Business, Economic and Financial data
a.y. 2022-23

Mariangela Guidolin

Department of Statistical Sciences
University of Padua

October 2022

About me

Instructor: prof. **Mariangela Guidolin**

email: **guidolin@stat.unipd.it, mariangela.guidolin@unipd.it**

Students' tutorial: by appointment

Course program

General introduction

- General introduction to business and economic problems and data
- Preliminary concepts and illustrative examples

Moving beyond the linear model

- Linear regression model: main ideas and assumptions
- Nonlinear regression models for new product growth → Bass model
- Beyond linearity: regression splines, local regression, generalized additive models (GAMs)
- Tree-based methods

Time series analysis

- Models for time series data: Exponential Smoothing and ARIMA models

Course outline

- **Course description**
  The course is composed of frontal lectures and technical-practical lessons with R
- **Final exam**
  1. Practical exam (working group with final presentation)
  2. Oral exam

→ only ppt no reports required

→ Select real world Time series Data for the projects

Reference materials

Materials: Course handouts

Textbooks:

- **Forecasting, Principles and Practice** (O Texts, 2018) R.J. Hyndman, G. Athanasopoulos
- **Data analysis and data mining: An introduction.** (Oxford University Press, 2012). Azzalini, A., Scarpa, B.
- **The elements of statistical learning** (Springer series in statistics, 2008). Hastie, T., Tibshirani, R., Friedman, J. → chapter 3
- **An Introduction to Statistical Learning, with applications in R** (Springer, 2013) G. James, D. Witten, T. Hastie and R. Tibshirani

Preliminary concepts
Illustrative examples

Forecasting, Goals, Planning

Forecasting is a common statistical task in business, where it helps to inform decisions about the scheduling of production, transportation and personnel, and provides a guide to long-term strategic planning. However, business forecasting is often done poorly, and is frequently confused with planning and goals. They are three different things.

- Forecasting: is about predicting the future as accurately as possible, given all of the information available, including historical data and knowledge of any future events that might impact the forecasts.
- Goals: are what you would like to have happen. Goals should be linked to forecasts and plans, but this does not always occur. Too often, goals are set without any plan for how to achieve them, and no forecasts for whether they are realistic.
- Planning: is a response to forecasts and goals. Planning involves determining the appropriate actions that are required to make your forecasts match your goals.

Forecasting, Goals, Planning

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organisations require short-term, medium-term and long-term forecasts, depending on the specific application.

- Short-term forecasts: are needed for the scheduling of personnel, production and transportation. As part of the scheduling process, forecasts of demand are often also required.
- Medium-term forecasts: are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment.
- Long-term forecasts: are used in strategic planning. Such decisions must take account of market opportunities, environmental factors and internal resources.

Determining what to forecast

In the early stages of a forecasting project, decisions need to be made about what should be forecast. For example, if forecasts are required for items in a manufacturing environment, it is necessary to ask whether forecasts are needed for:

- every product line, or for groups of products?
- every sales outlet, or for outlets grouped by region, or only for total sales?
- weekly data, monthly data or annual data?

It is also necessary to consider the forecasting horizon.
Will forecasts be required for one month in advance, for 6 months, or for ten years? Different types of models will be necessary, depending on what forecast horizon is most important.

Determining what to forecast

- How frequently are forecasts required? Forecasts that need to be produced frequently are better done using an automated system than with methods that require careful manual work.
- It is worth spending time talking to the people who will use the forecasts to ensure that you understand their needs, and how the forecasts are to be used, before embarking on extensive work in producing the forecasts.

Forecasting methods and data

The appropriate forecasting methods depend largely on what data are available.
If there are no data available, or if the data available are not relevant to the forecasts, then qualitative forecasting methods must be used. These methods are not purely guesswork: there are well-developed structured approaches to obtaining good forecasts without using historical data.
Quantitative forecasting can be applied when two conditions are satisfied:

- numerical information about the past is available;
- it is reasonable to assume that some aspects of the past patterns will continue into the future.

