Time Series Analysis

Lecture 4

Mixed Autoregressive Moving Average (ARMA) Models

Autoregressive Integrated Moving Average (ARIMA) Models

Seasonal ARIMA (SARIMA) Models

datascience@berkeley

Mathematical Formulation and Properties of ARMA Models

Mathematical Formulation of ARMA(p,q) Models

A time series $x_t : \cdots -2, -1, 0, 1, 2, \ldots$ is called a mixed autoregressive moving average process of order (p,q), **ARMA(p,q)**, if it is stationary and takes the following functional form

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t - \theta_1 w_{t-1} - \cdots - \theta_q w_{t-q} \qquad (4.1.1)$$

where $\phi_p \neq 0, \theta_q \neq 0$, and $\sigma_w^2 > 0$. Also, we implicitly assume that the series x_t is demeaned: $x_t - \mu$. To simplify notations, we do not use $\tilde{x}$ where $\tilde{x} = x_t - \mu$

The parameters p and q are called autoregressive and the moving average orders.

To incorporate a non-zero mean, μ into the model, we set $\alpha = \mu\,(1 - \phi_1 - \cdots - \phi_p)$ and re-write the model as

$$x_t = \alpha + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + \cdots + \theta_q w_{t-q} \qquad (4.1.2)$$

where w_t is assumed to be a Gaussian white noise series with mean zero and variance σ_w^2.

Mathematical Formulation and Properties of ARMA Models

Mathematical Formulation of ARMA Models (2)

Using AR and MA operators defined in the last lecture, the ARMA(p,q) model can be expressed concisely as

$$\phi(B)x_t = \theta(B)w_t \tag{4.1.3}$$

As the process is stationary, its mean is constant over time:

$$\boxed{\mu = \alpha + \phi_1\mu + \cdots + \phi_p\mu} \tag{4.1.4}$$

or

$$\mu = \frac{\alpha}{1 - \phi_1 - \cdots - \phi_p} \tag{4.1.5}$$

Properties of ARMA Models: The Case of ARMA(1,1)

To study the properties of **ARMA(p,q)** model, let's first consider the **ARMA(1,1)** model:

$$x_t = \phi x_{t-1} + \omega_t + \theta \omega_{t-1} \tag{4.1.6}$$

where $\{\omega_t\} \sim WN(0, \sigma^2)$ and $\phi + \theta \neq 0$. 4

Rewriting the model using the backward shift operator:

$$\phi(B) x_t = \theta(B) \omega_t \tag{4.1.7}$$

where $\phi(B)$ and $\theta(B)$ are the linear filters

$$\phi(B) = 1 - \phi B \text{ and } \theta(B) = 1 + \theta(B) \tag{4.1.8}$$

where the zeros of $\phi(z)$ lie outside the unit circle for the model to be stationary. In other words, in the simple ARMA(1,1) model, A stationary solution exists if and only if $\phi \neq 1 or -1$.

Mean, Variance, and Autocovariance of ARMA(1,1) Model

Let's write the process using the infinite MA representation:

$$x_t = \sum_{j=0}^{\infty} \psi_j \omega_{t-j} \tag{4.1.9}$$

It follows immediately that $E(x_t) = 0$.

The autocovariance function is

$$\gamma(h) = cov(x_{t+h}, x_t) = \sigma_\omega^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \qquad h \geq 0 \tag{4.1.10}$$

Autocorrelation Function of the ARMA(1,1) Model

ACF of ARMA(1,1):

The autocovariance function satisfies

$$\gamma(h) - \phi\gamma(h-1) = 0, \quad h = 2, 3, \ldots \tag{4.1.11}$$

With a few lines of algebra, we can arrive at the following form:

$$\gamma(h) = \frac{\gamma(1)}{\phi}\phi_h = \sigma_\omega^2 \frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2}\phi^{h-1} \tag{4.1.12}$$

As the autocorrelation function at time lag h is the ratio of the autocovariance function at time lag h divided by the variance:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} \tag{4.1.13}$$

$$\rho(h) = \frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2}\phi^{h-1} \qquad h \geq 1 \tag{4.1.14}$$

Autocorrelation Function of the ARMA(1,1) Model (2)

Recall that the ACF of an AR(1) model takes the form

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \qquad h \geq 0 \tag{4.1.15}$$

Note that the general form of the autocorrelation functions between the two models are not that different from each other. For this reason, it is very hard to distinguish an ARMA(1,1) and an AR(1) model based only on the sample ACF.

Comparing ARMA Models and AR Models Using Simulated Series, Part 1

ACF of ARMA and AR Models Recap

- Recall that the ACF of ARMA models and AR models are hard to distinguish. Their functional forms are:

ACF of ARMA(1,1) Model

$$\boxed{\rho(h) = \underbrace{\frac{(1+\theta\phi)(\phi+\theta)}{1-\phi^2}}\phi^{h-1} \qquad h \geq 1}$$

ACF of AR (1) Model

$$\boxed{\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^h, \qquad h \geq 0}$$

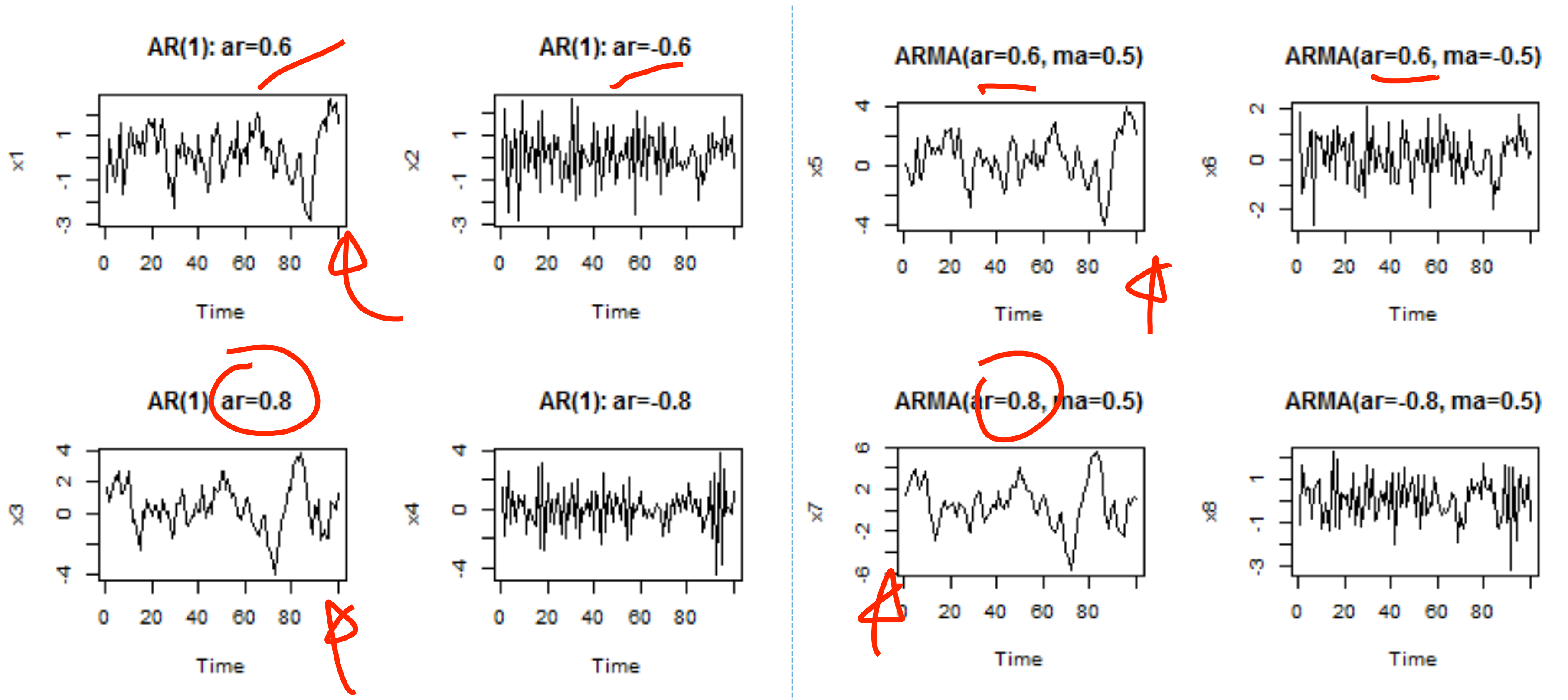
- Their functional forms differ only by a constant multiple.

- Therefore, using ACF alone is not sufficient for model identification.

Time-Series Plots of AR(1) and ARMA(1,1) Models

- All the simulations used in this module have 100 simulated points.

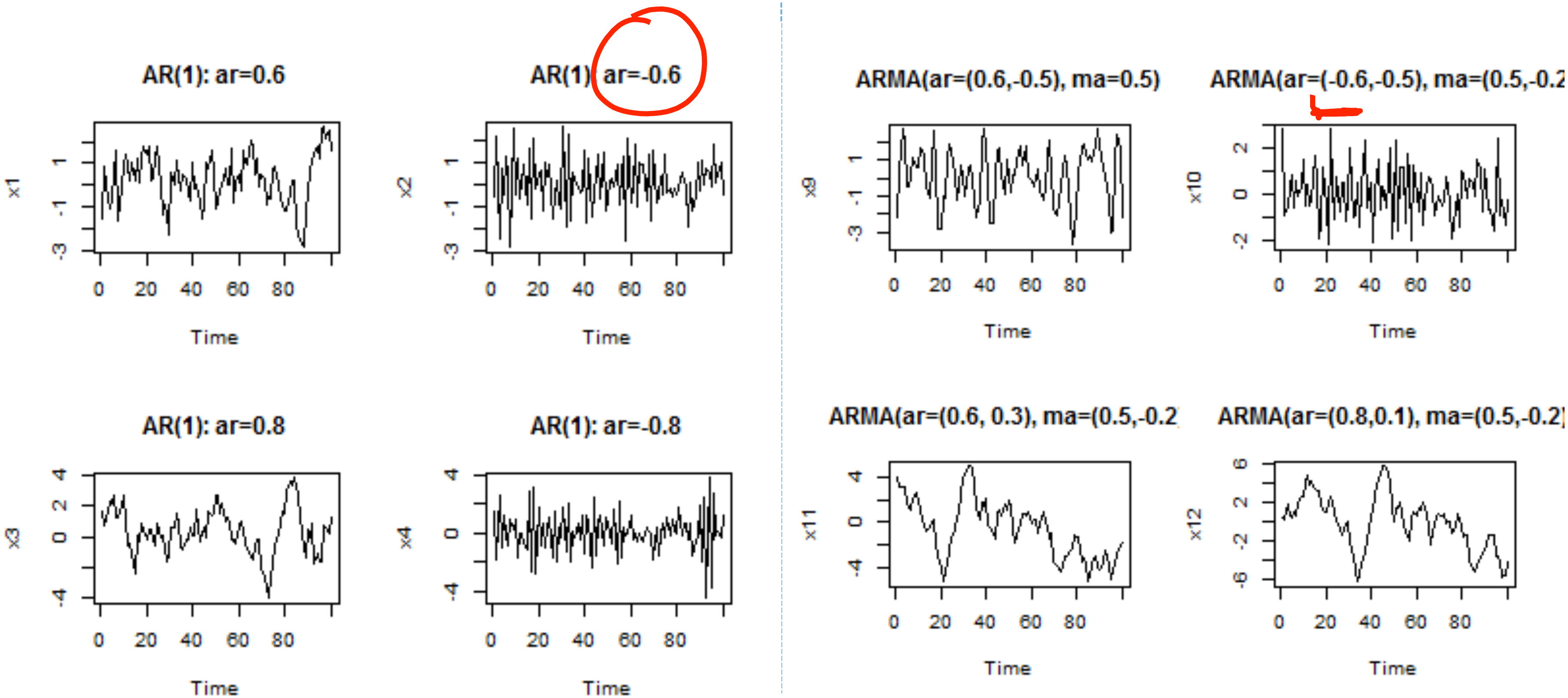
- If the AR parts of the AR(1) and ARMA(1,1) are identical, it is very difficult to distinguish between the two based only on the t-plots.



Time-Series Plots of AR(1) and ARMA(1,1) Models

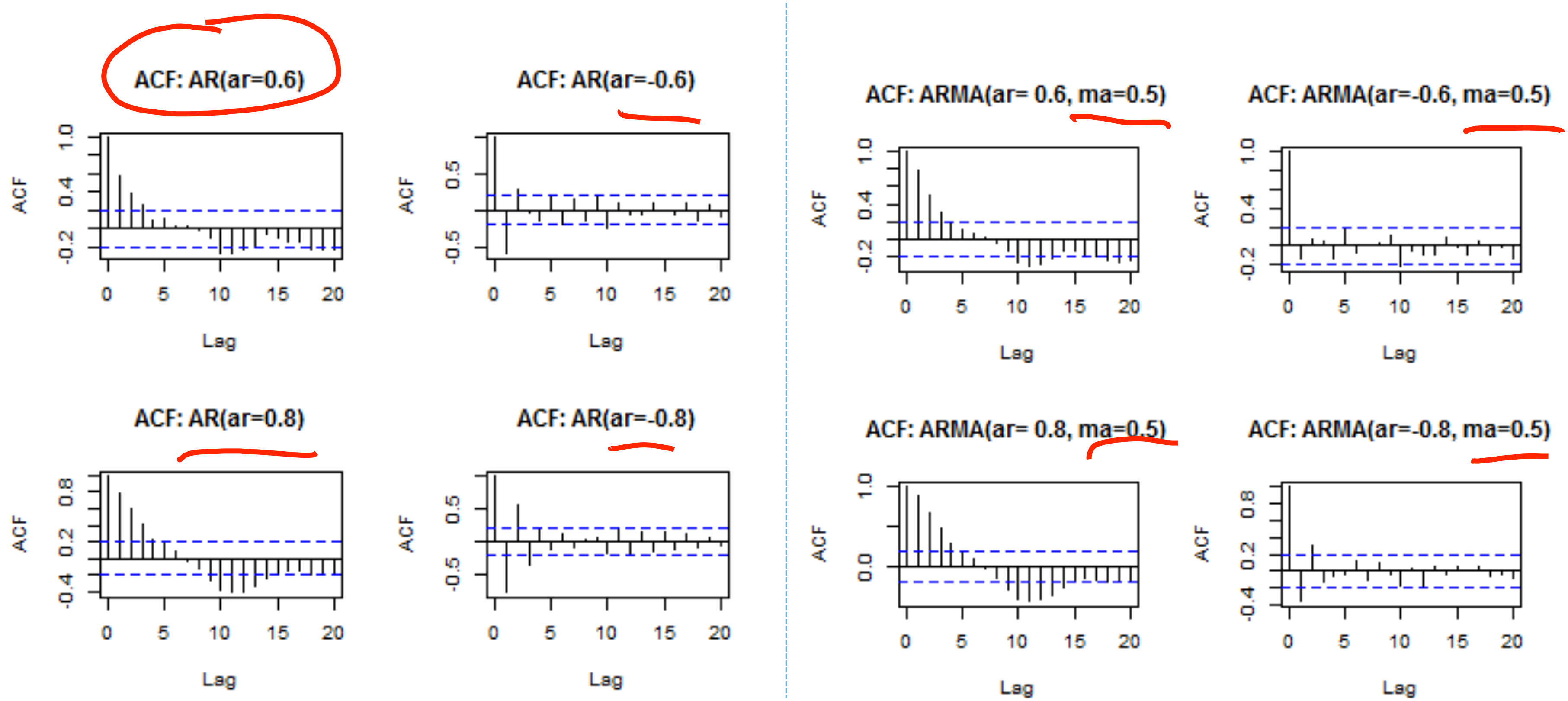
- Adding an AR and/or MA term does not help to distinguish the two types of models.

- Adding another AR term with negative coefficient makes the distinction even harder.



ACF of AR(1) and ARMA(1,1) Models

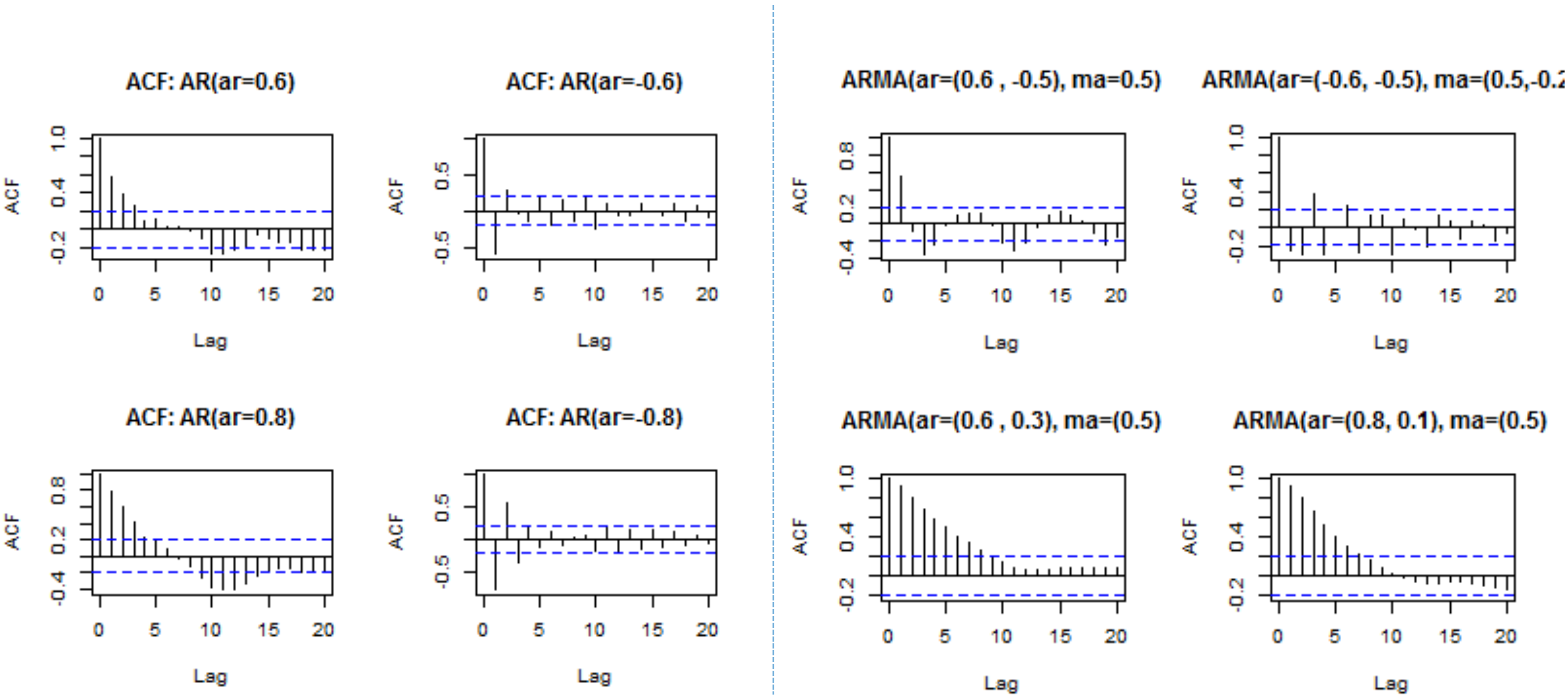
- As seen from the theoretical ACF, using a correlogram alone cannot distinguish AR and ARMA models.
- This can be seen from the empirical ACFs of the simulated AR and ARMA models.



