Time Series Forecasting

Henrika Langen

Central European University
Department of Economics and Business

Winter 2024

Structure of the Lecture

1. Introduction to time series
2. **Modeling time series and evaluating forecasting performance**
3. Exponential smoothing models
4. ARIMA models
5. Regression models and advanced forecasting models
6. Machine Learning and Deep Learning methods

Modeling Time Series and Evaluating Forecasting Performance

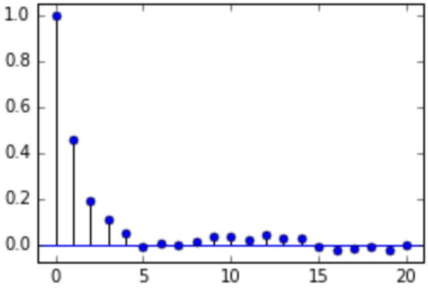
Topics

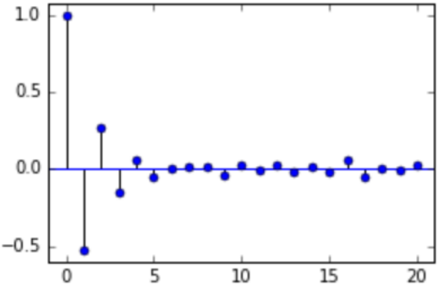
- Autocorrelation
- Time series modeling
- Evaluating model fit
- Evaluating forecasting accuracy
- Prediction intervals

Autocorrelation

- Autocorrelation quantifies the **linear relationship between lagged values of a time series**
- Notation: r_1 denotes the correlation between y_t and y_{t-1}, r_2 denotes the correlation between y_t and y_{t-2}, etc.
- Auto-correlation coefficient: $$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^{T}(y_t - \bar{y})^2}$$
- The graphical representation of autocorrelation coefficients at various lags is referred to as autocorrelation function (**ACF**)

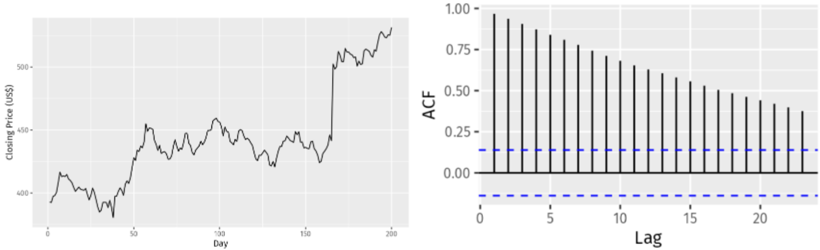
Autocorrelation





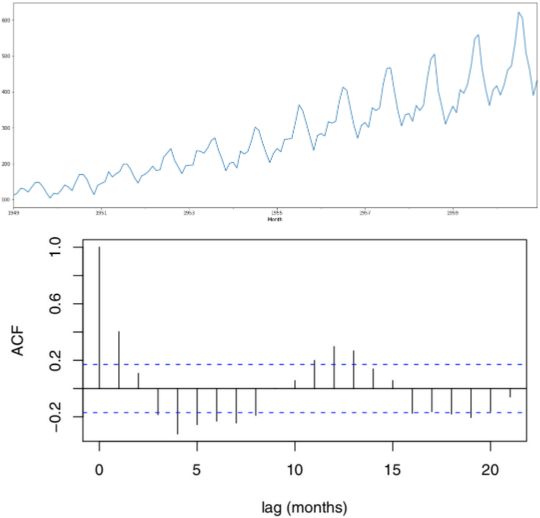
Two ACF plots of simulated data.

Autocorrelation



Google stock closing price on 200 consecutive days in 2013 and the corresponding ACF plot.

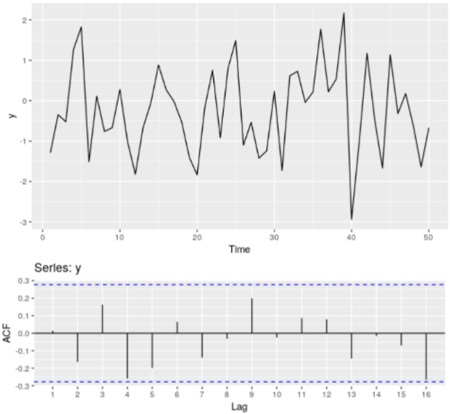
Source: FPP



Monthly international airline passengers and the corresponding ACF plot.

Source: FPP

Autocorrelation



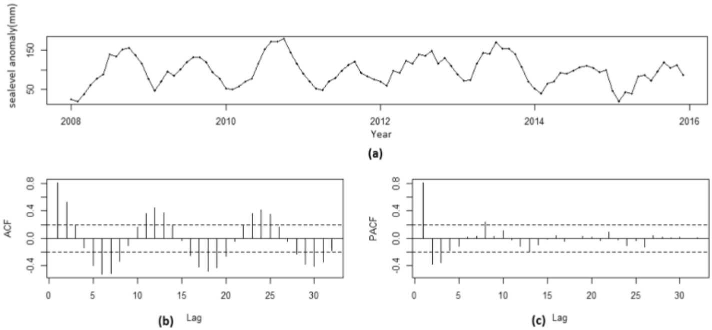
Time series with no autocorrelation and the corresponding ACF plot.

Source: FPP

Partial Autocorrelation

- Partial Autocorrelation quantifies the **linear relationship between y_t and y_{t-k} after removing the influence of the lags in between**, i.e., $y_{t-1}, y_{t-2}, .., y_{t-k+1}$
- Notation:
  - α_1 denotes the first partial correlation coefficient and is equal to r_1,
  - α_2 denotes the correlation between y_t and y_{t-2} after controlling for the influence of y_{t-1}
  - ...
- The graphical representation of partial autocorrelation coefficients at various lags is referred to as partial autocorrelation function (**PACF**)

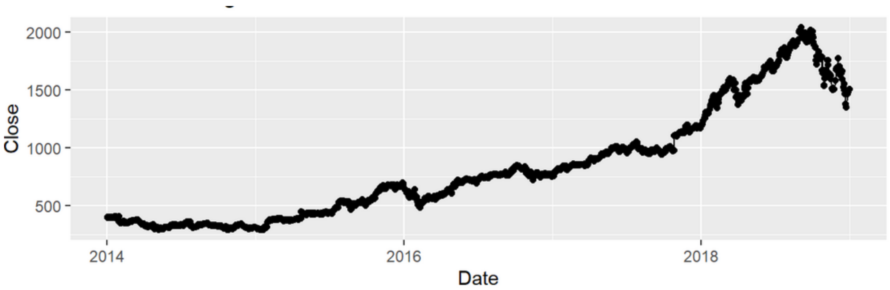
Partial Autocorrelation

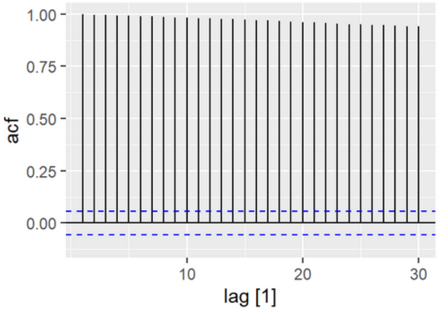


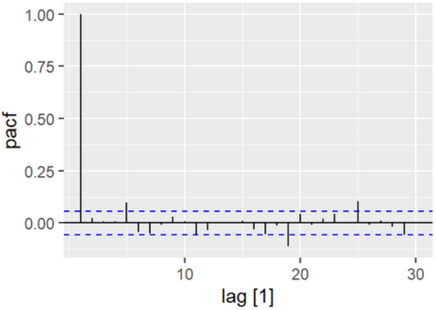
Monthly sea level anomalies in South China with the corresponding ACF and PACF.

Source: *https://doi.org/10.1063/1.5012237*

Partial Autocorrelation







Time Series Modeling

- A **time series process (or stochastic process)** is a sequence of random variables indexed by time t:

$$\{\ldots, Y_1, Y_2, \ldots, Y_t, Y_{t+1}, \ldots\} = \{Y_t\}_{t=-\infty}^{\infty}$$

- The sequence of data (=time series) we observe are T realizations of a process:

$$\{y_1, y_2, \ldots, y_T\} = \{y_t\}_{t=1}^{T}$$

- The goal of **time series modeling** is to obtain a **mathematical description of the process** that might have generated the observed time series.

Time Series Modeling

Time series models try to capture patterns in the series and temporal dependencies among the variables

- Decomposition models (Session 3) aim at disentangling trend and seasonality components in the data.
- ARIMA models (Session 4) seek to model temporal dependencies/autocorrelation within the series.
- Regression models (Session 5) aims at capturing relationships between the time series of interest and others.
- Several approaches (Session 5 & 6) combine some of the above.

Simple Time Series Models

- Time series with zero mean, constant variance and no autocorrelation are called **white noise**:
$$y_t = \varepsilon_t,$$
where $\varepsilon_t \sim iid(0, \sigma)$.
- A **random walk** is defined as:
$$y_t = y_{t-1} + \varepsilon_t,$$
where $\varepsilon_t \sim iid(0, \sigma)$ is white noise.
- A **random walk with a drift** is defined as:
$$y_t = c + y_{t-1} + \varepsilon_t,$$
where $\varepsilon_t \sim iid(0, \sigma)$ and c is a constant.
- Random walk series are characterized by periods of rise and fall, followed by sudden, unpredictable changes in direction.

Evaluating Model Fit

Residuals

- To evaluate the fit of a model to the observed time series, one can calculate the residuals of the fitted series:

$$e_t = \hat{y}_{t|t-1} - y_t$$

  where $\hat{y}_{t|t-1}$ are fitted values, i.e., the forecast of y_t based on the estimated model and using observations $y_1, y_2, \ldots, y_{t-1}$.
- **Careful:** Fitted values are no true forecasts because the model is estimated using all information available in the time series.

$\Rightarrow$ The residuals solely measure how well the model fits the data and not how well the model performs in forecasting

Residuals: A model fits the data well if the residuals e_t have certain properties:

- The residuals have zero mean. Else the model would be biased.
  - $\Rightarrow$ If $E(e_t) \neq 0$, simply correct the model for $E(e_t)$ to obtain unbiased estimates.
- The residuals are uncorrelated. If the residuals were correlated, there would still be information left in the residuals that could be used to better model the data.
  - $\Rightarrow$ To check whether this property is met, do checks for autocorrelation as outlined on next slide.

Evaluating Model Fit

Residuals: Test for autocorrelation in the residuals

- Visual checks: Check if there are large spikes in the residual ACF plot.
- Formal tests: test whether the first autocorrelations are significantly different from what would be expected from a white noise process using a formal test for autocorrelation, called a portmanteau test. E.g.:
  - Box-Pierce Test
  - Ljung-Box Test

  $\Rightarrow$ The H_0 of these tests is that the first autocorrelations are similar to what would be expected from a white noise process, i.e., a low p-value indicates that the residuals are autocorrelated.

Time Series Forecasting

- The model fitted to a series can usually be easily adapted for forecasting the value of y_t in $T+h$, denoted as y_{T+h}.
- Depending on the modeling approach, forecasting may involve iteratively predicting $\hat{y}_{T+1}$, then plugging this forecast into the model to obtain a forecast for $T+2$,...
- If the series is modeled considering related time series, these may be forecast first before incorporating them into the estimation of $\hat{y}_{T+1}$.
- Additionally, there are some simple forecasting methods that simply continue a series based on basic calculation, without the need for modeling the series first (see next slide).

Simple Forecasting Methods

- **Average method:** The forecasts are set to be the average of all observed data: $\hat{y}_{T+h|T} = (y_1 + ... + y_T)/T \quad \forall\, h$
- **Naïve method:** The forecasts are set to be the value of the last observation: $\hat{y}_{T+h|T} = y_T \quad \forall\, h$
- **Seasonal naïve method:** For seasonal data with frequency m, the forecasts are set to be the value of the last observation of the same season:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \quad \forall\, h,$$

  where k is the integer part of $(h-1)/m$
- **Drift method:** For a series with a trend, the forecasts can be derived by drawing a line connecting the earliest and latest observations and extrapolating it forward:

$$\hat{y}_{T+h|T} = y_T + h\,\frac{y_T - y_1}{T-1} \quad \forall h$$

Evaluating forecasting accuracy

- To evaluate the forecasting performance of a model, it is necessary to **predict data points that were not used for estimating the model**

$\Rightarrow$ Split the data into a training set for model estimation and a test set to evaluate forecast performance

- **Forecasting error:** For a model estimated based on $\{y_1, ..., y_T\}$, the forecasting error of the h-step forecast, $\hat{y}_{T+h|T}$ is given by:

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T}$$

Source (figure): FPP

Residuals vs. Forecasting Errors

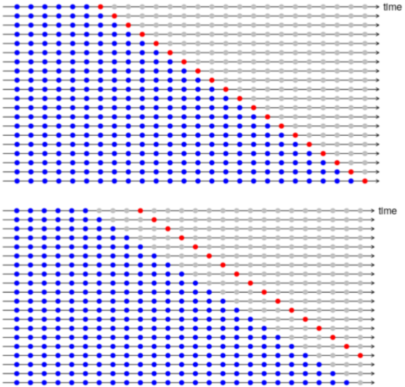
Residuals measure the goodness-of-fit

- Residuals quantify the model's fit to observed (historical) data, indicating whether a model utilizes all available information in the data.
- Low residuals do not guarantee accurate forecasting, as the model's fit can always be enhanced by increasing complexity, eventually leading to overfitting.

Forecasting errors quantify the predictive performance of the model

- Forecasting errors assist in assessing the model's ability to generalize to unseen data and accurately predict future observations.

Evaluating forecasting accuracy: Cross-validation



Cross-validation for evaluating 1-step-ahead forecasting performance (top) and 4-step-ahead forecasting performance (bottom).

Source (figure): FPP

Evaluating forecasting accuracy: cross-validation

- **Approach: 1-step-ahead forecasting**
  1. Define the training set as the first few observations of the time series.
  2. Estimate the model based on the training set and evaluate the forecasting accuracy using the subsequent data point as test data (=calculate forecasting error).
  3. Expand the training set by the test data point and repeat steps 2-3 until the training set contains all but one observations.
  4. Compute the forecast accuracy by averaging the errors from all test sets.
- To evaluate model performance in k**-step-ahead forecasting**, the procedure can be adapted by simply selecting test observations k steps ahead of the training sample.
- **Alternative approach:** Rolling window approach: A fixed-size training window is shifted across the time series. The model is trained on each window and assessed on the subsequent data point(s).

Evaluating Forecasting Accuracy

- **Information criteria** can be used to compare the goodness of fit of different candidate models while accounting for the complexity of the models, particularly the number of estimated parameters.
  - The **Akaike Information Criterion (AIC)** estimates the amount of information lost by a model and weighs it against the model's complexity.
  - For small T, a modified version of the AIC, the **"Corrected AIC"** denoted as AIC$_c$, is more suitable.
  - Another commonly used information criterion is the **Schwarz's Bayesian Information Criterion (BIC)**, which, compared to AIC, penalizes complexity more heavily.
- All information criteria become smaller, the better a model balances goodness-of-fit and complexity. $\Rightarrow$ Select the model that minimizes the chosen information criterion.
- The information criteria are calculated differently depending on the modeling approach used (exponential smoothing, ARIMA and regression).$\Rightarrow$ They cannot be used to compare between time series models obtained from different modeling approaches.

Prediction Intervals

- **Definition**: The prediction interval for an h-step forecast provides a range in which the true value of y_{T+h} is expected to fall with a specified probability.
- **Properties:** As the forecast horizon h increases, the forecasts become more uncertain, leading to wider prediction intervals.
- **Estimation:** The estimation of prediction intervals becomes comparably simple if the residuals
  - ... have constant variance and
  - ... are normally distributed,

  i.e., if the residuals e_t satisfy $e_t \sim \mathcal{N}(0, \sigma^2) \ \forall t$.

$\Rightarrow$ To check how the true observations are distributed around the model and verify if $e_t \sim \mathcal{N}(0, \sigma^2) \ \forall t$ holds approximately, examine the distribution of the residuals using histogram and residual time plot.

Prediction Intervals

- **Estimation:** If the residuals **have constant variance and are normally distributed**, the prediction intervals can be estimated as:

$$\hat{y}_{T+h|T} \pm z\,\hat{\sigma}_h,$$

  where $\hat{\sigma}_h$ is the estimated standard deviation of the h-step forecast distribution and z is a multiplier determining the range covered by the prediction interval (for the 95% prediction interval $z = 1.96$).

- **Estimation of $\hat{\sigma}_h$:** For $h = 1$, σ_h can be approximated by the residual standard deviation $sd(e_t)$. For $h > 1$, the estimation of σ_h depends on the time series model used (prediction intervals will be computed in Python, formulas are usually provided in documentation).

Prediction Intervals

Estimation with bootstrapping: If the assumption of normally distributed residuals is unreasonable, the prediction intervals can be estimated by bootstrapping.

Procedure:

1. Under the assumption that future errors are similar to past errors, the unknown future forecast error e_{T+1} in $y_{T+1} = \hat{y}_{T+1|T} + e_{T+1}$ can be replaced by sampling from the collection of residuals.
   - $\Rightarrow$ Do so repeatedly to obtain several different possible futures.
   - $\Rightarrow$ Estimate the prediction interval for $h = 1$ based on the percentiles of the distribution of these possible future values.
2. To obtain the prediction interval for $h = 2$, predict $\hat{y}_{T+2|T+1}$ based on each simulated future value obtained in step 1 as $y_{T+2} = \hat{y}_{T+2|T+1} + e_{T+2}$, where e_{T+2} is again replaced by sampling from the collection of residuals.
   - $\Rightarrow$ Do so repeatedly to obtain several different possible futures and estimate the prediction interval for $h = 2$.
3. For $h > 2$ proceed accordingly.

Strategies for handling missing values

- **Forward fill:** Carry forward the last known value prior to the missing one.
- **Moving average with forward-looking data:** Set the missing data point to the (weighted) mean/median of a certain number of values that come before it (or potentially also those following it).
- **Inter-/extrapolation:** Impute missing data based on a fitting a curve to the known values preceding it (or potentially also those following it).

When assessing prediction accuracy with 1-step-ahead forecasts, missing values should only be imputed based on previous values. For time series visualization tasks, values that follow the missing ones may also be included to improve the fit.