Ch6. Nonstationary and seasonal time series models

6.1 ARIMA(p, d, q) model for nonstationary TS

- Recall that we have used the decomposition

$$X_t = m(t) + Y_t$$

for trend parts $m(t)$ and stationary errors $\{Y_t\}$.

- By using polynomial regression

$$m(t) \approx \beta_0 + \beta_1 t + \ldots + \beta_p t^d$$

we were able to remove polynomial trend.

- Alternatively, d-th order differencing can also remove polynomial trend of d-th order.

$$(1 - B)^d X_t$$

- We model $\{Y_t\}$ by ARMA(p, q), so we can combine d-th order differencing in ARMA(p, q) model and this is called the ARIMA(p, d, q) model.

ARIMA(p, d, q) model for nonstationary TS

Definition (ARIMA(p, d, q))

If d is a non-negative **integer**, *then $\{X_t\}$ is an ARIMA(p, d, q) process if*

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t,$$

that is $(1-B)^d X_t$ follows usual ARMA(p, q).

- If $d = 0$, then it is usual ARMA(p, q)
- If $d \geq 1$, then it kills d-th order polynomial trend.
- Therefore, use ARIMA(p, d, q) model only if data looks like polynomial trend + ARMA(p, q) errors. Otherwise, consider other models!
- Main advantages of ARIMA model are its brief representation of model and forecasting (since they are still linear process!)

ARIMA(1,1,0)

ARIMA(1,1,0) is given by

$$(1-\phi B)(1-B)X_t = Z_t, \quad Z_t \sim WN(0,\sigma^2).$$

- For a causal stationary solution (after differencing) we need $|\phi| < 1$.
- Let $Y_t = (1-B)X_t = X_t - X_{t-1}$, and

$$(1-\phi B)Y_t = Z_t \Rightarrow Y_t \text{ is AR(1)}$$

- Also note that

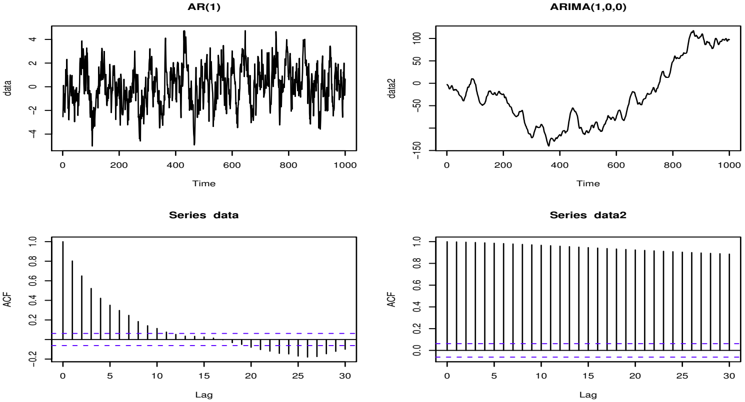
$$X_t = X_{t-1} + Y_t = (X_{t-2} + Y_{t-1}) + Y_t = ... = X_0 + \sum_{j=1}^{t} Y_j$$

  implies that observed series X_t is a cumulative sum of AR(1) proess, that is "random walk".

ARIMA(1,1,0) example

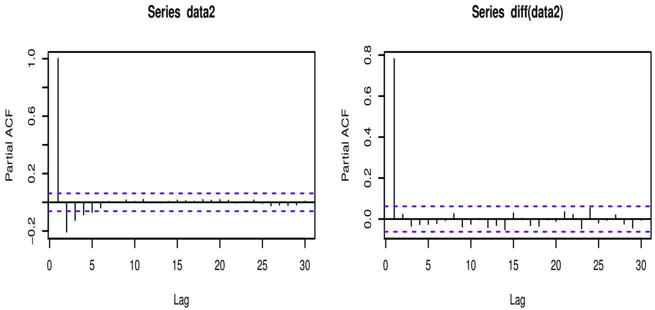
Consider

$$(1 - .8B)(1 - B)X_t = Z_t, \quad Z_t \sim \mathcal{N}(0, 1)$$



ARIMA(1,1,0) example

- Can observe linear trend and very slowly decaying SACF. Hence taking difference makes sense.
- Also, PACF from the observation indicates PACF(1) ≈ 1, hence also suggests differencing. After differencing PACF(1) $\approx .8$ and vanishes after lag 1, so AR(1) is suitable.



Estimation of ARIMA(p, d, q)

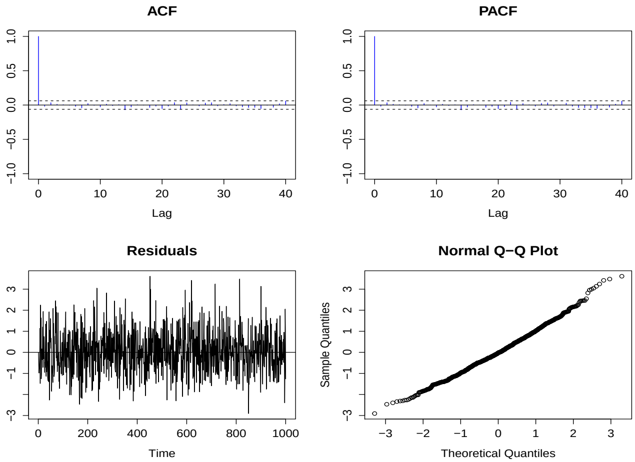
In principle, they are ARMA after d-th order differencing, so we can use MLE after $\nabla^d X_t$. In R, simply add order term, e.g, `arima(data2, order=c(1,1,0))`

```
Coefficients:
         ar1
      0.7825
s.e.  0.0196

sigma^2 estimated as 1.01:  log likelihood = -1422.81,  aic = 2849.63
```

ARIMA(1,1,0) diagnostics

Residual diagnostics: seems OK.



