STATS 207: Time Series Analysis Autumn 2020

Lecture 8: ARIMA/SARIMA

Dr. Alon Kipnis

October 7th 2020

Genera Info

- HW2 is out. Due on Monday October 19.
- **Midquarter Feedback Survey** on Canvas.
- Next Wednesday: Prof. David Donoho will talk about ARCH, GARCH, and stochastic volatility models.

Outline

So far:

- **ARMA**(p, q) is a useful model for **stationary processes**.
- We can express **ACF** and **optimal linear m-step forecast** in terms of model's parameters.
- We can **fit ARMA(p, q) to data** using several techniques. Leading to asymptotically normal estimators.

Next:

- ARMA + Seasonality and Trend (ARIMA, SARIMA)
- How to build ARMA, ARIMA, and SARIMA from data?
- Diagnostics.

ARIMA

IMA Models

- Motivation: Random walk

$$x_t = x_{t-1} + w_t, \qquad w_t \text{ is white noise.}$$

- The process

$$\nabla x_t = (1 - B)x_t = w_t$$

is stationary.

- Definition: **IMA**(d, q) process (Integrated **MA**)

$$\nabla^d x_t = \theta(B)w_t$$

where w_t is white noise, $\theta(B) = \sum_{j=0}^{q} \theta_j B^j$, and $\theta_0 = 1$.

- Examples:
  - **IMA**$(d = 0, q = 0)$ is white noise.
  - **IMA**$(0, q) =$ **MA**(q).
  - **IMA**$(1, 0)$ is random walk.
  - **IMA**$(2, 0)$ is Integrated random walk.
  - **IMA**$(1, 1)$ is random walk with MA correlated increments, aka **Exponential Weighted MA** (EWMA): "Most Frequently-used IMA Model."

EWMA

- Usually written as

$$x_t = x_{t-1} + w_t - \lambda w_{t-1}, \qquad |\lambda| < 1, \qquad x_0 = 0$$

(i.e., $\theta = -\lambda$).

- Because **MA polynomial is invertible**,

$$x_t = \sum_{u=1}^{\infty} (1-\lambda)\lambda^{u-1} x_{t-u} + w_t.$$

- One-step ahead prediction

$$x_{n+1}^n = \sum_{u=0}^{\infty} (1-\lambda)\lambda^u x_{n-u}.$$

- Neat updating formula as **weighted average of new data and old prediction**:

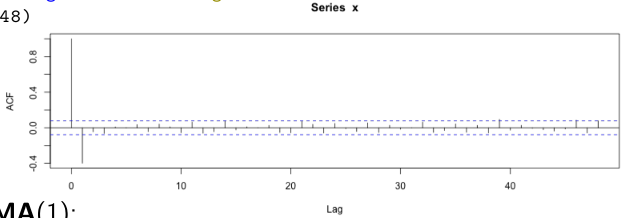
$$x_{n+1}^n = (1-\lambda)x_n + \lambda x_n^{n-1}.$$

larger values of λ lead to **smoother** estimate.

EWMA – Example 3.33

Example: 3.33 Fit **IMA**(1,1) to logged Glacial Varve Series

```
x = diff(log(varve)) # log transform + diff
acf(x, 48)
```



Fit an **MA**(1):

```
fit <- arima(log(varve), order=c(0,1,1))
fit$coef
```

```
ma1: -0.770539867652878
```

Suggested model:

$$\log(x_t) = \log(x_{t-1}) + w_t - 0.771 w_{t-1}.$$

Implies a **smoother** for Glacial Varve series

$$s_t = 0.229 \log(x_t) + 0.771 s_{t-1}.$$

ARIMA Process

- Definition: x_t is **ARIMA**(p, d, q) if

$$\nabla^d x_t = (1 - B)^d x_t$$

  is **ARMA**(p, q). We write

$$\phi(B)(1 - B)^d x_t = \theta(B) w_t.$$

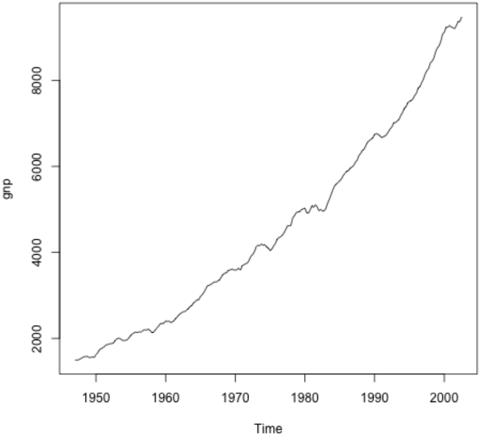
- Examples:
  - $d = 0$: classical **ARMA**(p, q)
  - $d = 1$: random walk with **ARMA**(p, q)-correlated increments.
- **Operationally**: We **difference** the time series d times to produce a **stationary time series**, then **use an ARMA model** of the result of differencing.

