**Time Series Forecasting**

Table of Contents

[1. Forecasting, goals and planning 4](#_Toc162857994)

[1.1. Forecasting 4](#_Toc162857995)

[1.2. Goals 4](#_Toc162857996)

[1.3. Planning 4](#_Toc162857997)

[2. One-step and Multi-step Forecasting 5](#_Toc162857998)

[2.1. One-Step Forecasting 5](#_Toc162857999)

[2.2. Multi-Step Forecasting 5](#_Toc162858000)

[3. Autocorrelation | Correlogram 6](#_Toc162858001)

[4. Trend and seasonality in ACF plots 6](#_Toc162858002)

[5. Transformations and Adjustments 7](#_Toc162858003)

[6. Decomposition 7](#_Toc162858004)

[6.1. Transformations and Adjustments 8](#_Toc162858005)

[**1.** **Calendar Adjustments** 9](#_Toc162858006)

[**2.** **Population Adjustments** 9](#_Toc162858007)

[**3.** **Inflation Adjustments** 9](#_Toc162858008)

[6.2. Decomposition Methods 9](#_Toc162858009)

[7. Some simple forecasting methods | Simple benchmark models 10](#_Toc162858010)

[a. Mean method 10](#_Toc162858011)

[b. Naïve method 10](#_Toc162858012)

[c. Seasonal naïve method 11](#_Toc162858013)

[d. Drift method 12](#_Toc162858014)

[e. Examples 13](#_Toc162858015)

[8. Residuals 14](#_Toc162858016)

[9. Forecast Distributions 15](#_Toc162858017)

[a. Prediction Interval 15](#_Toc162858018)

[b. One step prediction interval 16](#_Toc162858019)

[c. Multi step prediction interval 16](#_Toc162858020)

[10. Evaluating point forecast accuracy 17](#_Toc162858021)

[RMSE (Root Mean Squared Error) 17](#_Toc162858022)

[MAE (Mean Absolute Error) 18](#_Toc162858023)

[MAPE (Mean Absolute Percentage Error) 18](#_Toc162858024)

[11. Evaluating distribution forecast accuracy 18](#_Toc162858025)

[12. Time series Cross-validation 18](#_Toc162858026)

**Time Series Forecasting**

# Forecasting, goals and planning

## Forecasting

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

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

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 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***

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***

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***

Long-term forecasts are used in strategic planning. Such decisions must take account of market opportunities, environmental factors and internal resources.

# One-step and Multi-step Forecasting

Time series forecasting involves predicting future values based on historical data. One-step and multi-step forecasting are two common approaches to making predictions in time series analysis.

## One-Step Forecasting

In one-step forecasting, the goal is to predict the value of the time series only one step into the future.

* **Process:** The model is trained using historical data, and then it is used to predict the next value in the sequence. After making this prediction, the actual observed value for that time step becomes part of the training data for the next prediction.
* **Use Case:** This approach is suitable when the primary interest is in short-term predictions or when the model needs to be updated frequently with new data.

## Multi-Step Forecasting

Multi-step forecasting involves predicting multiple future values in the time series.

* **Process:** The model is trained using historical data, and then it is used to predict several future values at once. These predicted values are then used as inputs for subsequent predictions. This process is repeated until the desired forecast horizon is reached.
* **Use Case:** Multi-step forecasting is useful when a longer-term prediction is required, and it helps in understanding the trend of the time series over a more extended period.

In summary, one-step forecasting focuses on predicting the very next point in the time series, while multi-step forecasting involves predicting multiple future points in the sequence. The choice between these approaches depends on the specific requirements of the forecasting task and the nature of the data.

# **Autocorrelation | Correlogram**

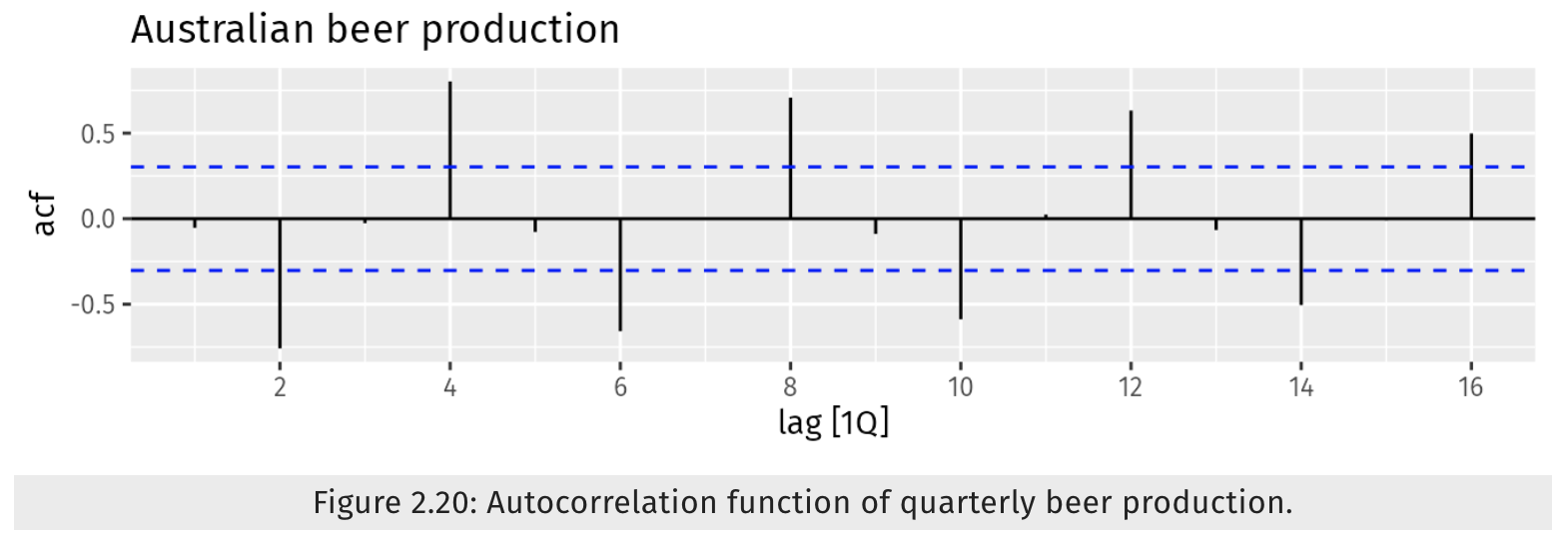
[Autocorrelation | Correlogram](https://otexts.com/fpp3/acf.html#acf)

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**.

There are several autocorrelation coefficients, corresponding to each panel in the lag plot. For example, measures the relationship between and , measures the relationship between and , and so on.