Forecasting process: phases

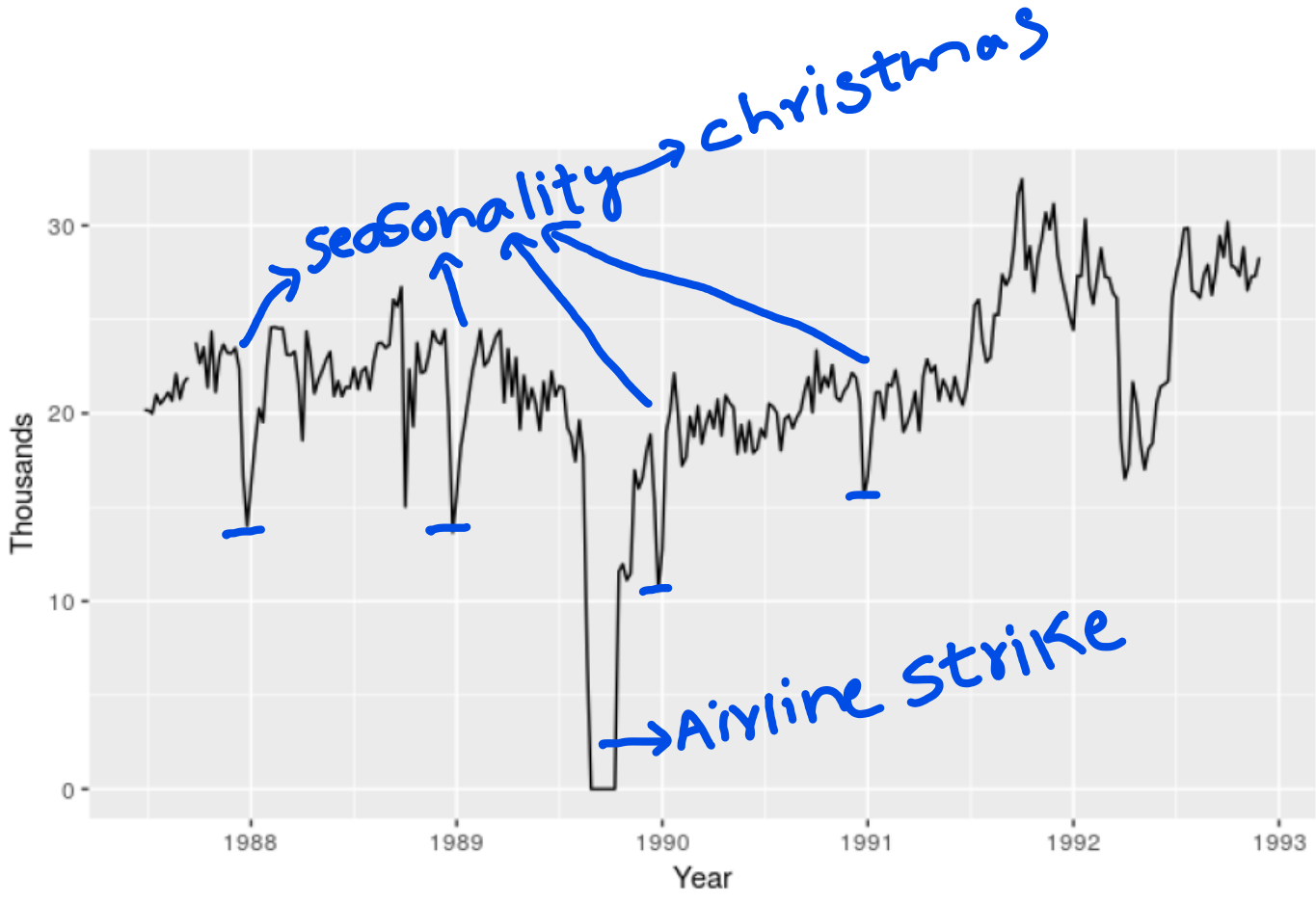
- Step 1: Problem definition Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organisation requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for future planning.
- Step 2: Gathering information There are always at least two kinds of information required: (a) statistical data, and (b) the accumulated expertise of the people who collect the data and use the forecasts. Often, it will be difficult to obtain enough historical data to be able to fit a good statistical model. In that case, the judgmental forecasting methods can be used. Occasionally, old data will be less useful due to structural changes in the system being forecast; then we may choose to use only the most recent data.

However, remember that good statistical models will handle evolutionary changes in the system;
don't throw away good data unnecessarily!

