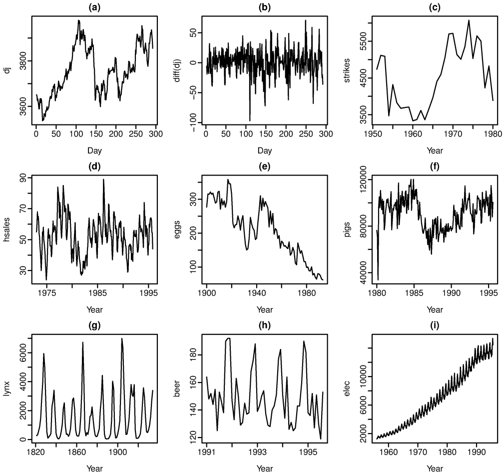
Time-series Forecasting - Parametric prediction models, ARIMA

Dr. Josif Grabocka

ISMLL, University of Hildesheim

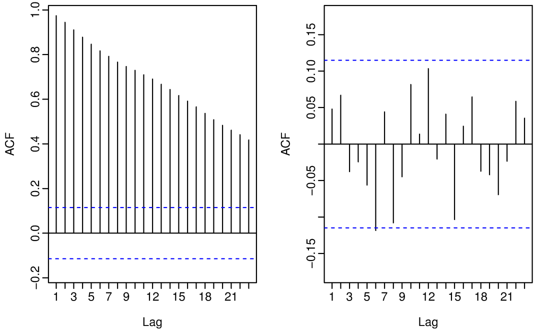
Business Analytics

Stationarity: Measurements independent on time



Seasonality: d,h,i; Trend: a,c,e,f,i; Stationarity: b,g. Hyndman et al. 2014

Stationarity inspection: Small ACF values



ACF of series from previous slide. Left (a): trend; Right (b): stationary. Hyndman et al. 2014

Remember: ACF=correlation with lagged values; $r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$

Differencing: Random walk model

- First difference series are based on changes between subsequent measurements:

$$y'_t = y_t - y_{t-1}$$

- The model for the series is written as a Random Walk (white noise) model:

$$y_t - y_{t-1} = e_t, \text{ or } y_t = y_{t-1} + e_t$$

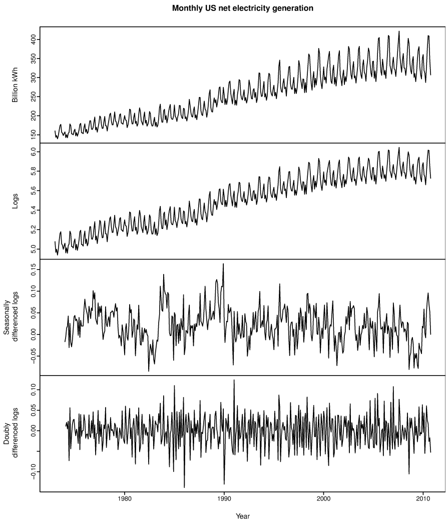
- Second difference:

$$y''_t = y'_t - y'_{t-1} = y_t - 2y_{t-1} + y_{t-2}$$

- Seasonal difference and white nosis:

$$y'_t = y_t - y_{t-m}, \text{ with model } y_t = y_{t-m} + e_t$$

Differencing leads to stationarity



US net electricity generation. Hyndman et al. 2014

Differencing test: Unit Root Test

How to know if differencing is needed? Fit the following regression model:

$$y'_t = \phi y_{t-1} + \beta_1 y'_{t-1} + \beta_2 y'_{t-2} + \cdots + \beta_k y'_{t-k}$$

- If $|\phi| < \epsilon$, the series is already stationary
- If $|\phi| > \epsilon$, the series needs differencing

- Known as Augmented Disckey-Fuller test, and available in multiple software packages
- When using R's "adf.test" check if the output is less then 0.05

Remember Regression - An example

A	MEDV: Median value of owner-occupied homes in $1000's
B1	CRIM: per capita crime rate by town
B2	ZN: proportion of residential land zoned for lots over 25,000 sq.ft.
B3	INDUS: proportion of non-retail business acres per town
B4	CHAS: Charles River ($=1$ if tract bounds river; 0 otherwise)
B5	NOX: nitric oxides concentration (parts per 10 million)
B6	RM: average number of rooms per dwelling
B7	AGE: proportion of owner-occupied units built prior to 1940
B8	DIS: weighted distances to five Boston employment centres
B9	RAD: index of accessibility to radial highways
B10	TAX: full-value property-tax rate per $10,000
B11	PTRATIO: pupil-teacher ratio by town
B12	B: $1000(Bk - 0.63)^2$ where Bk is the proportion of pupulation by town
B13	LSTAT: % lower status of the population

For each house i, estimate $\hat{A}_i$ using $B_{i,1}, B_{i,2}, \ldots, B_{i,13}$?