ARIMA(1,1,0) diagnostics

Formal tests give

```
Null hypothesis: Residuals are iid noise.
Test                        Distribution Statistic   p-value
Ljung-Box Q                Q ~ chisq(20)     15.02    0.7753
McLeod-Li Q                Q ~ chisq(20)     15.55    0.7439
Turning points T (T-665.3)/13.3 ~ N(0,1)       664    0.9203
Diff signs S      (S-499.5)/9.1 ~ N(0,1)       492    0.4115
Rank P        (P-249750)/5274.4 ~ N(0,1)    248388    0.7962
```

6.2 Identification Techniques

Sources of non-stationarity

- Deterministic trend: can be removed by regression / smoothing / differencing or ARIMA(p, d, q)
- Seasonality: can be removed by Harmonic regression / seasonal smoothing / seasonal differencing
- non-homogeneous variance: In general, we consider

$$X_t = m(t) + \sigma(t) Y_t, \quad Y_t \sim \text{ ARMA}(p, q).$$

This is beyond the scope of this course, so we will only consider basic ways to remove heteroscedasticity by simple transformation.

Variance Stablizing Transformation (VST)

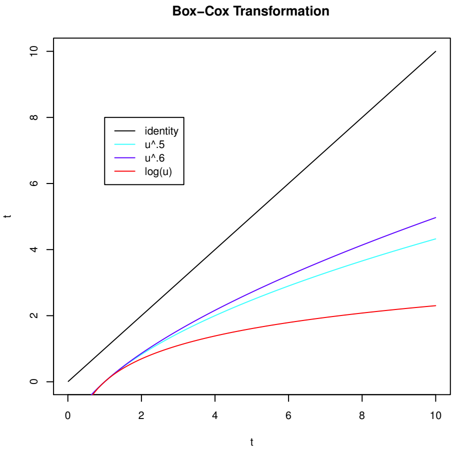
If the variance of $\{U_t\}$ is increasing, hence we apply the Box-Cox transformation

$$f_\lambda(U_t) = \begin{cases} \dfrac{U_t^\lambda - 1}{\lambda}, & U_t \geq 0, \lambda > 0, \\ \log U_t, & \lambda = 0 \end{cases}$$

(Note that $\lim_{\lambda \downarrow 0}(u^\lambda - 1)/\lambda = \log u$ by L'Hopital's rule.)

Idea: Approximately it equals to $f_\lambda(U_t) \approx U_t^\lambda$. (power transform.)
$\Rightarrow$ As $U_t \uparrow$, power transformation with $\lambda \in (0,1)$ downweights U_t, hence it reduces variance as $t \to \infty$.

Box-Cox transformation



Choice of λ? Appeal to residual plot which looks homogeneous variance or likelihood (or some other measures of goodness of fit).

Model estimation / identification

- General guideline for fitting ARMA(p, q) model and order selection is to assess goodness of fit by looking at
  - Residual plots to check (WN/IID/Gaussianity etc), or you can apply formal tests such as Portmanteau, rank, difference sign, Jacque-Bera etc.
  - Individual parameter estimates are significantly away from zero.
- What can be done if some of parameters are significantly away from zero but some are not?
- We can fit a subset of model by offsetting specific parameter estimates $= 0$. This is called the "constraint optimization".

Constraint optimization

- Example: ARMA(3,3) with

  $$X_t = \phi_3 X_{t-3} + 0X_{t-2} + 0X_{t-1} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \theta_3 Z_{t-3},$$

  that is,

  $$\phi_2 = \phi_1 = 0.$$

  In R, you can still use `arima()` with

  `arima(data, c(3,0,3), fixed =c(0, 0, NA, NA, NA, NA))`

- Constraint optimization can also be used to do model selection as in the regression (backward selection). For example, start with the largest model, say,

  AR(12)

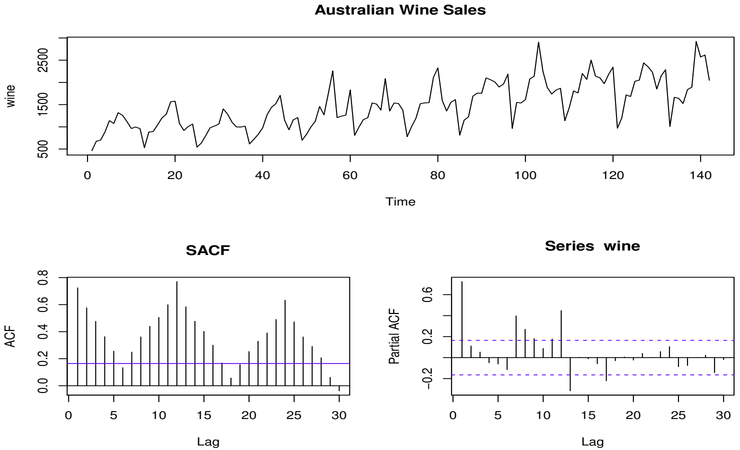
  Keep removing zero coefficients

  ⋮

  find the optimal model

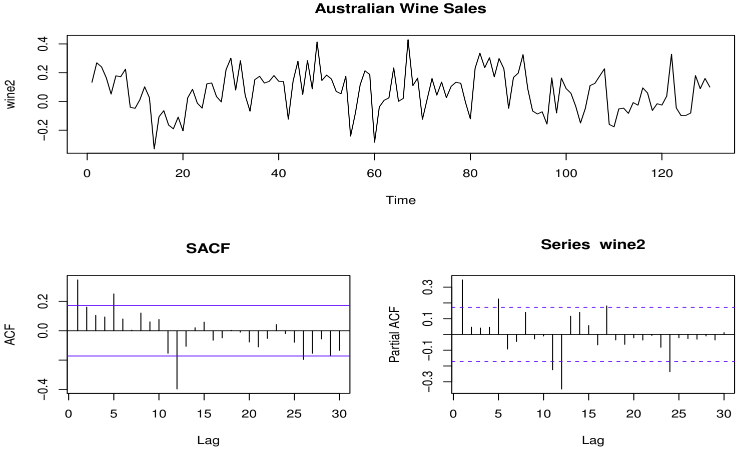
  However, you still need some diagnostics to ensure that the final model is the best model.