The value of can be written as

where is the length of the time series. The autocorrelation coefficients make up the **Auto Correlation Function or ACF**.



In this graph:

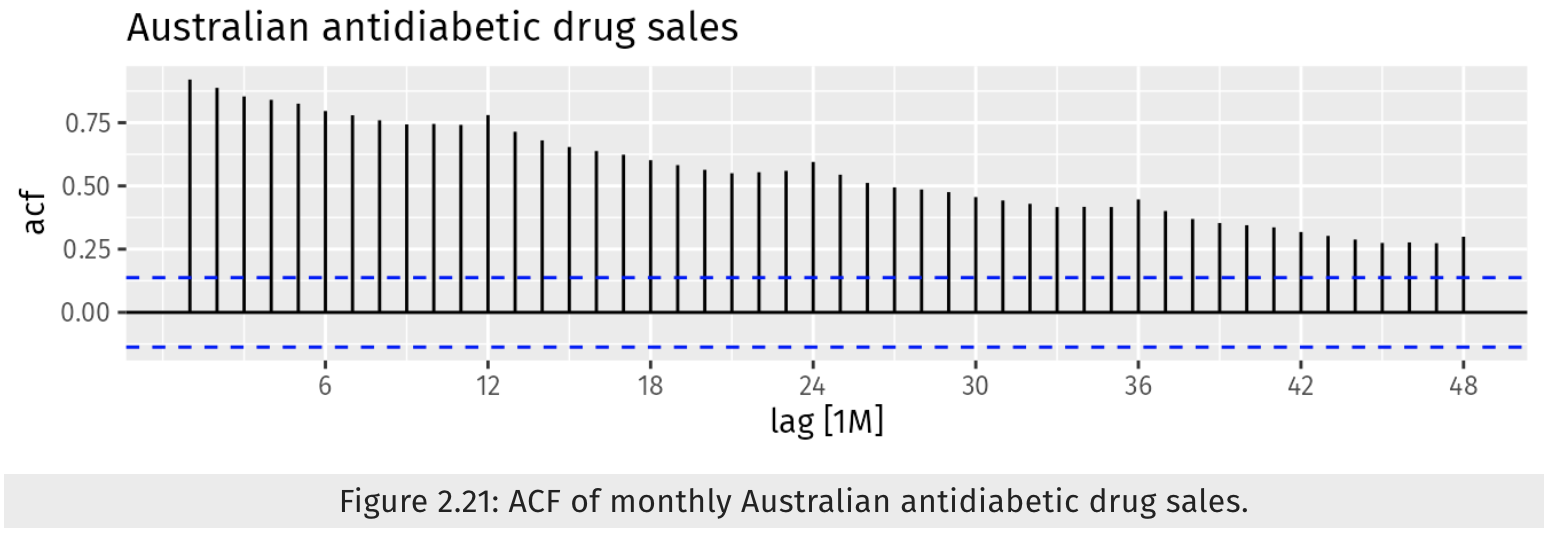
1. is higher than for the other lags. This is due to the seasonal pattern in the data: the peaks tend to be four quarters apart and the troughs tend to be four quarters apart.
2. is more negative than for the other lags because troughs tend to be two quarters behind peaks.
3. The dashed blue lines indicate whether the correlations are significantly different from zero

# **Trend and seasonality in ACF plots**

When data have a trend, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also nearby in value. So the ACF of a trended time series tends to have positive values that slowly decrease as the lags increase.

When data are seasonal, the autocorrelations will be larger for the seasonal lags (at multiples of the seasonal period) than for other lags.

When data are both trended and seasonal, you see a combination of these effects.



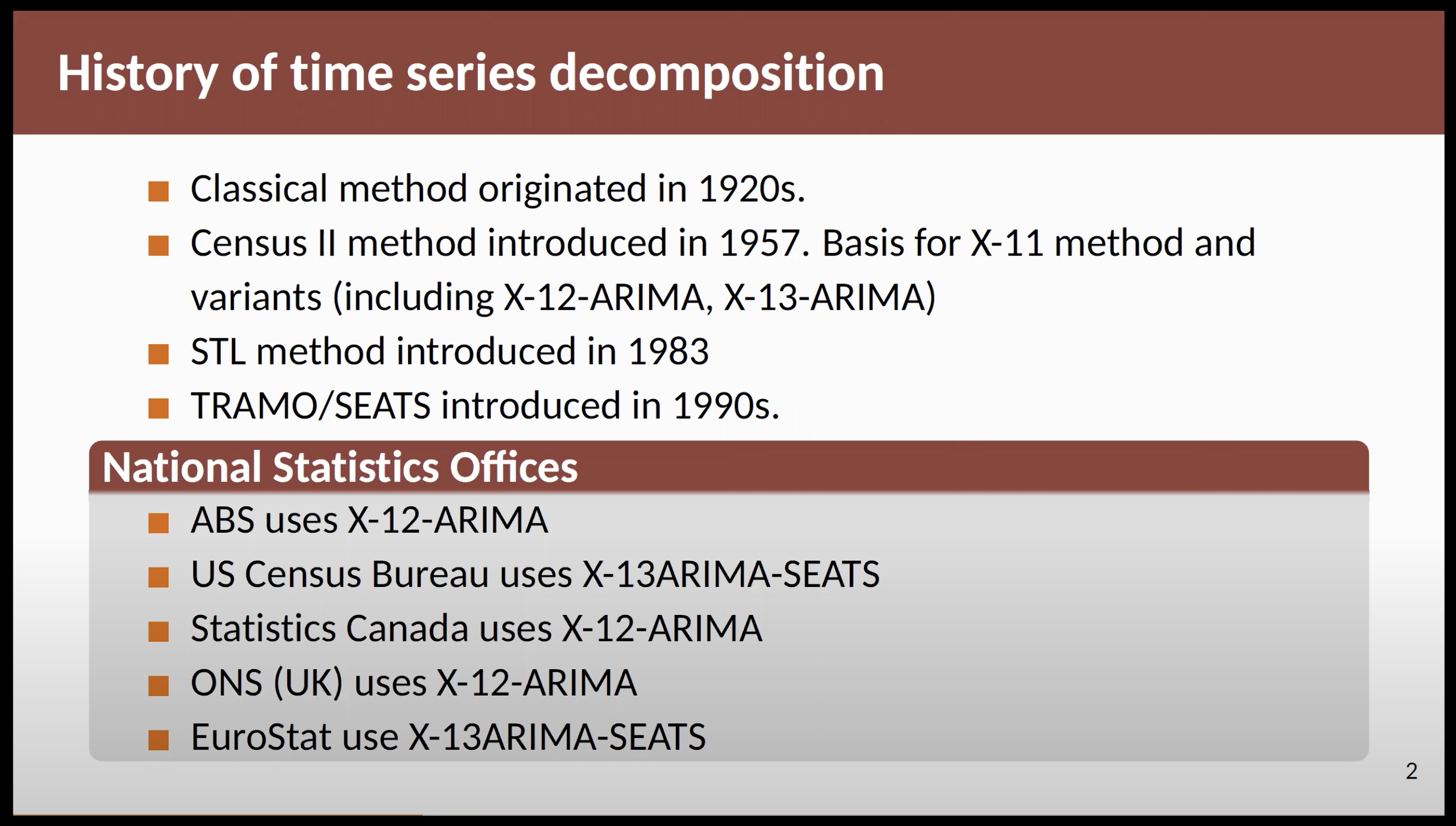
# Transformations and Adjustments

# Decomposition

Decomposition in time series forecasting is a technique for breaking down a complex time series signal into its constituent components:

1. **Trend**: The long-term underlying tendency of the data, indicating its overall direction (increasing, decreasing, or stable). Generally the trend and cycle is combined together into a single trend-cycle component (often just called the trend for simplicity)
2. **Seasonality**: The recurring patterns within the data that occur at regular intervals, such as daily, weekly, monthly, or yearly cycles.
3. **Remainder**: The random fluctuations or noise in the data that cannot be explained by the trend or seasonality.

Extracting these components from a time series is done to help improve understanding of the time series and it can also improve forecast accuracy.



**When decomposing a time series, it is sometimes helpful to first transform or adjust the series in order to make the decomposition (and later analysis) as simple as possible.**

## 6.1. Transformations and Adjustments

Transformations and adjustments are important techniques used in time series forecasting to improve the quality of the data or to make it more suitable for modeling.

|  |  |  |
| --- | --- | --- |
| Aspect | Transformations | Adjustments |
| What is it? | Applying mathematical functions to original data to achieve desirable properties or make it more suitable for analysis/modeling. | Modifying original data to account for known factors or biases affecting analysis or forecasting. |
| How is it done? | Common transformations include   * Logarithmic * Square root * Box-Cox transformations   Functions are applied to each observation in the time series. | Types include:   * Calendar adjustments * Population adjustments * Inflation adjustments. |
| When is it needed? | When data exhibits non-constant variance or non-linearity. | When raw data contains systematic biases or confounding factors affecting accuracy of forecasts. |
| Merits | - Stabilizes variance  - Linearizes relationships  - Comparability across different time periods, regions, or demographic groups. | Improves accuracy by removing biases or confounding effects. |
| Demerits | - Challenges in interpretation due to transformed values not directly corresponding to original scale.  - Risk of over-transformation distorting relationships. | - May require additional data sources or information, which may not always be readily available.  - Reliance on assumptions about data or underlying processes, which could introduce errors if assumptions are violated. |

### **Calendar Adjustments**

These are used to remove variation in seasonal data due to calendar effects. For example, the number of days in each month can vary, causing variation in monthly data. By considering average daily production instead of total monthly production, we can remove the variation due to different month lengths.

### **Population Adjustments**

These are used when data are affected by population changes. Instead of considering the total, we consider the data per person (or per thousand or million people). This helps in interpreting whether there have been real increases in a variable, or whether the increases are due to population increases.

### **Inflation Adjustments**

These are used when data are affected by the value of money. For example, the average cost of a new house will have increased over the last few decades due to inflation. Financial time series are usually adjusted so that all values are stated in dollar values from a particular year.

## 6.2. Decomposition Methods

1. Classical Decomposition [https://otexts.com/fpp3/classical-decomposition.html]
   1. Additive decomposition
   2. Multiplicative decomposition
2. ARIMA
3. Exponential smoothing
4. X-11 Method
5. X-12-ARIMA Method
6. X-13-ARIMA Method
7. Seasonal-Trend decomposition using Loess (STL)

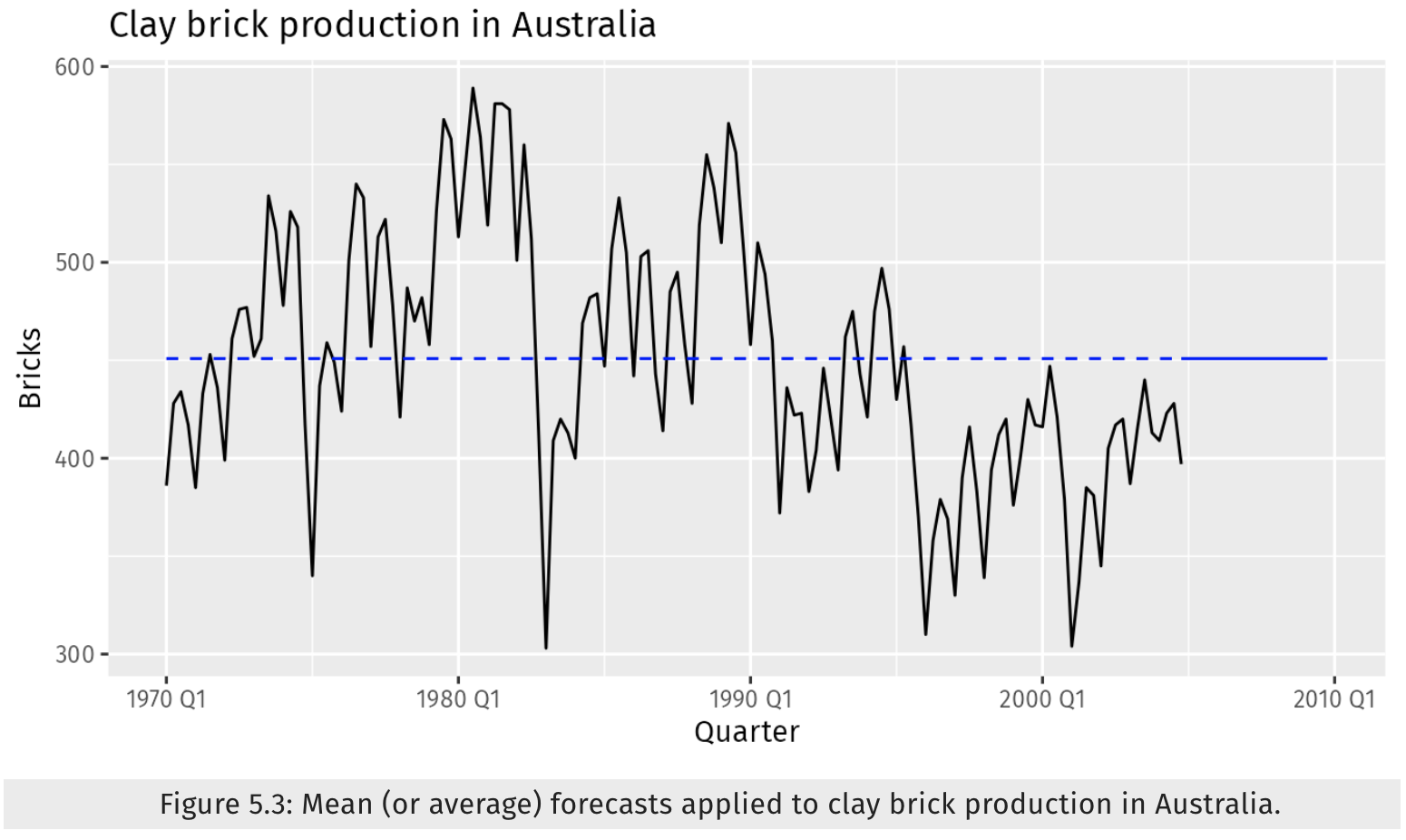
This method uses a localized smoothing technique to estimate the trend and seasonality components.

