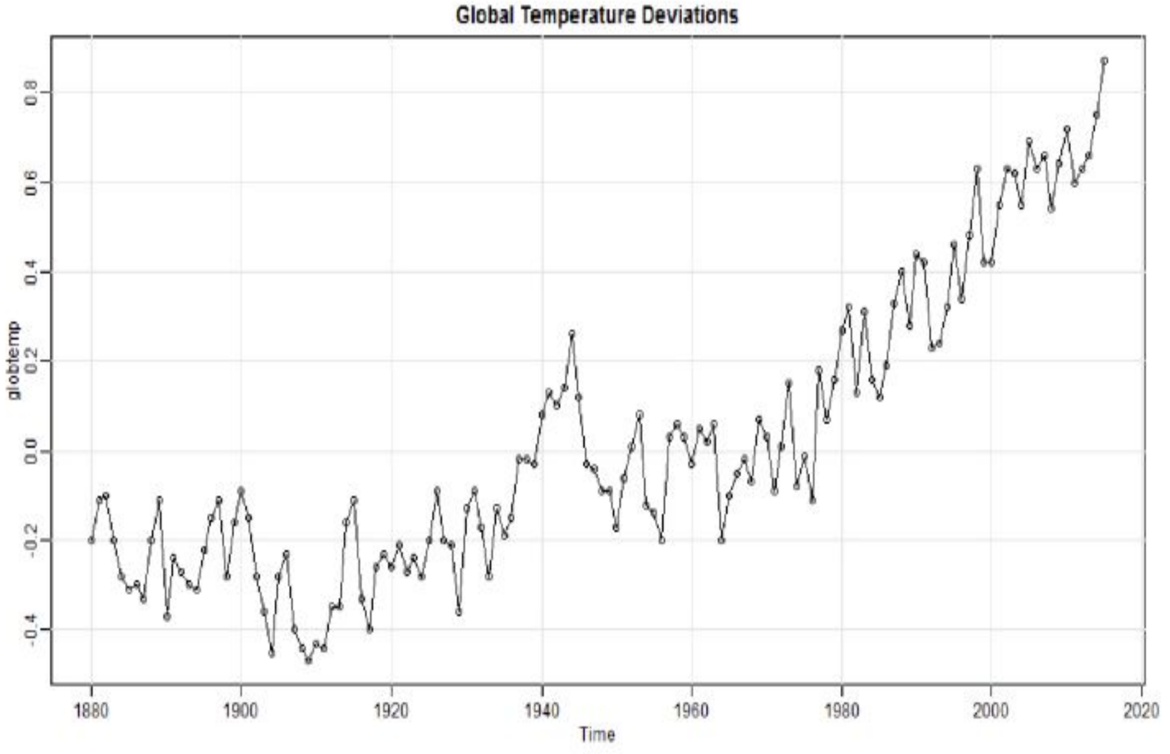
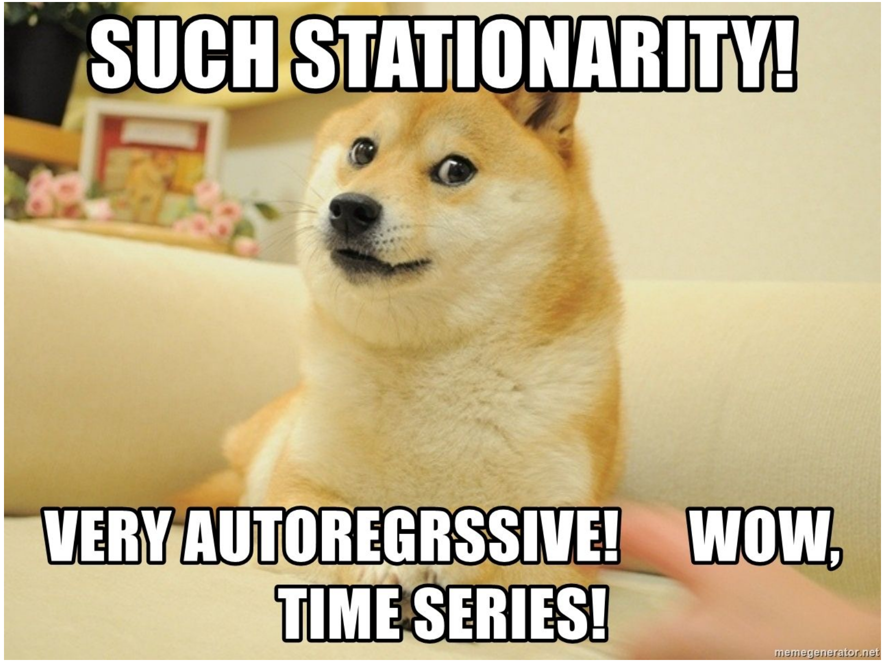
Introduction to Time Series with R