Example: Australian Wine



- Can observe increasing variance, and seasonality.
- Will take log-transformation and seasonal difference of lag 12.

Example: Australian Wine



- Major seasonality has disappeared, but still can observe some of them.
- Trend is not so obvious, so we will only subtract mean.

Example: Australian Wine

- From the PACF, it suggests AR(12) model:

```
Series: wine2
ARIMA(12,0,0) with zero mean

Coefficients:
         ar1     ar2     ar3     ar4     ar5      ar6      ar7
      0.2431  0.0695  0.0191  0.0260  0.1958  -0.0269  -0.0096
s.e.  0.0802  0.0827  0.0823  0.0818  0.0808   0.0844   0.0828
         ar8     ar9    ar10     ar11     ar12
      0.1355  0.0062  0.0931  -0.1203  -0.3733
s.e.  0.0817  0.0845  0.0840   0.0841   0.0815

sigma^2 estimated as 0.01346:  log likelihood=94
AIC=-162   AICc=-158.87   BIC=-124.73
```

- Some AR coefficients are quite close to zero, so we conducted test of

$$\mathrm{H}_0 : \phi_j = 0 \quad \mathrm{H}_1 : \phi_j \neq 0.$$

  In R, manually calculate them by

```
2*(1-pnorm(abs(ar12.out$coef/(sqrt(diag(ar12.out$var.coef))))))
```