ACF of AR(1) and ARMA(1,1) Models

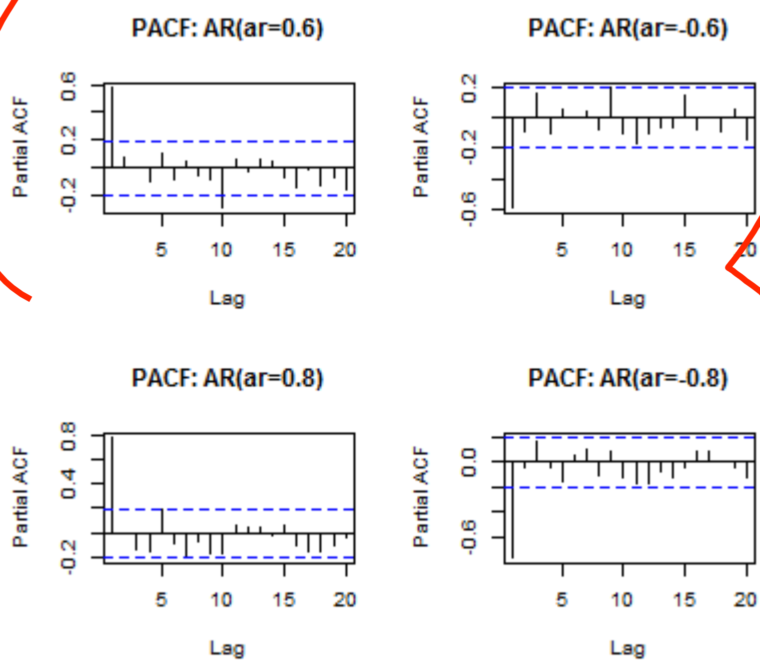
- The general patterns still look similar…

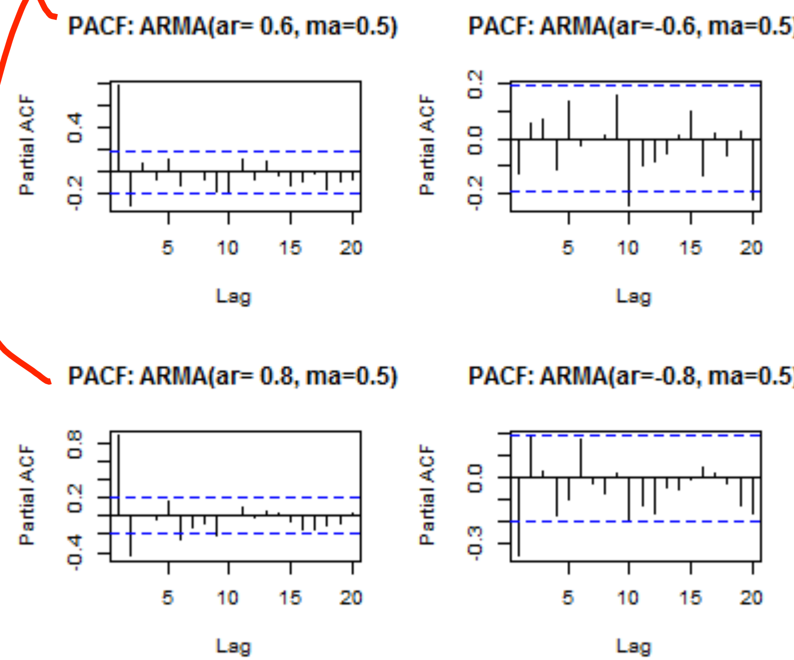
- The conclusion is that if we are analyzing the series and plot and the ACF and the ACF look like one of those below, it is not possible to tell which AR or ARMA (or even MA) models to use.



PACF of AR(1) and ARMA(1,1) Models

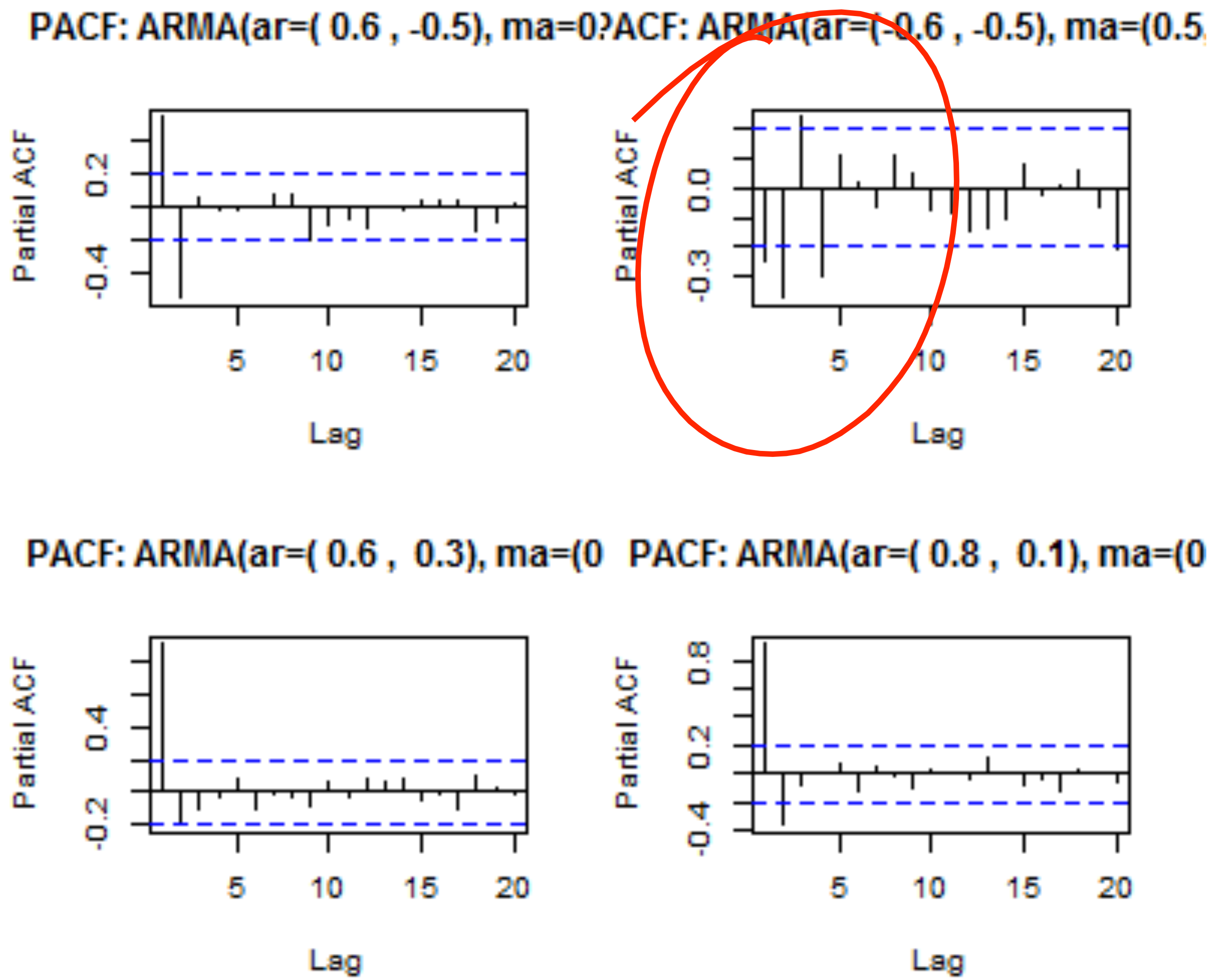
- The PACF of AR(p) models sharply drop off after p lags while the PACF of ARMA(p,q) models generally gradually decline to zero.
- Based on the graphs below, this feature is not apparent.
- Yet, I still want to show them to you because these graphs come from my simulation, and I don't want to cherry-pick graphs just to show you this feature.





PACF of AR(1) and ARMA(1,1) Models

- Similar conclusion, although the PACF do linger a little longer…



Comparing ARMA Models and AR Models Using Simulated Series Part 2: Model Identification

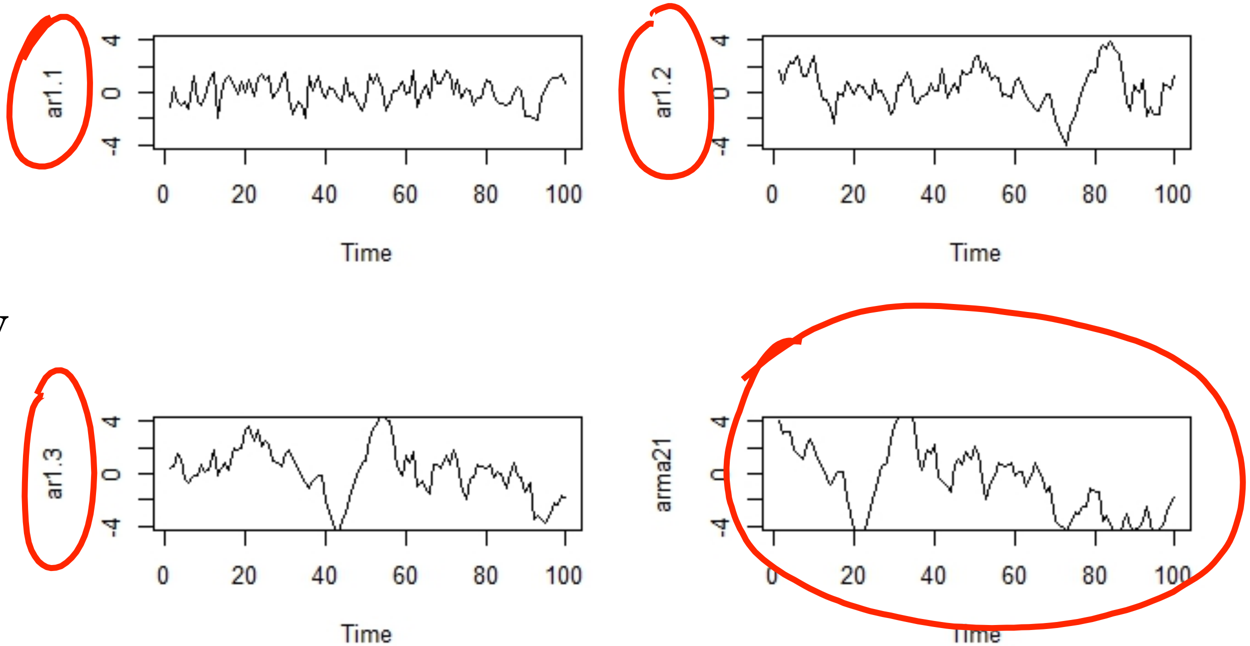
AR vs. ARMA Models: One More Example

Let's try one more example with three AR(1) models and one ARMA(2,1) model.

As the AR parameter gets closer to 1, the more persistent the series becomes.

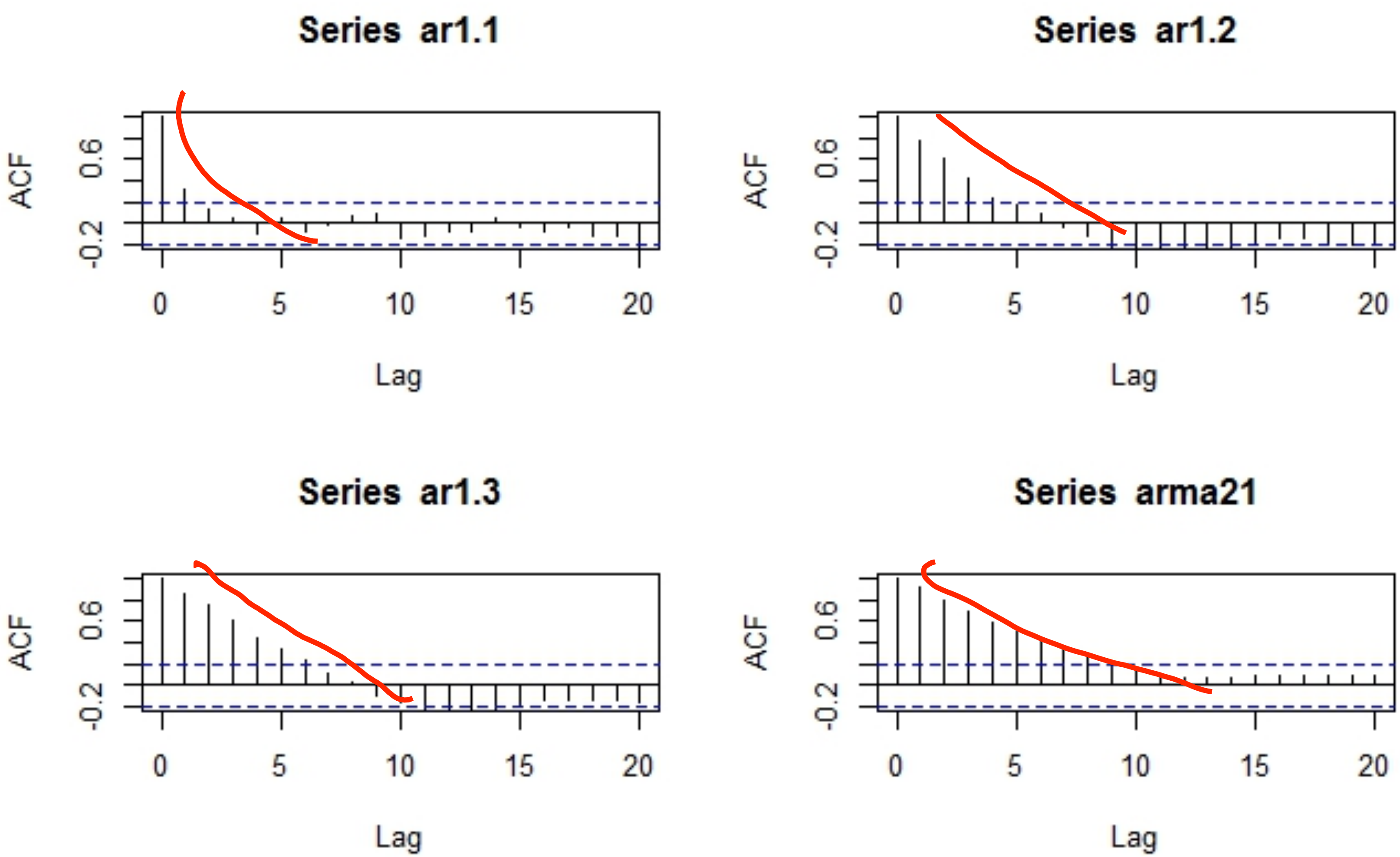
Both AR1(ar=0.9) and ARMA(ar=0.6,ma=0.3) models can produce very persistent series.

```
set.seed(898)
ar1.1 <- arima.sim(n = 100, list(ar = c(0.4)))
set.seed(898)
ar1.2 <- arima.sim(n = 100, list(ar = c(0.8)))
set.seed(898)
ar1.3 <- arima.sim(n = 100, list(ar = c(0.9)))
set.seed(898)
arma21 <- arima.sim(n = 100, list(ar = c(0.6, 0.3), ma = c(0.5)))
```



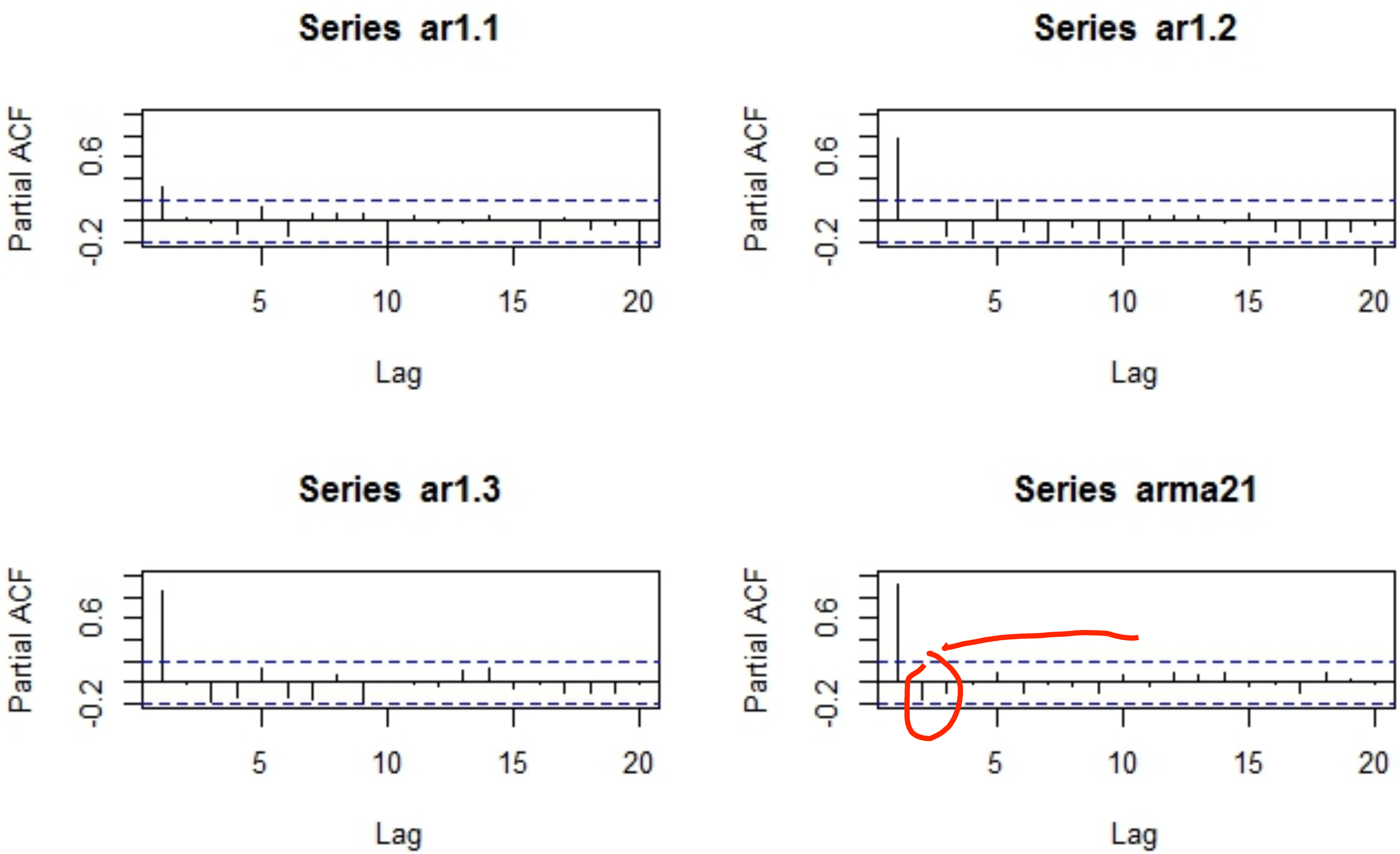
AR vs. ARMA Models: ACF

The ARMA model can produce a longer ACF series than AR models do.



AR vs. ARMA Models: PACF

The PACFs between the AR1.3 and the ARMA21 model are still not indistinguishable.



ARMA-Type Model Identification

In general, the ACF and PACF of AR, MA, and ARMA models have the following characteristics:

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag q p	Tails Off	Tails off

- As we have seen in the simulated examples, these features can be used only to narrow down the possibilities of processes underlying a realization we observe.

- It is typical that in practice we estimate a series of ARMA models of different orders and use various statistics, such as AIC and BICs, and perhaps even forecast performance to choose a model.

ARMA Models and MA Models

Modeling Using the British Pound–New Zealand Dollar Exchange Rate: Part 1

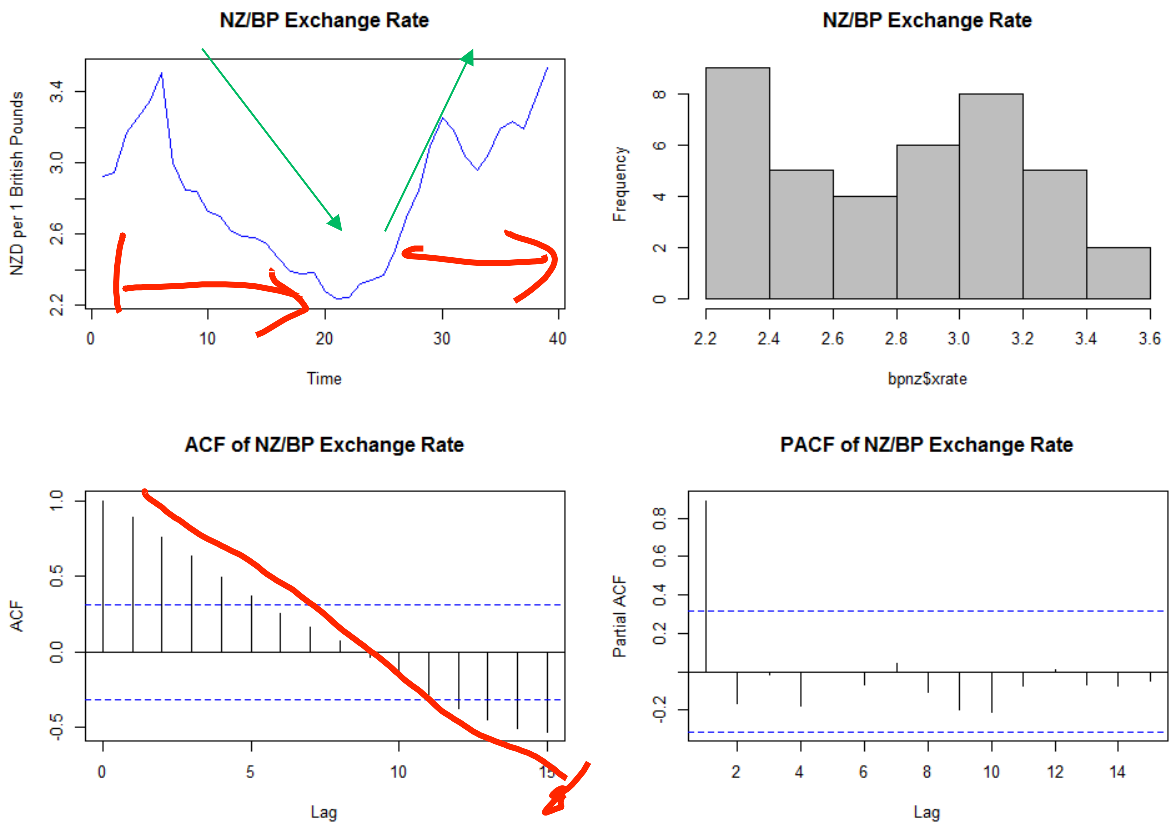
The Data: Basic Structure

- This is the British pound–New Zealand dollar exchange rate series provided by the book. The series can be obtained from the authors' website.
- This is an annual series with 39 observations; the data contains only the series itself (i.e., one variable).
- The summary statistics are displayed below.

```
'data.frame':  39 obs. of  1 variable:
 $ xrate: num  2.92 2.94 3.17 3.25 3.35 ...
> summary(bpnz)
     xrate
 Min.   :2.2
 1st Qu.:2.5
 Median :2.8
 Mean   :2.8
 3rd Qu.:3.2
 Max.   :3.5
```

Descriptive Statistics and Data Visualization

- Similar to the USD–NZD monthly exchange rate series, this series is very persistent, and it cannot be seen using only the histogram and descriptive statistics.
- It does not appear to be captured well by a MA model, although an ARMA model may provide a better model.