1. TRAMO
2. SEATS

# Some simple forecasting methods | Simple benchmark models

## Mean method

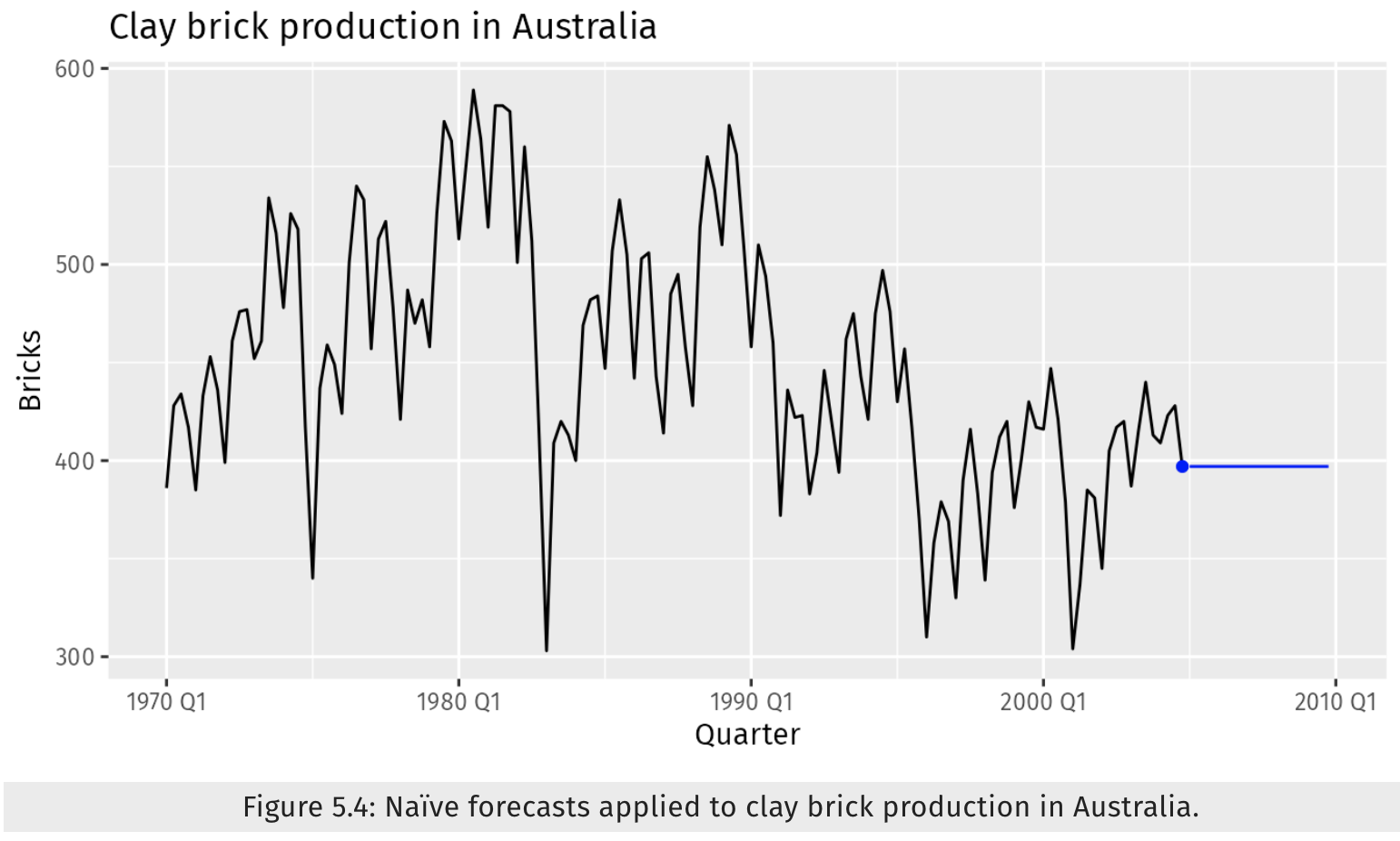
Here, the forecasts of all future values are equal to the average (or “mean”) of the historical data.



## Naïve method

For naïve forecasts, we simply set all forecasts to be the value of the last observation. That is,