Forecasting process: phases

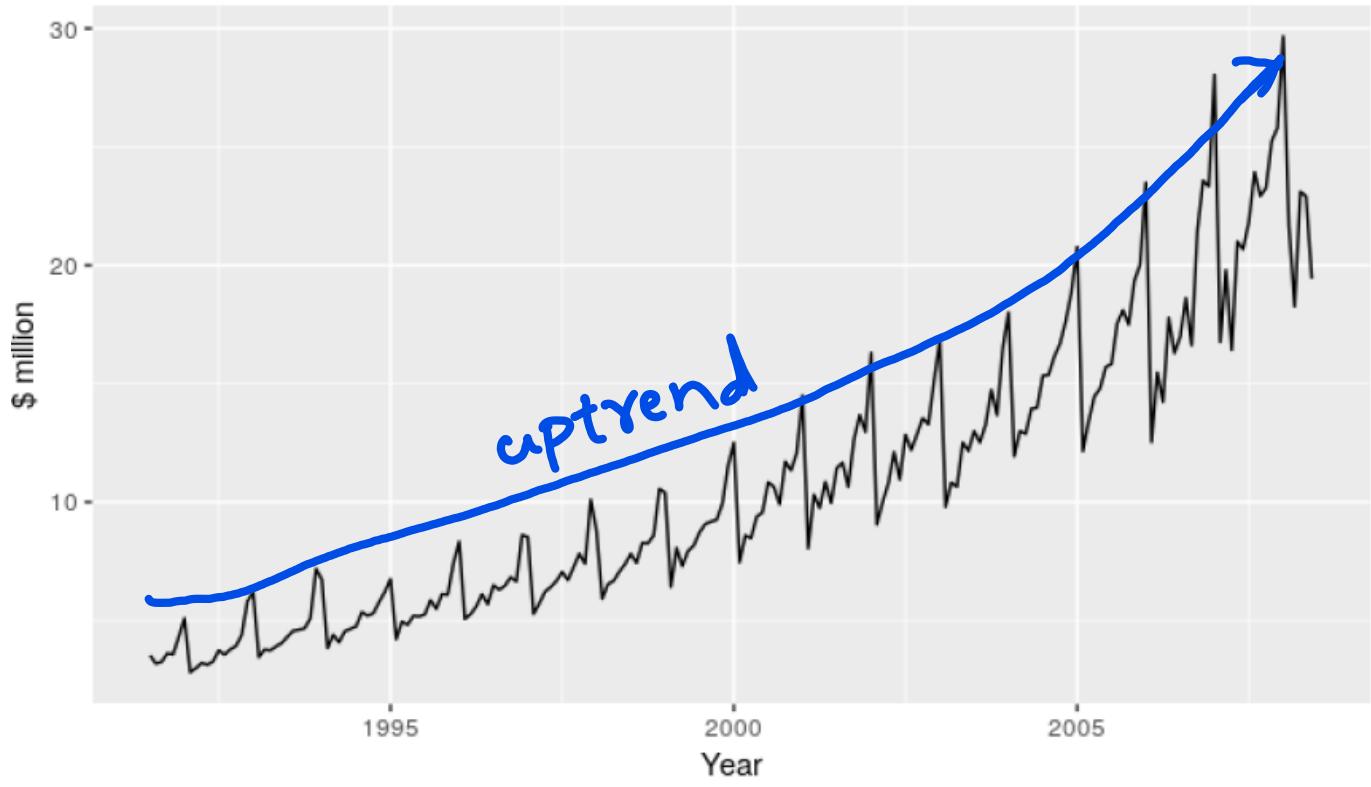
- Step 3: Preliminary (exploratory) analysis. Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis?
- Step 4: Choosing and fitting models. The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the forecasts are to be used. It is common to compare two or three potential models.
- Step 5: Using and evaluating a forecasting model. Once a model has been selected and its parameters estimated, the model is used to make forecasts. The performance of the model can only be properly evaluated after the data for the forecast period have become available.

Time series analysis: useful graphs



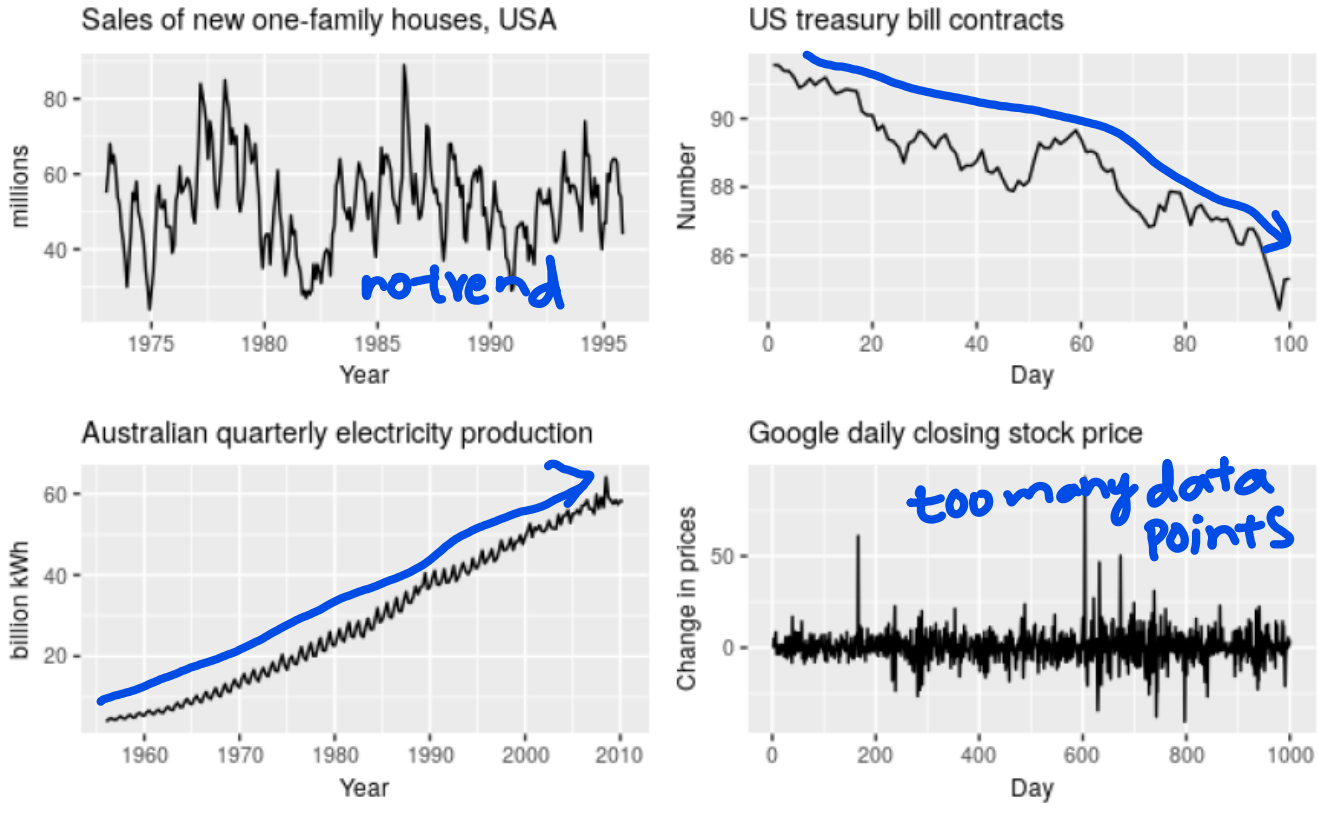
Weekly economy passenger load on an Airline Company