Estimation: MA(5) Model

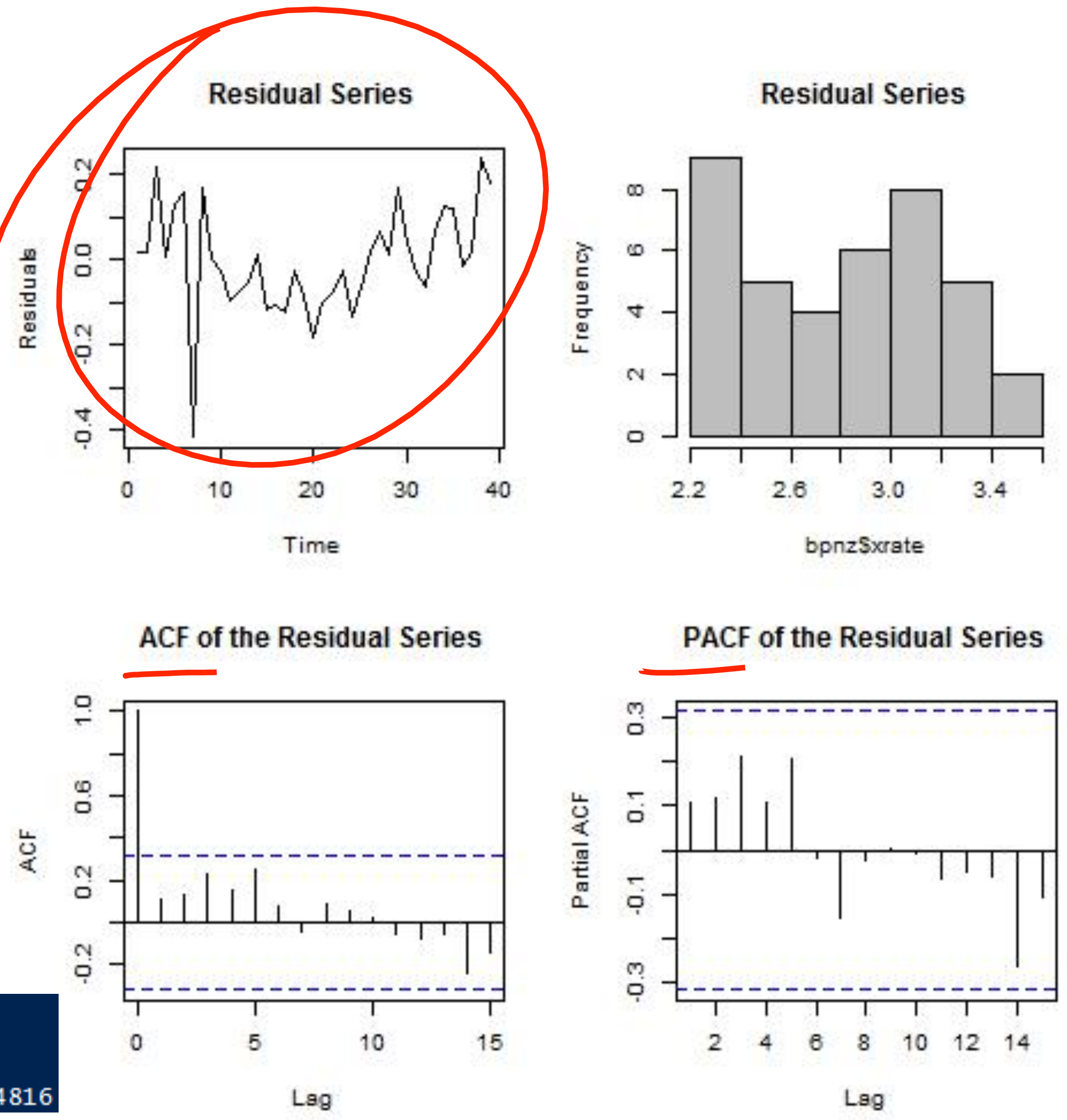
- We will estimate a MA model for comparison.
- The first four MA parameters and the intercept are all significant.
- Note that the last MA parameter is not significant.
- At this point, it is hard to judge how good the estimation is. We will have to examine the residuals, visualizing the in-sample fit and out-of-sample forecast.

```
Series: bpnz
ARIMA(0,0,5) with non-zero mean

Coefficients:
       ma1   ma2   ma3   ma4   ma5  intercept
       1.5   1.3  1.17  0.64  0.27       2.87
s.e.   0.2   0.3  0.28  0.26  0.17       0.11

sigma^2 estimated as 0.015:  log likelihood=24
AIC=-34   AICc=-31   BIC=-23
```

Model Diagnostics

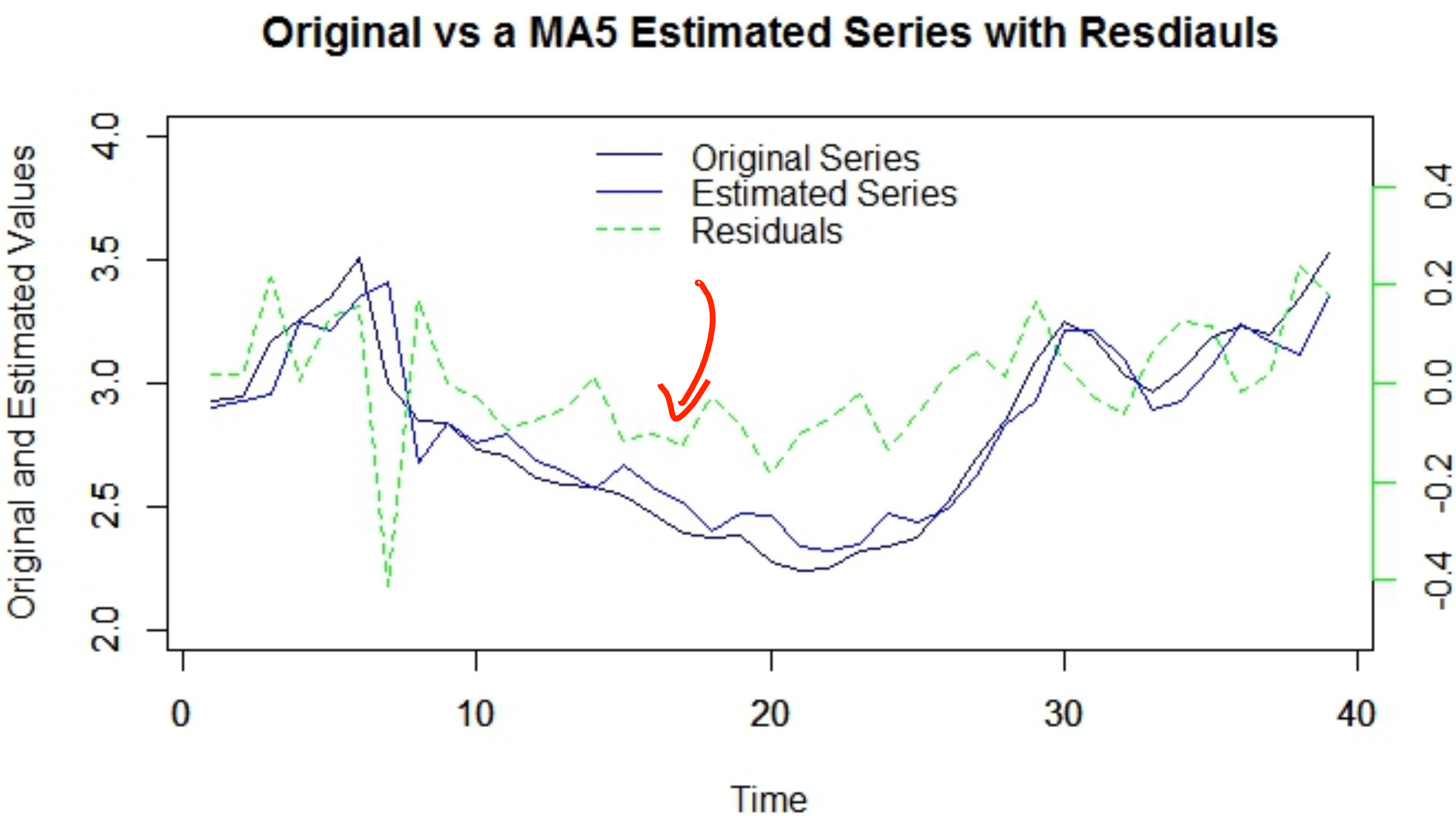
- The time series plot of the residuals does not suggest a white noise sample path.
- However, both the ACF and PACF do not show any statistical significant autocorrelations.
- The null hypothesis that the series is not correlated cannot be rejected based on the Ljung-Box statistic.

```
 Box-Ljung test

data:  ma5.bpnzfit$resid
X-squared = 0.5, df = 1, p-value = 0.4816
```



Model Performance Evaluation: In-Sample Fit

- The following time series plot displays the original series, which has 39 observations, the estimated values (in blue), overlaid with the residual series (in green and right y-axis).
- In-sample fit “looks reasonable.”

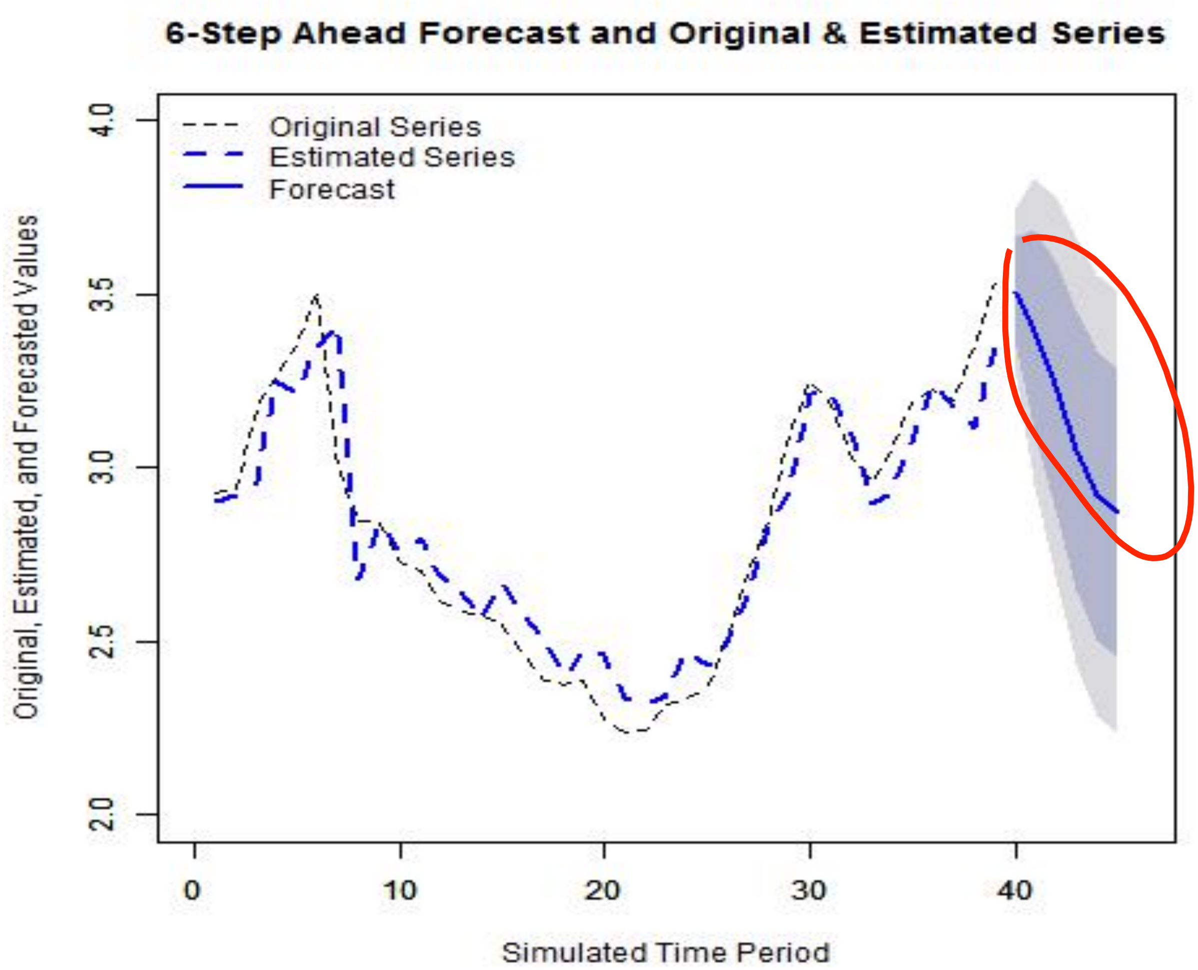


Forecasting

The following plot displays the original series, the estimated values, and a six-step-ahead forecast as well as forecast intervals.

Forecasted Values

	value
1	3.5
2	3.4
3	3.2
4	3.1
5	2.9
6	2.9



Back-Testing and Out-of-Sample Forecasting

- An alternative way to evaluate the model (or the class of model under considerations) is to use out-of-sample forecasting (or back-testing) by leaving out a subsample in the estimation.

- Leaving out a subsample from a time series has to be done very carefully, because we cannot just "randomly" drop observations from the series, as it will break the dependency structure embedded in the series.

- In this example, I excluded the last six observation (i.e., only 33 observations left) from the sample and re-estimate the model using a MA(5).

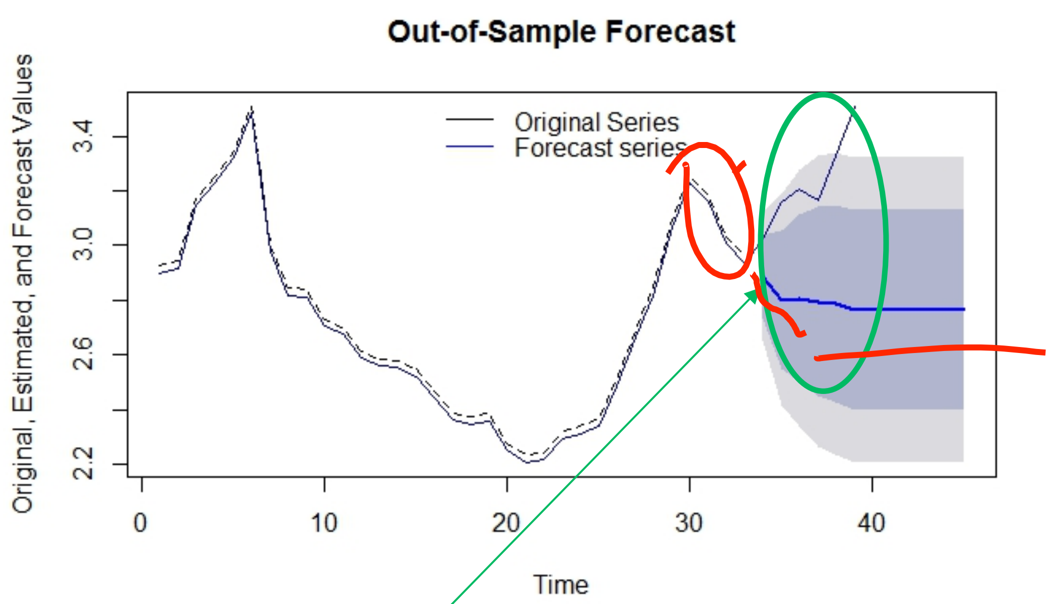
- Then, I produced a 12-step-ahead forecast. Because the last six observations are actually observed, we can compare the forecasts with the actual values. This practice is call back-testing.

Back-Testing and Out-of-Sample Forecasting

- This shows a very dramatic difference from the forecast produced using the entire series.
- The forecast continues with the downward trend observed from the series, although the confidence intervals, with the first few forecasts, do include the observed values.
- The forecast after the first five periods stay flat.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
34	2.9	2.7	3.0	2.7	3.1
35	2.8	2.5	3.1	2.4	3.2
36	2.8	2.5	3.1	2.3	3.3
37	2.8	2.4	3.1	2.3	3.3
38	2.8	2.4	3.1	2.2	3.3
39	2.8	2.4	3.1	2.2	3.3
40	2.8	2.4	3.1	2.2	3.3
41	2.8	2.4	3.1	2.2	3.3
42	2.8	2.4	3.1	2.2	3.3
43	2.8	2.4	3.1	2.2	3.3
44	2.8	2.4	3.1	2.2	3.3
45	2.8	2.4	3.1	2.2	3.3

Note that the slight difference between the forecast values in the graph and in the list is due to rounding.



Divergence between the actual and forecasted values, though the 95% confidence intervals, do include most of the actual values.

ARMA Models and MA Models

Modeling Using the British Pound–New Zealand Dollar Exchange Rate, Part 2

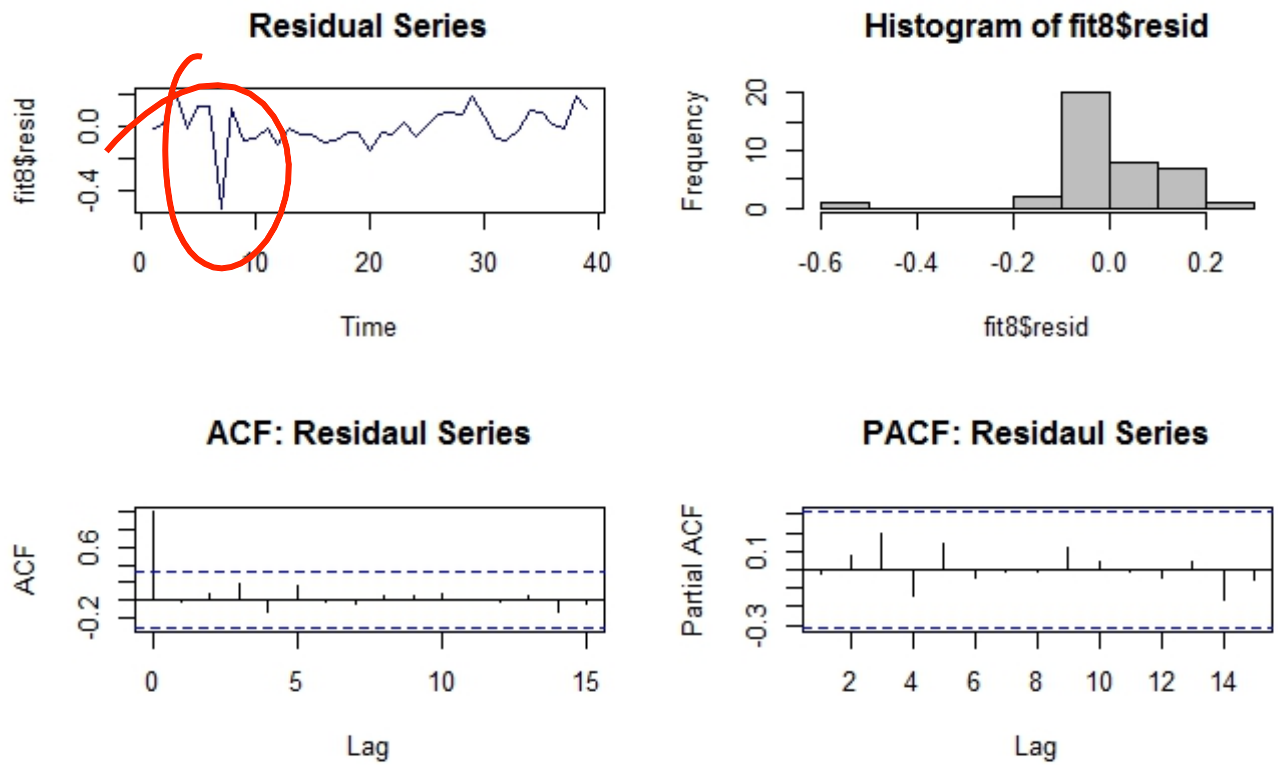
Estimation: ARMA(1,1) Model

- Based on the time series, ACF, and PACF plots, a low-order ARMA model may do a better job than does a high-order pure MA model.
- Let's estimate a ARMA(1,1) model.
- All the AR and MA parameters are statistically significant.
- Note that the estimated AR parameter is close 1.

```
> fit8 <- Arima(bpnz$xrate, order=c(1,0,1))
> summary(fit8)
Series: bpnz$xrate
ARIMA(1,0,1) with non-zero mean

Coefficients:
        ar1   ma1  intercept
      0.892  0.53       2.96
s.e.  0.076  0.20       0.24

sigma^2 estimated as 0.0151:  log likelihood=25
AIC=-42   AICc=-41   BIC=-36

Training set error measures:
                  ME RMSE   MAE   MPE MAPE MASE   ACF1
Training set 0.00029 0.12 0.085 -0.21    3 0.85 -0.018
```