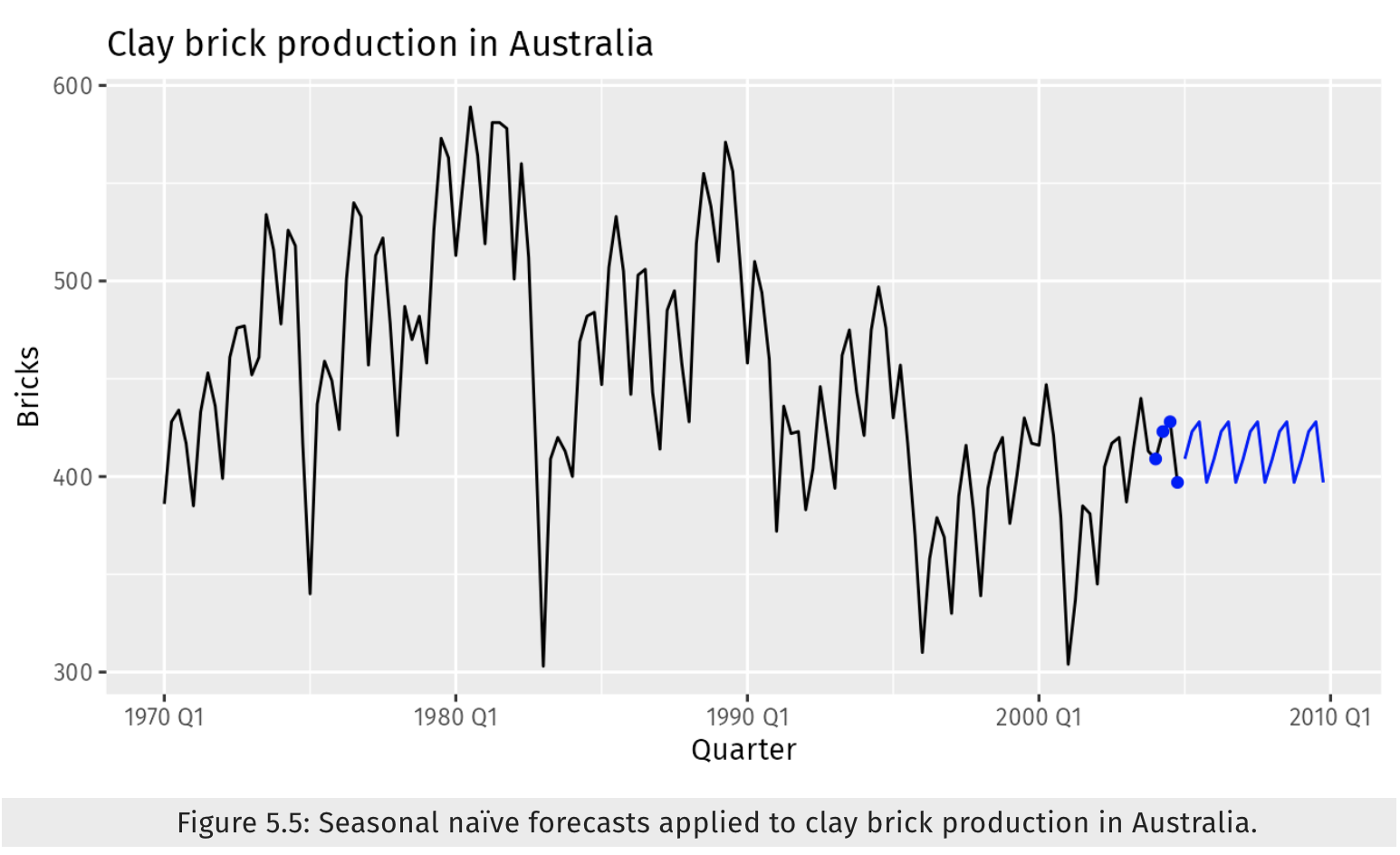
This method works remarkably well for many economic and financial time series. Because a naïve forecast is optimal when data follow a random walk these are also called **random walk forecasts**.



## Seasonal naïve method

A similar method is useful for highly seasonal data. In this case, we set each forecast to be equal to the last observed value from the same season (e.g., the same month of the previous year). Formally, the forecast for time is written as

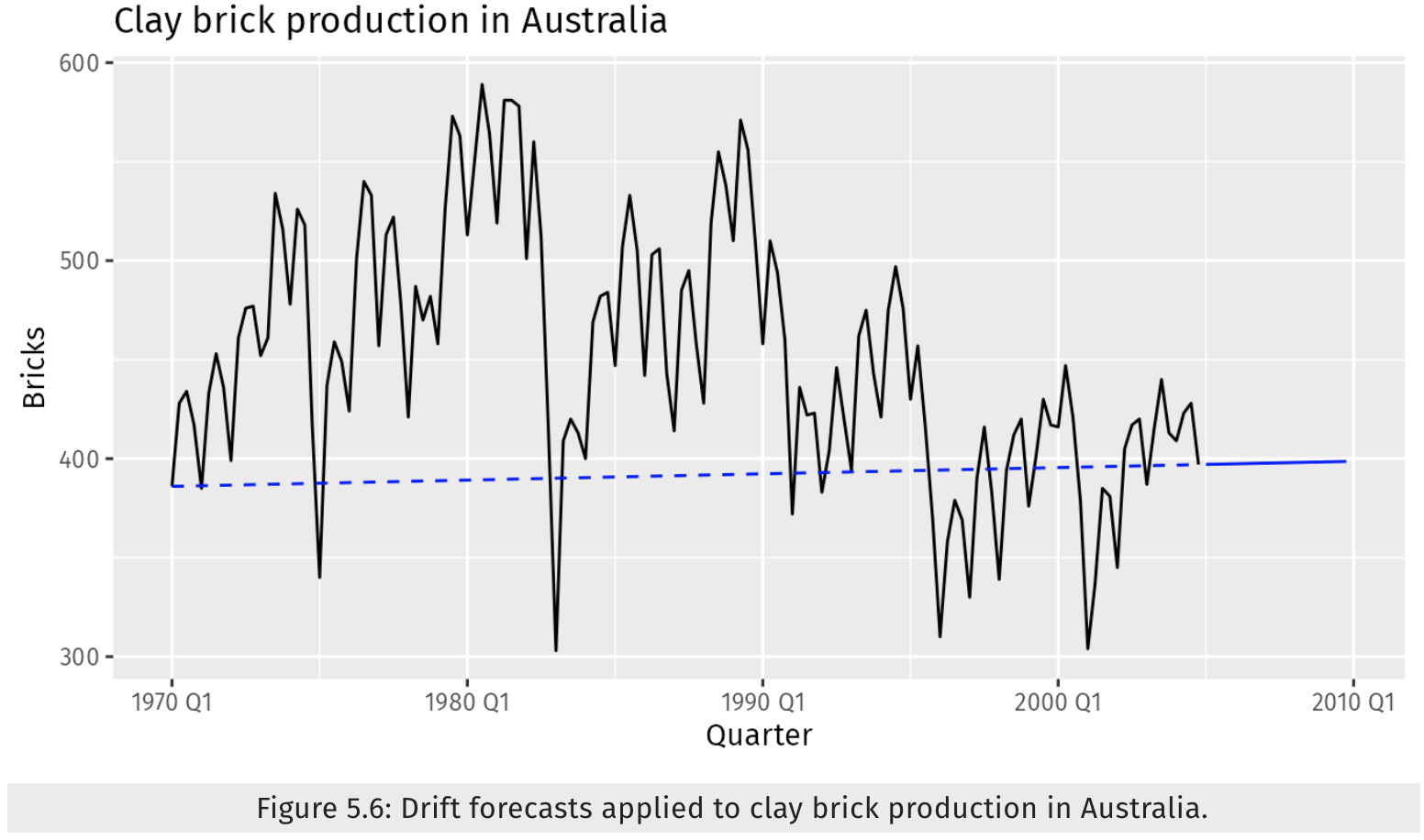
This looks more complicated than it really is. For example, with monthly data, the forecast for all future February values is equal to the last observed February value. With quarterly data, the forecast of all future Q2 values is equal to the last observed Q2 value (where Q2 means the second quarter). Similar rules apply for other months and quarters, and for other seasonal periods.