Example: Australian Wine

- ```
       ar1      ar2      ar3      ar4      ar5      ar6      ar7 
  2.45e-03 4.00e-01 8.17e-01 7.51e-01 1.54e-02 7.50e-01 9.08e-01 
       ar8      ar9     ar10     ar11     ar12 
  9.72e-02 9.42e-01 2.68e-01 1.52e-01 4.63e-06 
  ```
- coefficients can set zero for $2, 3, 4, 6, 7, 9, 10$. $(\alpha \approx .20)$

  ```
  ARIMA(12,0,0) with zero mean

  Coefficients:
          ar1  ar2  ar3  ar4    ar5  ar6  ar7    ar8  ar9  ar10     ar11
        0.270    0    0    0  0.224    0    0  0.150    0     0   -0.099
  s.e.  0.072    0    0    0  0.072    0    0  0.072    0     0    0.078
           ar12
        -0.352
  s.e.   0.078

  sigma^2 estimated as 0.0138:  log likelihood=92.58
  AIC=-173.17   AICc=-170.03   BIC=-135.89
  ```
- Can check that AIC has decreases from -158.87 to -170.03.
  ```

6.4 Forecasting ARIMA model

Want to find BLP (Best Linear Predictor) of X_{n+h} for ARIMA.

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

$$\phi(B) Y_t = \theta(B) Z_t, \quad Y_t = (1-B)^d X_t$$

Since we know BLP of Y_t, we can use this.

- For $d = 1$, note that

$$Y_t = X_t - X_{t-1} \Rightarrow X_{n+1} = X_n + Y_{n+1}$$

$$\boxed{\text{BLP of } X_{n+1} = X_n + \text{ BLP of } Y_{n+1}}$$

BLP of Y_{n+1} = linear combination of $\{1, Y_1, Y_2, \ldots, Y_n\}$ and this can be done exactly same as for ARMA(p, q).

Forecasting ARIMA model

- If $d = 2$,

  $$Y_t = (1-B)^2 X_t = X_t - 2X_{t-1} + X_{t-2} \Rightarrow X_{n+1} = 2X_n - X_{n-1} + Y_{n+1}$$

  $$\boxed{\text{BLP of } X_{n+1} = 2X_n - X_{n-1} + \text{BLP of } Y_{n+1}}$$

- However, differencing will only give you $\{Y_3, \ldots, Y_n\}$. Trick is to use X_{-1}, X_0 as zero (since it is zero-mean process) or use $\overline{Y}$.

Forecasting ARIMA model - Example

Consider Dow Jones Index (Aug 28 - Dec 18, 1972). The model was

$$X_t - .4471X_{t-1} = Z_t, \quad Z_t \sim WN(0, .1455),$$

where $X_t = D_t - D_{t-1} - .1336$. The most recent observation $D_{77} = 121.23$, and $D_{76} = 122$.

- Recall that for AR(1) model, h-step ahead forecast is given by $\phi^h X_t$. Thus, 1-step ahead forecast $\widehat{X}_{78} = P_{77}X_{78} = \hat{\phi}_1 X_{77} = .4471*(D_{77}-D_{76}-.1336) = -.404$
- Therefore, $D_t = X_t + D_{t-1} + .1336$ gives that

$$\widehat{D}_{78} = P_{77}D_{78} = \widehat{X}_{78} + D_{77} + .1336 = 120.96$$

- What will be $\widehat{D}_{79} = P_{77}D_{79}$? (120.91)

6.5 Seasonal ARIMA models

- ARMA model with seasonal components
- **Motivation** Consider a monthly TS $\{Y_t\}$ with $s = 12$.

	month 1	month2	...	month12
Year 1	Y_1	Y_2	...	Y_{12}