Time series analysis: useful graphs



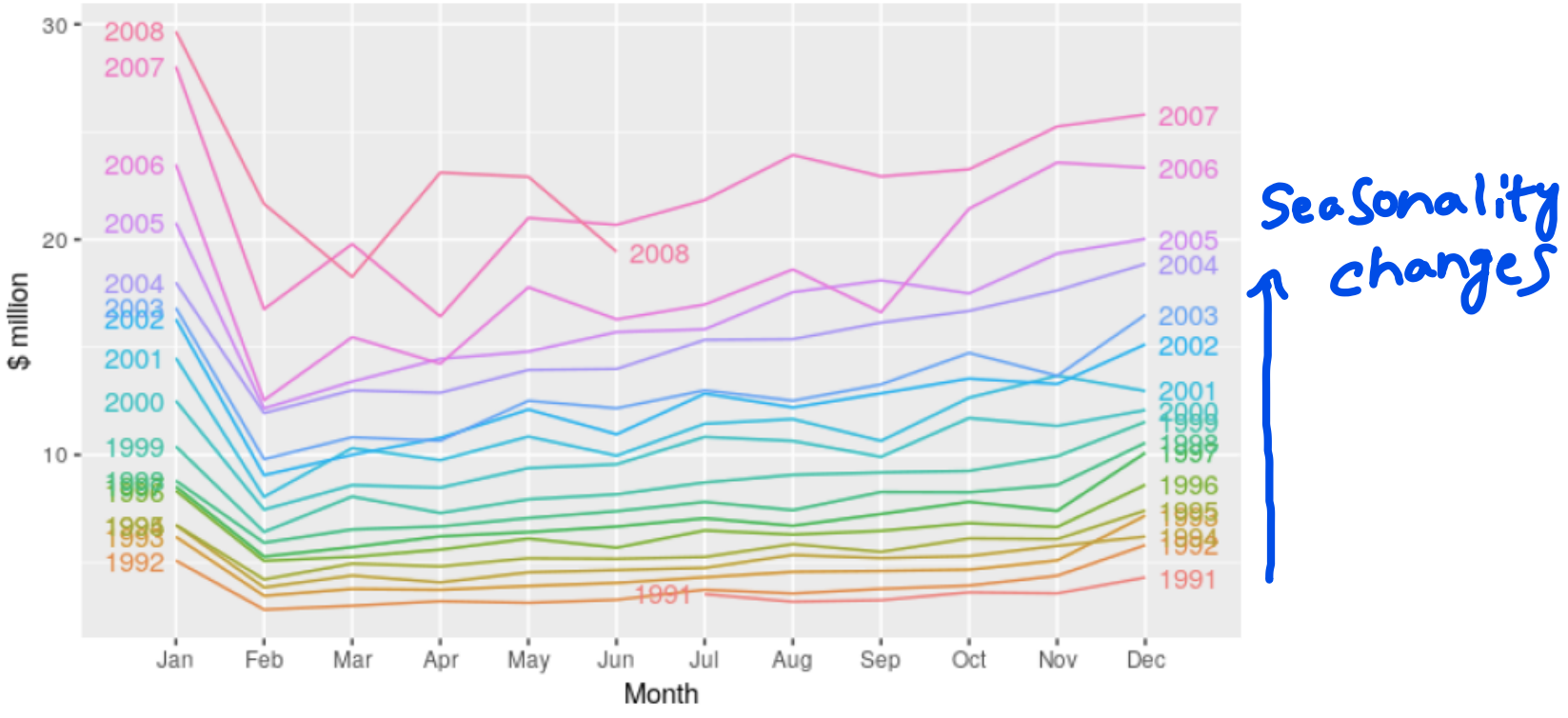
Monthly sales of antidiabetic drugs

Time series analysis: useful graphs



4 time series with different patterns

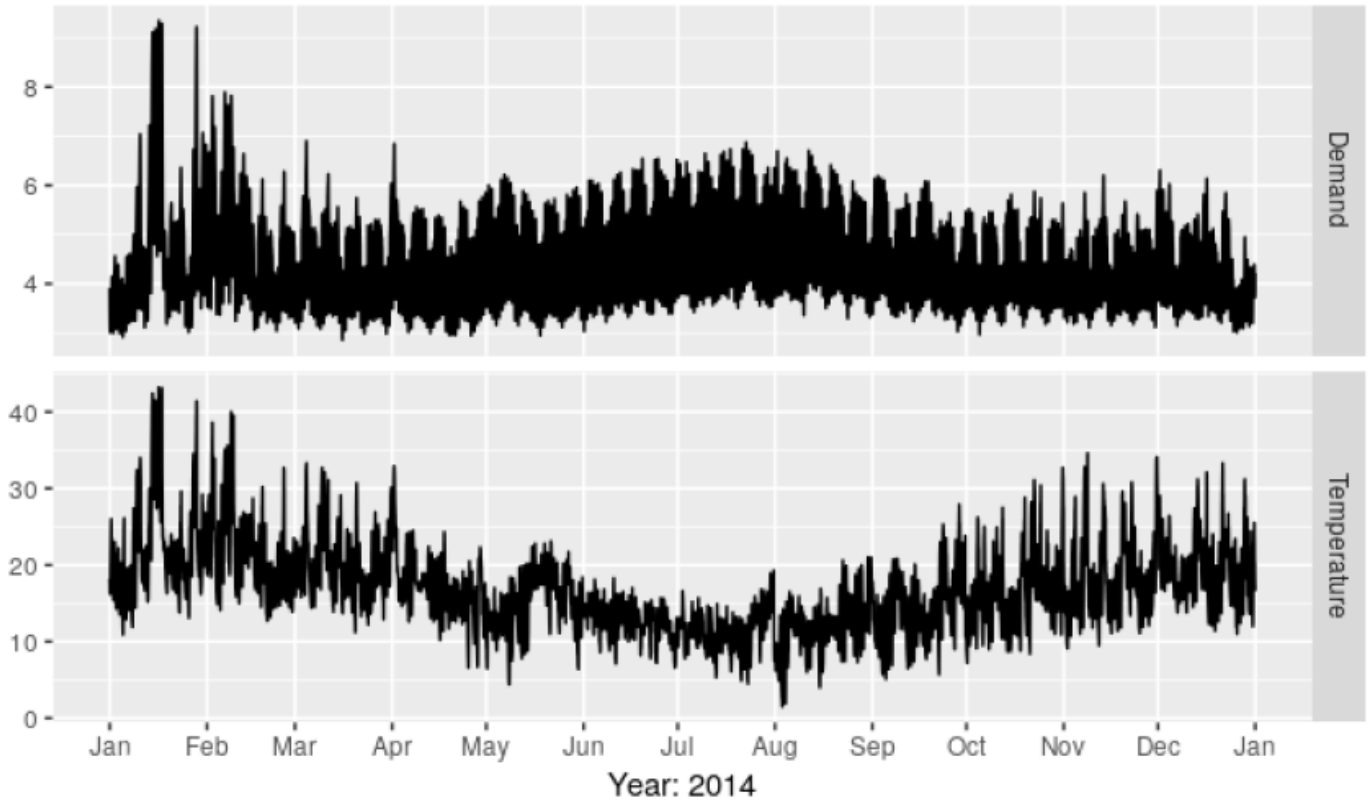
Time series analysis: useful graphs



Monthly sales of antidiabetic drugs: 'seasonal plot'