Model Diagnostics

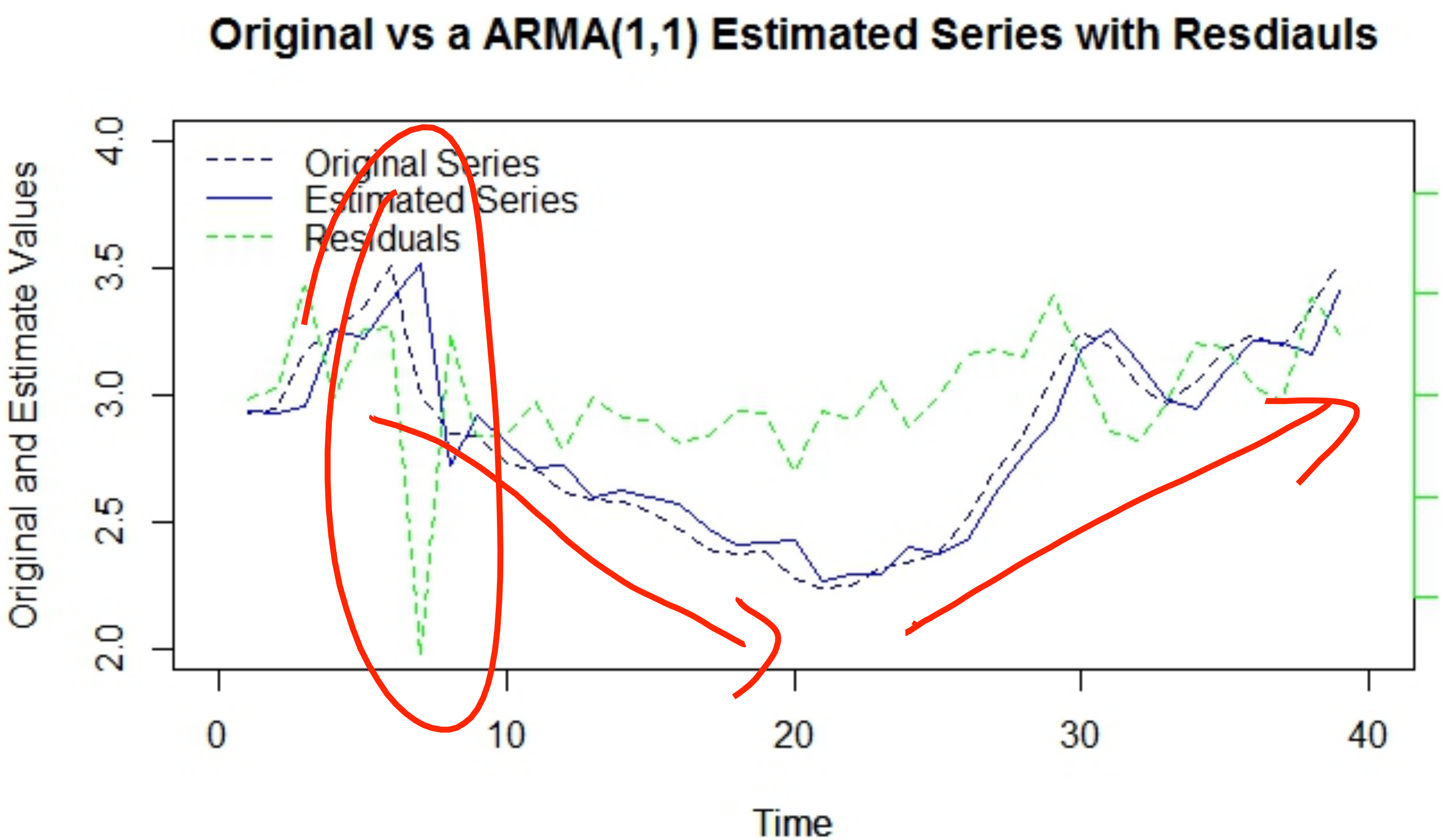
- The ACF, PACF, and Lyung-Box statistics cannot reject the hypothesis that the residual series comes from a white noise process.




```
> Box.test(fit8$resid, type="Ljung-Box")

 Box-Ljung test

data:  fit8$resid
X-squared = 0.014, df = 1, p-value = 0.9063
```

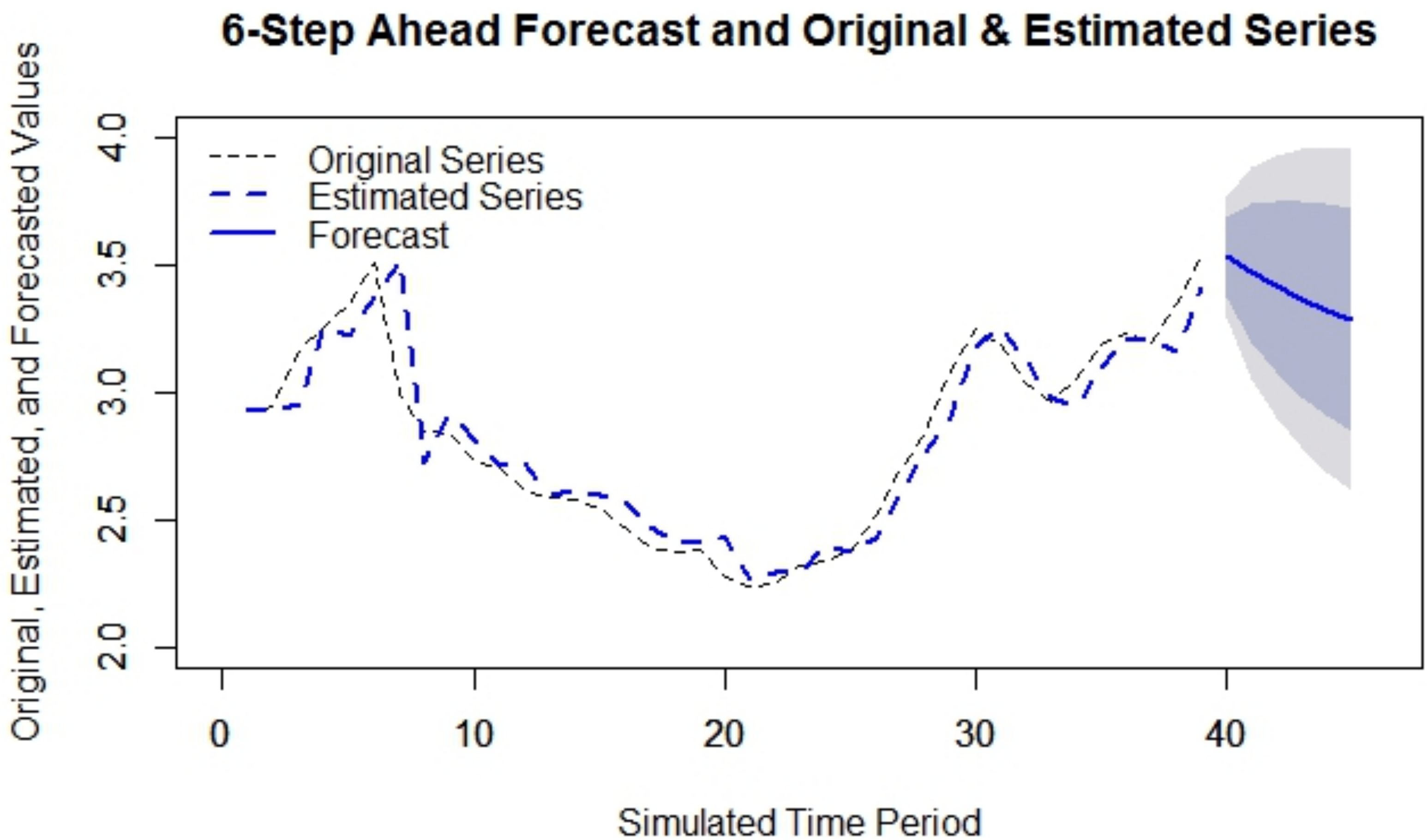
Model Performance Evaluation: In-Sample Fit

- Like that of the MA(5) model, the in-sample fit from the ARMA(1,1) model looks reasonable.



Forecasting

- Notice that the forecast still trends downward, although it does not declines as rapidly as that of the MA(5) model.



Back-Testing and Out-of-Sample Forecasting

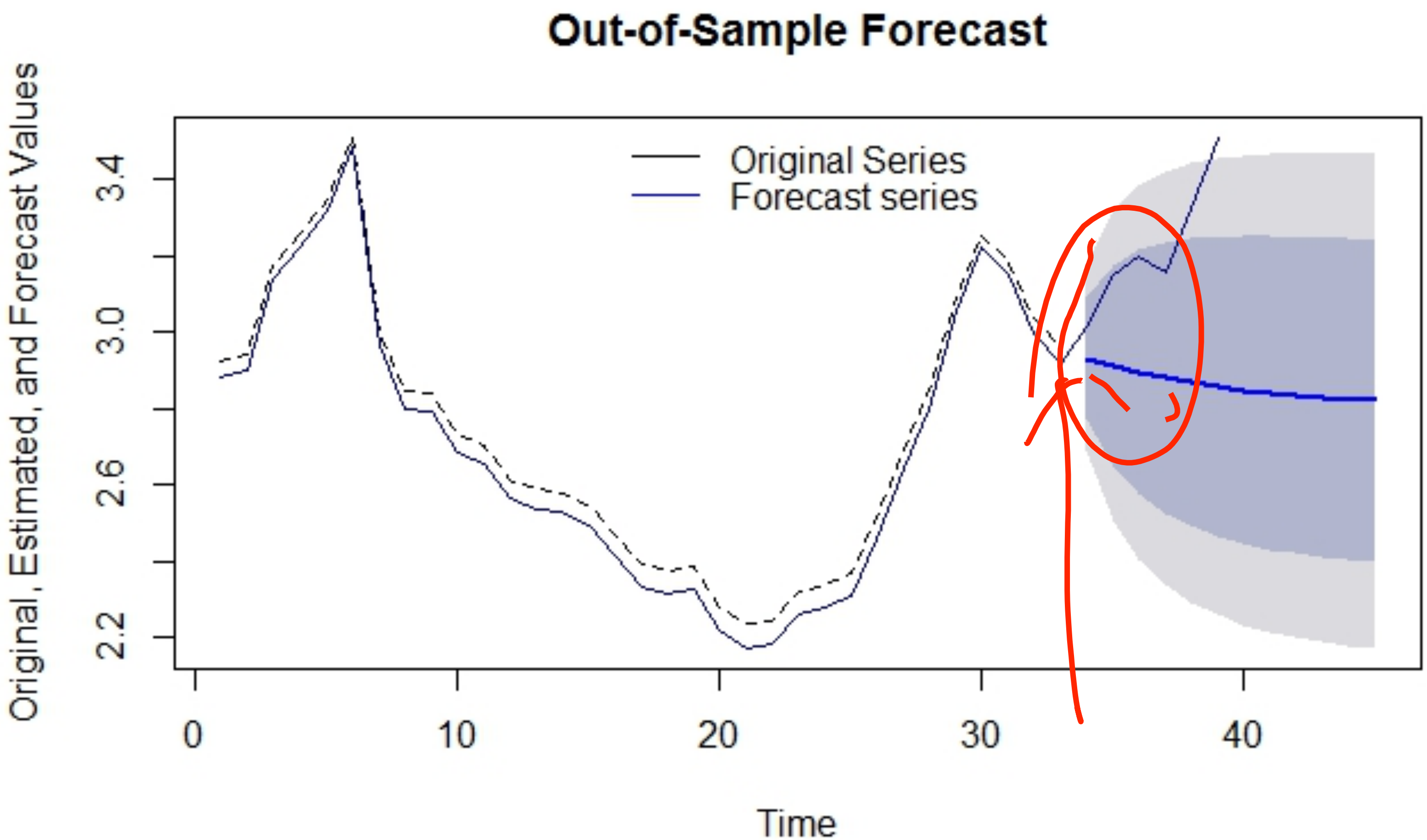
- Reestimate the ARMA(1,1) model using only the first 33 (of the 39) observations in the original series.

- All of the parameters continue to be significant.

```
Series: bpnz$xrate[1:(length(bpnz$xrate) - 6)]
ARIMA(1,0,1) with non-zero mean

Coefficients:
        ar1    ma1   intercept
      0.846   0.48        2.80
s.e.  0.091   0.23        0.18

sigma^2 estimated as 0.0155:  log likelihood=21
AIC=-34   AICc=-32   BIC=-28

Training set error measures:
                    ME RMSE   MAE    MPE MAPE MASE   ACF1
Training set -0.0043 0.12 0.087 -0.37  3.1 0.87 0.0048
```

Back-Testing and Out-of-Sample Forecasting

- The forecasts still deviate from the actual values.
- However, 95% confidence intervals still include the actual values, meaning that the difference is not statistically significant.



Review the Steps to Build ARIMA Time Series Models

Recall the General Steps to Analyze a Time Series

1. Based on the interaction of theory, subject matter expertise, and practice, consider a useful class of models.
2. Collect and cleanse the data.
3. Conduct exploratory time series data analysis (ETSDA) by plotting the series, and examine the main patterns and atypical observations of the graph, after collecting and "cleaning" the data :
   - Trend
   - The fluctuation around a trend
     - Seasonal variation (or seasonality)
     - Cyclical variation (that does not appear to be seasonal variation)
   - Sharp change in behavior (i.e., structural change or jumps)
   - Outliers
4. **Examine and (statistically) test whether the series is stationary (when applying a stationary model).**

General Steps to Analyze a Time Series (2)

5. **If the series is not stationary, transform the series to a stationary series, because the time series models covered in this course apply only to stationary or integrated times series. Common transformation techniques include trend removal (i.e., detrending), seasonality removal, logarithmic, and difference transformation.**
6. Model the transformed series with a stationary or integrated time series model.
7. Examine the validity of the model's underlying assumptions.
   - This is an important step, because if the model's underlying assumptions are not satisfied, one should not proceed to conducting statistical inference and forecasting.
8. Among the valid models, choose the one that perform "best" according to some pre-specified metrics.
9. Once a (statistically) valid model is chosen, conduct statistical inference and/or forecasting, if the underlying statistical assumptions are all satisfied.

Modeling Procedure of ARIMA Models

When estimating an ARMA model using a time series, we will use the following general steps:

1. Plot the data, and identify the key dynamics and any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. If the data are nonstationary: Take first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR(p) or MA(q) model appropriate?
5. Try your chosen model(s), and use the AICc to search for a better model.
6. Conduct diagnostic testing and formal model assumption testing:
    - Check the residuals on the chosen model by plotting the ACF of the residuals, and doing a portmanteau test of the residuals. If they do not look like white noise, the model need to be respecified, reestimated, and retested.
7. Once the hypothesis that the residuals follow a white noise process cannot be rejected, the model can be used for forecasting.

Note: One could attempt to write an algorithm to automate steps 3–5.

Nonstationary Models: An Introduction

Nonstationary Models

- In the previous lectures, we focused on stationary time series models and studied primarily series that are covariance stationary. In this lecture, we shift the focus to **nonstationary time series models**, as many time series encountered in practice are nonstationary due to the existence of trends or seasonal effects.

- Fortunately, many of the nonstationary series in the real world can be converted to a covariance stationary series using simple differencing, especially first-order differencing. This modeling strategy leads to the celebrated **Box-Jenkins Approach** to derive **A**uto**R**egressive **I**ntegrated **M**oving **A**verage, or ARIMA, models.

Nonstationary Model

- The term "integrated" comes from the fact that the differenced series need to be aggregated to recover the original series, the underlying process is called an "integrated" series.

- The ARIMA process can accommodate seasonal terms, giving rise to seasonal ARIMA (SARIMA) model, which will also be discussed in this lecture.

- Not all nonstationarity can be dealt with using differencing. For instance, volatility clustering, which occurs in many financial and macroeconomic time series, leading to **conditionally heteroskedaticity** is more appropriately modeled using an Autoregression Conditional Heteroskedastic (ARCH) or a Generalize ARCH (GARCH) model.

Nonstationary Model

- Without seasonal effects, first differencing can remove both (stochastic and deterministic) trends. An example of stochastic trends includes random walks, and an example of deterministic trend includes a linear trend:

Recall that a Random Walk takes the following form:

$$y_t = y_{t-1} + \omega_t$$

Taking a first difference transforms the model to the following form:

$$\nabla y_t = \underbrace{y_t - y_{t-1}} \cancel{= \omega_t} = \omega_t$$

which is a mean-zero, stationary white-noise series.

Nonstationary Model

On the other hand, a linear trend with white noise errors

$$y_t = a + bt + w_t$$

can be transformed into a stationary moving average (MA(1)) process with first differencing:

$$\nabla y_t = y_t - y_{t-1} = b + (w_t + w_{t-1})$$

Another transformation we can take is to subtract the trend from the series, which gives a white noise process, which may be a more sensible approach:

$$y_t - (a + bt) = w_t$$

In practice, do not just blindly take first differencing or first differencing in log or even higher-level differencing. Always first investigate the series using various graphical techniques and statistical tests before estimating a model on some transformation of the original series.

Also, when making transformations, always ask what the meaning/ interpretation is of the transformed series.

Random Walk Process

Random Walk Process: Introduction

Random walk is nothing more than an AR(1) process with the AR parameter being 1:

$$y_t = \phi_1 y_{t-1} + \epsilon_t$$
$$\epsilon_t \sim WN(0, \sigma^2)$$

- As mentioned a few lectures ago, random walk is a very important process. It forms the foundation of many important processes in continuous-time finance.

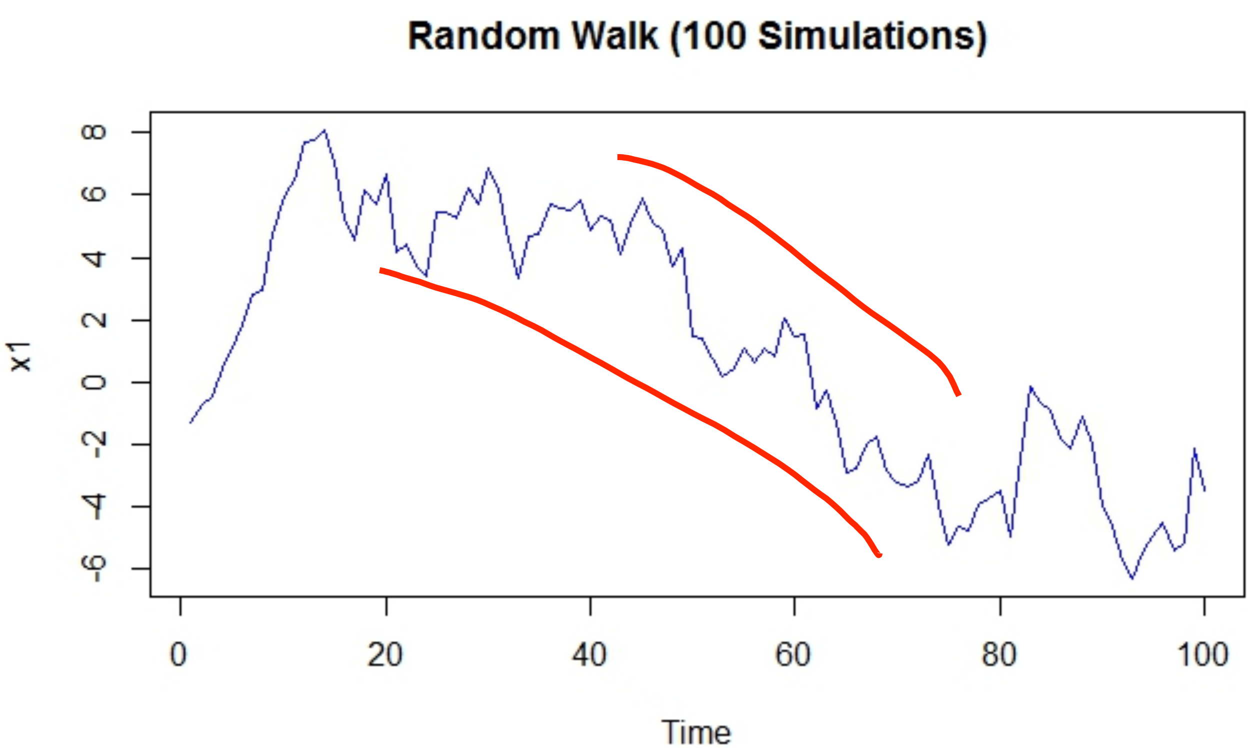
- Note that a random walk does not revert back to any constant level. In particular, it is not a mean-reversion process.

- It wanders up and down with no tendency to settle at any particular level.

- Although random walk is ill-behaved (in that it is not covariance stationary), its first difference becomes a stationary white noise.

