Fahrplan SARIMA

Khandelwal:

ARIMA models are well-known for their notable forecasting accuracy and

flexibility in representing several different types of time series

ARIMA models are based on the fundamental principle that the future values of a time series are

generated from a linear function of the past observations and white noise terms.

Andreoni:

a time series is a stochastic process, i.e. an ordered sequence of random variables, where the time index t takes on a finite or countable infinite set of value

\cite{Andreoni} give a definition of time series, where „a time series is a stochastic process, i.e. an ordered sequence of random variables, where the time index t takes on a finite or countable infinite set of value“.

the series must be stationary and ACF and PACF must be time-independent. Variance non-stationarity can be removed if the series is transformed with the logarithmic function. Mean non-stationarity can be removed by using the operator ∇ = 1-B applied d times in order to make the series stationary

The ACF is a measure of the correlation between two variables composing the stochastic process, which are k temporal lags far away; the PACF measures the net correlation between two variables, which are k temporal lags far away.

Box-Jenkins The method requires a preliminary data analysis to verify the presence of outliers and then the identification, estimation and diagnostic checking steps.

The identification stage provides an initial ARIMA model specified on the basis of the estimated ACF and PACF, starting from the original data; particularly, the characteristics of ACF and PACF allow the identification of the model order:

1. if the autocorrelations decrease slowly or do not vanish, there is non stationarity and the series should be differenced until stationarity is obtained; then, an ARIMA model can be identified for the differenced series;

2. if ACF ρk is zero for k>q and PACF is decreasing, then the process underlying the series is an MA(q);

3. if PACF πk is zero for k>p and ACF is decreasing, then the process underlying the series is an AR(p);

4. if there is no evidence for an MA or an AR then an ARMA model may be adequate.

generally, model parameters are estimated by using least squares or maximum likelihood methods.

Finally, different diagnostic tests can be performed.

For large sample size, if the order of the AR component is p, the estimate of the partial autocorrelations πk are approximately normally distributed with mean zero and variance 1/N for k>p, where N is the sample size. Then, it should be verified if the residuals of the calibrated model belong to a white noise process.

Starting from a univariate ARIMA model, some explanatory (or independent) variables can be inserted. In this case, the dependent variable Xt depends on lagged values of the independent variables. The lag length may sometimes be known a priori, but usually it is unknown and in some cases it is assumed to be infinite.

Arlt:

The principle and the application of the SARIMA models in the time series modelling has been well known for many years. Its practical applications can be found in many areas where empirical analyses are needed and it has become a basis standard tool of modern econometric analysis. The crucial phase of the practical application of the Box-Jenkins methodology is the identification and verification of the suitable model.

Mircetic:

Freeman finds no evidence of short-run cyclical responseof beer to economic variables.

S-ARIMA is regarded as one of the

best models for forecasting complicated seasonal time series,

and it represents the combination of differencing with

autoregression and a moving average model

Key question in using S-ARIMAmodel is selection of an appropriate

model order, that is the values p, q, P, Q, D, d. If d and D

are known, we can select the orders p, q, P and Q via an

information criterion such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC): (hier Formeln)

For evaluating the created models we used the root

mean square error (RMSE), mean absolute percentage error

(MAPE), mean absolute scaled error (MASE), AIC and

BIC

Hunt:

The time series considered in this article are the sequences

of observations observed at equally spaced intervals.

Being stationary means

that the statistical relationship between the observations

at time t and time t + l is the same as the statistical relationship

between the observations at time t + t and

t + z + 't .

the autocorrelation for an

autoregressive model gradually dies away while the

autocorrelation for a moving average model will cut off

at lag q.

the autocorrelation for an

autoregressive model gradually dies away while the

autocorrelation for a moving average model will cut off

at lag q.

In the situation where the series y1 is not stationary

it may be possible to make it stationary by applying

differencing or seasonal differencing.

Plotting the data will indicate if the series is stationary.

If not, the series can be transformed using standard functions

such as LOG or SQRT

The major concern here is that the residuals are systematically

distributed across the series (e.g., they could

be negative in the first part of the series and approach a

zero in the second part) or that they contain some serial

dependency which may suggest that the ARIMA model

is inadequate

Adhikari:

a timeseriesisadynamicalprocessthatkeepsonchanging continuously withhighdegreeofuncertaintyandmayeven exhibit regime switches. Thisjeopardizesthevalidityofthefore- casting modelwhennewobservationsareaddedtotheavailable data. Finally,aparticularmodelisalwayspronetofaultyassump- tions, implementationbiasesanderrorsinparameterestimation, which considerablyaffectthedesiredforecasts [5]. Thesedis- couraging factsofsinglemodelingapproachmotivatedthe explorationofvariousforecastcombinationtechniques.Acombi- nation offorecastsbenefits fromtheinter-modeldiversities, mitigates therisksofusinganisolatedmodel,andcompensates the drawbacksoftheindividualmodels.

First,atime series seldomhasthe independence andidenticaldistribution(i.i.d.) property thatisafundamentalrequirementofrealisticstatistical processes [4]. Asaresult,themodelthatperformedbestforthein- sample datasetmightnotalwaysprovidethebestforecastsforthe unseen futurevalues.Second,aforecastingmodelisspecific tothe nature ofthetimeseries,i.e.whethertheseriesisgeneratedfrom a linearornonlinearprocess,followsstationaryornonstationary distribution, containstrend,seasonal,orcyclicalpatterns,etc. Estimating theexactnaturerequireslargenumberofhistorical observations,butinpracticethereisonlyaverysmallsampleof availabledataandsothe fitted modelmaybeinappropriate

Univariatetimeserieswithdiscrete valuesaremostwidelystudiedintheliterature,wheresucha series isrepresentedas Y ¼ y1; y2;…; yN T, yt being theobserva- tion attime t.

Intihar:

Specifically,

the simpler forecasting models, which had

a leading role in the port economics for many years,

are no longer able to capture increased volatility and

complexity of the time series.

This interrelation is not surprising since

economic indicators, such as GDP (gross domestic

product) or import/export, can importantly influence

the trade flows and consequently the port throughput

volume. T

In the third stage (see Figure 5), the diagnostic

checking and GOF testing are applied to each model

candidate. Here, several statistical, information-based,

and residual-based criteria are computed to investigate

the statistical properties, goodness-of-fit, and

predictive performance of each model candidate. The

following criteria and measures have been considered

for every single ARIMAX model [50]:

–– The root mean square error (RMSE) and the mean absolute

estimation (MAE);

–– The percentage of model fit (FIT); and

–– The information-based Akaike’s corrected AICc criterion.

Besides these measures, the Jarque-Bera (JB) test,

as well as the Portmanteu (Ljung-Box) LB test [50]

have also been applied to investigate whether the

model residual holds the properties of the white noise

without any serious serial correlation

<https://otexts.com/fpp2/>

Wang:

Parameter estimation. Using the AIC criterion to

determine the model parameters, compare the AIC values

of each parameter model, select the model with the

smallest AIC value as the optimal model.

Wu:

However, they are easy to be disturbed

by noise data, and as a result, perform not so well when

applied to prediction of short term traffic flow with much

randomness.

Zhao:

Non-seasonal ARIMA models are generally denoted ARIMA

(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time

lags) of the autoregressive model, d is the degree of differencing (the number of times the data

have had past values subtracted), and q is the order of the moving-average model

Augmented Dickey-Fuller Test was used to

check the stationarity of series + detaillierte Beschreibung

Zhou:

AutoRegressive Integrated Moving Average (ARIMA) is a flexible and widely used

time series method.

The Box and Jenkins methodology (Box and Jenkins 1976) is an approach to build

ARIMA model from time series data. It has an iterative procedure involving (a) identification

of a tentative model, (b) estimation of model parameters, (c) diagnostic checking, and (d)

forecasting.

Box jenkins visual method - and the model got may not be the optimal solution.

According to our investigation from literature, it is proposed

that *p*/*q* between 0 and 10 and *P*/*Q* between 0 and 5 while *d*/*D* still from 0 to 2, since an extreme large number of models found in the literature belong to this range.

About AIC/BIC: These statistics include a penalty about model complexity in order to avoid overfitting



Note that ARIMA modeling doesn’t need a separate validation set,