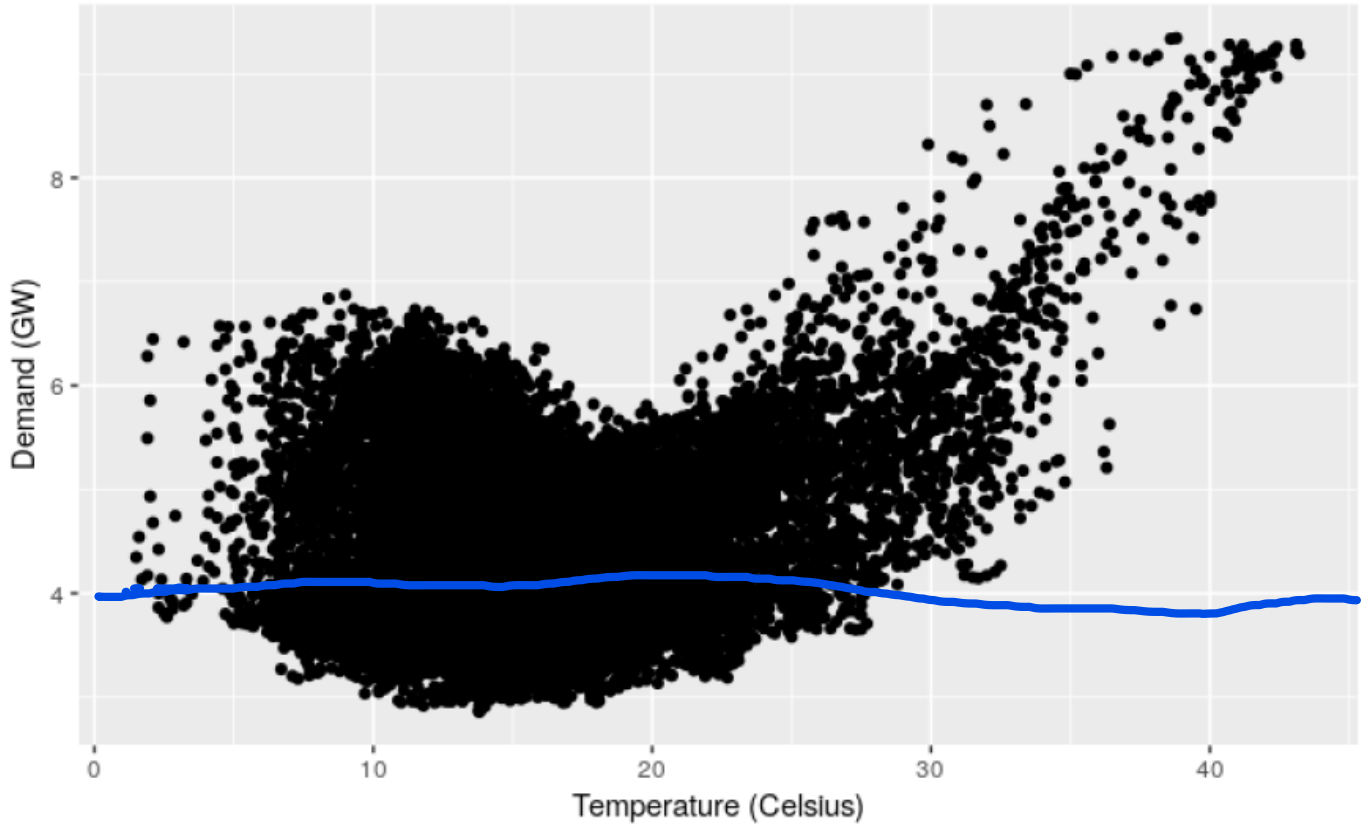
Time series analysis: useful graphs

electricity demand increased more with decrease in temp rather than increase



Electricity demand and temperatures in Australia (year 2014): relationship between series