Basic Steps in Building ARIMA Models

- Plotting the data.
- Possibly transforming the data.
- Identifying the dependence order of the model.
- Parameter estimation.
- Diagnostic.
- Model Choice.

Example – ARIMA Modelling of US GNP Data

Data: Quarterly U.S. GNP from 1947(1) to 2002(3), $n = 223$ observations.

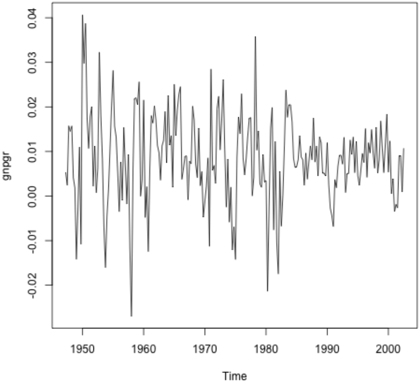


Data seem nonstationary – trending.

ARIMA Modelling of US GNP Data (cont'd)

Take logs and difference:

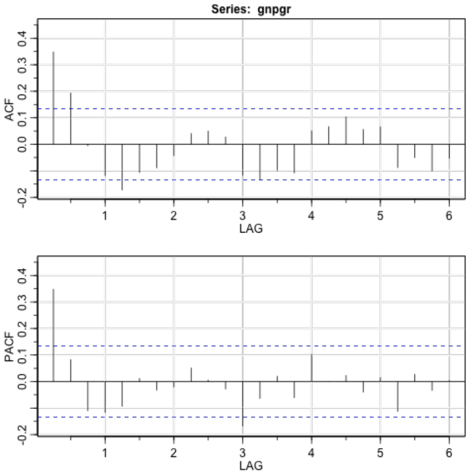
$$x_t \equiv \nabla \log(y_t)$$



- Differencing eliminates trend.
- Data seem more nearly stationary.
- Difference of Log = rate of growth ('natural' quantity).

ARIMA Modelling of US GNP Data (cont'd)

ACF and PACF of Differenced Log GNP



MA(2)?

MLE fit of **MA**(2):

```
arima(gnpgr, order=c(0, 0, 2))  # MA(2)
```

```
    Call:
arima(x = gnpgr, order = c(0, 0, 2))

Coefficients:
         ma1     ma2  intercept
      0.3028  0.2035     0.0083
s.e.  0.0654  0.0644     0.0010

sigma^2 estimated as 8.919e-05:  log likelihood = 719.96,  aic = -1431.93
```

MA(2) fit:

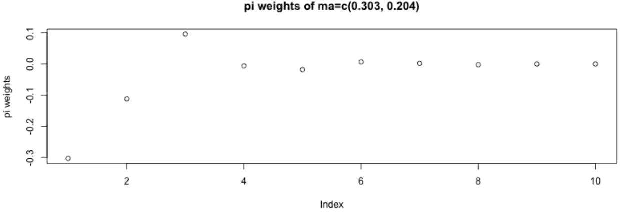
$$\hat{x}_t = 0.08 + 0.303\hat{w}_{t-1} + 0.204\hat{w}_{t-2} + \hat{w}_t, \qquad \hat{\sigma}_w = 0.00009$$

over 219 degrees of freedom.

ARIMA Modelling of US GNP Data (cont'd)

- Equivalent AR representation:

```
ARMAtoAR(ar=0, ma=c(0.303, 0.204), 10) # prints pi-weights
```



- Suggests that **AR**(1) May also fit well.
- Indeed:

```
arima(gnpgr, order=c(1, 0, 0))  # AR(1)
```

```
Coefficients:
         ar1   intercept
      0.3467      0.0083
s.e.  0.0627      0.0010
```

ARIMA Modelling of US GNP Data (cont'd)

Final Models for $y_t = \mathrm{GNP}_t$

- **ARIMA**$(0,1,2)$:

$$(1-B)\log(y_t) = .008 + (1 + .303B + .204B^2)w_t, \qquad \hat{\sigma}_w^2 = .0094$$

  on 219 degrees of freedom.