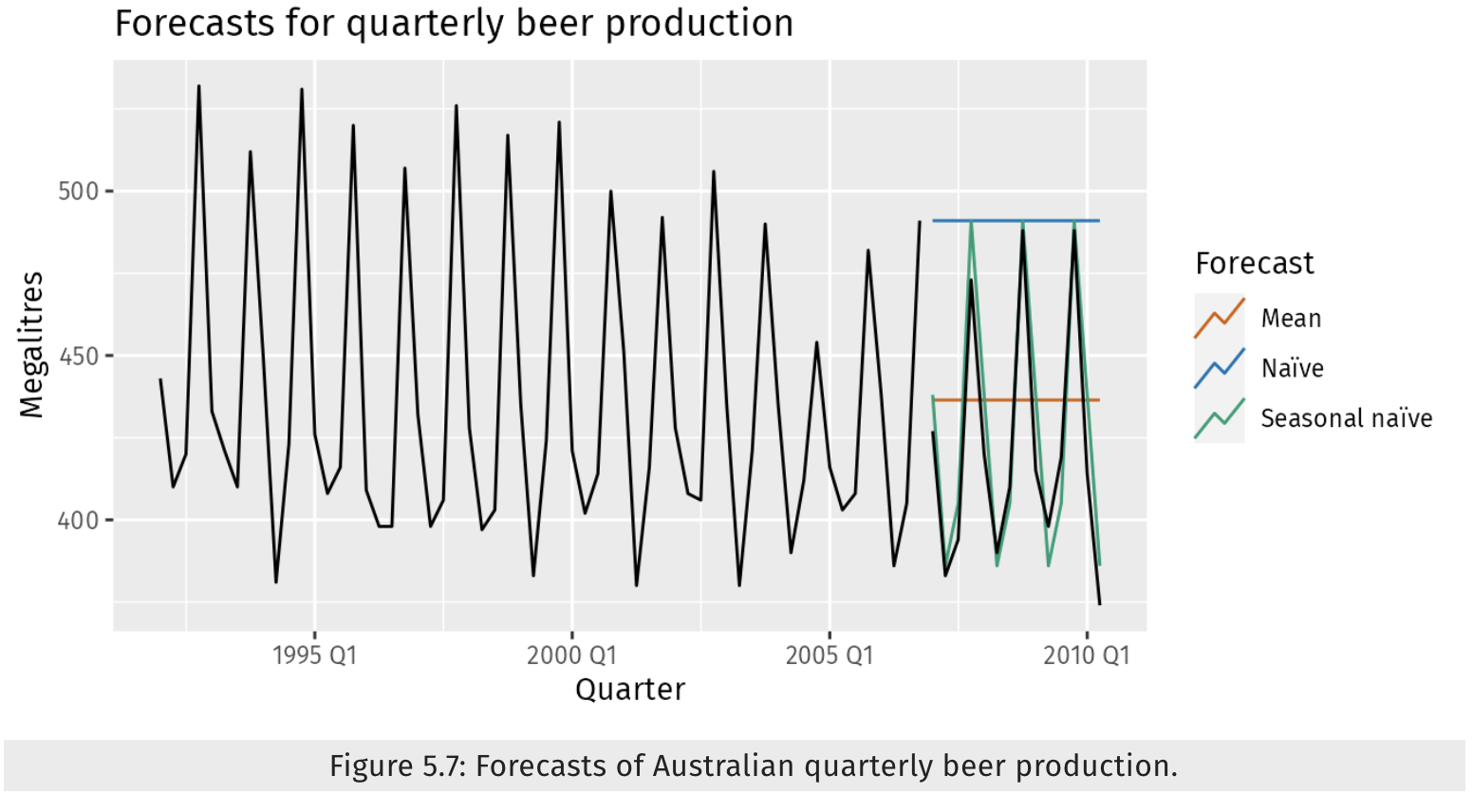
## Drift method

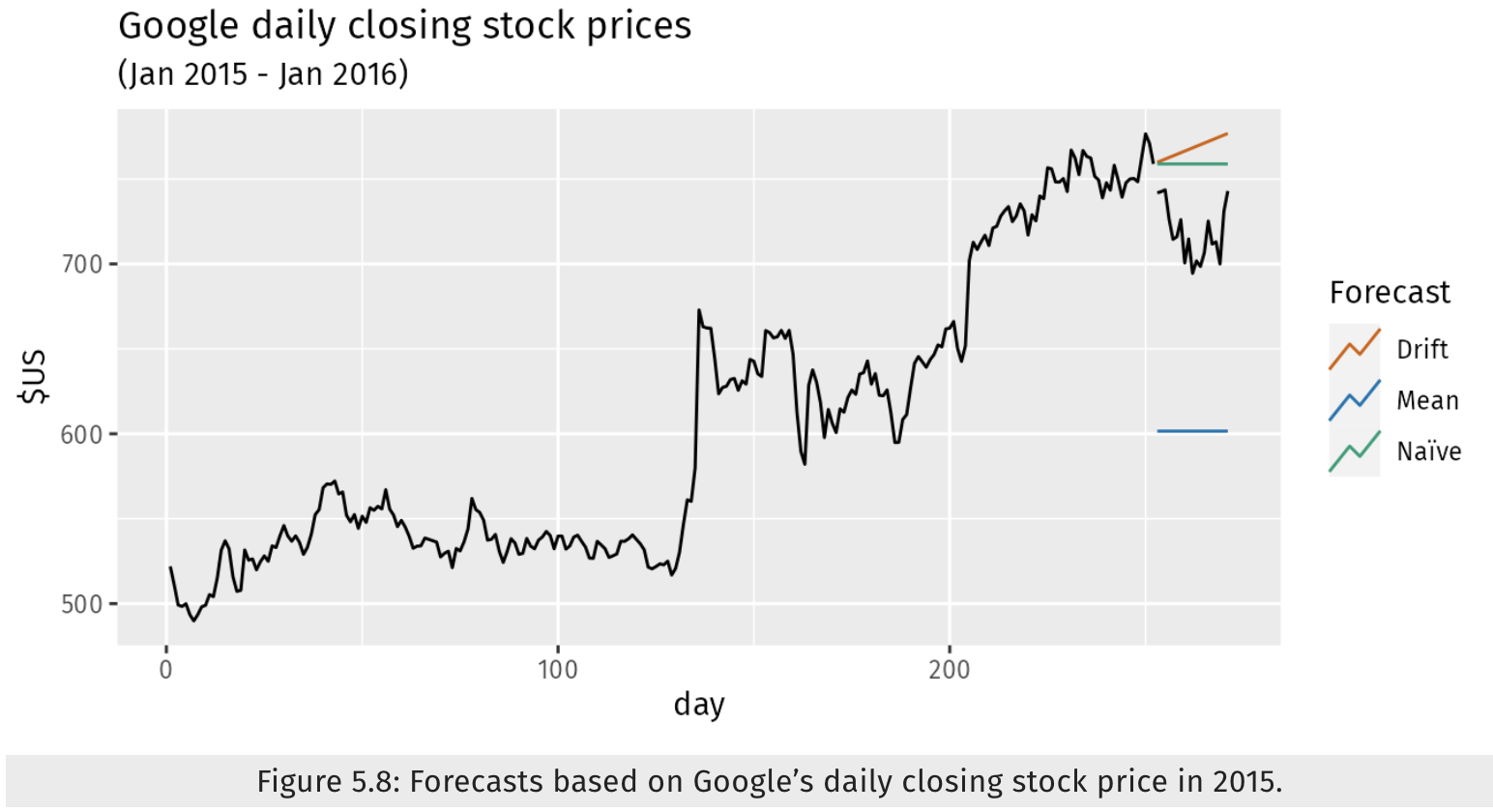
A variation on the naïve method is to allow the forecasts to increase or decrease over time, where the amount of change over time (called the **drift**) is set to be the average change seen in the historical data.

