Моделирование и предсказание временных рядов

А.Г. Трофимов

к.т.н., доцент, НИЯУ МИФИ

lab@neuroinfo.ru
http://datalearning.ru

Курс "Статистическая обработка временных рядов"

Сентябрь 2018

Box-Jenkins Methodology

The Box-Jenkins methodology is a procedure for identifying, selecting and estimating ARMA models for discrete univariate time series

Step 1. Establish the stationarity of your time series. If it is non-stationary try to transform it to be stationary
Detrending and deseasonalizing, unit root tests

Step 2. Identify a (stationary) ARMA model for your data
Estimation of model's type and its order

Step 3. Estimate the parameters of the chosen model
Fitting the model's parameters to the data

Step 4. Conduct goodness-of-fit checks to ensure the model describes your data adequately
Statistical analysis of residuals

Step 5. Use the model to forecasting

Integrated Process

A unit root process (or difference-stationary process, DSP) $\{Y_t\}$ is a stochastic process whose first difference is stationary:

$$Y_t = c + Y_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a stationary process, c is a drift

The process $\{Y_t\}$ can be transformed to stationary process $\{\varepsilon_t\}$ by differencing and conversely it can be obtained by integrating the stationary process $\{\varepsilon_t\}$

Definition

The process $\{Y_t\}$ is called integrated processes of order D (or $I(D)$ process) if it can be obtained by integrating some stationary process $\{\varepsilon_t\}$ D times

If $\{Y_t\}$ is $I(D)$ process then its D-th differenced process $\{\Delta^D Y_t\}$ is stationary

Differencing Operator

The previous value Y_{t-1} can be rewritten using the lag operator L:

$$Y_{t-1} = LY_t$$

The differenced process in operator form:

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - LY_t = (1 - L)Y_t$$

The differencing operator Δ is related to the lag operator L:

$$\Delta = 1 - L$$

and the D-th differencing operator:

$$\Delta^D = (1 - L)^D$$

$$\Delta^D Y_t = (1 - L)^D Y_t$$

ARIMA Process

Let $\{Y_t\}$ be $I(D)$ process. Then the differenced process $\{\Delta^D Y_t\}$ is stationary and it can be modelled as stationary ARMA(p,q) process

Definition

Discrete-time autoregressive integrated moving average process of AR order p, MA order q and differentiation order D (ARIMA(p,D,q) process) $\{Y_t, t = 0, 1, ...\}$ is defined as:

$$\Delta^D Y_t = c + \phi_1 \Delta^D Y_{t-1} + ... + \phi_p \Delta^D Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

where $\{\varepsilon_t\}$ is a discrete-time white noise and c, $\phi_1, ..., \phi_p$, $\theta_1, ..., \theta_q$ are constants

ARIMA models with differentiation order D are are applicable to model $I(D)$ processes

In a particular case, $ARIMA(p, 0, q) \equiv ARMA(p, q)$

ARIMA Process in Operator Form

ARIMA(p,D,q) process:

$$\Delta^D Y_t = c + \phi_1 \Delta^D Y_{t-1} + \ldots + \phi_p \Delta^D Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}$$

In operator form:

$$(1-L)^D Y_t = c + (1-L)^D (\phi_1 L + \ldots + \phi_p L^p) Y_t + (1 + \theta_1 L + \ldots + \theta_q L^p) \varepsilon_t$$

$$(1-L)^D (1 - \phi_1 L - \ldots - \phi_p L^p) Y_t = c + (1 + \theta_1 L + \ldots + \theta_q L^q) \varepsilon_t$$

$$\phi^*(L) Y_t = c + \theta(L) \varepsilon_t$$

where

$$\phi^*(L) = (1-L)^D \phi(L)$$

$$\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p \quad \text{(stable)}$$

$$\theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \quad \text{(invertible)}$$

The characteristic polynomial $\phi^*(z)$ of ARIMA(p,D,q) process has exactly D unit roots

ARMA Process vs ARIMA Process

The characteristic polynomial of ARMA(p,q) process:

$$\phi(L)Y_t = c + \theta(L)\varepsilon_t$$

The characteristic polynomial of ARIMA(p,D,q) process:

$$(1 - L)^D \phi(L)Y_t = c + \theta(L)\varepsilon_t$$

if ARMA(p,q) process has D unit roots, then it is ARIMA(p,D,q) process

ARIMA(p,D,q) process is a type of non-stationary random walk process with D unit roots

ARIMA(p,D,q) model of the time series $\{y_1, ..., y_T\}$ is equivalent to ARMA(p,q) model of D times differenced time series $\{\Delta^D y_1, ..., \Delta^D y_T\}$

Under-differencing and Over-differencing

If the unit root tests (e.g. augmented Dickey-Fuller test) reveal that the process $\{Y_t\}$ has a unit root it should be differenced to achieve stationarity

$I(D)$ process should be differenced D times

For highly persistent but stationary $I(0)$ processes the unit root tests tend to fail, so the $I(0)$ process will be considered as $I(1)$ process and will be wrongly differenced, i.e. over-differenced

If $I(1)$ process is wrongly considered as $I(0)$ process and so won't be differenced it results to be under-differenced

What is the cost of under-differencing and over-differencing in time series modelling and forecasting?

To Diff or Not to Diff?

If $\{Y_t\}$ is $I(0)$ ARMA process, then the differenced process will have a unit root in MA part:

$$\phi(L)Y_t = c + \theta(L)\varepsilon_t$$

$$(1 - L)\phi(L)Y_t = (1 - L)\theta(L)\varepsilon_t$$

$$\phi(L)\Delta Y_t = \theta^*(L)\varepsilon_t$$

where

$$\Delta Y_t = (1 - L)Y_t = Y_t - Y_{t-1}$$

is differenced process, and characteristic MA polynomial $\theta^*(L) = (1 - L)\theta(L)$ has a unit root

It means that differenced stationary ARMA process is non-invertible, i.e. it cannot be represented in stable AR(∞) form, the problems in its coefficients estimation and time series forecasting occur

Fear of Over-Differencing

If the data is really stationary then differencing the data can result in a misspecified model

Over-differencing tremendously emphasizes the small specification inaccuracies and measurement errors in the data, relative to the signal under modelling

Iterated differentiation of a time series makes it more memoryless but a time series can be both memoryless and non-stationary

Many researchers avoid over-differencing, it is the fear of over-differencing

Choosing ARIMA Structural Parameters

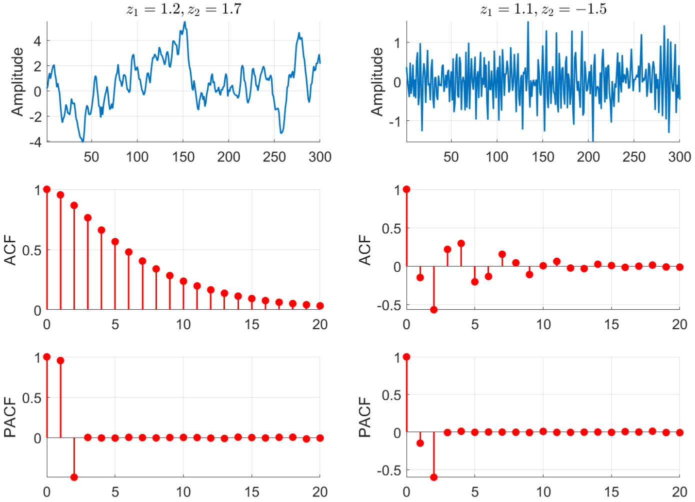
The parameters p, D, q or ARIMA(p,D,q) model are structural and must be specified. To estimate them the qualitative and quantitative analysis of ACF and PACF and unit root tests are used

Rules to choose p, D, q:

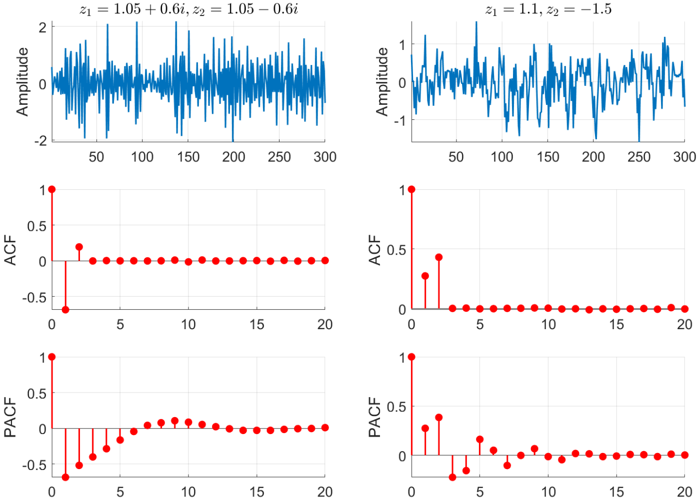
- If the PACF displays a sharp cutoff while the ACF decays more slowly (the time series displays so called "AR signature"), this autocorrelation pattern can be explained more easily by adding AR terms than by adding MA terms. The lag at which the PACF cuts off is the estimate of p

- If the ACF displays a sharp cutoff while the PACF decays more slowly (the time series displays so called "MA signature"), this autocorrelation pattern can be explained more easily by adding MA terms than by adding AR terms. The lag at which the ACF cuts off is the estimate of q

AR Pattern. Illustration



MA Pattern. Illustration



ARIMA Parameters Estimation

The parameters p, D, q completely determine the model structure and must be specified. All other parameters (coefficients $\phi_1, ..., \phi_p$, $\theta_1, ..., \theta_q$, variance of innovations σ^2) are estimable

Maximum likelihood estimation (MLE) method is commonly used for ARIMA parameters estimation

Under assumption that random vector of innovations has Gaussian distribution:

$$\varepsilon = (\varepsilon_1, ..., \varepsilon_T)^T \sim N(0, \sigma^2 I)$$

we write the likelihood function:

$$\mathscr{L}(\varepsilon) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^T} \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p+1}^{T} \varepsilon_t^2\right)$$

For AR(p) processes the ordinary least square (OLS) estimation can also be used (since the errors $\varepsilon_1, ..., \varepsilon_T$ can be calculated directly)

MLE Estimation

Given the time series $y_1, ..., y_T$, the likelihood function $\mathscr{L}(\varepsilon)$ depends on unknown parameters $\phi_1, ..., \phi_p$, $\theta_1, ..., \theta_q$ since

$$\phi^*(L)y_t = c + \theta(L)\varepsilon_t$$

$$\varepsilon_t = \theta(L)^{-1}(\phi^*(L)y_t - c)$$

and $\phi(L)$ depends on $\phi_1, ..., \phi_p$ and $\theta(L)$ depends on $\theta_1, ..., \theta_q$

MLE estimation consists in solving the optimization problem

$$\mathscr{L}(\varepsilon) \to \max_{\phi_1,...,\phi_p,\theta_1,...,\theta_q}$$

To solve it the iterative methods of non-linear optimization are used. They require initial values of estimated parameters. They can be set to zeros or special techniques to estimate them can be applied

Presample Data Initialization

Sample data is observed time series $y_1, ..., y_T$. Presample data comes from time points before the beginning of the observation period. For example, for AR(2) model:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

$$\varepsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2}$$

the innovation ε_2 explicitly depends on y_1 and y_0, and the innovation ε_1 explicitly depends on unobservable y_0 and y_{-1}

The amount of presample data depends on the AR degree p and the amount of presample innovations depends on the MA degree q

Approaches to presample data initialization:

- Use first data as presample and fit model to remaining data
- Set custom presample data and innovations
- Generate presample data by backward forecasting and set presample innovations to zero

Check for Unit Root

Let $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q$ are the parameters of ARMA(p,q) model fitted to the time series data $y_1, ..., y_T$

Then the time series data $y_1, ..., y_T$ can be considered as a sample path of ARMA(p,q) process $\{Y_t\}$, where

$$Y_t = c + \phi_1 Y_{t-1} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q}$$

How to check if this process have unit roots?

The characteristic polynomial:

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$$

has p roots $z_1, ..., z_p$. It is time consuming to calculate them

Fast rule: if the sum $\phi_1 + ... + \phi_p = 1$, then the characteristic polynomial has a unit root

Indeed, $\phi(1) = 1 - \phi_1 - ... - \phi_p = 1 - 1 = 0 \Rightarrow$ the characteristic polynomial $\phi(z)$ has a root $z = 1$

Iterative Correction of ARIMA Structural Parameters

After fitting the model to the data:

- If there is a unit root in the AR part of the fitted model (the sum of the AR coefficients is almost 1) you should reduce the number of AR terms by one and increase the order of differencing by one
- If there is a unit root in the MA part of the model (the sum of the MA coefficients is almost 1) you should reduce the number of MA terms by one and reduce the order of differencing by one
- Look at long-term dynamics. If it is erratic or unstable, modify structural parameters of the model

To identify the best values of structural parameters, fit a set of models with different parameters to the same data and choose the best one

What criterion should be used to choose the best model?

AIC and BIC

Criteria to take into account when choosing p, D, q:

- Accuracy of model (goodness-of-fit)
- Complexity of model (number of estimable parameters)

The measures that combine accuracy and complexity of the model are informational measures:

- Akaike Information Criterion (AIC)

$$AIC = -2 \ln \mathscr{L}^* + 2k$$

- Bayesian Information Criterion (BIC)

$$BIC = -2 \ln \mathscr{L}^* + k \ln T$$

Here $\ln \mathscr{L}^*$ is the value of the maximized log-likelihood objective function for a model with k parameters fitted to T data points

When comparing AIC and BIC values for multiple models, the smaller values are better

AIC and BIC. Notes

- With AIC the penalty for model's complexity is $2k$, with BIC the penalty is $k \ln T$
- For ARIMA(p,D,q) the number of parameters is

$$k = p + q + 1$$

- Some simulation studies demonstrate that AIC selects the "true model" better than BIC*
- Normalized AIC:

$$nAIC = \frac{1}{T} AIC$$

- Small sample-size corrected AIC:

$$AICc = AIC + 2k \frac{k+1}{T-k-1}$$

Vrieze S.I. Model selection and psychological theory: a discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Psychological Methods. 2012. Vol. 17(2), pp. 228–243.

Likelihood Ratio Test

The primary goal of model selection is choosing the most parsimonious model that adequately fits your data

Given the ARIMA(p, D, q) model, is it possible to reduce its complexity by imposing some restrictions on its parameters (such as assign them to zero)?

Three asymptotically equivalent tests compare a restricted model (the null model) against an unrestricted model (the alternative model), fitted to the same data:

- Likelihood ratio test (LR-test)
- Lagrange multiplier test (LM-test)
- Wald test (W-test)

Likelihood Ratio Test

Does the restriction bring significant changes in optimal value of log-likelihood function?

Various statistical tests can be used to check it. One of them is likelihood ratio (LR) test

Assume $\mathscr{L}^*$ and $\mathscr{L}_0^*$ is the value of the maximized log-likelihood objective function for unrestricted model (e.g. ARIMA(p,D,q)) and restricted model (e.g. ARIMA($p-1$,D,q))

Null hypothesis H_0: $\mathscr{L}^*$ and $\mathscr{L}_0^*$ differs insignificantly

Test statistic:

$$Z = 2(\mathscr{L}^* - \mathscr{L}_0^*), \quad Z|_{H_0} \sim \chi^2(r)$$

where r is the number of restricted parameters

Critical region: right-sided

If H_0 is accepted, then the null model can be restricted

Goodness of Fit

After specifying a model and estimating its parameters, it is good practice to perform goodness-of-fit checks to diagnose the adequacy of your fitted model

When assessing model adequacy, areas of primary concern are:

- Violations of model assumptions
- Poor predictive performance
- Missing explanatory variables

Goodness-of-fit checks can help you identify areas of model inadequacy and suggest ways to improve your model

They include the analysis of model's residuals and estimating the model performance on unseen data

Residual Diagnostics

- Normality (innovations have Gaussian distribution)

  How to check: histogram, box plot, goodness-of-fit tests, etc.
  If non-normal: specify other distribution for innovations and fit the model again

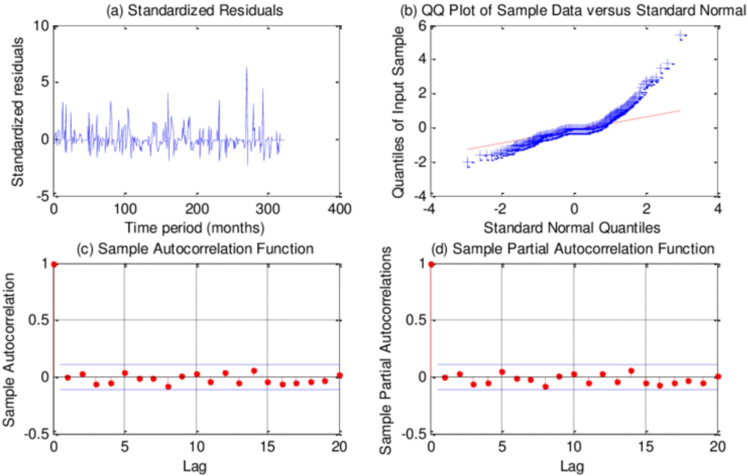
- Autocorrelation (innovation process is assumed to be uncorrelated)

  How to check: ACF, PACF, Ljung-Box Q-test, etc.
  If correlated: include additional AR or MA terms to the model

- Conditional heteroscedasticity (innovation process has constant variance)

  How to check: ACF, PACF, Ljung-Box Q-test for squared residual series, Engle's test etc.
  If heteroscedastic: include a conditional variance process to the model (e.g. GARCH model)

Residual Diagnostics. Illustration



Prediction Mean Squared Error

To estimate the predictive performance divide your time series into two parts: a training set and a test set. Fit the model to the training data and simulate the fitted model over the test period to estimate possible overfitting

Prediction mean squared error (PMSE):

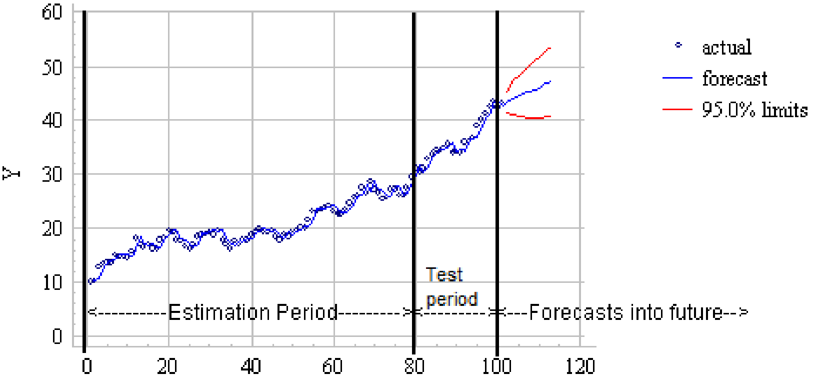
$$PMSE = \frac{1}{T - T_{train}} \sum_{t=T_{train}+1}^{T} (y_t - \tilde{y}_t)^2$$

where y_t and $\tilde{y}_t$ are observed and predicted value at time moment t

It's good practice to calculate PMSE for various T_{train} to verify the robustness of your results

The unseen period can be used to test a great number of models and choose the model whose errors are smallest in it. In this case the third, validation sample is needed

Training and Validation Periods



Optimal Forecasting

Assume the ARMA model is fitted on the sample time series $y_1, ..., y_T$ and $\varepsilon_1, ..., \varepsilon_T$ are residuals, $\varepsilon_t = y_t - \tilde{y}_t$, where $\tilde{y}_t$ is modelled value at time t, $t = 1, ..., T$

How to obtain the future values $\tilde{y}_{T+h}$, $h = 1, 2, ...$?

Criterion of forecast optimality in mean squared sense:

$$\mathrm{M}\left[(Y_{T+h} - \tilde{y}_{T+h})^2 | T\right] \to \min_{\tilde{y}_{T+h}}$$

The expectation is conditioned by all known values $\{y_t\}$, $\{\varepsilon_t\}$

$$\begin{aligned} \mathrm{M}[(Y_{T+h} - \tilde{y}_{T+h})^2 | T] &= \mathrm{M}[Y_{T+h}^2 | T] - 2\mathrm{M}[Y_{T+h}|T]\tilde{y}_{T+h} + \tilde{y}_{T+h}^2 \\ &= \mathrm{D}[Y_{T+h}|T] + \mathrm{M}[Y_{T+h}|T]^2 - 2\mathrm{M}[Y_{T+h}|T]\tilde{y}_{T+h} + \tilde{y}_{T+h}^2 \\ &= \mathrm{D}[Y_{T+h}|T] + (\mathrm{M}[Y_{T+h}|T] - \tilde{y}_{T+h})^2 \end{aligned}$$

Thus, the optimal forecast $\tilde{y}_{T+h}$ is forecast by regression:

$$\tilde{y}_{T+h} = \mathrm{M}[Y_{T+h}|T] = \mathrm{M}[Y_{T+h}|y_1, ..., y_T, \varepsilon_1, ..., \varepsilon_T]$$

AR(1) Model Forecasting

Optimal forecast:

$$\tilde{y}_{t+h} = \mathrm{M}[Y_{t+h}|t] = \mathrm{M}[Y_{t+h}|y_1, ..., y_t, \varepsilon_1, ..., \varepsilon_t]$$

AR(1) model:

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t$$

$$\mathrm{M}[Y_{t+1}|t] = \mathrm{M}[c+\phi_1 Y_t+\varepsilon_{t+1}|t] = c+\phi_1 \mathrm{M}[Y_t|t]+\mathrm{M}[\varepsilon_{t+1}|t] = c+\phi_1 y_t$$

$$\begin{aligned}\mathrm{M}[Y_{t+2}|t] &= \mathrm{M}[c + \phi_1 Y_{t+1} + \varepsilon_{t+2}|t] = c + \phi_1 \mathrm{M}[Y_{t+1}|t] + \mathrm{M}[\varepsilon_{t+2}|t] \\ &= c + \phi_1(c + \phi_1 y_t) = c(1 + \phi_1) + \phi_1^2 y_t\end{aligned}$$

$$\begin{aligned}\mathrm{M}[Y_{t+h}|t] &= \mathrm{M}[c + \phi_1 Y_{t+h-1} + \varepsilon_{t+h}|t] = c + \phi_1 \mathrm{M}[Y_{t+h-1}|t] + \mathrm{M}[\varepsilon_{t+h}|t] \\ &= c + \phi_1 \mathrm{M}[Y_{t+h-1}|t]\end{aligned}$$

$$\mathrm{M}[Y_{t+h}|t] = c\sum_{i=0}^{h-1} \phi_1^i + \phi_1^h y_t$$

AR(1) Model Forecasting

Optimal forecast for AR(1) model:

$$\mathrm{M}[Y_{t+h}|t] = c\sum_{i=0}^{h-1}\phi_1^i + \phi_1^h y_t$$

For $|\phi_1| < 1$:

$$\mathrm{M}[Y_{t+h}|t] \to \frac{c}{1-\phi_1}, \quad \text{as } h \to \infty$$

The forecast converges to the unconditional mean of AR(1) process $\mathrm{M}[Y_t] = \frac{c}{1-\phi_1}$ and forgets the last value y_t

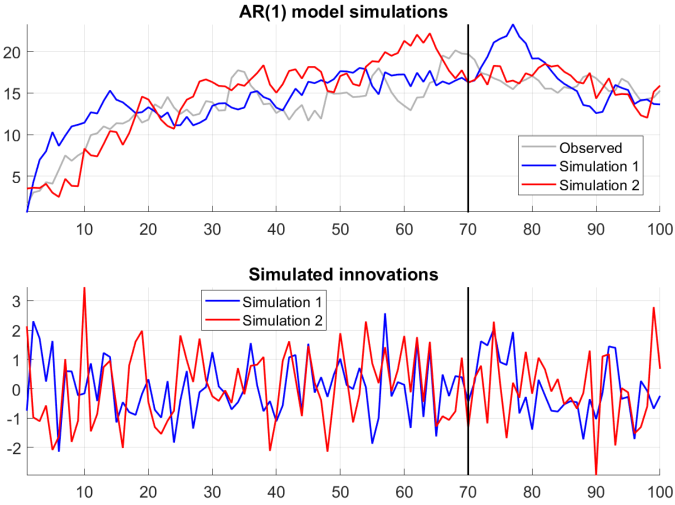
For $\phi_1 = 1$:

$$\mathrm{M}[Y_{t+h}|t] = ch + y_t$$

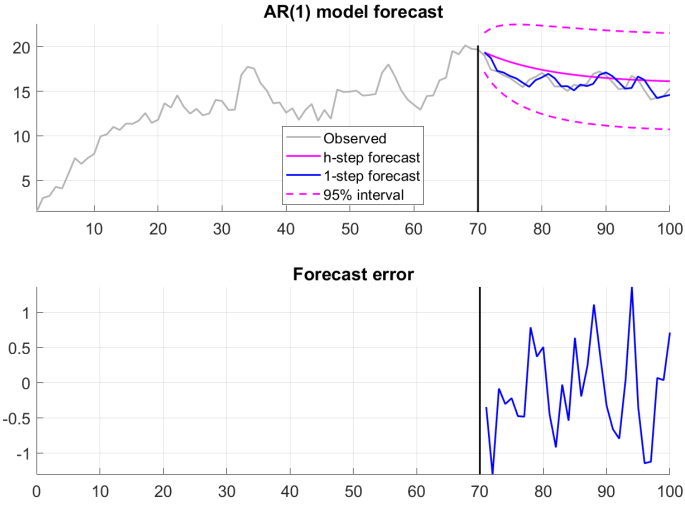
The forecast is linear function starting from the last value y_t

For $|\phi_1| > 1$: the forecast is explosive

AR(1) Model Estimation and Simulation. Illustration



AR(1) Model Forecast. Illustration



MA(2) Model Forecasting

MA(2) model:

$$Y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

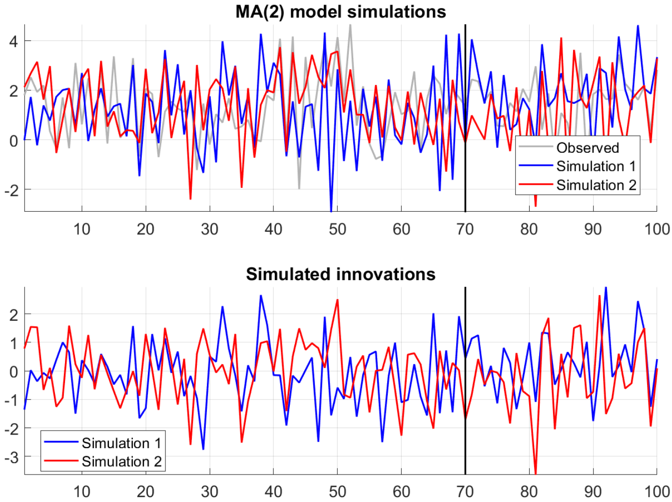
$$\begin{aligned} \mathrm{M}[Y_{t+1}|t] &= \mathrm{M}[c + \varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}|t] \\ &= c + \mathrm{M}[\varepsilon_{t+1}|t] + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} = c + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} \end{aligned}$$

$$\begin{aligned} \mathrm{M}[Y_{t+2}|t] &= \mathrm{M}[c + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t|t] \\ &= c + \mathrm{M}[\varepsilon_{t+2}|t] + \theta_1 \mathrm{M}[\varepsilon_{t+1}] + \theta_2 \varepsilon_t = c + \theta_2 \varepsilon_t \end{aligned}$$

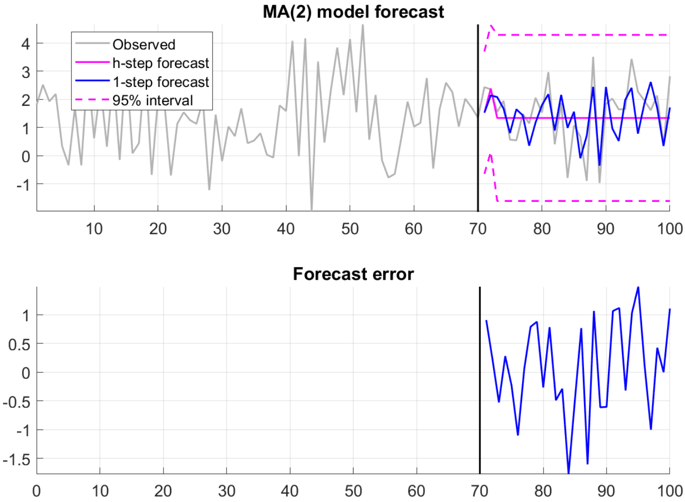
$$\mathrm{M}[Y_{t+h}|t] = \mathrm{M}[c + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \theta_2 \varepsilon_{t+h-2}|t] = c, \quad h > 2$$

The forecast converges to the unconditional mean of MA(2) process $\mathrm{M}[Y_t] = c$ and forgets the last innovations $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}$ after 2 steps

MA(2) Model Estimation and Simulation. Illustration



MA(2) Model Forecast. Illustration



ARMA(1,2) Model Forecasting

ARMA(1,2) model:

$$Y_t = c + \phi_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$$

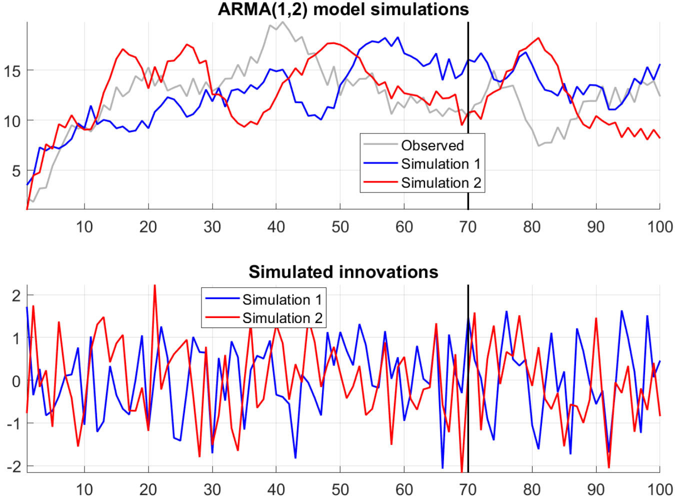
$$\begin{aligned} \mathrm{M}[Y_{t+1}|t] &= \mathrm{M}[c + \phi_1 Y_t + \varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}|t] \\ &= c + \phi_1 y_t + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} \end{aligned}$$

$$\begin{aligned} \mathrm{M}[Y_{t+2}|t] &= \mathrm{M}[c + \phi_1 Y_{t+1} + \varepsilon_{t+2} + \theta_1 \varepsilon_{t+1} + \theta_2 \varepsilon_t|t] \\ &= c + \phi_1 \mathrm{M}[Y_{t+1}|t] + \theta_2 \varepsilon_t \\ &= c + \phi_1 (c + \phi_1 y_t + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1}) + \theta_2 \varepsilon_t \\ &= c(1 + \phi_1) + \phi_1^2 y_t + (\phi_1 \theta_1 + \theta_2) \varepsilon_t + \phi_1 \theta_2 \varepsilon_{t-1} \end{aligned}$$

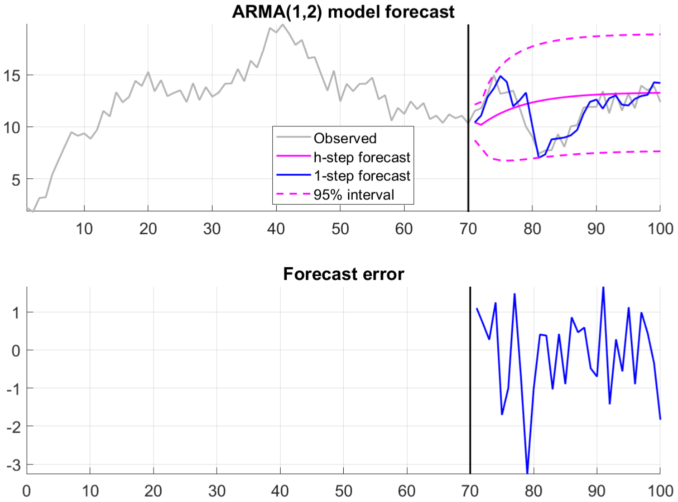
$$\begin{aligned} \mathrm{M}[Y_{t+h}|t] &= \mathrm{M}[c + \phi_1 Y_{t+h-1} + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \theta_2 \varepsilon_{t+h-2}|t] \\ &= c + \phi_1 \mathrm{M}[Y_{t+h-1}|t], \quad h > 2 \end{aligned}$$

For steps $h > 2$ the forecast follows AR(1) pattern

ARMA(1,2) Model Estimation and Simulation. Illustration



ARMA(1,2) Model Forecast. Illustration



Prediction Intervals for MA(q) Model

To estimate the prediction interval for predicted value $\tilde{y}_{t+h}$ we need its standard deviation:

$$\mathrm{D}[\tilde{y}_{t+1}|t] = \mathrm{D}\left[c + \varepsilon_{t+1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i+1}|t\right] = \mathrm{D}[\varepsilon_{t+1}] = \sigma^2$$

$$\mathrm{D}[\tilde{y}_{t+2}|t] = \mathrm{D}\left[c + \varepsilon_{t+2} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i+2}|t\right] = \mathrm{D}[\varepsilon_{t+2}] + \theta_1^2 \mathrm{D}[\varepsilon_{t+1}]$$
$$= \sigma^2(1 + \theta_1^2)$$

$$\mathrm{D}[\tilde{y}_{t+h}|t] = \mathrm{D}\left[c + \varepsilon_{t+h} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i+h}|t\right] = \sigma^2\left(1 + \sum_{i=1}^{h-1} \theta_i^2\right)$$

(under assumption that residuals are uncorrelated)

Prediction Intervals for MA(q) Model

The estimation of $\mathrm{D}[\tilde{y}_{t+h}|t]$:

$$\hat{\sigma}_h^2 = \hat{\sigma}^2[\tilde{y}_{t+h}|t] = \hat{\sigma}^2 \left(1 + \sum_{i=1}^{h-1} \hat{\theta}_i^2\right)$$

where $\hat{\sigma}^2$ is estimated variance of residuals, $\hat{\theta}_i$ is estimation of θ_i

The standartized Y_{t+h}:

$$Y_{t+h} = \frac{Y_{t+h} - \mathrm{M}[Y_{t+h}]}{\sigma[Y_{t+h}]} \sim N(0,1)$$

(under assumption that residuals are normal)

The confidence interval for $\mathrm{M}[Y_{t+h}]$:

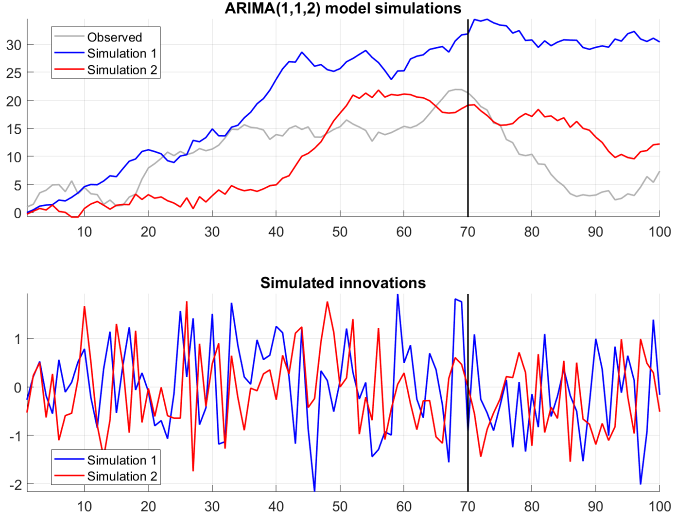
$$Y_{t+h} - u_{1-\alpha/2}\sigma[Y_{t+h}] < \mathrm{M}[Y_{t+h}] < Y_{t+h} + u_{1-\alpha/2}\sigma[Y_{t+h}]$$

$$\tilde{y}_{t+h} - u_{1-\alpha/2}\tilde{\sigma}_h < \mathrm{M}[Y_{t+h}] < \tilde{y}_{t+h} + u_{1-\alpha/2}\tilde{\sigma}_h$$

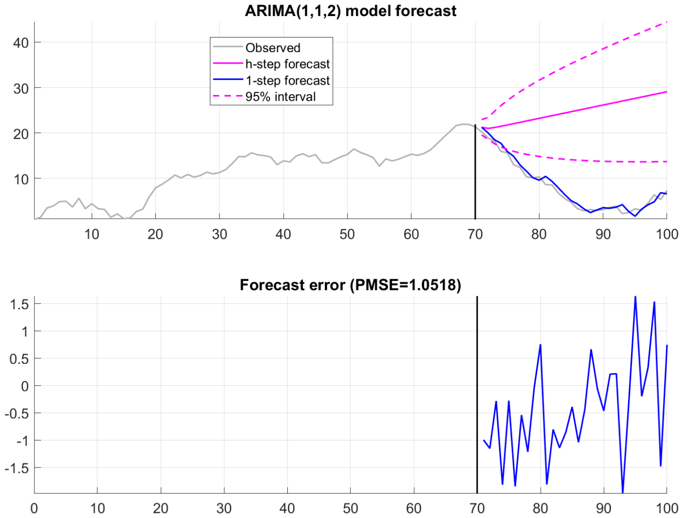
Prediction Interval for ARIMA Model

- The width of prediction intervals for MA(q) model grows up to step q, then it becomes constant
- To calculate prediction intervals for stable ARMA(p, q) model it should be rewritten in MA(∞) form
- The width of prediction intervals for stable ARMA(p, q) model grows with step h, but converges to some constant value
- The width of prediction intervals for ARMA(p, q) model with unit roots (i.e. for ARIMA(p, D, q) model) grows infititely with step h
- Usually the prediction intervals tend to be too narrow because only the variation in the errors has been accounted for to calculate them. There is also variation in the parameter estimates

ARIMA(1,1,2) Model Estimation and Simulation. Illustration



ARIMA(1,1,2) Model Forecast. Illustration



Naive Forecast

The 1-step ahead forecast seems to be accurate but really is it accurate?

Consider the naive 1-step ahead forecast:

$$\tilde{y}_{t+1} = y_t$$

In naive forecast the next predicted value is equal to the current value. It leads to 1-step delayed time series

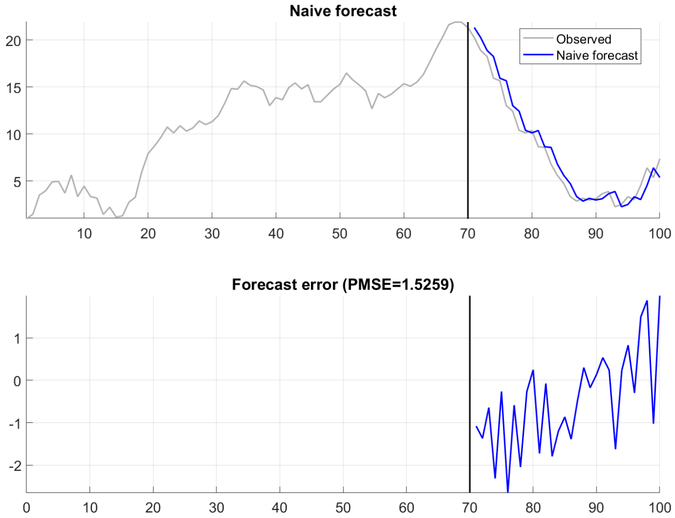
Naive forecast corresponds to a persistent prediction model that is often used as a reference for determining the efficiency of the constructed model

The value

$$\eta = \frac{PMSE_{naive} - PMSE}{PMSE_{naive}}$$

can be used as a measure of the constructed model's efficiency over the persistent model

Naive Forecast. Illustration



Predictive Performance Estimation

Let y_t and $\tilde{y}_t$ are observed and predicted values at time moment t

There are a lot of measures of model's predictive performance: MSE, RMSE, MAE, MAPE, RMSLE, etc.

Mean squared error (MSE):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \tilde{y}_t)^2}$$

Mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{y_t - \tilde{y}_t}{y_t}\right|$$

All measures are calculated over test time interval

Time Series Modelling and Forecasting. Overview

Step 1. Data preprocessing

- Visual analysis of time series, ACF and PACF
- Trend and seasonality estimation
- Detrending and deseasonalizing
- Unit root tests (ADF, PP, KPSS tests)
- Time series transformations

Step 2. Model identification

- Visual analysis of ACF and PACF
- Maximum likelihood estimation
- Compare model with alternatives (likelihood ratio test)
- AIC and BIC

Step 3. Model diagnostics

- Residual analysis
- Validation and test data

Step 4. Time series forecasting