Backshift notation

- The backward shift operator B is useful when working with lags:

$$By_t = y_{t-1}, \quad \text{and} \quad B(By_t) = B^2 y_t = y_{t-2}$$

- Backward shift helps describe differencing:

$$\begin{aligned} y'_t &= y_t - y_{t-1} \\ &= y_t - By_t = (1-B)y_t \\ y''_t &= y_t - 2y_{t-1} + y_{t-2} \\ &= (1 - 2B + B^2)y_t = (1-B)^2 y_t \end{aligned}$$

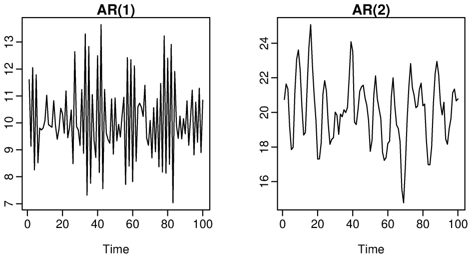
- Example: A difference followed by a seasonal difference:

$$\begin{aligned} (1-B)(1-B^m)y_t &= (1 - B - B^m + B^{m+1})y_t \\ &= y_t - y_{t-1} - y_{t-m} + y_{t-m-1} \end{aligned}$$

Autoregressive Model for Stationary Series

An autoregression of order p, denoted AR(p), forecasts using a linear combination of the past p values:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t$$

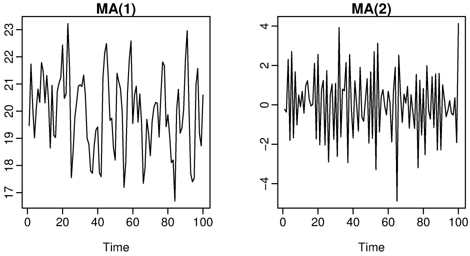


E.g. AR(1) $y_t = 18 - 0.8 y_{t-1} + e_t$, AR(2) $y_t = 8 + 1.3 y_{t-1} - 0.7 y_{t-2} + e_t$
Hyndman et al. 2014

Moving Average Model

A moving average model of order q, denoted MA(q), forecasts using a linear combination of the past q residuals:

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$



E.g. MA(1) $y_t = 20 + e_t + 0.8e_{t-1}$, AR(2) $y_t = e_t - e_{t-1} + 0.8e_{t-2}$ Hyndman et al. 2014

Equivalence of AR and MA models

- A stationary AR(p) can be represented as a MA(∞) model
- Let us demonstrate it for AR(1):

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + e_t \\ &= \phi_1(\phi_1 y_{t-2} + e_{t-1}) + e_t \\ &= \phi_1^2 y_{t-2} + \phi_1 e_{t-1} + e_t \\ &= \phi_1^3 y_{t-3} + \phi_1^2 e_{t-2} + \phi_1 e_{t-1} + e_t \end{aligned}$$

- Provided $-1 \leq \phi_1 \leq 1$, then $\phi_1^k \rightarrow 0$ as $k \rightarrow \infty$. Therefore:

$$y_t = e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \phi_1^4 e_{t-4} \ldots$$

Non-Seasonal ARIMA model

- ARIMA (AutoRegressive Integrated Moving Average) model combines differencing with autoregression and moving averages.

$$\begin{aligned} y'_t &= c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} \\ &\quad + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \end{aligned}$$

- Differencing reduces non-stationarity (note y'_t, not y_t is predicted)
- The difference of a series is predicted as a weighted average:
  - Lagged/differenced values generalized up to the d-th difference
  - Residual errors, arising from the MA model

Non-Seasonal ARIMA model - Generalized

- A generic ARIMA(p,d,q) model is defined as:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)e_t$$

- Decomposed as:
  - The AR(p) component: $(1 - \phi_1 B - \ldots \phi_p B^p)$
  - The d-th difference component: $(1 - B)^d y_t$
  - The MA(q) component: $(1 + \theta_1 B + \cdots + \theta_q B^q)e_t$
- E.g. ARIMA(2,1,3):