Year 2	Y_{13}	Y_{14}	...	Y_{24}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$	...	$Y_{12+12(r-1)}$

Classical decomposition models

$$Y_t = s_t + Z_t,$$

where s_t is a deterministic seasonal component. However, there is no reason to assume $\{s_t\}$ is fixed. We want a model allows seasonal component varies randomly from one cycle to the next.

Seasonal ARIMA models

STEP1 Fix column, and assume it follows ARMA(P, Q) (between year model). For lag $s = 12$,

$$Y_{j+12t} - \Phi_1 Y_{j+12(t-1)} - \ldots - \Phi_P Y_{j+12(t-P)} =$$

$$U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \ldots + \Theta_Q U_{j+12(t-Q)}, \quad t = 0, 1, \ldots, 11.$$

$$\iff \Phi(B^{12}) Y_t = \Theta(B^{12}) U_t, \quad U_t \sim WN(0, \sigma_U^2).$$

STEP2 Relax $U_t \sim WN(0, \sigma_U^2)$ by assuming ARMA(p, q) model.
Why?
Following by each row,

$$\{Y_1, Y_2, \ldots, Y_{12}\} \iff \{U_1, U_2, \ldots, U_{12}\}$$

can have some dependence structure! (Between month model)

$$\phi(B) U_t = \theta(B) Z_t, \quad Z_t \sim WN(0, \sigma^2)$$

SARIMA$(p, d, q) \times (P, D, Q)$

STEP3 Finally combining two gives

$$\Phi(B^{12})Y_t = \Theta(B^{12})\phi^{-1}(B)\theta(B)Z_t$$

$$\boxed{\phi(B)\Phi(B^{12})Y_t = \theta(B)\Theta(B^{12})Z_t, \quad Z_t \sim WN(0, \sigma^2)}$$

STEP4 Even, we can include differencing

$$Y_t = (1 - B)^d(1 - B^{12})^D X_t,$$

hence SARIMA$(p, d, q) \times (P, D, Q)$ is given by

$$\boxed{\phi(B)\Phi(B^{12})(1 - B)^d(1 - B^{12})^D X_t = \theta(B)\Theta(B^{12})Z_t,}$$

where $Z_t \sim WN(0, \sigma^2)$

- SARIMA is a linear process removes trend, seasonality in one model!

SARIMA - practical guideline

As in the case of ARMA model, order selection is important!

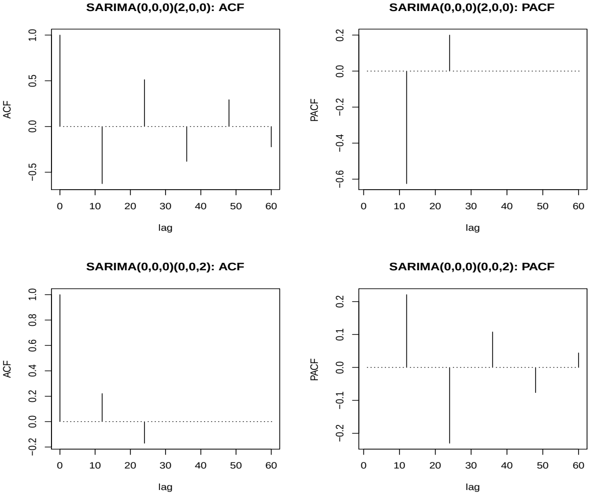
1. Transform data if necessary (take log, Box-Cox transformation)
2. Determine whether it need differencing or seasonal differencing

$$Y_t = (1 - B)^d (1 - B^{12})^D X_t$$

3. Examine ACF/PACF to determine order.
   For seasonal (P, Q): focus on SACF(ks), SPACF(ks)
   For ARMA(p, q): focus on SACF(k), SPACF(k), $k = 1, \dots, s - 1$. See example in the next to guess P, Q.

SARIMA - practical guideline

$$(1 + .5B^{12} - .2B^{24})X_t = Z_t, \quad X_t = (1 + 1.3B^{12} - .5B^{24})Z_t$$



SARIMA - practical guideline

4. Fit SARIMA model. In R, add seasonal term, e.g, `arima(data, order=c(p,d,q), seasonal= list(order=(P,D,Q),period=s), include.mean=FALSE)`

5. Do order selection by AICC/BIC criteria.
6. Do forecasting. Forecasting is essentially based on linear process

$$Z_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j},$$

so in R, use `forecast(model, n.ahead=h)`.

Difference between SARIMA and ARIMA

Conisder SARIMA(1,1,0)x(1,0,0)