December 2-3, 2020

Ozan Evkaya



Audience



- All grades
- Standard introduction with Visual Examples
- Short Introduction for everyone

Main Objective

➔ Understanding the nature of any time series by its plot

* Main properties of a time series

* Stationary versus Non-stationary

➔ Main difference between AR(p) and MA(q) processes

* Describing a suitable model in advance

➔ Combining AR and MA: ARMA

Materials

- Brief Presentation
- R codes
- Book suggestions

R packages,
Plotting TS

Main Concepts for TS

Simple TS Modeling

40 mins

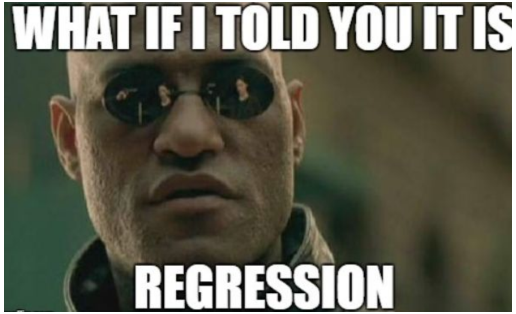
5 mins

40 mins

5 mins

30 mins

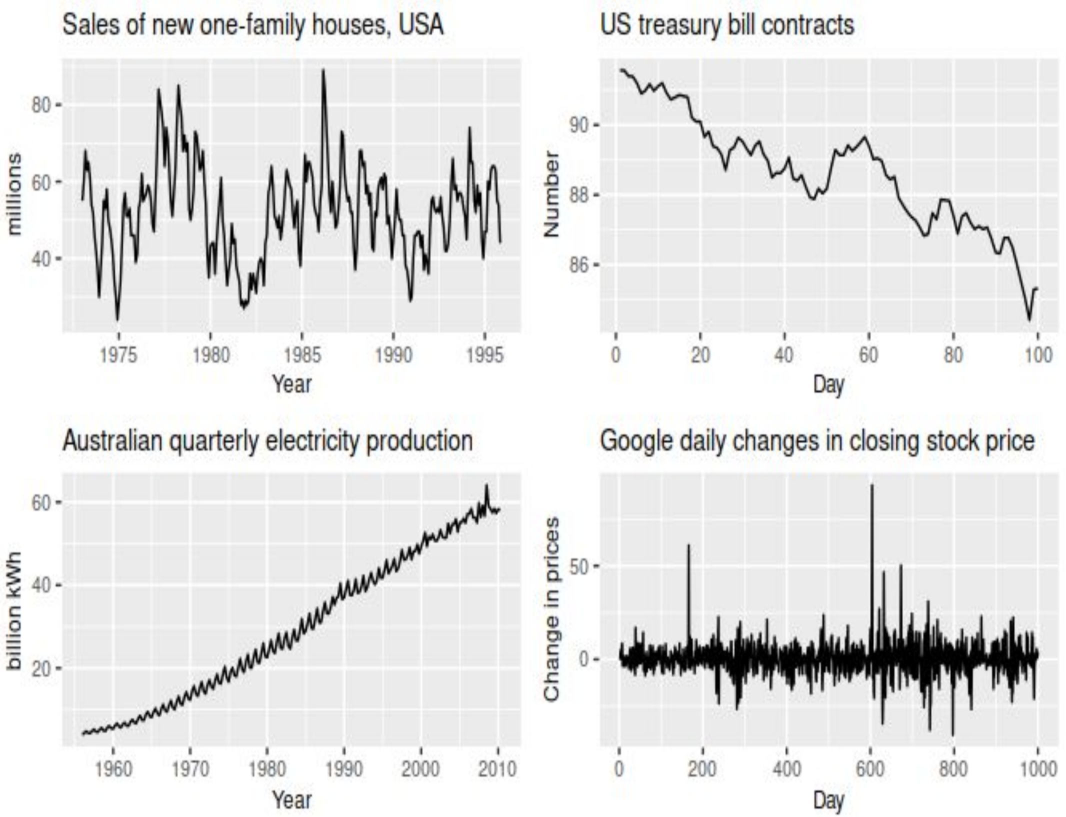
Relaxation

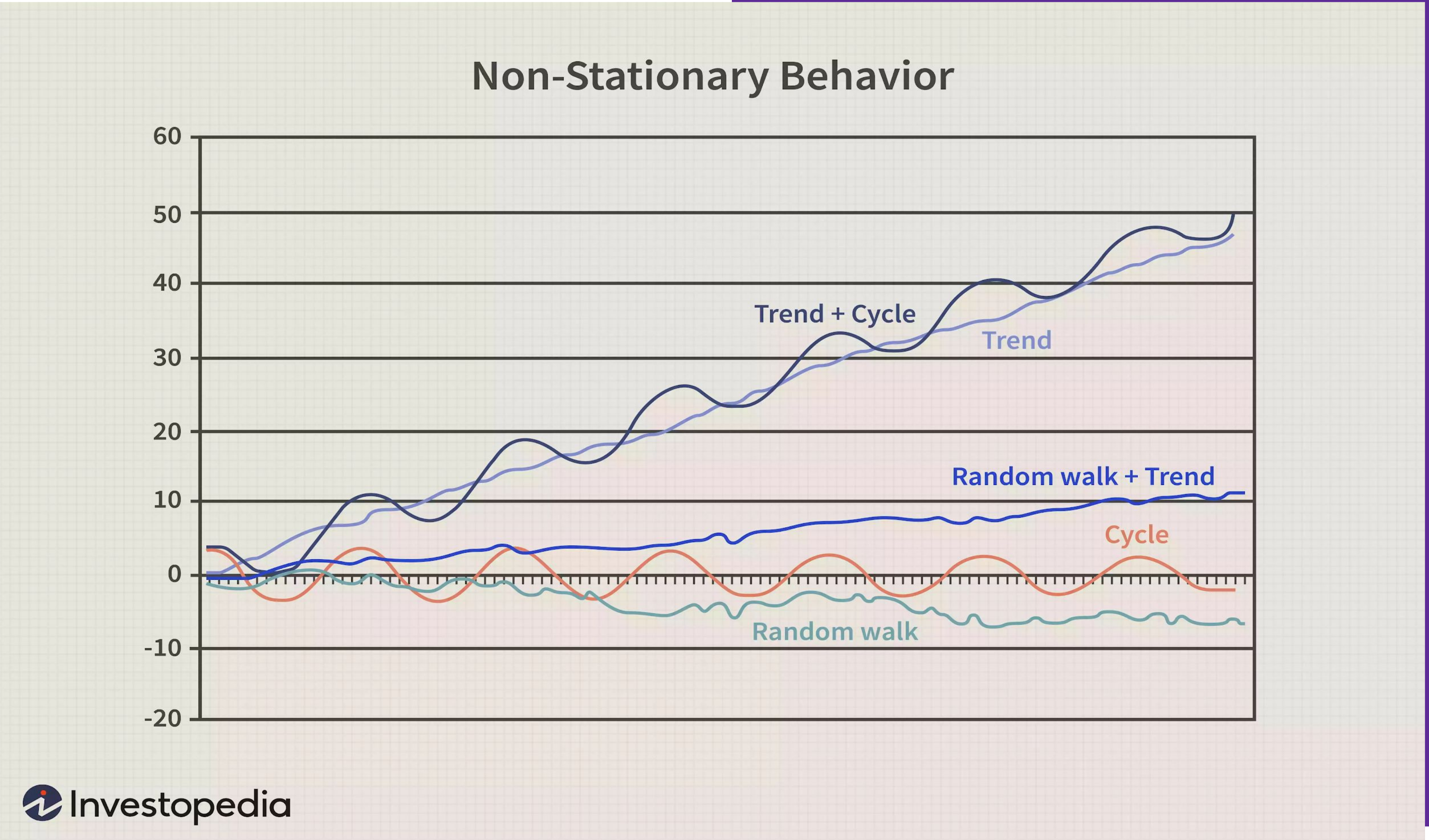


Relaxation

Basic Time Series Models

- White Noise (WN)
- Random Walk (RW)
- Autoregression (AR)
- Simple Moving Average (MA)