This is equivalent to drawing a line between the first and last observations, and extrapolating it into the future.

****

## Examples





**Note**: Sometimes one of these simple methods will be the best forecasting method available; but in many cases, these methods will serve as benchmarks rather than the method of choice. That is, any forecasting methods we develop will be compared to these simple methods to ensure that the new method is better than these simple alternatives. If not, the new method is not worth considering.

# Residuals

The “residuals” in a time series model are what is left over after fitting a model. The residuals are equal to the difference between the observations and the corresponding fitted values.

If a transformation has been used in the model, then it is often useful to look at residuals on the transformed scale. We call these “**innovation residuals**”. For example, suppose we modelled the logarithms of the data . Then the innovation residuals are given by  whereas the regular residuals are given by . If no transformation has been used then the innovation residuals are identical to the regular residuals, and in such cases we will simply call them “residuals”.

Residuals are useful in checking whether a model has adequately captured the information in the data. For this purpose, we use innovation residuals.

If patterns are observable in the innovation residuals, the model can probably be improved.

A good forecasting method will yield innovation residuals with the following properties:

1. The innovation residuals are uncorrelated. If there are correlations between innovation residuals, then there is information left in the residuals which should be used in computing forecasts.
2. The innovation residuals have zero mean. If they have a mean other than zero, then the forecasts are biased.

Any forecasting method that does not satisfy these properties can be improved. However, that does not mean that forecasting methods that satisfy these properties cannot be improved.

If either of these properties is not satisfied, then the forecasting method can be modified to give better forecasts. Adjusting for bias is easy: if the residuals have mean , then simply add to all forecasts and the bias problem is solved. **Fixing the correlation problem is harder.**

In addition to these essential properties, it is useful (but not necessary) for the residuals to also have the following two properties.

1. The innovation residuals have constant variance. This is known as “**homoscedasticity**”.
2. The innovation residuals are normally distributed.

These two properties make the calculation of prediction intervals easier. However, a forecasting method that does not satisfy these properties cannot necessarily be improved. Sometimes applying a **Box-Cox transformation** may assist with these properties, but otherwise there is usually little that you can do to ensure that your innovation residuals have constant variance and a normal distribution. Instead, an alternative approach to obtaining prediction intervals is necessary.

# Forecast Distributions

We express the uncertainty in our forecasts using a **probability distribution**. It describes the probability of observing possible future values using the fitted model. The point forecast is the mean of this distribution. Most time series models produce normally distributed forecasts.

## Prediction Interval

A prediction interval gives an interval within which we expect to lie with a specified probability. For example, assuming that distribution of future observations is normal, a 95% prediction interval for the - step forecast is

where is an estimate of the standard deviation of the - step forecast distribution.

More generally, a prediction interval can be written as

where the multiplier depends on the coverage probability.

Value of for a range of coverage probabilities assuming a normal forecast distribution.

|  |  |
| --- | --- |
| Percentage | Multiplier |
| 50 | 0.67 |
| 55 | 0.76 |
| 60 | 0.84 |
| 65 | 0.93 |
| 70 | 1.04 |
| 75 | 1.15 |
| 80 | 1.28 |
| 85 | 1.44 |
| 90 | 1.64 |
| 95 | 1.96 |
| 96 | 2.05 |
| 97 | 2.17 |
| 98 | 2.33 |
| 99 | 2.58 |

The value of prediction intervals is that they express the uncertainty in the forecasts. If we only produce point forecasts, there is no way of telling how accurate the forecasts are. However, if we also produce prediction intervals, then it is clear how much uncertainty is associated with each forecast. For this reason, point forecasts can be of almost no value without the accompanying prediction intervals.