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)X_t = Z_t$$

and

$$(1-\phi_1 B)(1-\Phi_1 B^{12}) = 1-\phi_1 B - \Phi_1 B^{12} + \phi_1 \Phi_1 B^{13}$$

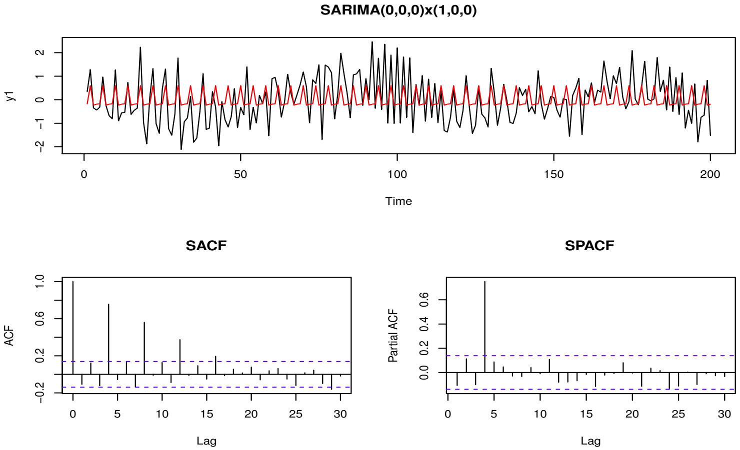
Therefore, SARIMA(1,1,0)x(1,0,0) is equivalent to ARIMA(13, 1,0) with constraint $\phi_{13} = \phi_1 \Theta_1$.

- SARIMA model gives parsimonious representation of ARIMA model with lots of zeros in the middle. Thus, it gives easier interpretation and stable parameter estimation.

Sample path of SARIMA(0,0,0)x(1,0,0) model

One realization of SARIMA(0,0,0)x(1,0,0) with period $d = 4$,

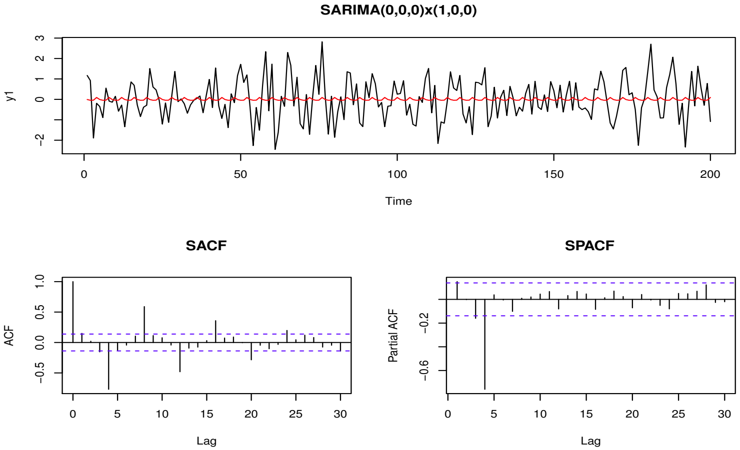
$$(1 - .8B^4)X_t = Z_t.$$



Sample path of SARIMA(0,0,0)x(1,0,0) model

One realization of SARIMA(0,0,0)x(1,0,0) with period $d = 4$,

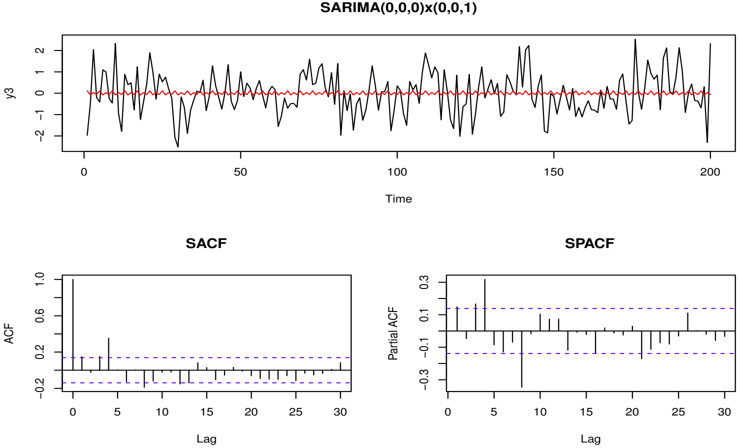
$$(1 + .8B^4)X_t = Z_t.$$



Sample path of SARIMA(0,0,0)x(0,0,1) model

One realization of SARIMA(0,0,0)x(0,0,1) with period $d = 4$,

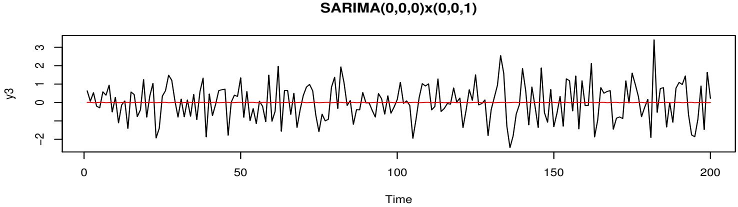
$$X_t = (1 + .7B^4)Z_t.$$

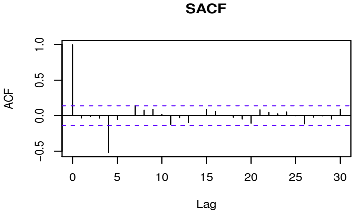


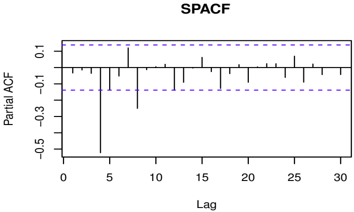
Sample path of SARIMA(0,0,0)x(0,0,1) model

One realization of SARIMA(0,0,0)x(0,0,1) with period $d = 4$,

$$X_t = (1 - .7B^4)Z_t.$$



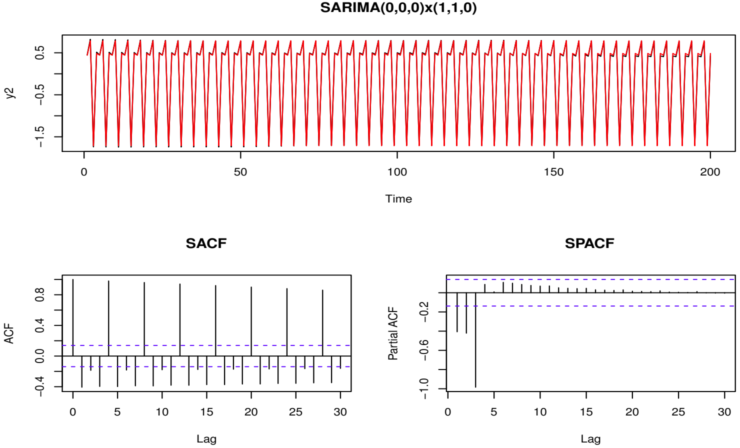




Sample path of SARIMA(0,0,0)x(1,1,0) model

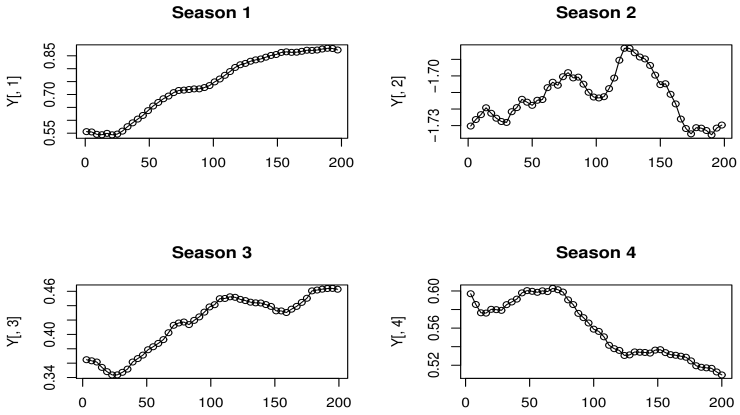
One realization of SARIMA(0,0,0)x(1,1,0) with period $d = 4$,

$$(1 - B^4)(1 - .8B^4)X_t = Z_t.$$



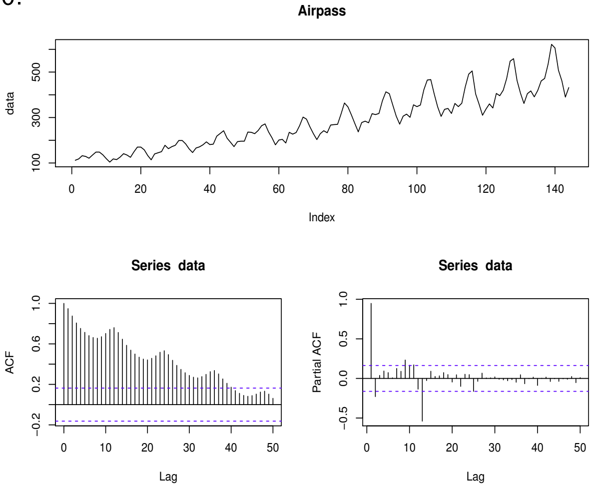
Sample path of SARIMA(0,0,0)x(1,1,0) model

It is better to see by each season. It is then ARIMA(1,1,0)



SARIMA: Airpass data example

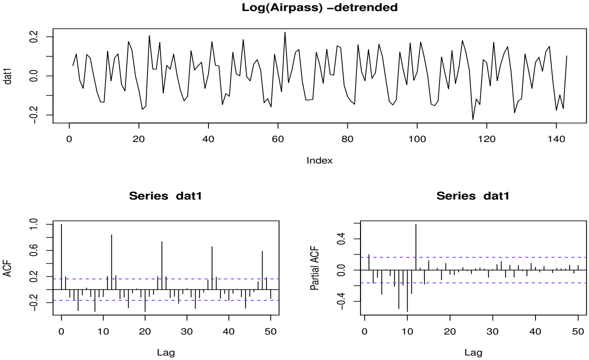
Monthly total international airline passengers from Jan 1949 to Dec 1960.



- Increasing variance, linear trend and seasonality!
- Will take log-transformation and lag-1 differencing first.

SARIMA: Airpass data example

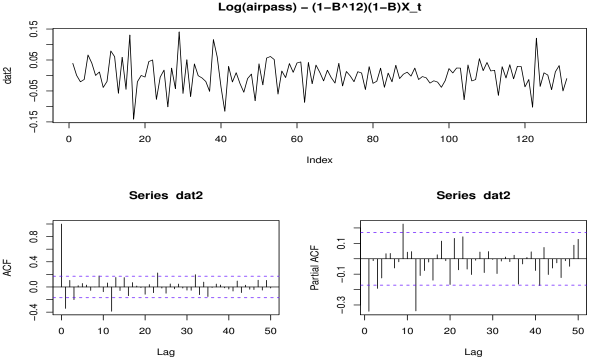
$$(1-B)X_t, \quad X_t = \log Y_t$$



- Trend is removed, but it still has some seasonality because ACF(12), ACF(24), ACF(36) and PACF(12) is surviving.

SARIMA: Airpass data example

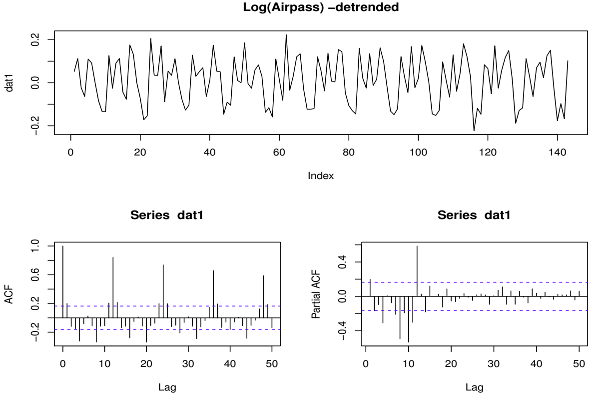
Do we need seasonal differencing or SAR(1) is enough? After seasonal differencing you would have



- Well, in the middle of TS plot, they don't look stationary. Maybe better to only differencing for trend.

SARIMA: Airpass data example

Determine the order of SARIMA? After trend differencing you would have



- PACF(12) is very significant, and PACF(1) is moderate. Consider SARIMA(1,1,0)x(1,0,0) and SARIMA(0,1,0)x(1,0,0)

SARIMA: Airpass data example

- SARIMA(1,1,0)x(1,0,0)