Non-Stationary Behavior
60
50
40
30
20
10
0
-10
-20
Trend + Cycle
Trend
Random walk + Trend
Cycle
Random walk
Investopedia

The Concepts and examples of
Unit-root tests and stationarity tests

Concept of Unit-root tests:
Null hypothesis: Unit-root
Alternative hypothesis: Process has root outside the unit circle, which is usually equivalent to stationarity or trend stationarity

Concept of Stationarity tests
Null hypothesis: (Trend) Stationarity
Alternative hypothesis: There is a unit root.

Some Unit root tests:

- Dickey-Fuller test
- Augmented Dickey Fuller test
- Phillipps-Perron test
- Zivot-Andrews test
- ADF-GLS test

The most simple test is the DF-test. The ADF and the PP test are similar to the Dickey-Fuller test, but they correct for lags. The ADF does so by including them the PP test does so by adjusting the test statistics.

Some Stationarity tests:

- KPSS
- Leybourne-McCabe

In practice KPSS test is used far more often. The main difference of both tests is that KPSS is a non-parametric test and Leybourne-McCabe is a parametric test.

Case 1 ADF test: you can't reject H_0;
KPSS test: reject H_0
Both imply that series has unit root.

Case 2 ADF test: Reject H_0
KPSS test: don't reject H_0
Both imply that series is stationary.

Case 3 If we can't reject both test: data give not enough observations.

Case 4 Reject unit root, reject stationarity: both hypotheses are component hypotheses – heteroskedasticity in a series may make a big difference; if there is structural break it will affect inference.

How unit-root test and stationarity-test complement each other

If you have a time series data

apply both a Unit root test: (Augmented) Dickey Fuller or Phillips-Perron depending on the structure of the underlying data and a KPSS test.

Definition 1.6 *A* **strictly stationary** *time series is one for which the probabilistic behavior of every collection of values*

$$\{x_{t_1}, x_{t_2}, \ldots, x_{t_k}\}$$

is identical to that of the time shifted set

$$\{x_{t_1+h}, x_{t_2+h}, \ldots, x_{t_k+h}\}.$$

That is,

$$P\{x_{t_1} \leq c_1, \ldots, x_{t_k} \leq c_k\} = P\{x_{t_1+h} \leq c_1, \ldots, x_{t_k+h} \leq c_k\} \tag{1.18}$$

for all $k = 1, 2, ...,$ *all time points* $t_1, t_2, \ldots, t_k$, *all numbers* $c_1, c_2, \ldots, c_k$, *and all time shifts* $h = 0, \pm 1, \pm 2, ...$.

Source: Time Series Analysis and Its Applications with R Examples Book

The version of stationarity in Definition 1.6 is too strong for most applications. Moreover, it is difficult to assess strict stationarity from a single data set. Rather than imposing conditions on all possible distributions of a time series, we will use a milder version that imposes conditions only on the first two moments of the series. We now have the following definition.

Definition 1.7 *A* **weakly stationary** *time series, x_t, is a finite variance process such that*

(i) the mean value function, μ_t, defined in (1.9) is constant and does not depend on time t, and
(ii) the autocovariance function, $\gamma(s,t)$, defined in (1.10) depends on s and t only through their difference $|s-t|$.

Henceforth, we will use the term **stationary** *to mean weakly stationary; if a process is stationary in the strict sense, we will use the term strictly stationary.*

Source: Time Series Analysis and Its Applications with R Examples Book

4.7 Autoregressive (AR) models

Autoregressive models of order p, abbreviated AR(p), are commonly used in time series analyses. In particular, AR(1) models (and their multivariate extensions) see considerable use in ecology as we will see later in the course. Recall from lecture that an AR(p) model is written as

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t, \qquad (4.22)$$

where $\{w_t\}$ is a white noise sequence with zero mean and some variance σ^2. For our purposes we usually assume that $w_t \sim \text{N}(0, q)$. Note that the random walk in Equation (4.18) is a special case of an AR(1) model where $\phi_1 = 1$ and $\phi_k = 0$ for $k \geq 2$.