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)y_t = c + (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3)e_t$$
$$y'_t - \phi_1 y'_{t-1} - \phi_2 y'_{t-2} = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \theta_3 e_{t-3}$$

How to optimize ARIMA's parameters

- For given p, q, d the parameters $c, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ are optimized through Maximum Likelihood Estimation (MLE), similar to minimizing the least-squares loss: $L \approx \sum_{t=1}^{T} e_t^2$.
  - math.unice.fr/~frapetti/CorsoP/Chapitre_4_IMEA_1.pdf
- The model order p, q, d is computed using the Akaike Information Criteria
- Overall an Automatic ARIMA modeling procedure is conducted using the Hyndman-Khandakar Algorithm
- Optimization libraries can be used, such as R's **auto.arima()**

Point Estimation using ARIMA (1)

How to predict $\hat{y}_{T+h|T}$ using the learned ARIMA model:

1. Get y_t on the left side
2. Rewrite by replacing t with $T+h$
3. On the right side replace:
    - Future observations with their forecasts: $y_{T+h} \leftarrow \hat{y}_{T+h|T}$
    - Future errors by zero: $e_t \leftarrow 0, \quad T < t < T+h$
    - Past errors by the residual $e_t \leftarrow \hat{e}_t = y_t - \hat{y}_t, \quad T < t < T+h$

Point Estimation using ARIMA (2)

Let us illustrate using a learned ARIMA(3,1,1) (constant dropped for ease):

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)y_t = (1 + \theta_1 B)e_t$$

Step 1: Get y_t on the left side

$$\begin{aligned} \left[1 - (1+\phi_1)B + (\phi_1 - \phi_2)B^2 + (\phi_2 - \phi_3)B^3 + \phi_3 B^4\right] y_t &= (1 + \theta_1 B)e_t \\ y_t - (1+\phi_1)y_{t-1} + (\phi_1 - \phi_2)y_{t-2} + (\phi_2 - \phi_3)y_{t-3} + \phi_3 y_{t-4} &= e_t + \theta_1 e_{t-1} \end{aligned}$$

Yielding:

$$y_t = (1+\phi_1)y_{t-1} - (\phi_1 - \phi_2)y_{t-2} - (\phi_2 - \phi_3)y_{t-3} - \phi_3 y_{t-4} + e_t + \theta_1 e_{t-1}$$

Step 2: Rewrite by replacing t with $T + h$, starting with $h = 1$:

$$y_{T+1} = (1+\phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} - \phi_3 y_{T-3} + e_{T+1} + \theta_1 e_T$$

Point Estimation using ARIMA (3)

Step 3: Replace $e_{T+1} \leftarrow 0$ and $e_T \leftarrow \hat{e}_t = y_T - \hat{y}_T$:

$$\hat{\mathbf{y}}_{\mathbf{T+1|T}} = (1+\phi_1)y_T - (\phi_1 - \phi_2)y_{T-1} - (\phi_2 - \phi_3)y_{T-2} - \phi_3 y_{T-3} + \theta_1 \hat{e}_T$$

For t=T+2, start from **Step 2**:

$$y_{T+2} = (1+\phi_1)y_{T+1} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi^3 y_{T-2} + e_{T+2} + \theta_1 e_{T+1}$$

Go to **Step 3** and replace $y_{T+1} \leftarrow \hat{y}_{T+1|T}$, and $e_{T+1} \leftarrow 0, e_{T+2} \leftarrow 0$:

$$\hat{y}_{T+2|T} = (1+\phi_1)\hat{y}_{T+1|T} - (\phi_1 - \phi_2)y_T - (\phi_2 - \phi_3)y_{T-1} - \phi^3 y_{T-2}$$

Forecast Intervals

- For predicting $T+1$, the interval is easy:
  - If $\hat{\sigma}$ is the standard deviation of the residuals, then a 95% forecast interval is given by $\hat{y}_{T+1|T} \pm 1.96\hat{\sigma}$
- For ARIMA(0,0,q), i.e. MA(1)
  - The model can be re-written as:

$$\hat{y}_t = e_t + \sum_{i=1}^{q} \theta_i e_{t-i}$$

  - Then the forecast variance is:

$$\hat{\vartheta}_{T+h|T} = \sigma^2 \left[1 + \sum_{i=1}^{h-1} \theta_i^2\right], \text{ for } h = 2, 3, \ldots$$