```
Call:
arima(x = dat, order = c(1, 1, 0),
seasonal = list(order = c(1, 0, 0), period = 12))
Coefficients:
          ar1    sar1
      -0.2905  0.9287
s.e.   0.0822  0.0229

sigma^2 estimated as 0.001777:  log likelihood = 237.94,  aic = -469.89
```

- SARIMA(0,1,0)x(1,0,0)

```
arima(x = dat, order = c(0, 1, 0),
seasonal = list(order = c(1, 0, 0), period = 12))

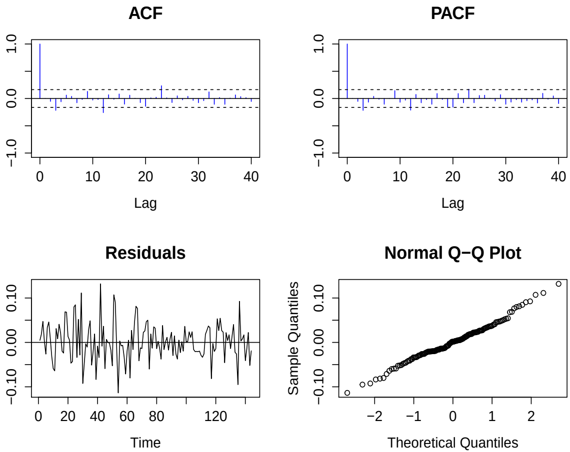
Coefficients:
        sar1
      0.9032
s.e.  0.0278

sigma^2 estimated as 0.001978:  log likelihood = 232.08,  aic = -460.17
```

- SARIMA(1,1,0)x(1,0,0) seems better based on aic.

SARIMA: Airpass data example

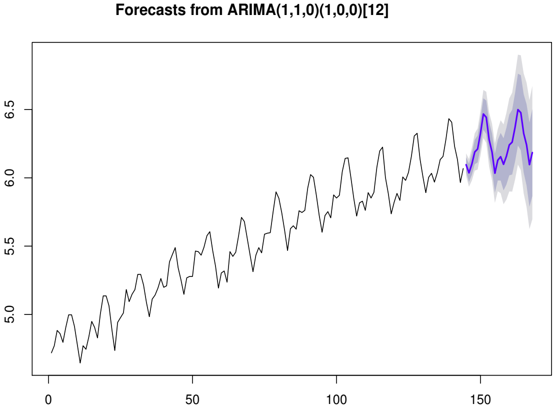
Residuals seems fine:



SARIMA: Airpass data example

Forecasting, for example lag 12 gives

```
> plot(forecast(fit.1, n.ahead=12))
```



6.6 Regression with ARMA errors

- Recall classical decomposition

$$Y_t = m(t) + s(t) + W_t$$

  - Trend

$$m(t) \approx \beta_0 + \beta_1 t + \ldots + \beta_d t^d$$

  - Seasonality

$$s(t) = a_0 + \sum_{j=1}^{k} a_j \cos(\lambda_j t) + b_j \sin(\lambda_j t)$$

- Therefore, in general, we can represent it by

$$\boxed{Y_t = \beta_0 + \sum_{j=1}^{p} \beta_j x_{tj} + W_t}$$

- Regression with "dependent" errors $\{W_t\}$.

Regression - OLS

- In a matrix form, write it as

$$\boxed{Y = X\beta + W, \quad E(W) = 0, \quad \mathrm{Cov}(W) = \Gamma}$$

- OLS(Ordinary Least Squares) estimator

$$\widehat{\beta}^{OLS} = \underset{\beta}{\mathrm{argmin}} (Y - X\beta)'(Y - X\beta)$$

$$\widehat{\beta}^{OLS} = (X'X)^{-1}X'Y$$

- We will argue that OLS estimation is not good for inference of β when errors are correlated.

Properties of OLS with IID errors

If $\mathrm{Cov}(W) = \sigma^2 I$, that is in IID case, then OLS estimator enjoys

- **Gauss-Markov Theorem**
  - $\widehat{\beta}^{OLS}$ is BLUE(Best Linear Unbiased Estimator) of β
  - Amongst all linear UE of β, $\widehat{\beta}^{OLS}$ achieves the smallest variance

$$\mathrm{Var}(\widehat{\beta}^{OLS}) \leq \mathrm{Var}(\widehat{b}), \quad E(\widehat{b}) = \beta$$

- **Asymptotic Normality**

$$\widehat{\beta}^{OLS} \approx \mathcal{N}(\beta, \sigma^2 (X'X)^{-1})$$

- Estimation of σ^2

$$\widehat{\sigma}^2 = \frac{(Y - X\widehat{\beta}^{OLS})'(Y - X\widehat{\beta}^{OLS})}{n - p - 1}$$

$$(n - p - 1)\frac{\widehat{\sigma}^2}{\sigma^2} \sim \chi^2(n - p - 1)$$

OLS with dependent errors

If erros are correlated $\mathrm{Cov}(W) = \Gamma$, note that

- Still unbiased

$$E(\widehat{\beta}^{OLS}) = E((X'X)^{-1}X'Y) = E((X'X)^{-1}X'(X\beta + W))$$

$$= \beta + E((X'X)^{-1}X'W) = \beta$$

- However,

$$\mathrm{Cov}(\widehat{\beta}^{OLS}) = \mathrm{Cov}((X'X)^{-1}X'Y)$$

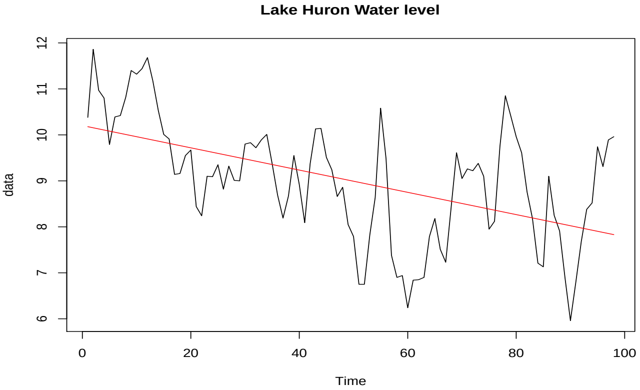
$$= (X'X)^{-1}X' \,\mathrm{Cov}(Y)((X'X)^{-1}X')'$$

$$= (X'X)^{-1}X'\Gamma X(X'X)^{-1} \neq (X'X)^{-1}\sigma^2$$

Recall $\mathrm{Cov}(AY) = A\,\mathrm{Cov}(Y)A'$.

Practical implication

You shouldn't believe the inference from `lm()` output when errors are correlated. For example, lake data



Practical implication

Estimates are consistent, but DO NOT perform inference on coefficients!