A Simulated Random Walk Series

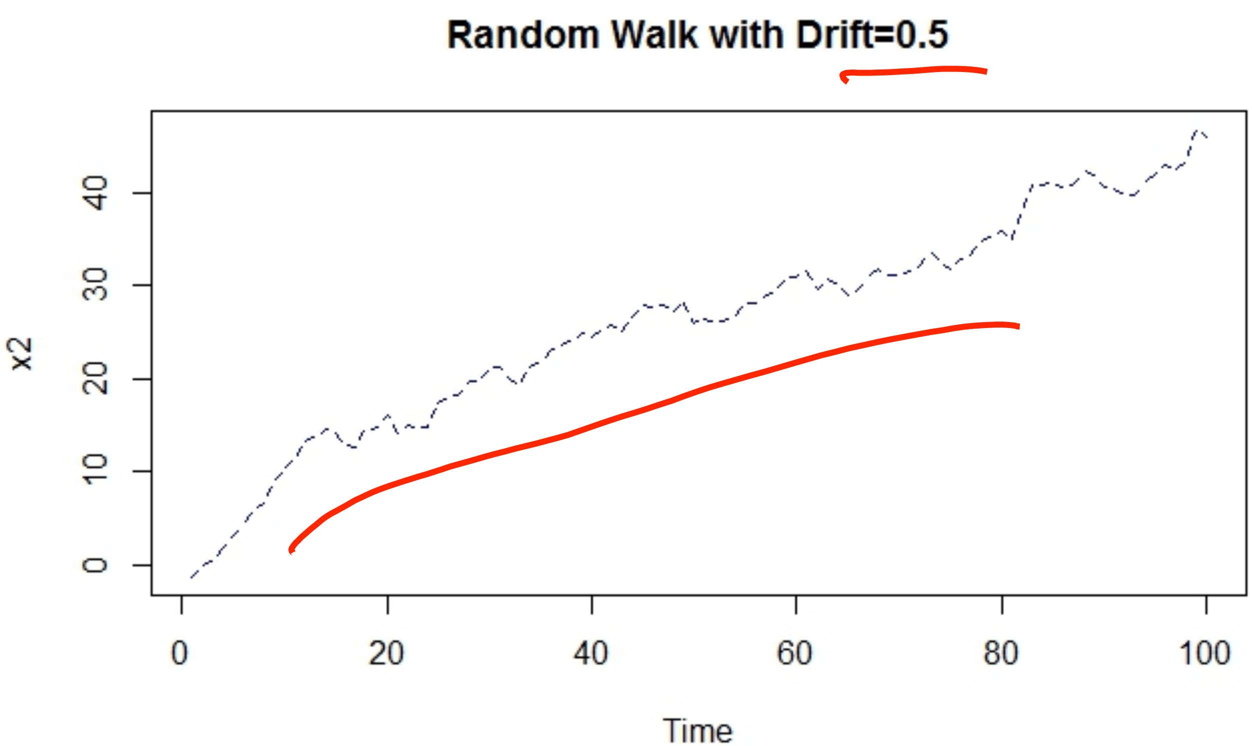
$$y_t = \phi_1 y_{t-1} + \epsilon_t$$
$$\epsilon_t \sim WN(0, \sigma^2)$$



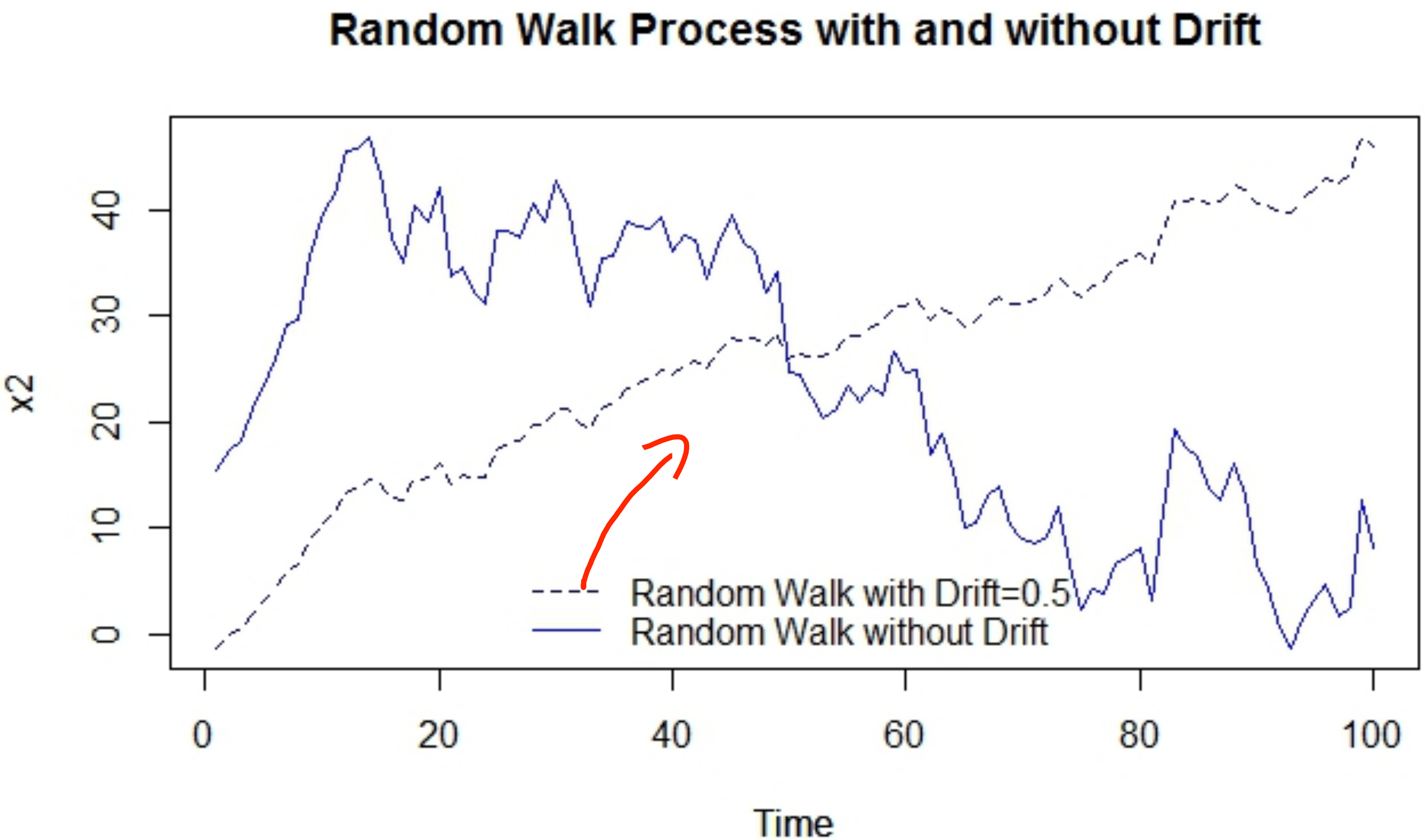
Random Walk With Drift Process

- Random walk with drift is essentially a model of trend.
- On average, the process "grows" by the drift in each period.

$$y_t = \delta + y_{t-1} + \epsilon_t$$
$$\epsilon_t \sim WN(0, \sigma^2)$$



Random Walk With Drift Process



Random Walk Process

- In fact, the drift parameter plays the same role as the slope parameter in deterministic linear trend models.

- On average, the process "grows" by the drift in each period.

- The random walk with drift is also called a model of **stochastic trend**, as its trend is driven by stochastic shocks.

- The most distinctive feature of random walk is that shocks affect the series permanently: If a shock lowers the value of the series, the random walk has no tendency to rise again, and it would stay lower permanently (until a new shock comes).

- That is, a unit shock moves the expected future path of the series by one unit!

Random Walk, Integrated Process, an Introduction to ARIMA Process

The Variance of Random Walk Process

- Assuming the process started at some time 0 with value y0, we can write the random walk process as

$$y_t = y_0 + \sum_{i=1}^{t} \epsilon_i$$
$$E(y_t) = y_0$$
$$var(y_t) = t\sigma^2$$

- Note that the variance grows without bounds over time.
- Likewise, for random walk with drift, we can express the process as

$$y_t = t\delta + y_0 + \sum_{i=1}^{t} \epsilon_i$$
$$E(y_t) = y_0 + t\delta$$
$$var(y_t) = t\sigma^2$$

- The mean grows by the speed of drift term, and the variance grows without bounds over time.

Integrated Process: An Introduction

- However, a first differencing can transform the nonstationary random walk process to a stationary white noise process.

- White noise is the simplest I(0) process, and the random walk is the simplest I(1) process, where I(1) means the process is a differenced one. This is called integrated process of order 1.

- In practice, I(0) and I(1) cases find themselves having the most applications. One reason is that the results are hard to explain once a series is differenced too many times.

Random Walk Process

A time series y_t follows an ARIMA(p, d, q) process if the d^{th} differences of the y_t series is an ARMA(p, q) process. Mathematically, using lag operator, it can expressed as

$$\phi_p(B)(1 - B)^d y_t = \theta_q(B)\omega_t$$

where ϕ_p and θ_q are polynomials of orders p and q discussed in the previous lectures.

Writing an **ARIMA(p,d,q)** may seem too abstract, and whenever a model is presented this way, you may get a feel of the model by making simple cases, such as a low-order **ARIMA(p,d,q)** model.

Next, two such examples are shown, but you should create more examples of your own. Once an example is created, use R (or Python) to simulate some realizations of the model.

ARIMA Model: Simulation

ARIMA: Algebraic Manipulation Before Simulation

- Following the approach in the last few lectures, we simulate an ARIMA model and examine its patterns exhibited in a time series plot, ACF, and PACF.
- The simulated model acts as a "true" model, so we can estimate the model and examine how close the parameter estimates are to the "true" model parameters.
- Consider a model taken from an example in (CM2009) section 7.2.4 on page 140:

$$y_t = 0.5y_{t-1} + y_{t-1} - 0.5y_{t-2} + \omega_t + 0.3\omega_{t-1}$$

$$y_t - y_{t-1} = 0.5(y_{t-1} - y_{t-2}) + \omega_t + 0.3\omega_{t-1}$$

$$(y_t - y_{t-1}) - 0.5(y_{t-1} - y_{t-2}) = \omega_t + 0.3\omega_{t-1}$$

$$\nabla y_t - 0.5 \nabla y_{t-1} = \omega_t + 0.3\omega_{t-1}$$

ARIMA: Algebraic Manipulation Before Simulation

- The equation can be rearranged and factored as an ARIMA(1,1,1) model:

$$(1 - 0.5B) \nabla y_t = (1 + 0.3B)\omega_t$$

$$\left(\nabla y_t = 0.5 \underline{\nabla y_{t-1}} + \omega_t + \underline{0.3\omega_{t-1}} \right)$$

- Note that after the first difference, the model becomes a stationary ARMA(1,1) model.

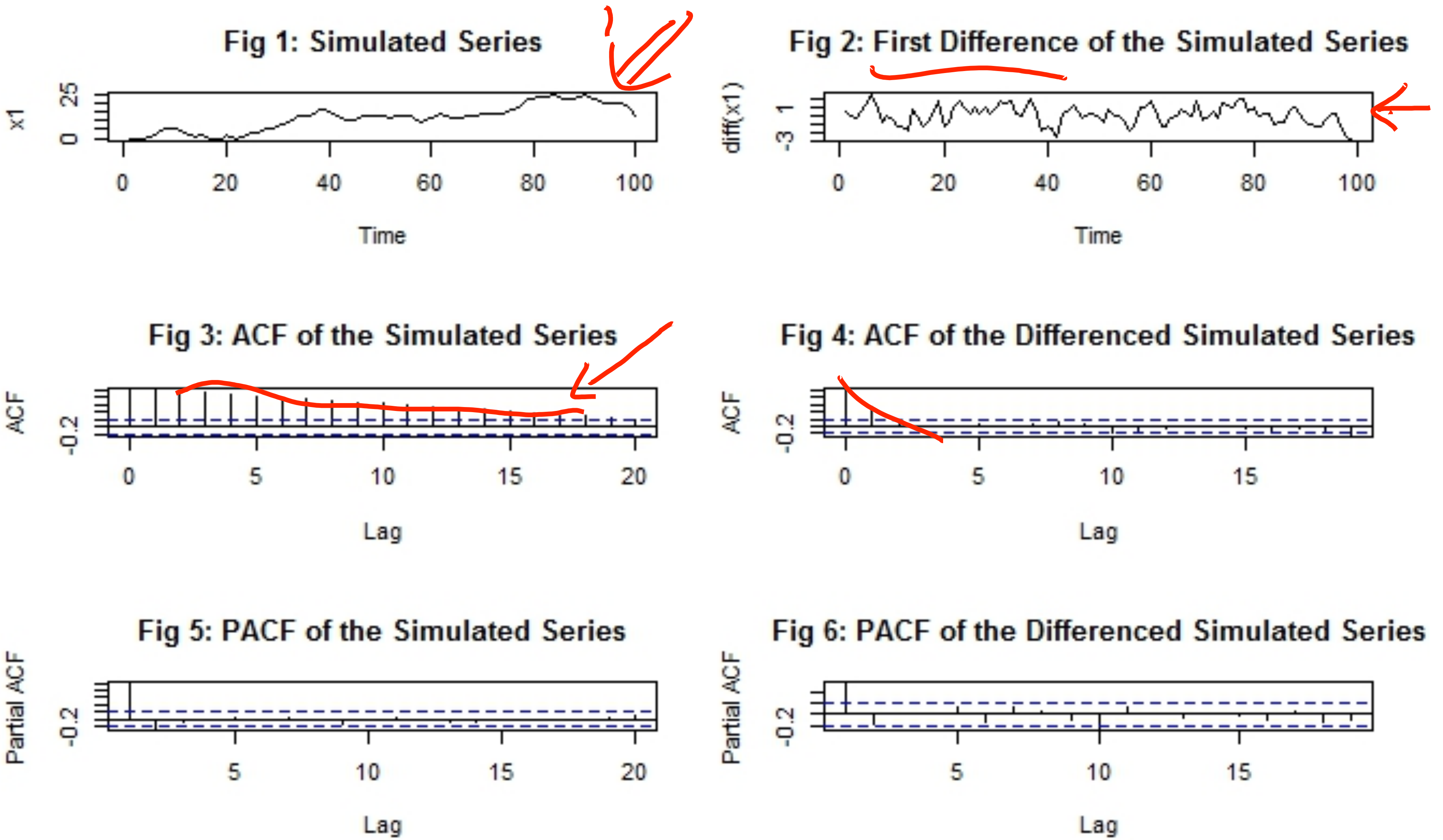
Simulated Data

```
set.seed(898)
x1 <- w <- rnorm(100)
for (i in 3:100) x1[i] <- 0.5*x1[i-1] + x1[i-1] - 0.5*x1[i-2] + w[i] + 0.3*w[i-1]
```

Basic structure and descriptive statistics of the data

```
> str(x1)
 num [1:100] -0.579 -0.0823 0.0254 -0.1728 0.254 ...
> summary(x1)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   -0.9     5.0    12.5    11.5    15.9    24.3
```

Data Visualization



ARIMA Model: Modeling With the Simulated Data

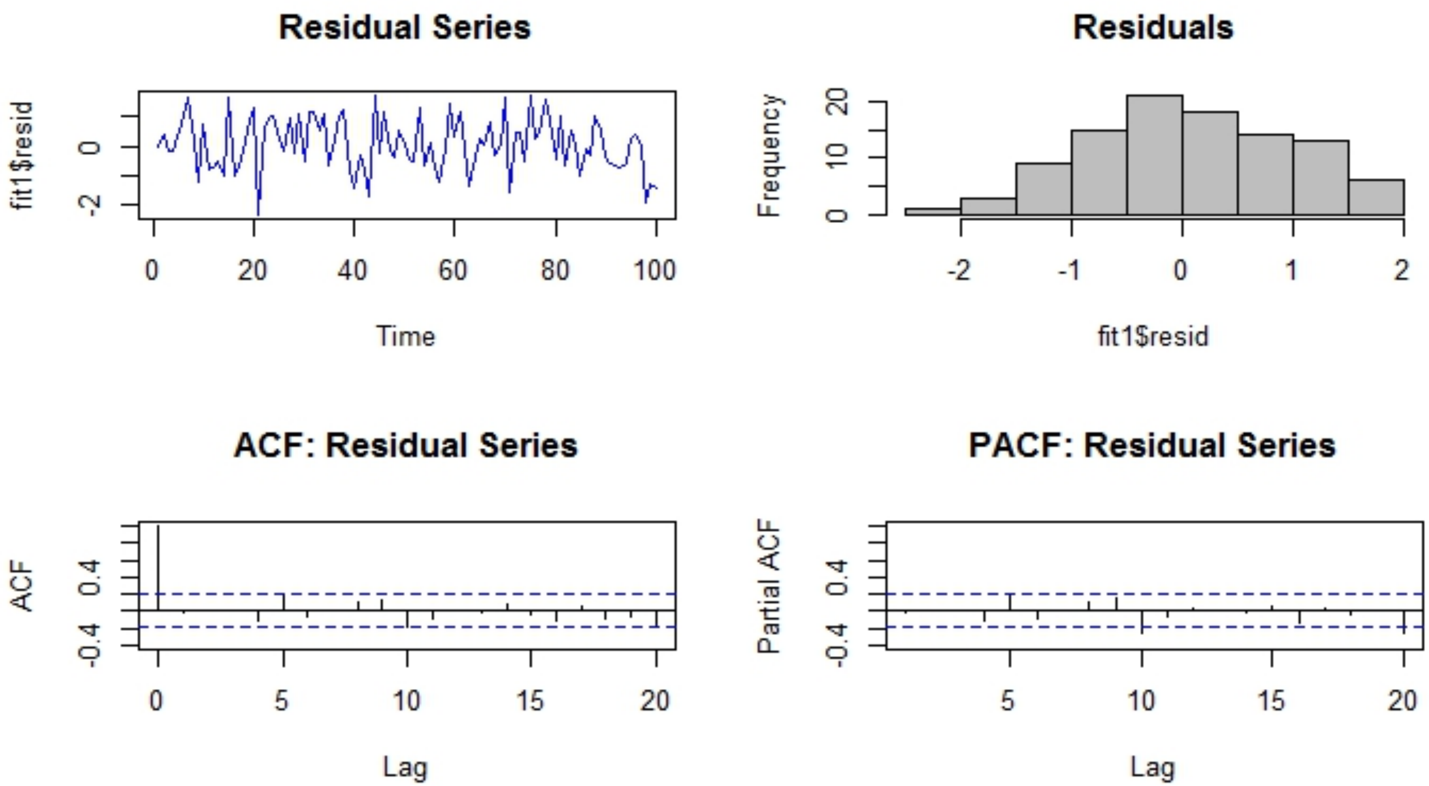
Estimation Using the Simulated Data

- Note that both of the estimated coefficients are not statistically different from their "true" values.

```
> fit1 <- arima(x1, order=c(1,1,1))
> summary(fit1)
Series: x1
ARIMA(1,1,1)

Coefficients:
         ar1    ma1
        0.36   0.32
s.e.    0.18   0.19

sigma^2 estimated as 0.827:  log likelihood=-131
AIC=269   AICc=269   BIC=276

Training set error measures:
                ME RMSE  MAE MPE MAPE MASE    ACF1
Training set 0.054  0.9 0.75 -23   61 0.82 -0.0028
```

Model Diagnostics Using Residuals

Both the graphical evidence and the Ljung-Box statistic do not reject the residual series as a realization of a white noise process.