  - And a 95% interval is $\hat{y}_{T+h|T} \pm 1.96\sqrt{\vartheta_{T+h|T}}$
- Intervals for the general ARIMA models are not covered here

Seasonal ARIMA model

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\uparrow} \quad \underbrace{(P, D, Q)_m}_{\uparrow}$$

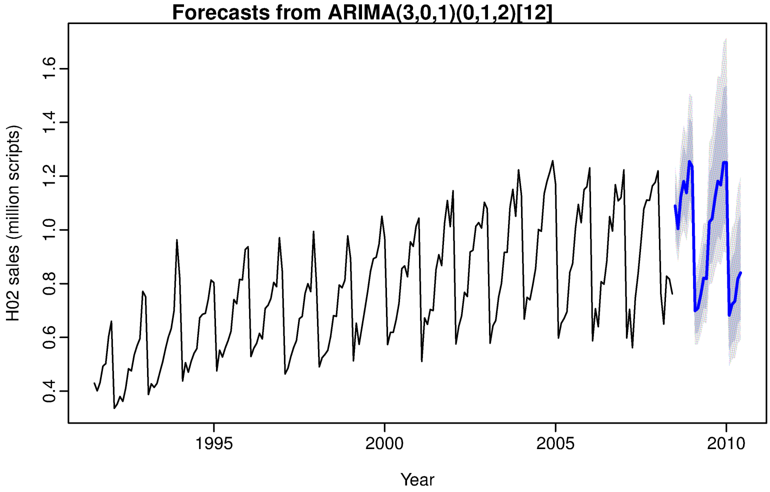
- (p, d, q): Non-seasonal part of the model
- $(P, D, Q)_m$: Seasonal part of the model

A seasonal ARIMA(1,1,1)(1,1,1)$_4$, where $m = 4$ for quarterly data:

$$(1 - \phi_1 B)\ (1 - \Phi_1 B^4)\ (1 - B)\ (1 - B^4) y_t \ = \ (1 + \theta_1 B)\ (1 + \Theta_1 B^4) e_t.$$

- $(1 - \phi_1 B)$: Non-seasonal AR(1)
- $(1 - \Phi_1 B^4)$: Seasonal AR(1)
- $(1 - B)$: Non-seasonal difference
- $(1 - B^4)$: Seasonal difference
- $(1 + \theta_1 B)$: Non-seasonal MA(1)
- $(1 + \Theta_1 B^4)$: Seasonal MA(1)

Example - H02 Sales



ARIMA(3,0,1)(0,1,2)$_{12}$, Hyndman et al. 2014

Nonlinear models - One-step ahead prediction

- Treat the problem as a regression of the last w measurements and a seasonal influence of the past k of length m:

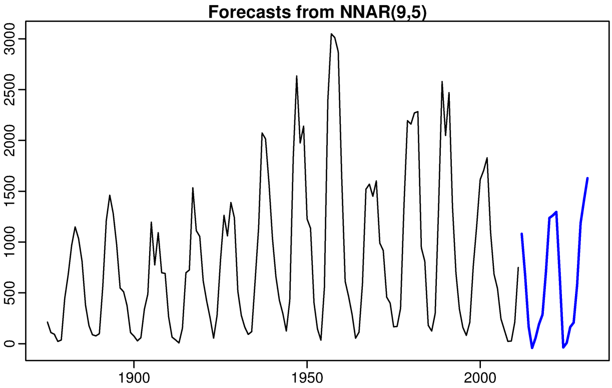
$$
\begin{array}{rcl}
\text{Target} & ; & \text{Predictors: autoregressive; seasonal} \\
y_t & ; & [y_{t-1}, y_{t-2}, \ldots, y_{t-w}, y_{t-m}, y_{t-2m}, \ldots, y_{t-km}] \\
y_{t-1} & ; & [y_{t-2}, y_{t-3}, \ldots, y_{t-w-1}, y_{t-m-1}, y_{t-2m-1}, \ldots, y_{t-km-1}] \\
 & \vdots &
\end{array}
$$

- Then learn the optimal parameters θ^{opt} of a prediction model $\mathcal{M}$ that minimizes the squared-error:

$$\theta^{\text{opt}} = \operatorname*{argmin}_{\theta^*} \sum_t \left(y_t - f([y_{t-1}, \ldots, y_{t-w}, y_{t-m}, \ldots, y_{t-km}], \theta^*\right)^2$$

- Successful models include Neural Networks, Gaussian process

Example - Neural-Network



Hyndman et al. 2014