- **ARIMA**$(1,1,0)$:

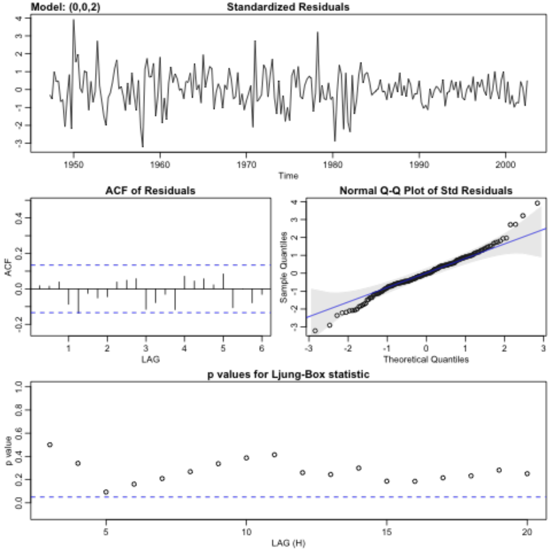
$$(1-.347B)(1-B)\log(y_t) = .008(1-.347) + w_t, \qquad \hat{\sigma}_w^2 = .0095$$

  on 220 degrees of freedom.

Next step: Diagnostics

Diagnostics for GNP Growth Rate (Example 3.40)

```
sarima(gnpgr, 0, 0, 2)  # MA(2)
```



Diagnostics

- Standardized residuals (innovations)

$$e_t = \frac{x_t - \hat{x}_t^{t-1}}{\sqrt{\hat{P}_t^{t-1}}}$$

$\hat{x}_t^{t-1}$ is the one-step ahead **prediction** of x_t based on the fitted model. $\hat{P}_t^{t-1}$ is the estimated one-step-ahead **error variance**.

- Inspect **marginal normality**: Q-Q plot of residuals

$$\Phi^{-1}\left(\frac{i - 1/2}{n}\right) \text{ vs } e_{(i)}, \qquad i = 1, \ldots, n$$

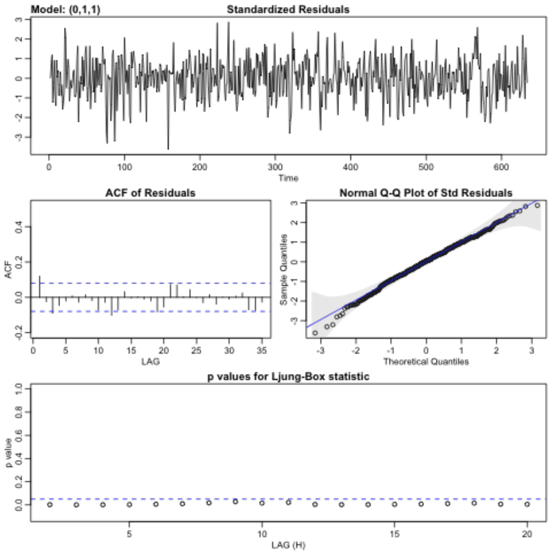
$e_{(i)}$ is the residuals i-th order statistics.

- Inspect $\hat{\rho}_e$ (the sample ACF of e_t) for **patterns or large values**.
- **Ljung-Box Test** (Portmanteau Test): Compare

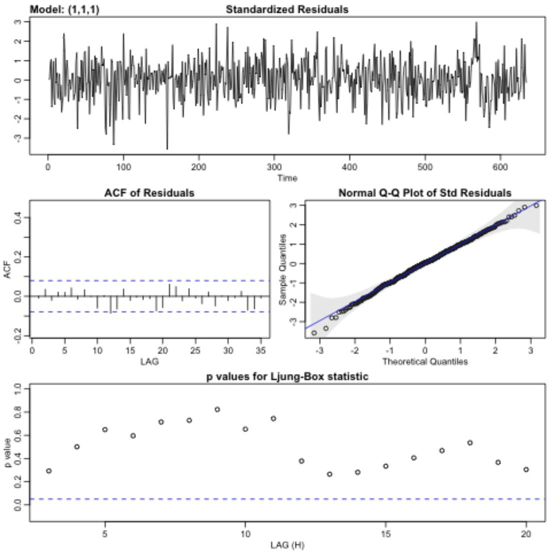
$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{\rho}_e(h)^2}{n-h}$$

to χ^2_{H-p-q}.