Source:https://nwfsc-timeseries.github.io/atsa-labs/

4.8 Moving-average (MA) models

A moving-averge process of order q, or MA(q), is a weighted sum of the current random error plus the q most recent errors, and can be written as

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}, \tag{4.23}$$

where $\{w_t\}$ is a white noise sequence with zero mean and some variance σ^2; for our purposes we usually assume that $w_t \sim \mathrm{N}(0, q)$. Of particular note is that because MA processes are finite sums of stationary errors, they themselves are stationary.

Source:https://nwfsc-timeseries.github.io/atsa-labs/

4.9 Autoregressive moving-average (ARMA) models

ARMA(p, q) models have a rich history in the time series literature, but they are not nearly as common in ecology as plain AR(p) models. As we discussed in lecture, both the ACF and PACF are important tools when trying to identify the appropriate order of p and q. Here we will see how to simulate time series from AR(p), MA(q), and ARMA(p, q) processes, as well as fit time series models to data based on insights gathered from the ACF and PACF.

We can write an ARMA(p, q) as a mixture of AR(p) and MA(q) models, such that

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + w_t + \theta w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q x_{t-q}, \qquad (4.25)$$

and the w_t are white noise.

Source:https://nwfsc-timeseries.github.io/atsa-labs/

ARIMA (p,d,q) Modeling

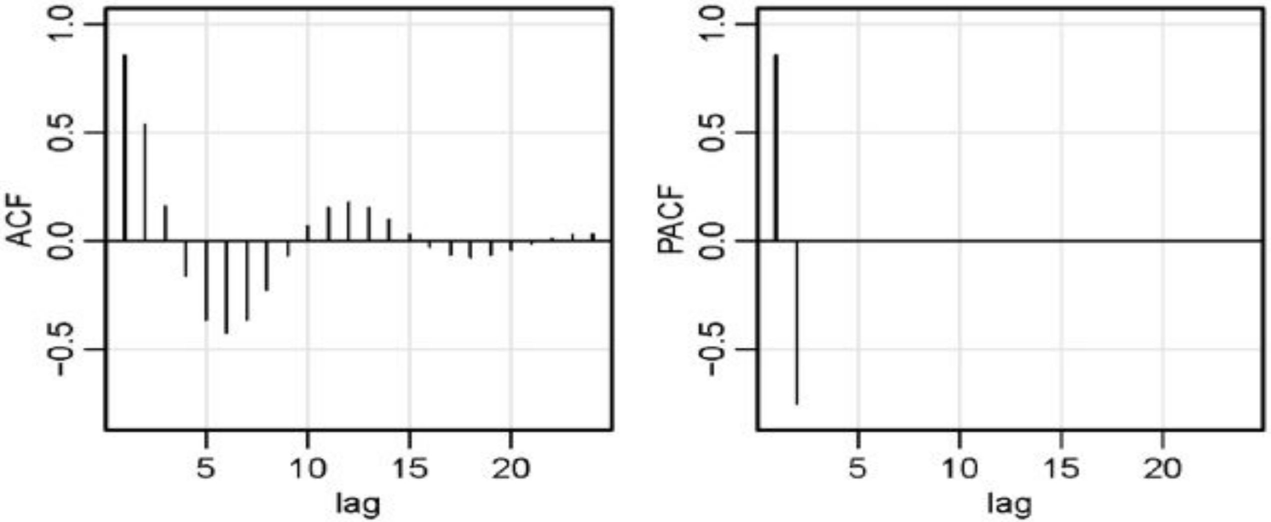
- To build a time series model issuing ARIMA, we need to study the time series and identify p,d,q.

1. **Ensuring Stationarity:**
   - Determine the appropriate values of d.
2. **Identification:**
   - Determine the appropriate values of p & q using the ACF, PACF.
3. **Diagnostic checking:**
   - pick best model with well behaved residuals.
4. **Forecasting:**
   - Produce out of sample forecasts or set aside last few data points for in-sample forecasting.

BY JOUD KHATTAB

Source:https://www.slideshare.net/joudkhattab/seasonal-arima

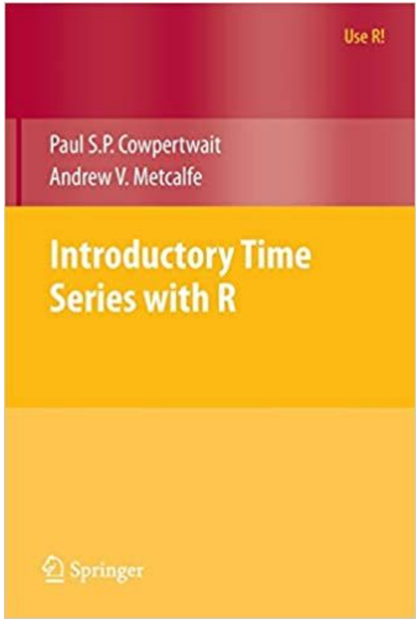
	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off lag q	Tails off
PACF	Cuts off lag p	Tails off	Tails off

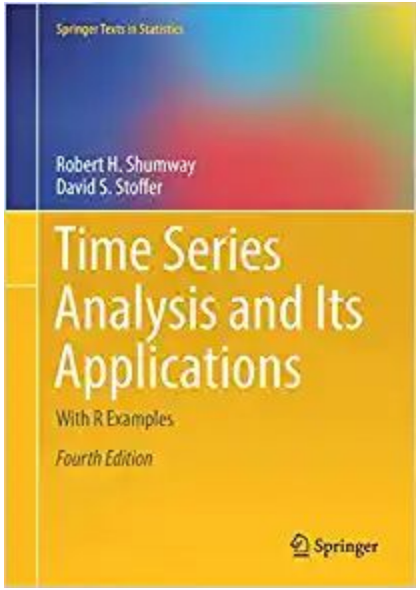


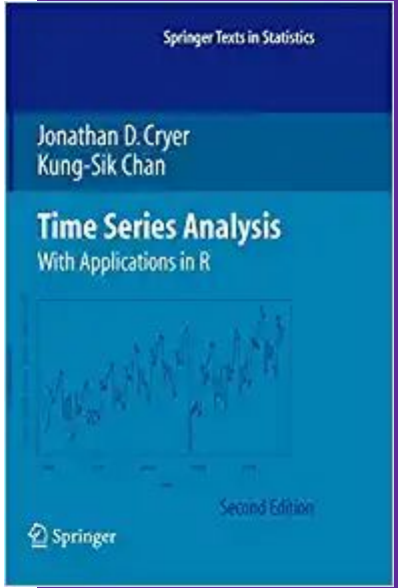
AR(2)

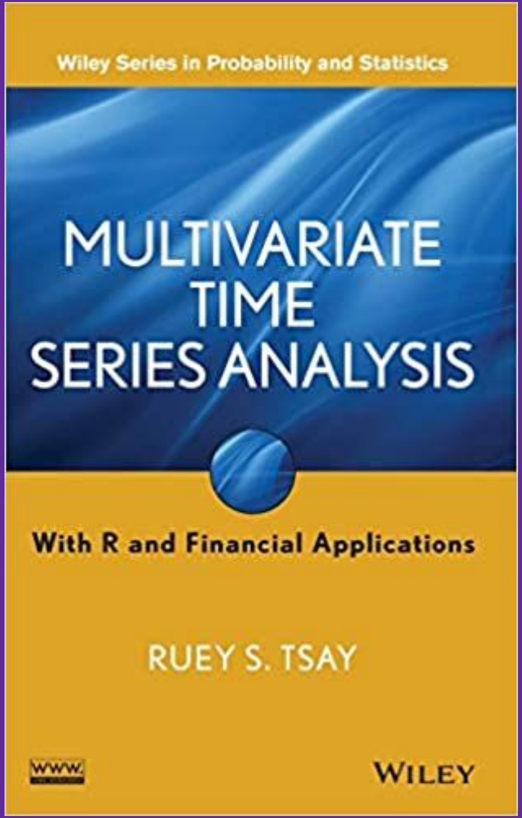
Source: DATACAMP Arima Models in R Lecture Notes

Useful References

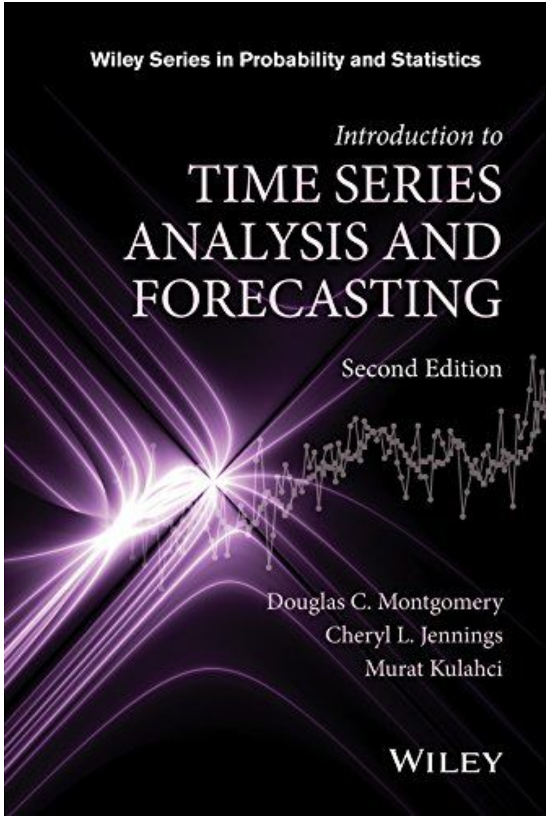


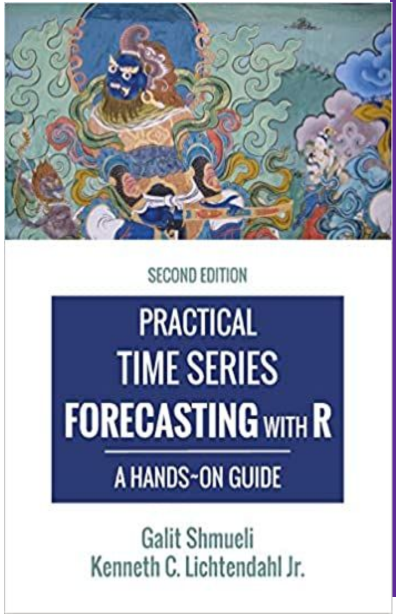


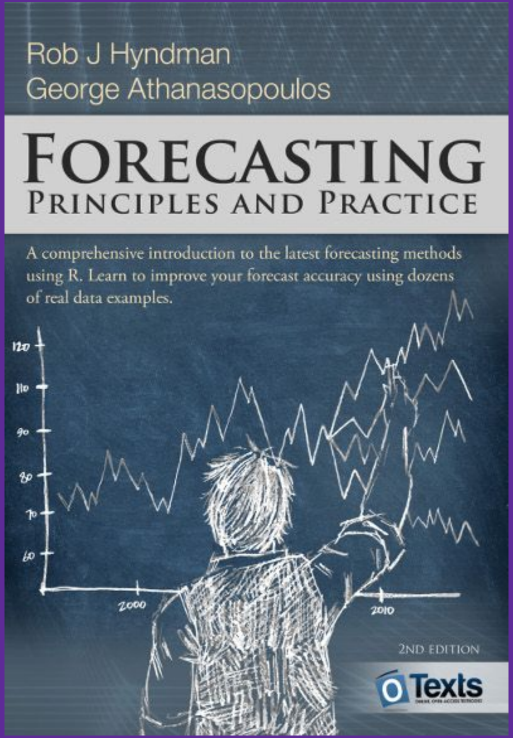


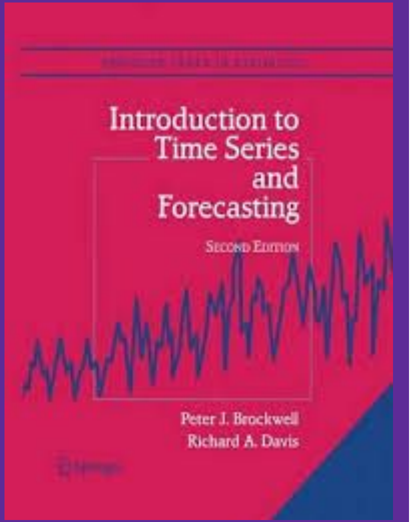


Useful References









KEEP in TOUCH

- https://www.linkedin.com/in/ozanevkaya/
- https://www.researchgate.net/profile/Ozan_Evkaya
- https://www.instagram.com/ozanevkaya/
- ozanevkaya@gmail.com