Time series analysis: useful graphs



Electricity demand and temperatures in Australia (year 2014): relationship between series

Autocorrelation

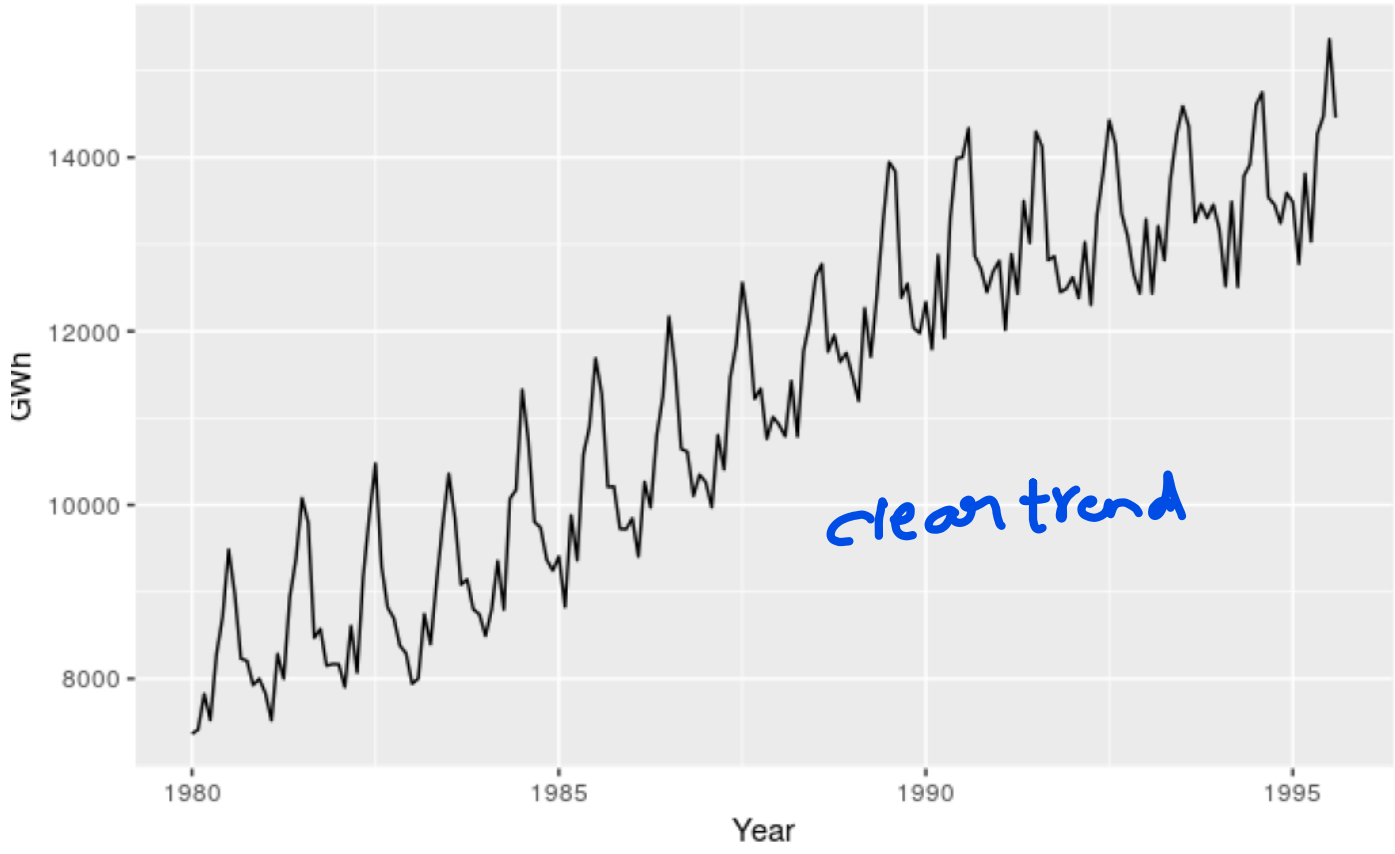
Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between lagged values of a time series.