Diagnostics for Glacial Varve ARIMA$(0, 1, 1)$ (Example 3.41)



Diagnostics for Glacial Varve ARIMA$(1,1,1)$ (Example 3.41)



SARIMA

Seasonal ARMA Models

- Definition: Seasonal **ARMA**$(P, Q)_s$:

$$\Phi_P(B^s)x_t = \Theta_Q(B^s)w_t,$$

where

$$\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps},$$
$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \ldots + \Theta_Q B^{Qs},$$

(**seasonal AR operator** of order P and **seasonal MA operator** of order Q, with **seasonal period** s).

- Examples:
  - Annual **SARMA**$(1, 1)_{12}$:

$$x_t = \Phi x_{t-12} + w_t + \Theta w_{t-12}.$$

  - Quarterly **SMA**$(2)_4$:

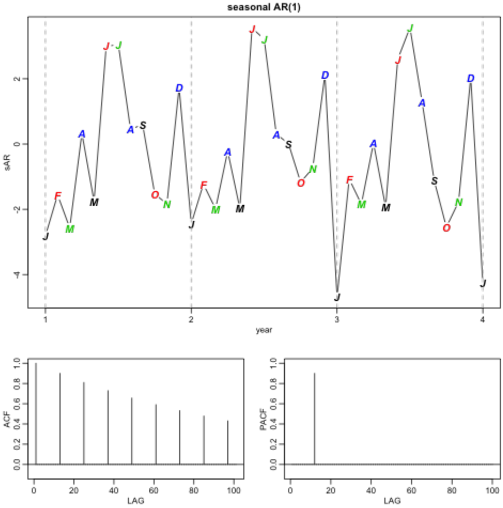
$$x_t = w_t + \Theta_1 w_{t-4} + \Theta_2 w_{t-8}.$$

  - Weekly **SAR**$(3)_{52}$:

$$x_t = w_t + \Phi_1 x_{t-52} + \Phi_2 x_{t-104} + \Phi_3 x_{t-156}.$$

Example 3.46: SAR(1)

$x_t = \Phi x_{t-12} + w_t$ (yearly seasonal period $s = 12$ months)



Contrasting the PACF and ACF

	SAR$(P)_s$	**SMA**$(Q)_s$	**SARMA**$(p, q)_s$
ACF(ks)	Decays	Cutoff $k > Q$	Decays
PACF(ks)	Cutoff $k > P$	Decays	Decays

Seasonal ARMA Models

- Definition: Seasonal **ARMA**$(p, q) \times (P, Q)_s$:

$$\Phi_P(B^s)\phi(B)x_t = \Theta_Q(B^s)\Theta(B)w_t,$$

where:

  - $\Phi_P(B^s) = 1 - \sum_{i=1}^{P} \Phi_i B^{si}$
  - $\Theta_Q(B^s) = 1 + \sum_{i=1}^{Q} \Theta_i \Phi_i B^{si}$
  - $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$
  - $\theta(B) = 1 + \sum_{i=1}^{q} \theta_i B^i$
  - w_t is white noise

- Definition: Seasonal **ARIMA**$(p, d, q) \times (P, D, Q)_s$:

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d x_t = \Theta_Q(B^s)\Theta(B)w_t,$$

where $\nabla_s^D = (1 - B^s)^D$.