```
Call:
lm(formula = data ~ 1 + x)

Residuals:
     Min       1Q   Median       3Q      Max
-2.50997 -0.72726  0.00083  0.74402  2.53565

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.202037   0.230111  44.335  < 2e-16 ***
x           -0.024201   0.004036  -5.996 3.55e-08 ***
---
Signif. codes:  0 ??***?? 0.001 ??**?? 0.01 ??*?? 0.05 ??.?? 0.1 ?? ?? 1

Residual standard error: 1.13 on 96 degrees of freedom
Multiple R-squared: 0.2725,     Adjusted R-squared: 0.2649
F-statistic: 35.95 on 1 and 96 DF,  p-value: 3.545e-08
```

GLS with correlated errors

Since Γ is a covariance of (weakly stationary process) $\{W_t\}$, it is **symmetric** and **non-negative definite**.

Theorem (Spectral Theorem)

Every real symmetric matrix $A_{p\times p}$ is diagonalizable

$$A = PDP'$$

$$PP' = P'P = I, \quad D = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$$

Furthermore, all eigenvalues are real

- If A is further postive definite, all eigenvalues are strictly positive and

$$A^{-1} = PD^{-1}P'$$

- By taking $D^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_p})$

$$A = PD^{1/2}P'PD^{1/2}P' = A^{1/2}A^{1/2}.$$

GLS with correlated errors

Therefore, if we further assume that Γ is positive definite, then there exists $\Gamma^{1/2}$ such that

$$\Gamma = \Gamma^{1/2}\Gamma^{1/2}$$

By multiplying $\Gamma^{-1/2}$ on both sides

Properties of GLS

- Note that OLS minimizes

$$(Y - X\beta)'(Y - X\beta),$$

  but GLS minimizes

$$(Y - X\beta)'\Gamma^{-1}(Y - X\beta),$$

  which is a distance measure incorporating Γ (effect of correlations). It is called Mahalanobis distance.
- $\widehat{\beta}^{GLS} = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}Y.$
- $E(\widehat{\beta}^{GLS}) =$
- $\mathrm{Cov}(\widehat{\beta}^{GLS}) =$

Properties of GLS

- **Gauss-Markov Theorem**: $\widehat{\beta}^{GLS}$ is BLUE for β minimizing

$$(Y - X\beta)'\Gamma^{-1}(Y - X\beta)$$

- For model

$$Y = X\beta + W, \quad W \sim WN(0, \Gamma),$$

  $\widehat{\beta}^{GLS}$ is more efficient estimator in the sense that $\mathrm{Cov}(\widehat{\beta}^{OLS}) - \mathrm{Cov}(\widehat{\beta}^{GLS})$ is non-negative definite. That is,

$$\mathrm{Var}(c'\widehat{\beta}^{GLS}) \leq \mathrm{Var}(c'\widehat{\beta}^{OLS})$$

- Thus, inference should be based on

$$\boxed{\widehat{\beta}^{GLS} \sim \mathcal{N}\left(\beta, (X'\Gamma^{-1}X)^{-1}\right)}$$

Numerical algorithm for GLS

- Iterative algorithm is used

STEP1 Estimate $\widehat{\beta}$ by OLS (initial estimator).

STEP2 Calculate residuals $Y - X\widehat{\beta}$, then find the best ARMA(p, q) model.

STEP3 For estimated Γ from the model in STEP2, re-estimate β now by $\widehat{\beta}^{GLS}$.

STEP4 Iterate STEP2-STEP4 until it converges, say

$$|\widehat{\beta}^{new} - \widehat{\beta}^{old}| \leq \epsilon$$

- Alternatively, fit MLE by including $X\beta$ term in the likelihood.

$$L(\eta, \beta) = |2\pi\Gamma(\eta)|^{-1/2} \exp\left(-\frac{1}{2}(Y - X\beta)'\Gamma^{-1}(Y - X\beta)\right).$$

  Take $Y^* = Y - X\beta$ then apply approach discussed in Ch5.

- Thus, together with ARMA(p, q) errors, MLE is more widely used in practice. Asymptotics are the same as GLS.

Lake Huron: GLS/MLE

```
resi = out.lm$residuals
const = rep(1,n);
time = 1:n;
X = cbind(const, time);
fit.reg = arima(data, order=c(2,0,0), xreg=X, include.mean=FALSE)

Series: data
ARIMA(2,0,0) with zero mean

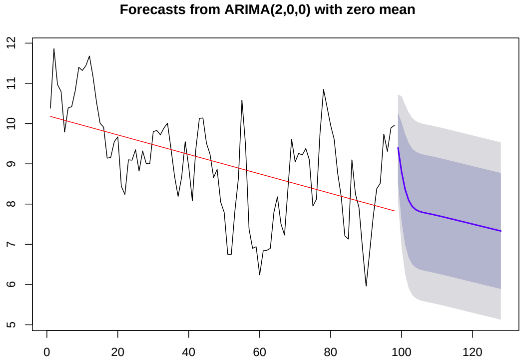
Coefficients:
         ar1      ar2    const     time
      1.0048  -0.2913  10.0915  -0.0216
s.e.  0.0976   0.1004   0.4636   0.0081

sigma^2 estimated as 0.4566:  log likelihood=-101.2
AIC=212.4   AICc=213.05   BIC=225.32
```

- What is your conclusion about the linear trend? Do inference based on $t = -.0216/.0081 = -2.667$.

Lake Huron: Forecasting

```
library(forecast)
h=30;
newx = (n+1):(n+h);
plot(forecast(fit.reg, h=30, xreg = cbind(rep(1, h), newx)))
lines(out.lm$fitted.values, col="red")
```



Lake Huron: Forecasting

Comparison with OLS and GLS with dependent error: