Summary/Review

**ARMA Models**

ARMA models combine two models:

* The first is an autoregressive (AR) model. Autoregressive models anticipate series’ dependence on **its own past values**.
* The second is a moving average (MA) model. Moving average models anticipate series’ dependence on **past forecast errors**.
* The combination (ARMA) is also known as the **Box - Jenkins approach**.

ARMA models are often expressed using orders ***p*** and ***q*** for the ***AR*** and ***MA*** components.

For a time series variable *X* that we want to predict for time *t*, the last few observations are:

X\_{t-3},\;X\_{t-2},\;X\_{t-1}*Xt*−3​,*Xt*−2​,*Xt*−1​

* ***AR(p)*** models are assumed to depend on the ***last p values*** of the time series. For ***p****=2,* the forecast has the form:

X\_t=\phi\_1X\_{t-1}+\phi\_2X\_{t-2}+\omega\_t*Xt*​=*ϕ*1​*Xt*−1​+*ϕ*2​*Xt*−2​+*ωt*​

Here, \omega\_t*ωt*​ is the ***forecast error***; \phi\_1*ϕ*1​ and \phi\_2*ϕ*2​ are the (***p****=2*) parameters (estimated by regression).

* ***MA(q)*** models are assumed to depend on the ***last q values*** of the forecast error. For ***q****=2,* the forecast has the form:

X\_t=\theta\_2\omega\_{t-2}+\theta\_1\omega\_{t-1}+\omega\_t*Xt*​=*θ*2​*ωt*−2​+*θ*1​*ωt*−1​+*ωt*​

Here, \omega\_t*ωt*​ is the forecast error, \omega\_{t-1}*ωt*−1​ is the previous forecast error, etc. \theta\_1*θ*1​ and \theta\_2*θ*2​ are the (***q****=2*) parameters.

Combining the ***AR(p)*** and ***MA(q)*** models yields the ***ARMA(p, q)*** model. For ***p****=2*, ***q****=2,* the ARMA(2, 2) forecast has the form:

X\_t=\phi\_1X\_{t-1}+\phi\_2X\_{t-2}+\theta\_2\omega\_{t-2}+\theta\_1\omega\_{t-1}+\omega\_t*Xt*​=*ϕ*1​*Xt*−1​+*ϕ*2​*Xt*−2​+*θ*2​*ωt*−2​+*θ*1​*ωt*−1​+*ωt*​

\omega\_t*ωt*​ is the forecast error, \phi\_1*ϕ*1​ , \phi\_2*ϕ*2​ , \theta\_1*θ*1​ , and \theta\_2*θ*2​ are the (***p*** + ***q*** *= 4*) parameters.

**ARMA Models Considerations**

These are important considerations to keep in mind when dealing with ARMA models:

* The time series is assumed to be stationary.
* A good rule of thumb is to have at least 100 observations when fitting an ARMA model.

There are three stages in building an ARMA model:

**Identification**

At this stage you:

* Validate that the time series is stationary.
* Confirm whether the time series contains a seasonal component.

You can determine if seasonality is present by using **autocorrelation** **and** **partial autocorrelation plots**, **seasonal subseries plots**, and **intuition** (possible in some cases, i.e. seasonal sales of consumer products, holidays, etc.).

An **Autocorrelation Plot** is commonly used to detect dependence on prior observations.

It summarizes total (2-way) correlation between the variable and its past values.

The **Partial Autocorrelation Plot** also summarizes dependence on past observations.

However, it measures partial results (including all lags)

**Seasonal Subseries Plot** is one approach for measuring seasonality. This chart shows the average level for each seasonal period and illustrates how individual observations relate to this level.

**Estimation**

Once we have a stationary series, we can estimate AR and MA models. We need to determine ***p*** and ***q***, the order of the AR and MA models.

One approach here is to look at autocorrelation and partial autocorrelation plots. Another approach is to treat ***p*** and ***q*** as hyperparameters and apply standard approaches (grid search, cross validation, etc.)

How do we determine the order p of the AR model?

* Plot confidence intervals on the Partial Autocorrelation Plot.
* Choose lag ***p*** such that partial autocorrelation becomes insignificant for ***p + 1*** and beyond

How can we determine the order ***q*** of the MA model?

* Plot confidence intervals on the Autocorrelation Plot
* Choose lag ***q*** such that autocorrelation becomes insignificant for ***q + 1*** and beyond.

**Evaluation**

You can assess your ARMA model by making sure that the residuals will approximate a Gaussian distribution (aka white noise). Otherwise, you need to iterate to obtain a better model.

These are guidelines to choose between an AR and a MA model based on the shape of the autocorrelation and partial autocorrelation plots.

| **SHAPE** | **MODEL** |
| --- | --- |
| Exponential Decaying to zero | AR models |
| Alternating positive and negative decaying to zero | AR models |
| One or more spikes, the rest are close to zero | MA model |
| Decay after a few lags | Mixed AR and MA |
| All zero or close to zero | Data is random |
| High values at fixed intervals | Include seasonal AR term |
| No decay to zero | Series is not stationary |

**ARIMA Models**

ARIMA stands for Auto-Regressive Integrated Moving Average.

ARIMA models have three components:

* AR Model
* Integrated Component
* MA Model

**SARIMA Models**

SARIMA is short for **Seasonal ARIMA**, an extension of ARIMA models to address seasonality.

This model is used to remove seasonal components.

* The SARIMA model is denoted **SARIMA (p, d, q) (P, D, Q)**.
* **P, D, Q** represent the same as p, d, q but they are applied across a season.
* **M** = one season

**ARIMA and SARIMA Estimation**

These are the steps to estimate p, d, q and P, D, Q?

* Visually inspect a run sequence plot for trend and seasonality.
* Generate an ACF Plot.
* Generate a PACF Plot.
* Treat as hyperparameters (cross validate).
* Examine ***information criteria*** (*AIC*, *BIC*) which penalize the number of parameters the model uses.