$$r_k = \frac{\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t}^{T}(Y_t - \bar{Y})^2}$$

lag steps

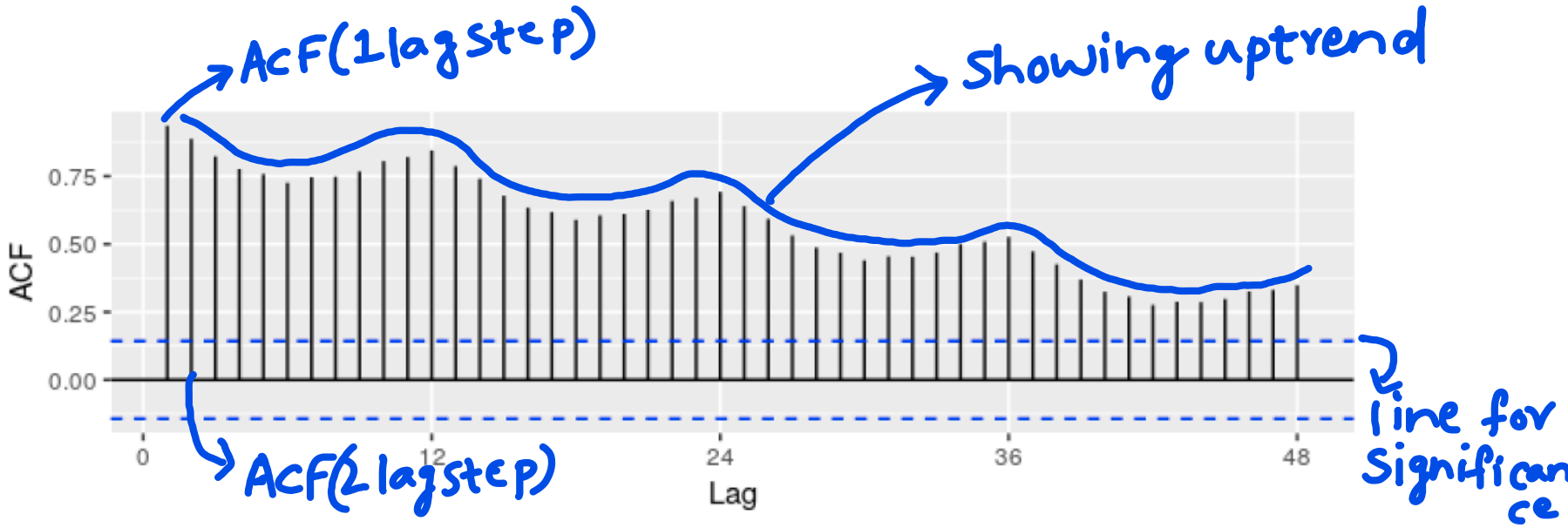
The autocorrelation coefficients are plotted to show the autocorrelation function or ACF.
The plot is also known as a correlogram.

Autocorrelation



Monthly electricity demand in the period 1980-1995

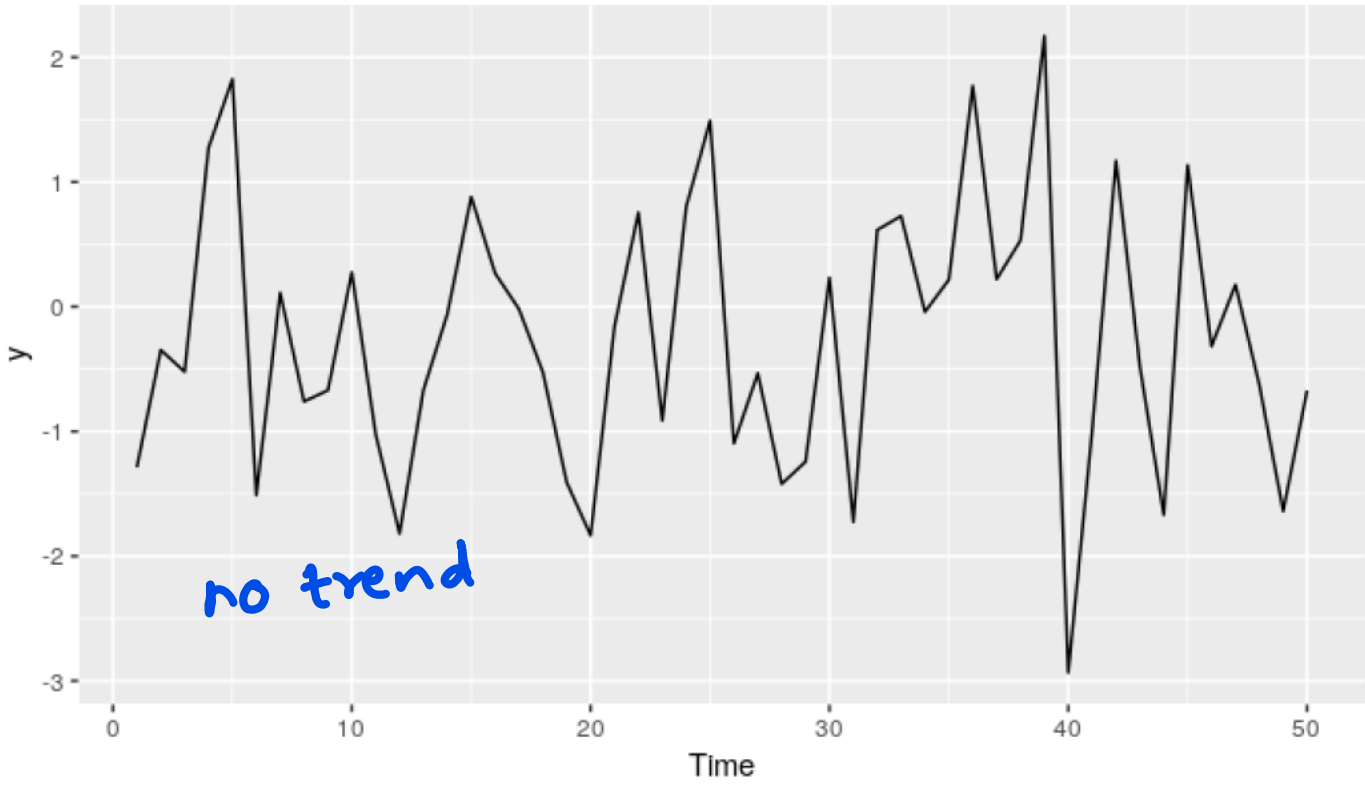
Autocorrelation



Monthly electricity demand in the period 1980-1995: correlogram

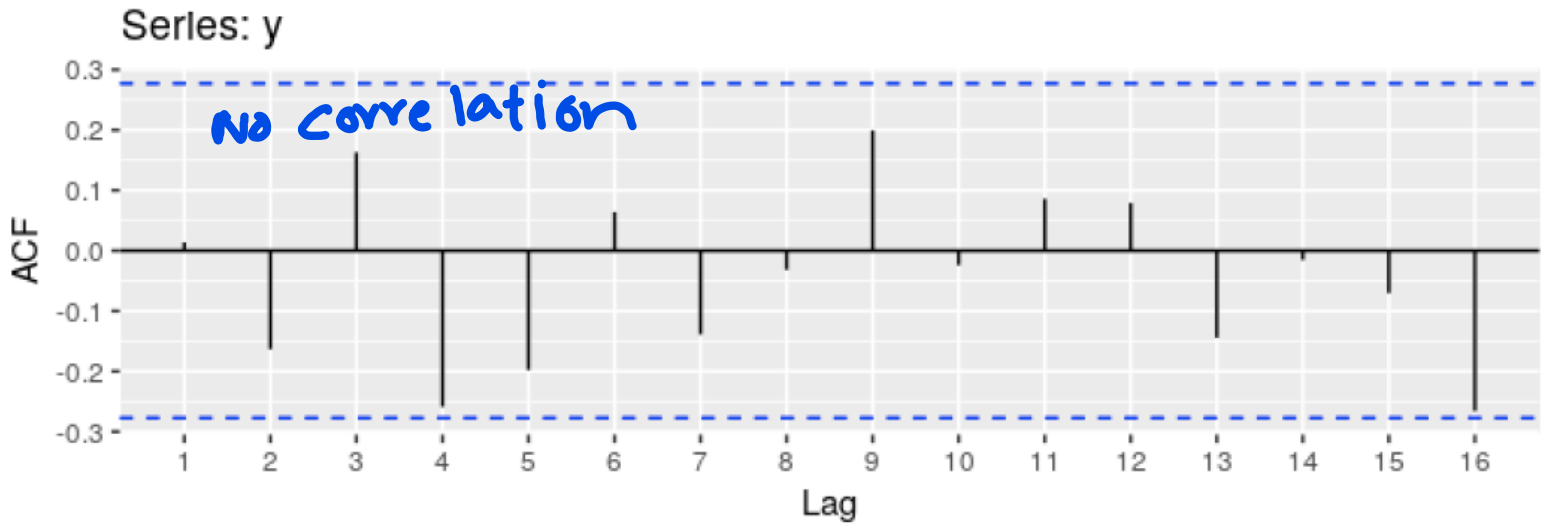
ACF range between -1 and +1

Autocorrelation



Time series that show no autocorrelation are called *White Noise* process

Autocorrelation



White Noise process: autocorrelation