Example 3.48: SARIMA$(0,1,1) \times (0,1,1)_{12}$

$$\nabla_{12}\nabla x_t = \Theta(B^{12})\theta(B)w_t$$

or

$$(1 - B^{12})(1 - B)x_t = (1 + \Theta B^{12})(1 + \theta B)w_t,$$

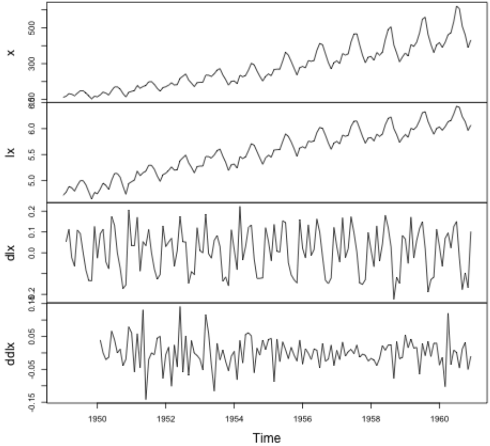
or

$$x_t = x_{t-1} + x_{t-12} - x_{t-13} + w_t + \theta w_{t-1} + \Theta w_{t-12} + \Theta\theta w_{t-13}.$$

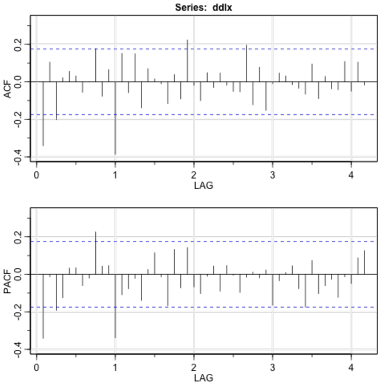
Example 3.49: Air Passengers

Monthly totals of international airline passengers between 1949 to 1960.

```
x = AirPassengers; lx = log(x);
dlx = diff(lx); ddlx  = diff(dlx, 12)
plot.ts(cbind(x, lx, dlx, ddlx), main="")
```



Example 3.49: Air Passengers (cont'd)



- **Seasonal components**: suggest $SMA(1)$, $P = 0$, $Q = 1$, in the season $s = 12$.
- **Non-Seasonal components**: suggest $\textbf{ARMA}(1, 1)$ within the seasons.

Example 3.49: Air Passengers (cont'd)

- Try to fit **SARIMA**$(1,1,1) \times (0,1,1)_{12}$ on the logged data:

```
sarima(lx, 1,1,1, 0,1,1,12)
```

```
      Estimate     SE t.value p.value
ar1     0.1960 0.2475  0.7921  0.4298
ma1    -0.5784 0.2132 -2.7127  0.0076
sma1   -0.5643 0.0747 -7.5544  0.0000
```

AR parameter is **not significant**.

- Try to fit **SARIMA**$(0,1,1) \times (0,1,1)_{12}$:

```
sarima(lx, 0,1,1, 0,1,1, 12)
```

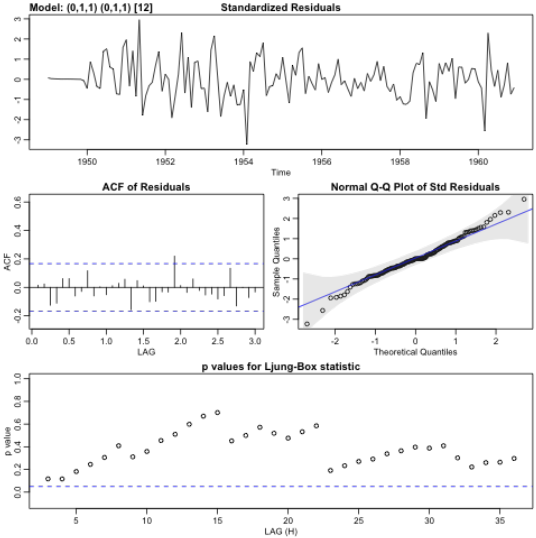
```
      Estimate     SE t.value p.value
ma1    -0.4018 0.0896 -4.4825       0
sma1   -0.5569 0.0731 -7.6190       0
$AIC -5.58133 $AICc -5.56625 $BIC -6.540082
```

- Try to fit **SARIMA**$(1,1,0) \times (0,1,1)_{12}$:

```
sarima(lx, 1,1,0, 0,1,1, 12)
```

```
     Estimate     SE t.value p.value
ar1    -0.3395 0.0822 -4.1295   1e-04
sma1   -0.5619 0.0748 -7.5109   0e+00
$AIC -5.567081 $AICc -5.552002 $BIC -6.525834
```

We pick **ARIMA**$(0,1,1) \times (0,1,1)_{12}$.



Example 3.49: Air Passengers (cont'd)

```
sarima.for(lx, 24, 0,1,1, 0,1,1,12)  # 24 months forecast
```