
         ar1  intercept     xreg
      0.9461   208.0000  13.6925
s.e.  0.0142   263.1221   0.9016

sigma^2 = 29641:  log likelihood = -3257.05
AIC=6522.1   AICc=6522.18   BIC=6538.93
```

$$y = PM$$
$$\hat{y}_t = 208.0000 + 13.6925t + \eta_t$$
$$\eta_t = 0.9461\eta_{t-1} + e_t$$

e) Fit an *MA*(1) model augmented with a deterministic linear trend to *PM.*

$$y_t = \varphi_0 + \lambda t + \varepsilon_t + \theta_1 \varepsilon_{t-1} \quad , \quad \varepsilon_t \sim iid(0, \sigma^2)$$

Unlike *AR*(*p*) models, *MA*(*q*) and *ARMA*(*p*,*q*) models cannot be estimated directly (i.e., in one step) by OLS because (some of) the independent variables are unobservable lagged error terms. Instead, they can be estimated with an iterative non-linear least-squares method (*EViews*), or with the maximum likelihood method (*R*).

ar0ma1_PM = Arima(PM, c(0,0,1), xreg = tm)
summary(ar0ma1_PM)

```
Series: PM
Regression with ARIMA(0,0,1) errors

Coefficients:
          ma1  intercept     xreg
      0.8345   209.5076  13.5826
s.e.  0.0184    52.9361   0.1845

sigma^2 = 104036:  log likelihood = -3567.9
AIC=7143.81   AICc=7143.89   BIC=7160.63
```

$$y = PM$$
$$\hat{y}_t = 209.5076 + 13.5826t + \eta_t$$
$$\eta_t = e_t + 0.8345 e_{t-1}$$

These are three popular model selection criteria: the *Akaike Information Criterion* (*AIC*), the *Corrected Akaike Information Criterion* (*AICc*), and the *Bayesian information criterion* (*BIC*, also known as the *Schwartz Information Criterion, SIC*).

These statistics are not informative on their own, but when alternative models are estimated for the same dependent variable from the same sample data, the model with the smallest *AIC* or *AICc* or *BIC* value can be considered the 'best'.

In our example all three statistics are in favour of the *AR*(1) model so far.

ARMA model selection based on *SACF* and *SPACF* is often subjective to some degree. For this reason, it is worth to experiment with some 'automated' model selection algorithm, like the *auto.arima()* function of the *forecast R* package, which tries to find the 'best' *AR*(*I*)*MA* model by combining a stepwise search with unit root tests (see later) and the minimisation of a model selection criterion.

f) Find the 'best' *ARMA* model for *PM* by running *auto.arima().*

```
best.arima = auto.arima(PM, seasonal = FALSE,
                        ic = "aicc", xreg = tm,
                        approximation = FALSE, …)
summary(best.arima)
```

← In general, *AICc* works well.

auto.arima() estimated 43 specifications and *ARMA*(2,1) produced the smallest *AICc* (see the details on the next slide).

```
Regression with ARIMA(0,0,0)            errors : 7657.213
Regression with ARIMA(0,0,0)            errors : 7640.349
Regression with ARIMA(0,0,1)            errors : 7157.251
Regression with ARIMA(0,0,1)            errors : 7143.887
Regression with ARIMA(0,0,2)            errors : 6919.192
Regression with ARIMA(0,0,2)            errors : 6909.596
Regression with ARIMA(0,0,3)            errors : 6792.697
Regression with ARIMA(0,0,3)            errors : 6785.841
Regression with ARIMA(0,0,4)            errors : 6705.096
Regression with ARIMA(0,0,4)            errors : 6699.872
Regression with ARIMA(0,0,5)            errors : 6650.286
Regression with ARIMA(0,0,5)            errors : 6646.807
Regression with ARIMA(1,0,0)            errors : 6520.754
Regression with ARIMA(1,0,0)            errors : 6522.184
Regression with ARIMA(1,0,1)            errors : 6521.912
Regression with ARIMA(1,0,1)            errors : 6523.304
Regression with ARIMA(1,0,2)            errors : 6522.09
Regression with ARIMA(1,0,2)            errors : 6523.542
Regression with ARIMA(1,0,3)            errors : 6520.719
Regression with ARIMA(1,0,3)            errors : 6522.112
Regression with ARIMA(1,0,4)            errors : 6522.753
Regression with ARIMA(1,0,4)            errors : 6524.147
Regression with ARIMA(2,0,0)            errors : 6522.02
Regression with ARIMA(2,0,0)            errors : 6523.417
Regression with ARIMA(2,0,1)            errors : 6520.378
Regression with ARIMA(2,0,1)            errors : 6521.79
Regression with ARIMA(2,0,2)            errors : 6521.895
Regression with ARIMA(2,0,2)            errors : 6523.351
Regression with ARIMA(2,0,3)            errors : 6521.834
Regression with ARIMA(2,0,3)            errors : Inf
Regression with ARIMA(3,0,0)            errors : 6522.494
Regression with ARIMA(3,0,0)            errors : 6523.952
Regression with ARIMA(3,0,1)            errors : 6521.848
Regression with ARIMA(3,0,1)            errors : 6523.303
ARIMA(3,0,2) with zero mean     : Inf
ARIMA(3,0,2) with non-zero mean : Inf
Regression with ARIMA(4,0,0)            errors : 6520.641
Regression with ARIMA(4,0,0)            errors : 6522.015
Regression with ARIMA(4,0,1)            errors : 6522.194
Regression with ARIMA(4,0,1)            errors : 6523.415
Regression with ARIMA(5,0,0)            errors : 6522.63
Regression with ARIMA(5,0,0)            errors : 6523.995
```

Each model was estimated twice, first with zero mean and then with non-zero mean.