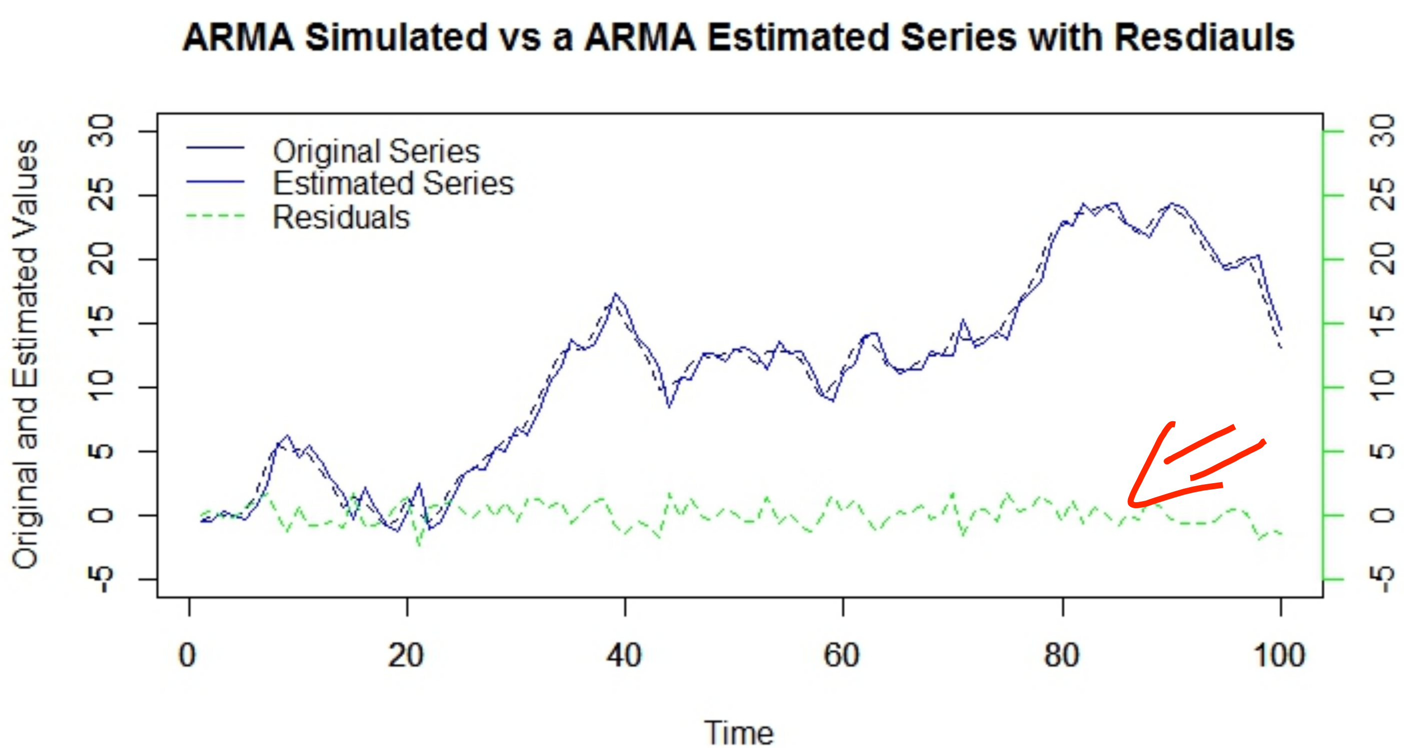



```
 Box-Ljung test

data:  fit1$resid
X-squared = 0.0008, df = 1, p-value = 0.9775
```

Model Performance Evaluation

The model fit is good; it can even capture some of the "turns" in the series.

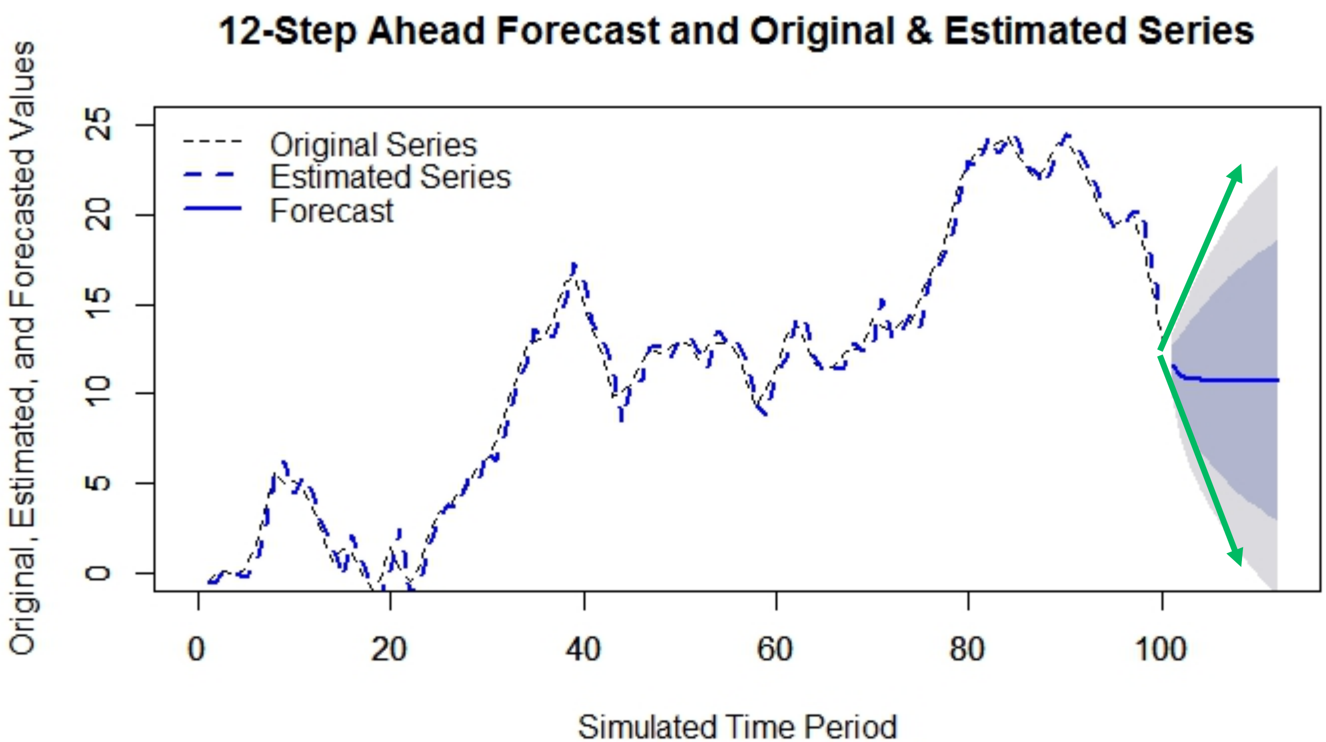
```
Descriptive Stat
==========================================
Statistic       N   Mean  St. Dev.  Min   Max
------------------------------------------
x1             100  12.0    7.4    -0.9  24.0
fitted.fit1.   100  12.0    7.5    -1.3  24.0
fit1.resid     100  0.1     0.9    -2.4  1.7
------------------------------------------
```



Forecasting

While the model fit is good, this model produces only a forecast that is a “flat” line.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
101	12	10.4	13	9.78	13
102	11	8.7	13	7.53	15
103	11	7.6	14	5.93	16
104	11	6.8	15	4.68	17
105	11	6.1	15	3.64	18
106	11	5.5	16	2.73	19
107	11	5.0	16	1.92	20
108	11	4.5	17	1.18	20
109	11	4.0	17	0.49	21
110	11	3.6	18	-0.16	22
111	11	3.2	18	-0.77	22



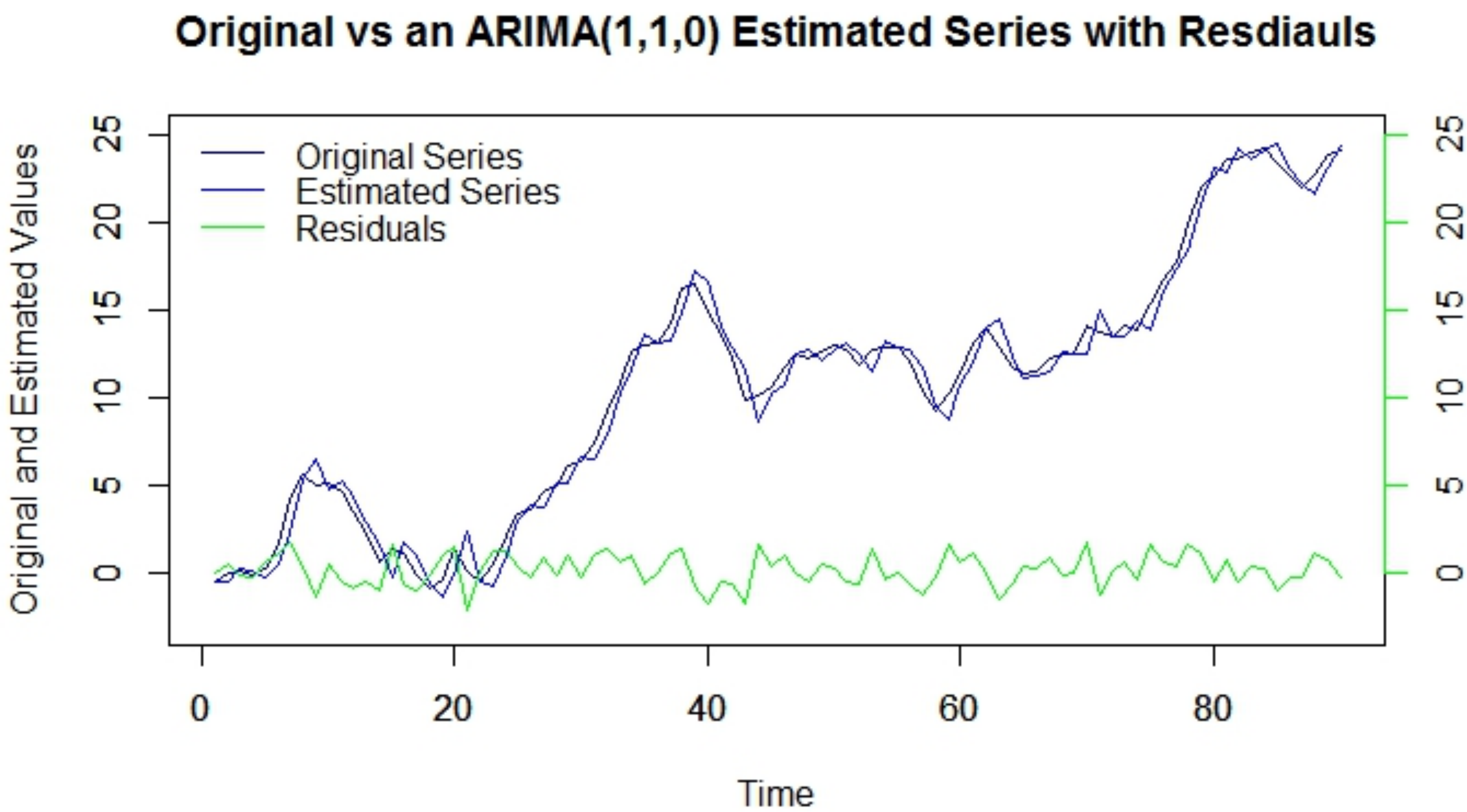
Back-Testing/Out-of-Sample Forecasting

- Re-estimate the model using an ARIMA(1,1,0).
- The AR coefficient is highly significant.

```
Series: x1[1:(length(x1) - 10)]
ARIMA(1,1,0)

Coefficients:
       ar1
      0.52
s.e.  0.09

sigma^2 estimated as 0.83:  log likelihood=-118
AIC=240   AICc=240   BIC=245

Training set error measures:
               ME RMSE  MAE MPE MAPE MASE  ACF1
Training set 0.13 0.91 0.74 -18   68 0.85 0.053
```

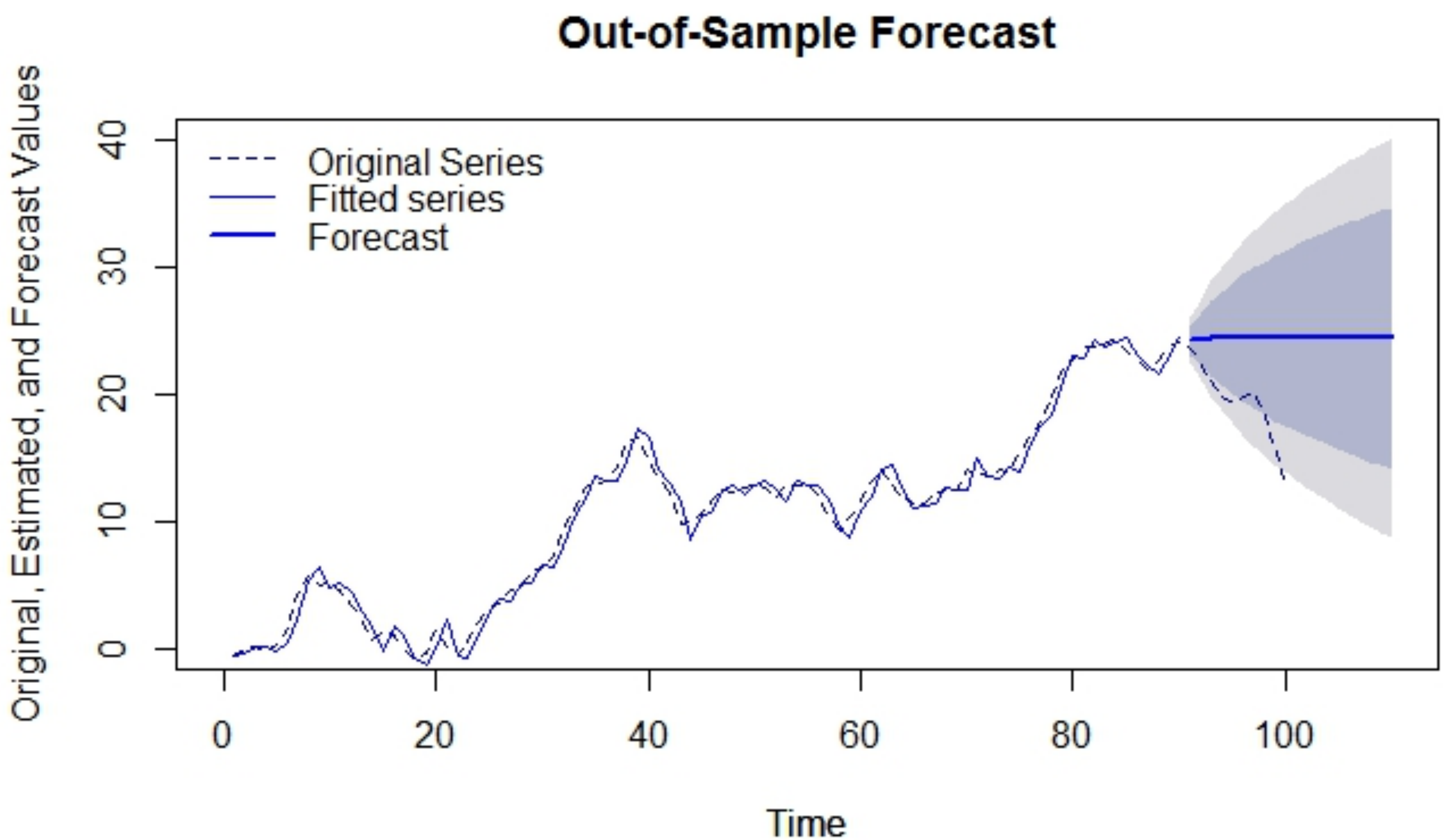
Back-Testing/Out-of-Sample Forecasting

- The model fit is very good.
- The residuals appear to be white noise.



Back-Testing/Out-of-Sample Forecasting

- Although the fit is good, only the 1-step ahead forecast is close for reasons we explained before



Seasonal ARIMA
Introduction and Mathematical Formulation

Seasonal ARIMA: Introduction

- So far we have been side-stepping the issue of seasonal effects.
- The dependence on the past often tends to occur most strongly at multiples of some underlying seasonal lag *s*.
- ARIMA models can be extended to include seasonal effects.
- A seasonal ARIMA (SARIMA) model uses differencing at a lag equal to the number of season(s) to remove additive seasonal effects.
- A seasonal ARIMA model is formed by including <u>additional seasonal terms</u> in the ARIMA models, as written below.

$$\text{ARIMA}\quad \underbrace{(p,d,q)}_{\substack{\uparrow \\ \left(\text{Non-seasonal part of the model}\right)}}\ \underbrace{(P,D,Q)_m}_{\substack{\uparrow \\ \left(\text{Seasonal part of the model}\right)}} \leftarrow S$$

where m is the number of periods per season.

Seasonal ARIMA: Mathematical Formulation

- Uppercase notation is for the <u>seasonal parts</u> of the model.
- Lowercase notation for the <u>nonseasonal parts</u> of the model.
- The seasonal part of the model consists of terms that are very similar to the nonseasonal components of the model, but they involve backshifts of the seasonal period.
- The general-form seasonal ARIMA model can be written as below.

The seasonal ARIMA$(p, d, q)(P, D, Q)_s$ model can be most succinctly expressed using the backward shift operator

$$\boxed{\Theta_P(\mathrm{B}^s)\theta_p(\mathrm{B})(1-\mathrm{B}^s)^D(1-\mathrm{B})^d x_t = \Phi_Q(\mathrm{B}^s)\phi_q(\mathrm{B})w_t}$$

where Θ_P, θ_p, Φ_Q, and ϕ_q are polynomials of orders P, p, Q, and q.

Seasonal ARIMA: Explanation of the Formula

The seasonal ARIMA$(p, d, q)(P, D, Q)_s$ model can be most succinctly expressed using the backward shift operator

$$\Theta_P(\mathrm{B}^s)\theta_p(\mathrm{B})(1-\mathrm{B}^s)^D(1-\mathrm{B})^d x_t = \Phi_Q(\mathrm{B}^s)\phi_q(\mathrm{B})w_t$$

where Θ_P, θ_p, Φ_Q, and ϕ_q are polynomials of orders P, p, Q, and q.

For example, an ARIMA(1,1,1)(1,1,1)$_4$ model (without a constant) is for quarterly data (i.e., m=4) and can be written as:

$$\underbrace{(1-\phi_1 B)}_{\text{Non-seasonal AR(1)}}\ \underbrace{(1-\Phi_1 B^4)}_{\boxed{\text{Seasonal AR(1)}}}\ \underbrace{(1-B)}_{\text{Non-seasonal difference}}\ \underbrace{(1-B^4)}_{\boxed{\text{Seasonal difference}}}y_t = \underbrace{(1+\theta_1 B)}_{\text{Non-seasonal MA(1)}}\ \underbrace{(1+\Theta_1 B^4)}_{\boxed{\text{Seasonal MA(1)}}}e_t.$$

Seasonal ARIMA: Examples

(a) A simple AR model with a seasonal period of 12 units, denoted as ARIMA(0, 0, 0)(1, 0, 0)$_{12}$, is $x_t = \alpha x_{t-12} + w_t$. Such a model would be appropriate for monthly data when only the value in the month of the previous year influences the current monthly value. The model is stationary when $|\alpha^{-1/12}| > 1$.

(b) It is common to find series with stochastic trends that nevertheless have seasonal influences. The model in (a) above could be extended to $x_t = x_{t-1} + \alpha x_{t-12} - \alpha x_{t-13} + w_t$. Rearranging and factorising gives $(1 - \alpha \mathbf{B}^{12})\boxed{(1 - \mathbf{B})}x_t = w_t$ or $\Theta_1(\mathbf{B}^{12})(1 - \mathbf{B})x_t = w_t$, which, on comparing with Equation (7.3), is ARIMA(0, 1, 0)(1, 0, 0)$_{12}$. Note that this model could also be written $\nabla x_t = \alpha \nabla x_{t-12} + w_t$, which emphasises that the *change* at time t depends on the change at the same time (i.e., month) of the previous year. The model is non-stationary since the polynomial on the left-hand side contains the term $(1 - \mathbf{B})$, which implies that there exists a unit root $B = 1$.

Seasonal ARIMA: Examples (2)

(c) A simple quarterly seasonal moving average model is $x_t = (1 - \beta B^4)w_t = w_t - \beta w_{t-4}$. This is stationary and only suitable for data without a trend. If the data also contain a stochastic trend, the model could be extended to include first-order differences, $x_t = x_{t-1} + w_t - \beta w_{t-4}$, which is an ARIMA(0, 1, 0)(0, 0, 1)$_4$ process. Alternatively, if the seasonal terms contain a stochastic trend, differencing can be applied at the seasonal period to give $x_t = x_{t-4} + w_t - \beta w_{t-4}$, which is ARIMA(0, 0, 0)(0, 1, 1)$_4$.

Seasonal ARIMA
An Example: European Quarterly Retail Trade

European Retail Trade: 1996–2011

- This is an example from the book *Forecasting: Principles and Practice* by Rob J. Hyndman and George Athanasopoulous.
- Rob Hyndman also wrote the R package *forecast*, which we use extensively in this course.
- The series is clearly nonstationary and appears to have some seasonality.
- First, take seasonal difference.

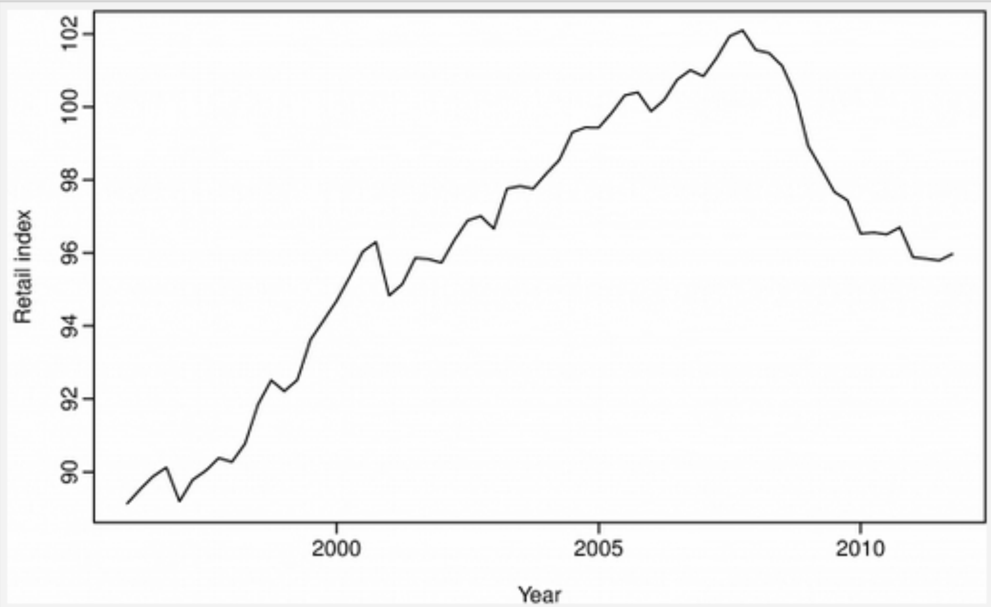


Figure 8.14: Quarterly retail trade index in the Euro area (17 countries), 1996–2011, covering wholesale and retail trade, and repair of motor vehicles and motorcycles. (Index: 2005 = 100).