## One step prediction interval

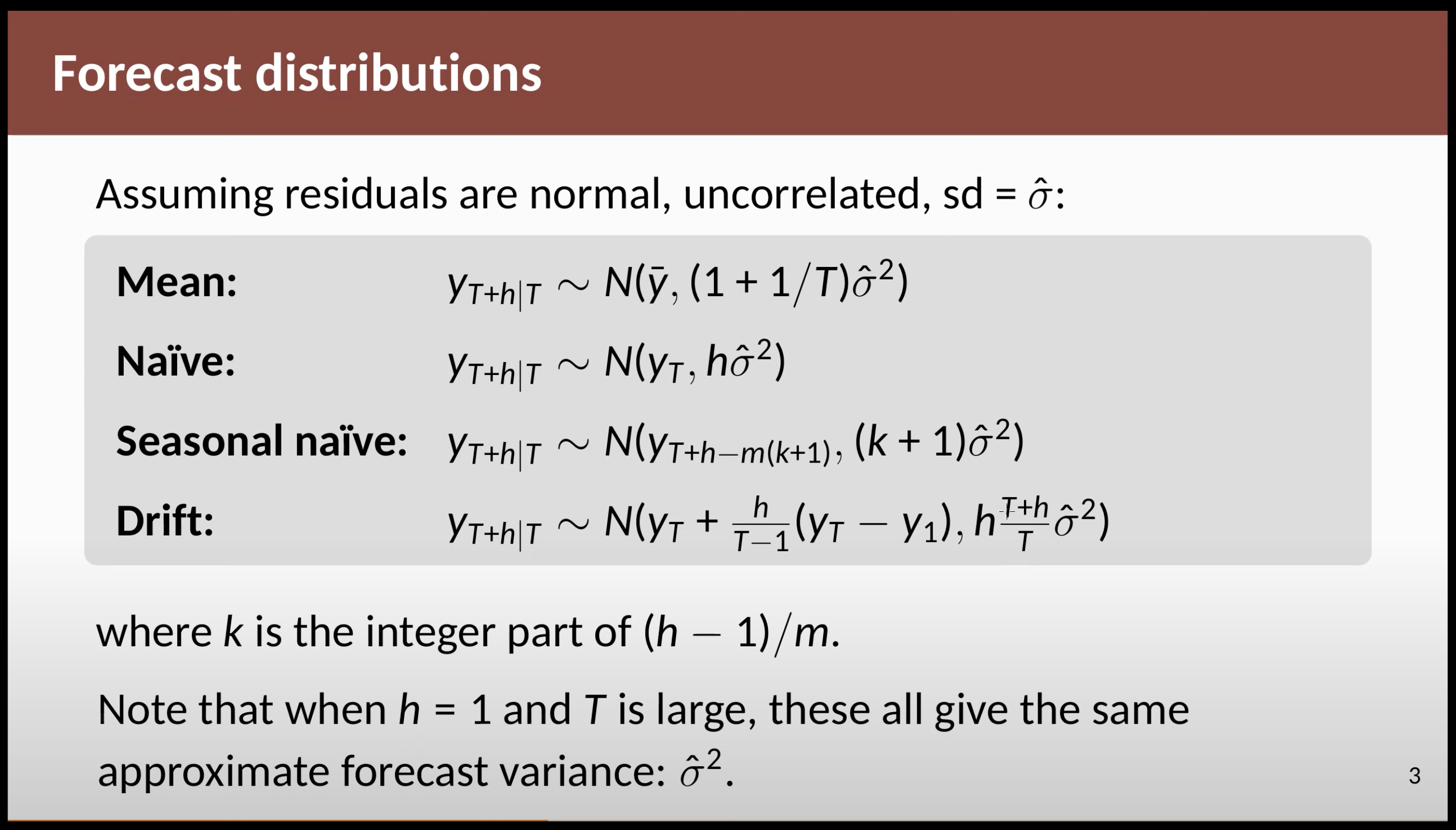
When forecasting one step ahead, the standard deviation of the forecast distribution can be estimated using the standard deviation of the residuals given by

where  is the number of parameters estimated in the forecasting method, and  is the number of missing values in the residuals.

## Multi step prediction interval

A common feature of prediction intervals is that they usually increase in length as the forecast horizon increases. The further ahead we forecast, the more uncertainty is associated with the forecast, and thus the wider the prediction intervals. That is,  usually increases with (although there are some non-linear forecasting methods which do not have this property).

To produce a prediction interval, it is necessary to have an estimate of . As already noted, for one-step forecasts , Equation provided above is a good estimate of the forecast standard deviation . For multi-step forecasts, a more complicated method of calculation is required. These calculations assume that the residuals are uncorrelated.



# Evaluating point forecast accuracy

A forecast “error” is the difference between an observed value and its forecast. Here “error” does not mean a mistake, it means the **unpredictable part of an observation.**

Residual are calculated on the training set.

Forecast error are calculated on the test set.

## **RMSE (Root Mean Squared Error)**

RMSE is a measure that tells us how much our predictions (or guesses) differ, on average, from the actual values in our dataset. It is calculated by taking the square root of the mean of the squared differences between the predicted values and the actual values.

Lower values indicate better accuracy. 0 indicates perfect predictions. Sensitive to large errors due to squaring.

## **MAE (Mean Absolute Error)**

MAE measures the average absolute difference between predicted and observed values in a time series, providing a more interpretable measure of forecast accuracy compared to RMSE.

Like RMSE, lower MAE values indicate better accuracy, with 0 indicating perfect predictions. MAE is less sensitive to large errors compared to RMSE because it does not involve squaring the errors.

## **MAPE (Mean Absolute Percentage Error)**

MAPE measures the average percentage difference between predicted and observed values in a time series, making it useful for comparing forecast accuracy across different scales.

MAPE provides the average percentage error of the forecasts. However, it has limitations when dealing with zero or close-to-zero actual values (because of in denominator).

# **Evaluating distribution forecast accuracy**

# **Time series Cross-validation**

A more sophisticated version of training/test sets is time series cross-validation. It involves splitting a time series dataset into multiple consecutive segments, or "folds", while maintaining the temporal order of the data. Each fold consists of a training set and a corresponding test set. The training set contains data preceding the test set, ensuring that the model only learns from past observations. After training on the training set, the model is evaluated on the test set to assess its predictive performance. This process is repeated for each fold, allowing for an unbiased evaluation of the model's ability to forecast future values.

The following diagram illustrates the series of training and test sets, where the blue observations from the training sets, and the orange observations from the test sets.



# Judgemental Forecast

Forecasting using judgment is common in practice. In many cases, judgmental forecasting is the only option, such as when there is a complete lack of historical data, or when a new product is being launched, or when a new competitor enters the market, or during completely new and unique market conditions. For example, in December 2012, the Australian government was the first in the world to pass legislation that banned the use of company logos on cigarette packets, and required all cigarette packets to be a dark green colour. Judgment must be applied in order to forecast the effect of such a policy, as there are no historical precedents.

There are also situations where the data are incomplete, or only become available after some delay. For example, central banks include judgment when forecasting the current level of economic activity, a procedure known as nowcasting, as GDP is only available on a quarterly basis.

Research in this area6 has shown that the accuracy of judgmental forecasting improves when the forecaster has

1. Important domain knowledge,
2. More timely, up-to-date information.

A judgmental approach can be quick to adjust to such changes, information or events.

Over the years, the acceptance of judgmental forecasting as a science has increased, as has the recognition of its need. More importantly, the quality of judgmental forecasts has also improved, as a direct result of recognising that improvements in judgmental forecasting can be achieved by implementing well-structured and systematic approaches. It is important to recognise that judgmental forecasting is subjective and comes with limitations. However, implementing systematic and well-structured approaches can confine these limitations and markedly improve forecast accuracy.

There are three general settings in which judgmental forecasting is used:

1. There are no available data, so that statistical methods are not applicable and judgmental forecasting is the only feasible approach;
2. Data are available, statistical forecasts are generated, and these are then adjusted using judgment;
3. Data are available and statistical and judgmental forecasts are generated independently and then combined.

We should clarify that when data are available, applying statistical methods (such as those discussed in other chapters of this book), is preferable and should always be used as a starting point. Statistical forecasts are generally superior to generating forecasts using only judgment. For the majority of the chapter, we focus on the first setting where no data are available, and in the last section we discuss the judgmental adjustment of statistical forecasts.