```
Best model: Regression with ARIMA(2,0,1)

Series: PM
Regression with ARIMA(2,0,1) errors

Coefficients:
         ar1     ar2     ma1     xreg
      0.2495  0.6563  0.7641  14.2972
s.e.  0.1639  0.1573  0.1472   0.4772

sigma^2 = 29468:  log likelihood = -3255.13
AIC=6520.26   AICc=6520.38   BIC=6541.29
```

$$y = PM$$

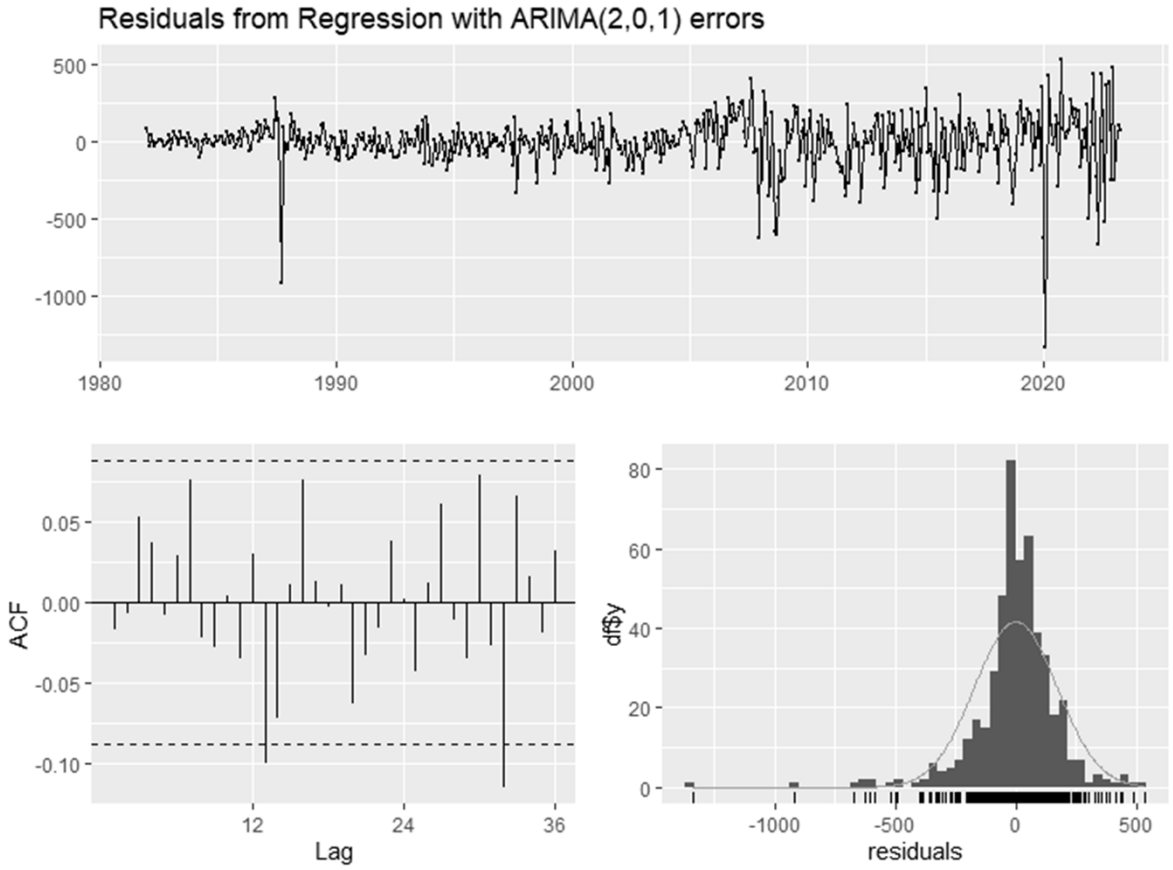
$$\hat{y}_t = 14.2972t + \eta_t$$

$$\eta_t = 0.2495\eta_{t-1} + 0.6563\eta_{t-2} + e_t + 0.7641e_{t-1}$$

g) Correctly specified *ARMA* models should have white noise residuals. Check whether the model selected by *auto.arima()* satisfies this requirement.

This can be done with the *checkresiduals()* function of the forecast package.

checkresiduals(best.arima)




```
Ljung-Box test

data:  Residuals from Regression with ARIMA(2,0,1) errors
Q* = 21.84, df = 21, p-value = 0.4088

Model df: 3.   Total lags used: 24
```

Recall the three properties of a white noise.

i. $E(\varepsilon_t) = 0$

A non-zero but constant mean would be absorbed by the intercept.

ii. $Var(\varepsilon_t) = \sigma^2$

The residual plot seems to suggest otherwise (we shall return to this issue later).

iii. $\rho_k = 0$ $(k > 0)$

Apart from two 'random' spikes *SACF* looks fine, and the *LB* test maintains the null hypothesis of no autocorrelation.

WHAT SHOULD YOU KNOW?

- Strong and weak stationarity
- White noise
- Autocorrelation and partial autocorrelation, *ACF*, *SACF*, *PACF*, *SPACF*, correlogram
- Bartlett's test, Box-Pierce test, Ljung-Box test
- *MA*(q), *AR*(p), and *ARMA*(p,q) models

BOARD OF FAME

George Edward Pelham Box (1919-2013)

British statistician

Statistician for Imperial Chemical Industries

Director of the Statistical Research Group at Princeton University

Professor of statistics at University of Wisconsin–Madison

Time series analysis, Quality control, Design of experiments, Bayesian inference

Gwilym Meirion Jenkins (1932-1982):

Welsh statistician and systems engineer

Professor at Imperial College London, Stanford University, Princeton University, University of Wisconsin–Madison, Lancaster University

Time series analysis, Forecasting