European Retail Trade: 1996–2011

- The seasonal differenced series, along with its ACF and PACF, is shown below.

- The series still appears to be nonstationary.

- Take an additional first difference to convert it into a stationary series.

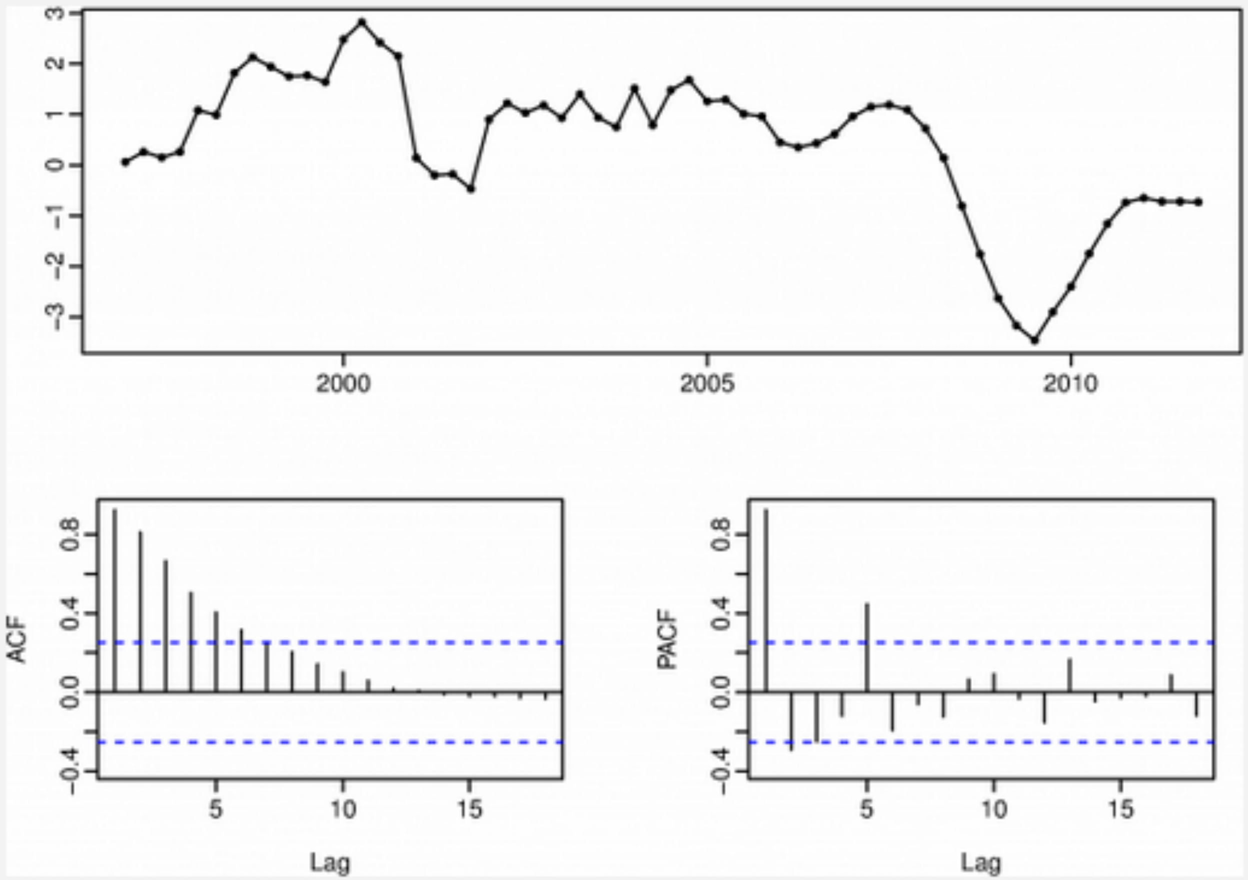


Figure 8.15: Seasonally differenced European retail trade index.

European Retail Trade: 1996–2011

- The seasonal differenced and first differenced series, along with its ACF and PACF, is shown below.

- After the differenced transformation, model the transformed series using ARMA model based on the ACF and PACF of the transformed series.
- The spike at lag 1 in the ACF suggests a nonseasonal MA(1) component.
- The spike at lag 4 in the ACF suggests a seasonal MA(1) component.

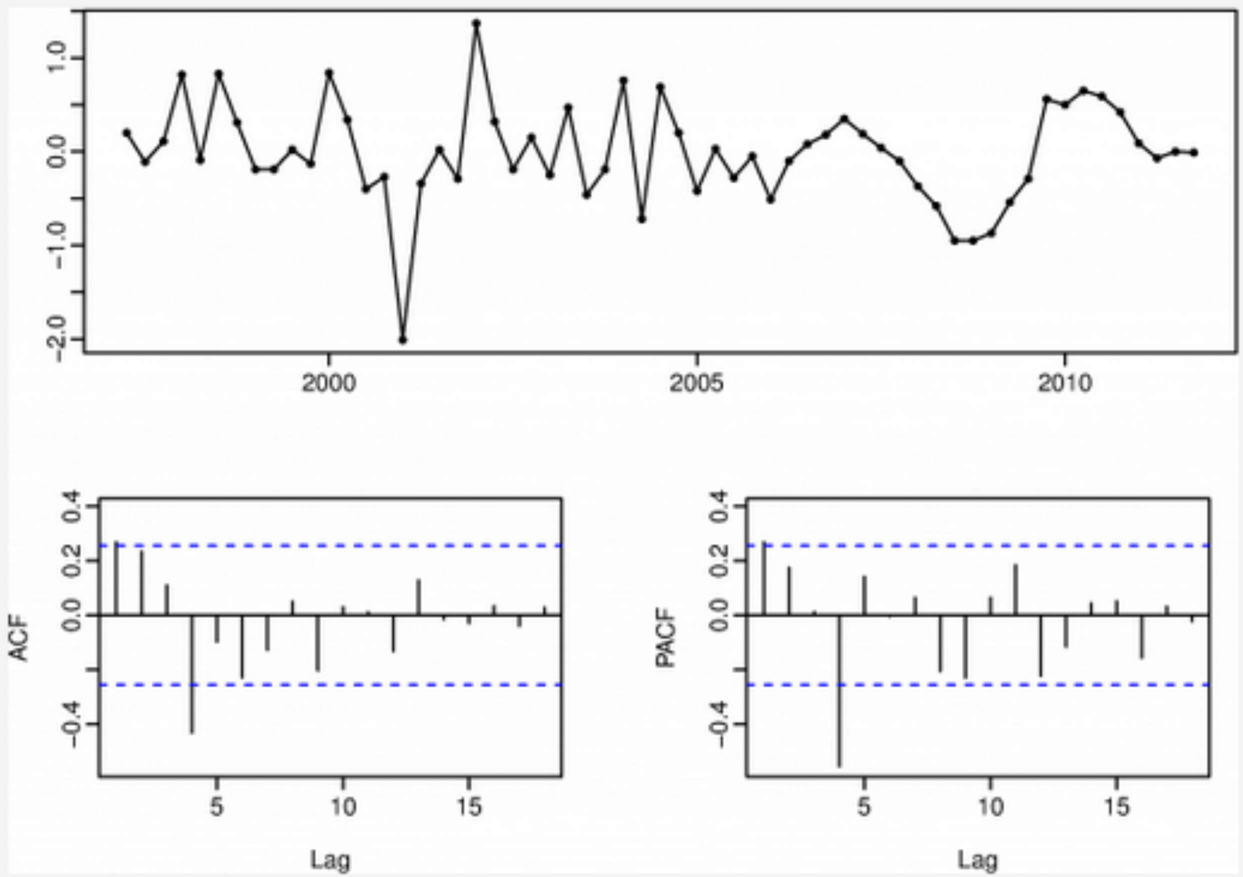


Figure 8.16: Double differenced European retail trade index.

European Retail Trade: 1996–2011

- We will start with a model of the form ARIMA(0,1,1)(0,1,1)$_4$.
- The residuals of the estimated model is shown below.

A first and seasonal difference

Nonseasonal and seasonal MA(1) components

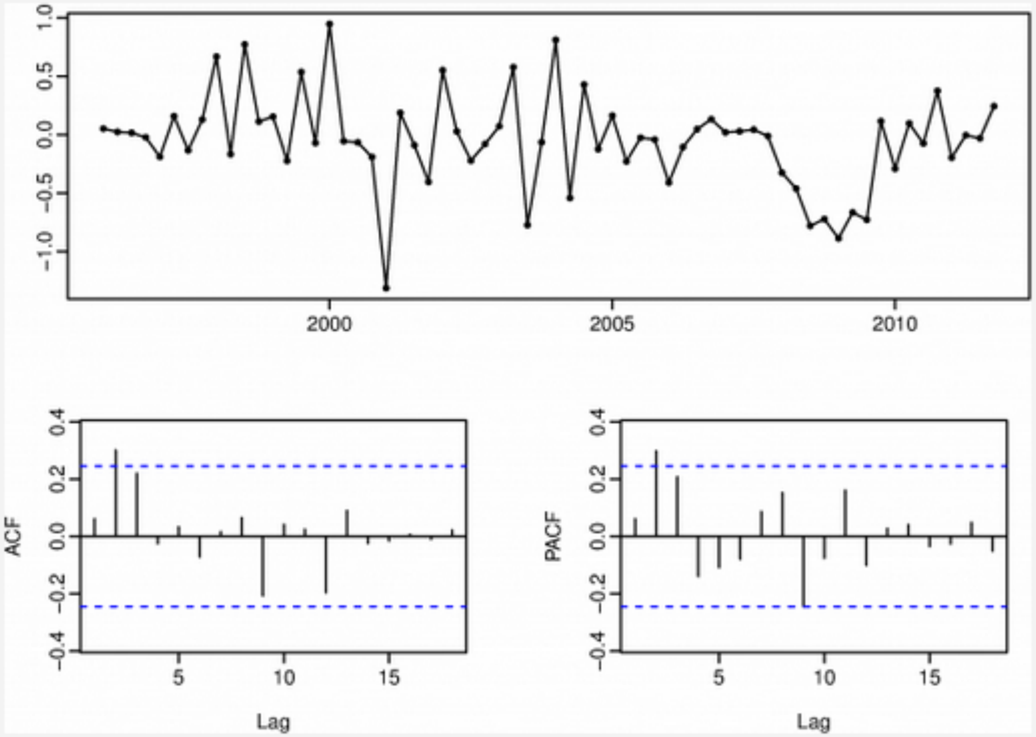


Figure 8.17: Residuals from the fitted ARIMA(0,1,1)(0,1,1)$_4$ model for the European retail trade index data.

Both the ACFs and PACFs show significant spikes at lag 2, suggesting the need to include additional nonseasonal terms.

European Retail Trade: 1996–2011

- Different models are estimated, including ARIMA$(0,1,2)(0,1,1)_4$, ARIMA$(0,1,3)(0,1,1)_4$, and other models with AR terms, but ARIMA$(0,1,3)(0,1,1)_4$ gave the smallest AIC value.
- Remember that AIC is not the only model fit measure. Also, we may want to use a recursive method and use out-of-sample fit measure to select a model.

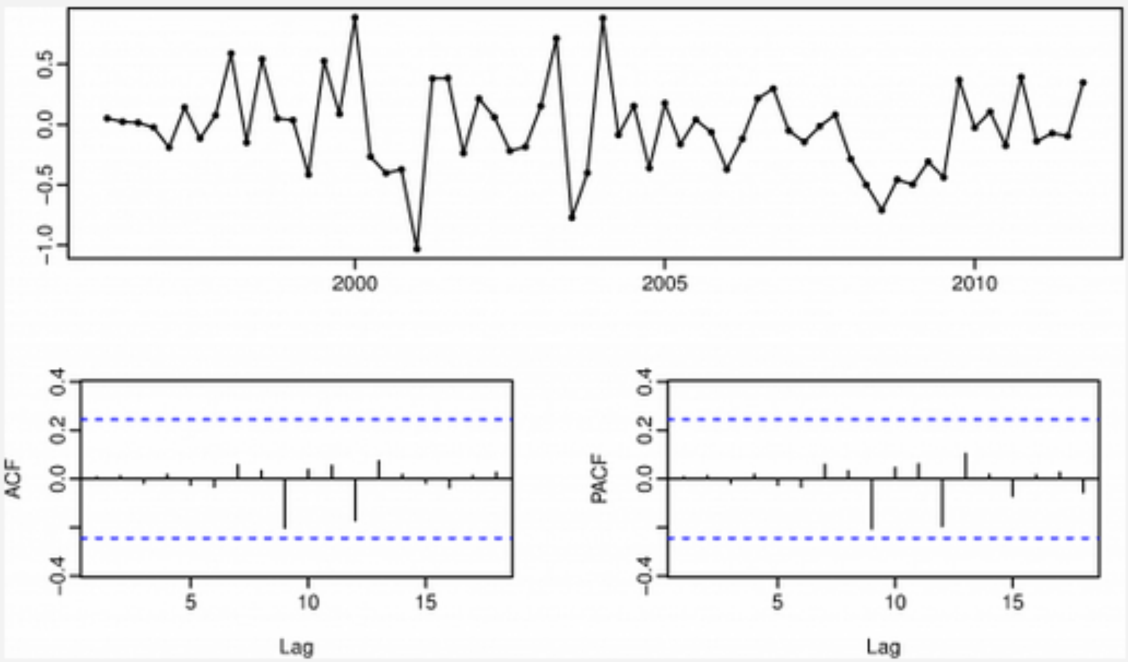


Figure 8.18: Residuals from the fitted ARIMA$(0,1,3)(0,1,1)_4$ model for the European retail trade index data.

European Retail Trade (1996–2011): Forecasting

- With a candidate model that appears to satisfy the statistical assumptions of the model, let's proceed to forecasting.
- A 12-step- (i.e., 12 quarters or three years) ahead forecast is shown below.

- Note 1: The forecast follows the "most recent trend" (due to the double differencing transformation applied to the series we model).

- Note 2: The large and increasing prediction interval allows for an increase retail trade index; in fact, both increasing and decreasing trends cannot be ruled out while the mean forecast points to a decreasing trend.

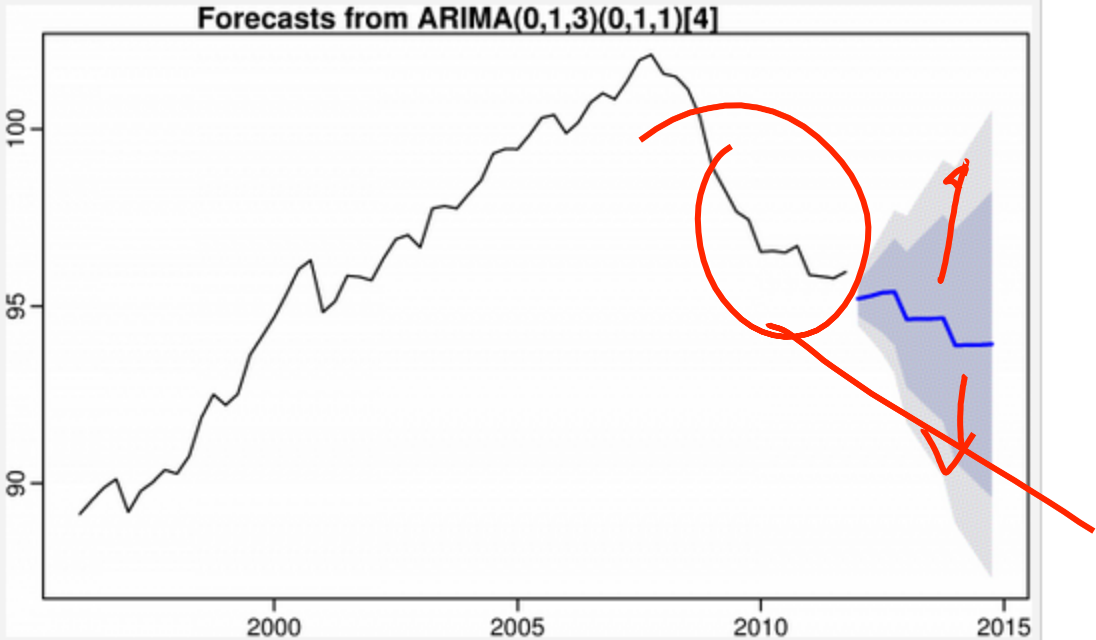


Figure 8.19: Forecasts of the European retail trade index data using the ARIMA(0,1,3)$(0,1,1)_4$ model. 80% and 95% prediction intervals are shown.

Putting Everything Together: ARIMA Modeling

Part 1: The Data

The Data: Australian Electricity Production Series

- This example uses the Australian Electricity Production series that comes with the textbook. This corresponds to the example on page 138 of the textbook.
- The data can be downloaded from the authors' website.
- The data contain three variables, but we will use only the electricity production series (elec); summary statistics are provided below.

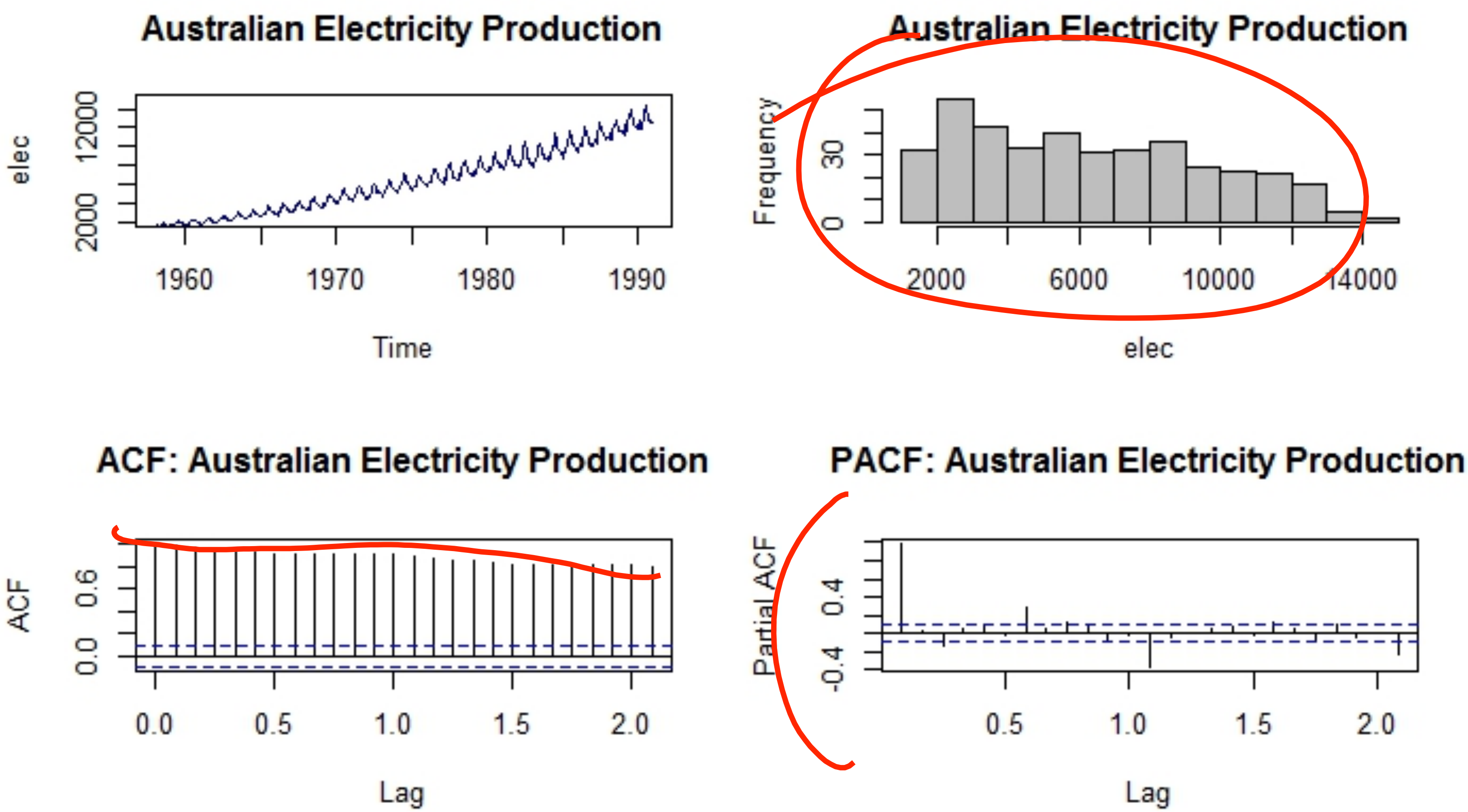
```
> str(cbe)
'data.frame':	396 obs. of  3 variables:
 $ choc: int  1451 2037 2477 2785 2994 2681 3098 2708 2517 2445 ...
 $ beer: num  96.3 84.4 91.2 81.9 80.5 70.4 74.8 75.9 86.3 98.7 ...
 $ elec: int  1497 1463 1648 1595 1777 1824 1994 1835 1787 1699 ...
```

```
> elec <- ts(cbe$elec,start=1958,freq=12)
> str(elec)
 Time-Series [1:396] from 1958 to 1991: 1497
```

```
> summary(elec)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1460    3240    5890    6310    8820   14300
> quantile(elec, c(0.01,0.05,0.1,0.25,0.5,0.75,0.9,0.95,0.99))
   1%    5%   10%   25%   50%   75%   90%   95%   99%
 1597  1848  2108  3239  5891  8820 11332 12192 13844
```

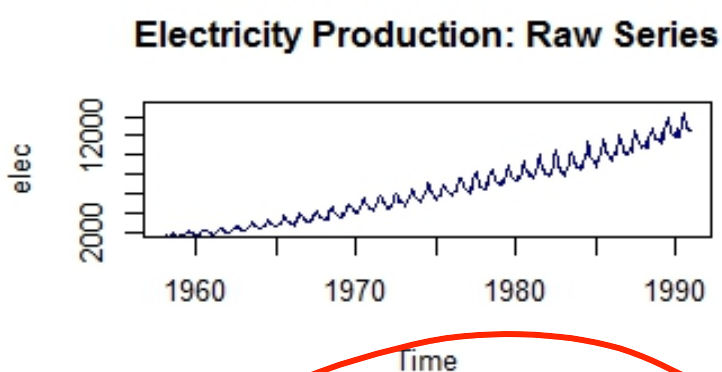
The Raw Series

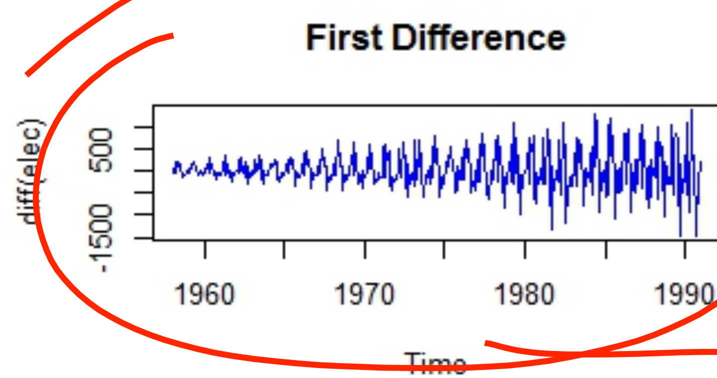
- The raw series appears to have both an upward trend and seasonality.
- The ACF graph shows a very strong autocorrelation even after 20 months, suggesting a strong candidate for first differencing.

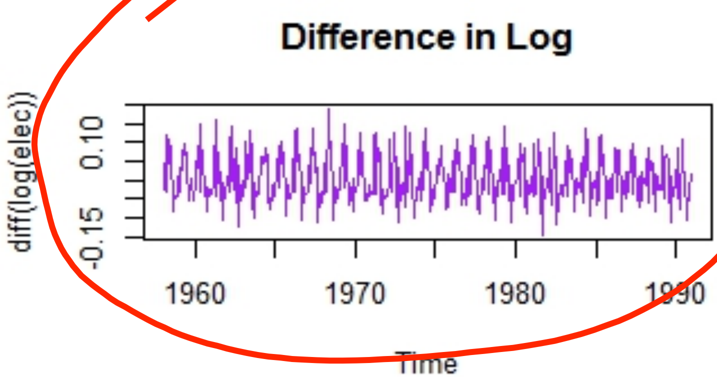


The Transformed Series

- The first differenced series takes out the trend, but the volatility increases over time.
- The difference in log, which approximates percentage change, also takes out the changing volatility.





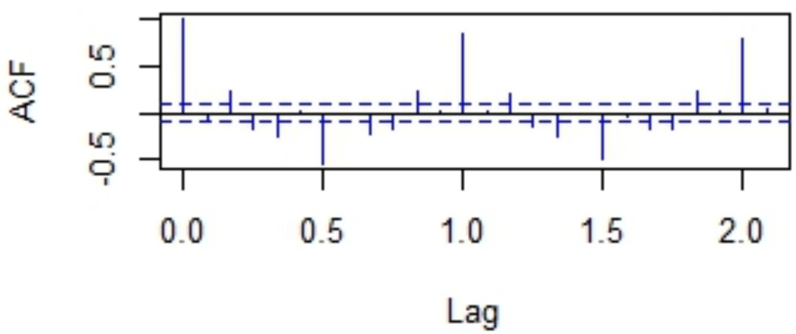



```
> summary(diff(elec))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  -1470    -197     -30      28     258    1380
> summary(diff(log(elec)))
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 -0.145  -0.038  -0.007   0.005   0.051   0.188
```

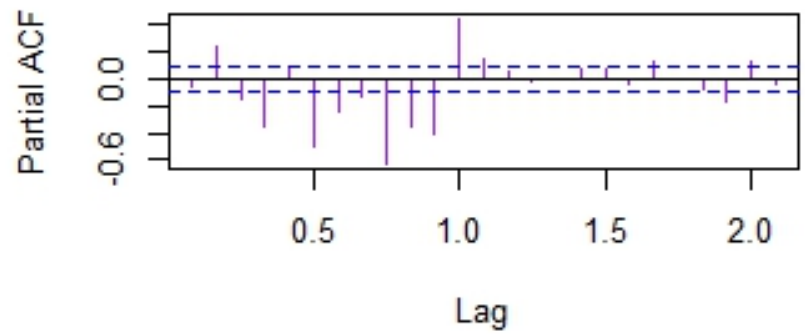
The Transformed Series

- The ACFs and the PACFs of the transformed series show seasonal effects.

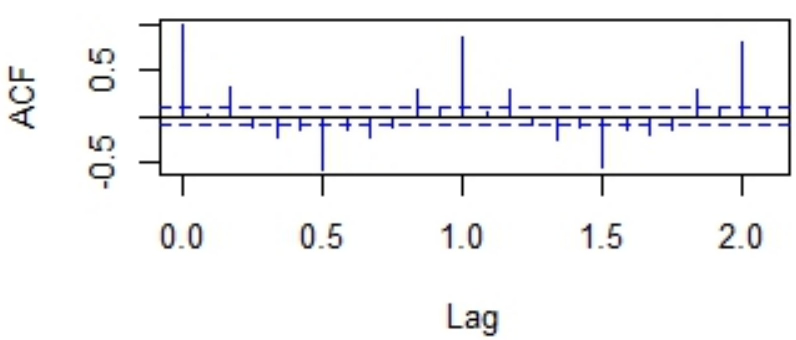
ACF: First Difference Series



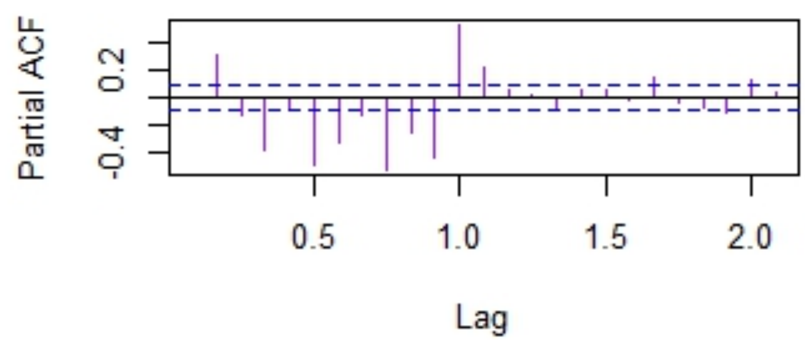
PACF: First Difference Series



ACF: First Difference in Log Series



PACF: First Difference in Log Series



Putting Everything Together: ARIMA Modeling

Part 2: Modeling

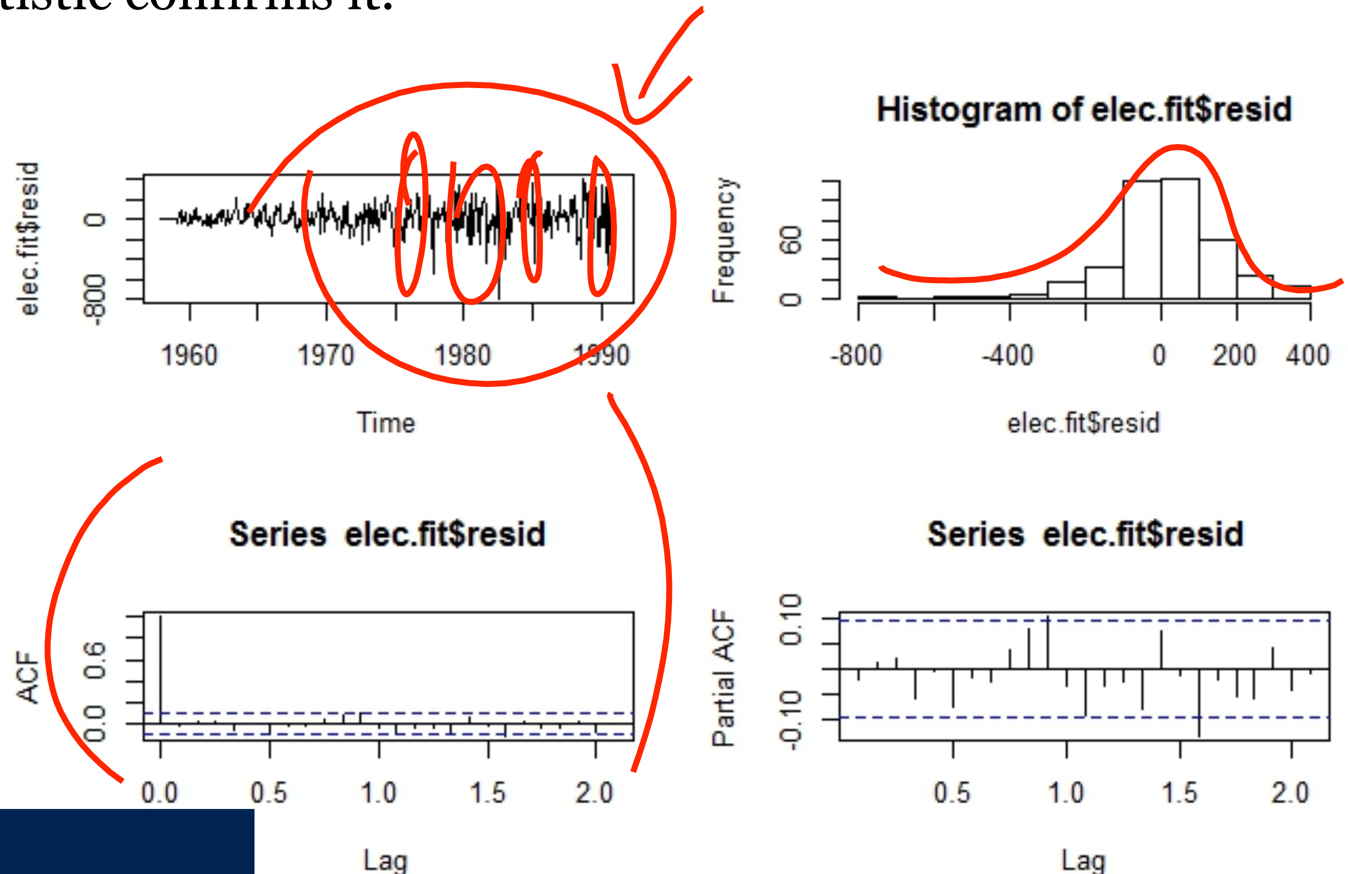
Estimation

- We will model on the first differenced series, using a seasonal differenced with a MA(1) component and a first differenced with both AR(1) and MA(2) components. We could also run a series of models of small variants of this model.
- All the components are highly significant.

```
> elec.fit <- Arima(elec, order=c(1,1,2), seasonal=c(0,1,1))
> summary(elec.fit)
Series: elec
ARIMA(1,1,2)(0,1,1)[12]

Coefficients:
       ar1     ma1    ma2     sma1
      0.83   -1.48   0.51   -0.537
s.e.  0.11    0.13   0.11    0.044

sigma^2 estimated as 22156:  log likelihood=-2462
AIC=4935   AICc=4935   BIC=4954

Training set error measures:
             ME RMSE MAE MPE MAPE MASE   ACF1
Training set 13  146 105 0.2  1.7 0.33 -0.022
```

Model Diagnostics Using Residuals

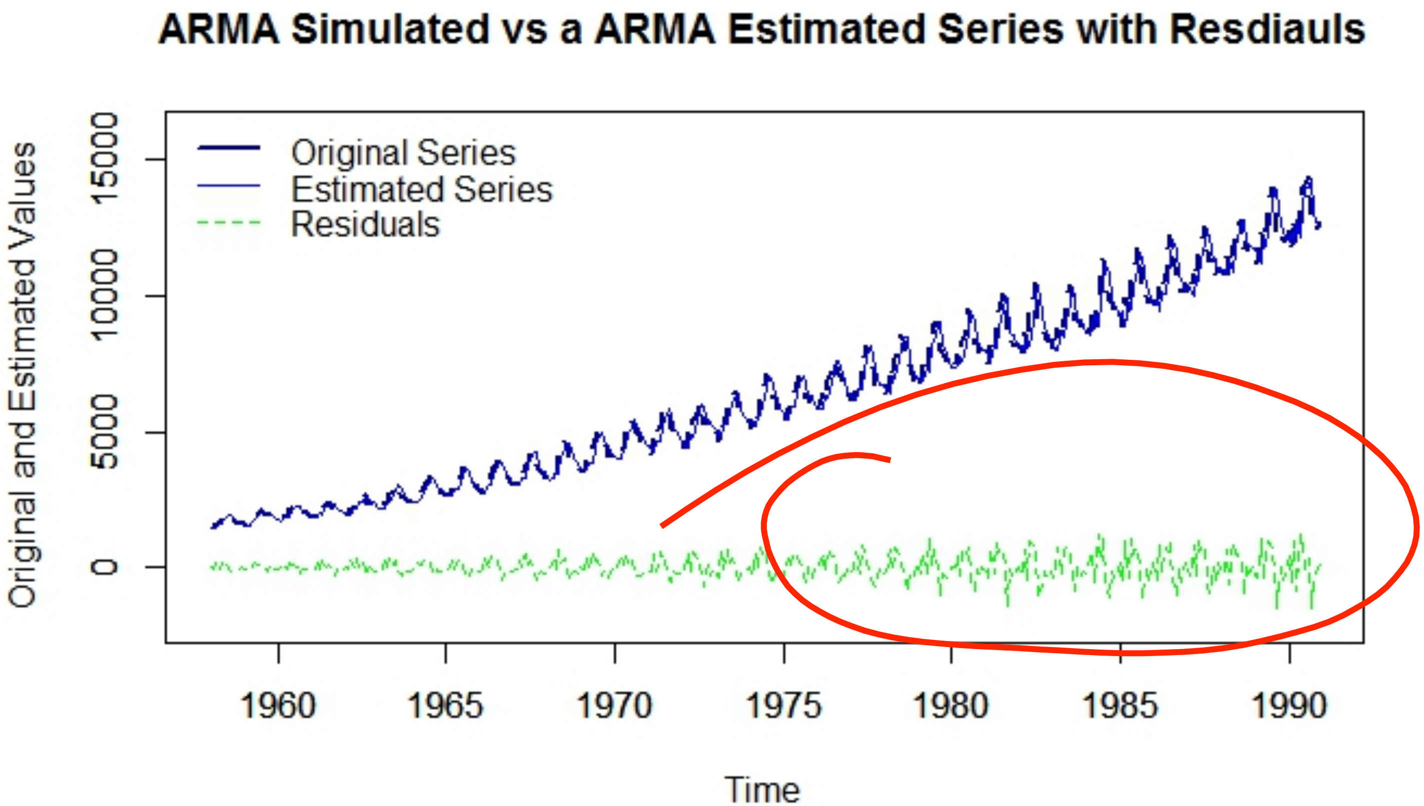
- The residuals clearly increase in volatility over time.
- The distribution of the residuals is skewed.
- However, both the ACFs and PACFs display no significant correlation; Ljung-Box statistic confirms it.




```
 Box-Ljung test

data:  elec.fit$resid
X-squared = 2.6, df = 1, p-value = 0.1062
```

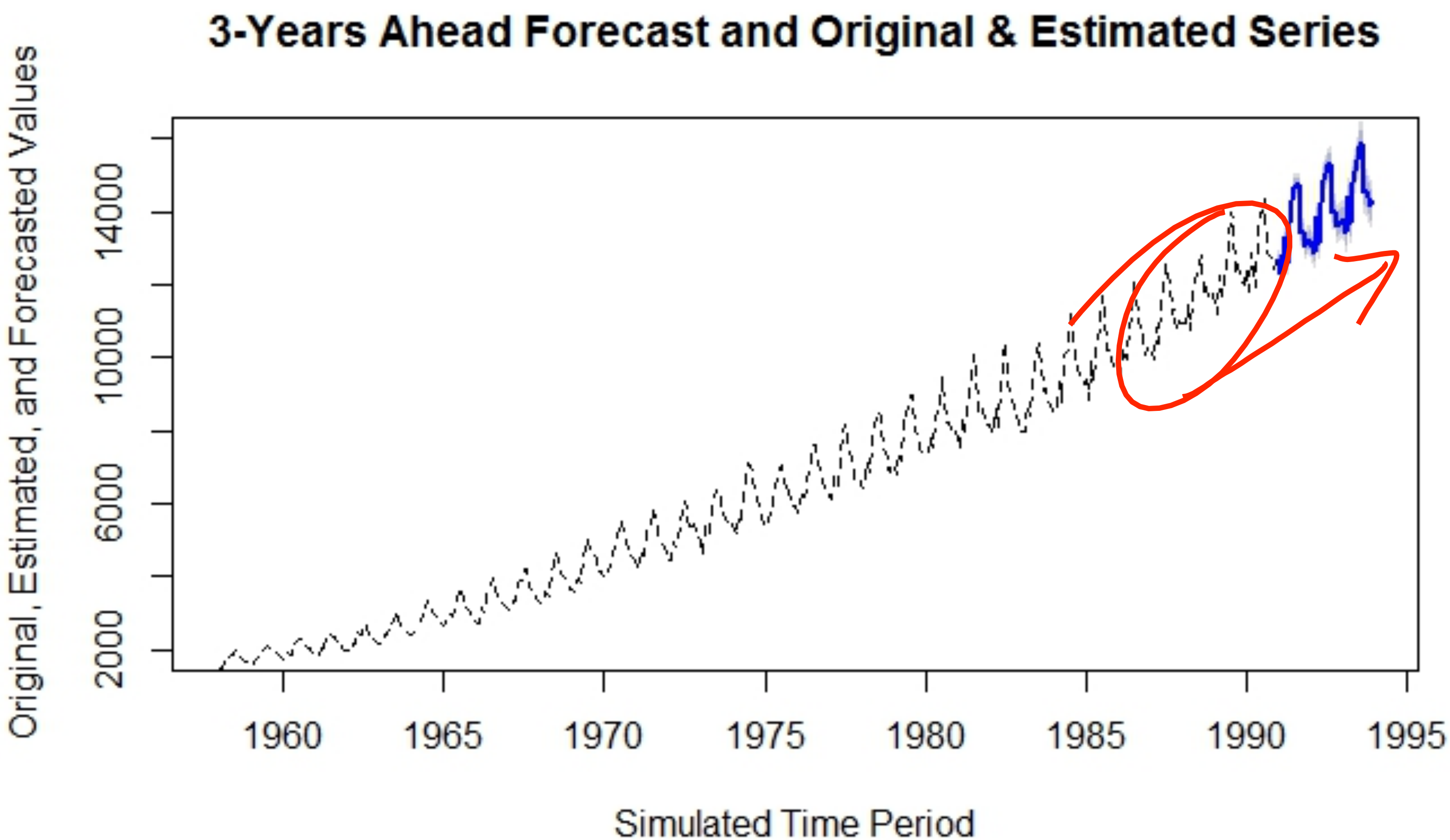
Model Performance Evaluation: In-Sample Fit

- Regardless, we proceed to examine the in-sample fit and forecast using this model, and we will try a series of variants of this model in the R-session.
- The graph below shows an excellent in-sample fit produced by the model.



Forecasting

- The three-year-ahead forecast continues the upward trend and reproduces the seasonal effect embedded in the original series.



Berkeley SCHOOL OF INFORMATION