Time Series Regression - Part II

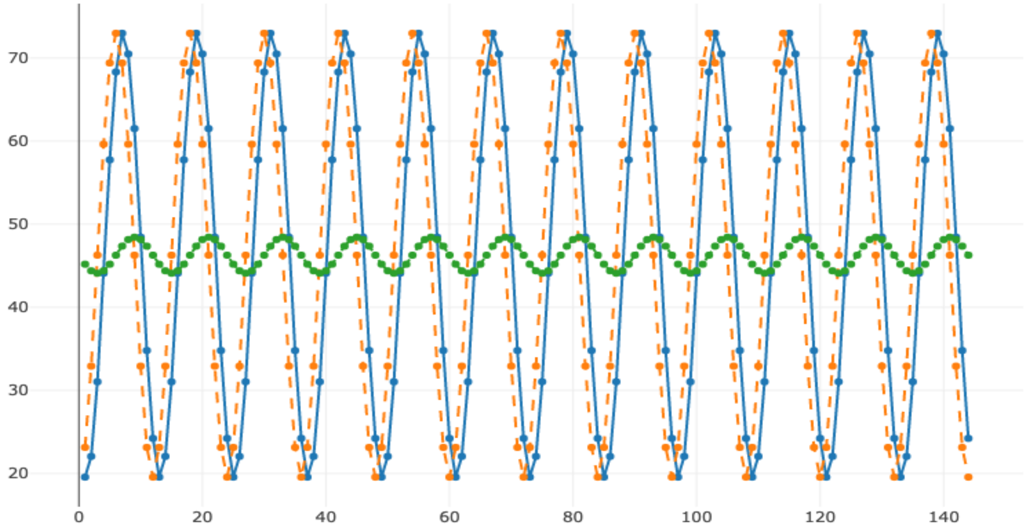
STAT 1321/2320

Kiran Nihlani

Department of Statistics
University of Pittsburgh

Outline

Harmonic Regression - Dubuque, Iowa



Harmonic Regression - Dubuque, Iowa

- The blue curve is the fitted values.
- The orange (dashed) curve is $\beta_0 + \beta_1 cos(2\pi ft)$
- The green (dotted) curve is $\beta_0 + \beta_2 sin(2\pi ft)$

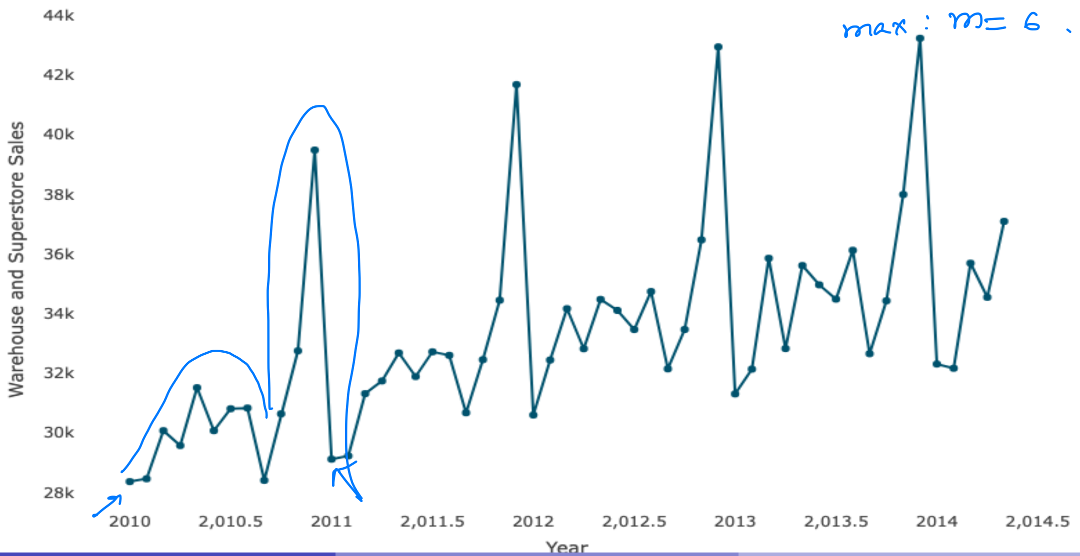
Harmonic Regression - Sales

- The temperature data for Dubuque is pretty clean - an annual cycle with not much noise. We were able to extract most information considering just one frequency (annual).
- Most real life data are not so clean and may contain observations at many frequencies and thus need to be analyzed with more harmonic components.
- Consider the sales data for Wholesale and Superstores from January 2010 to May 2014.

Harmonic Regression - Sales

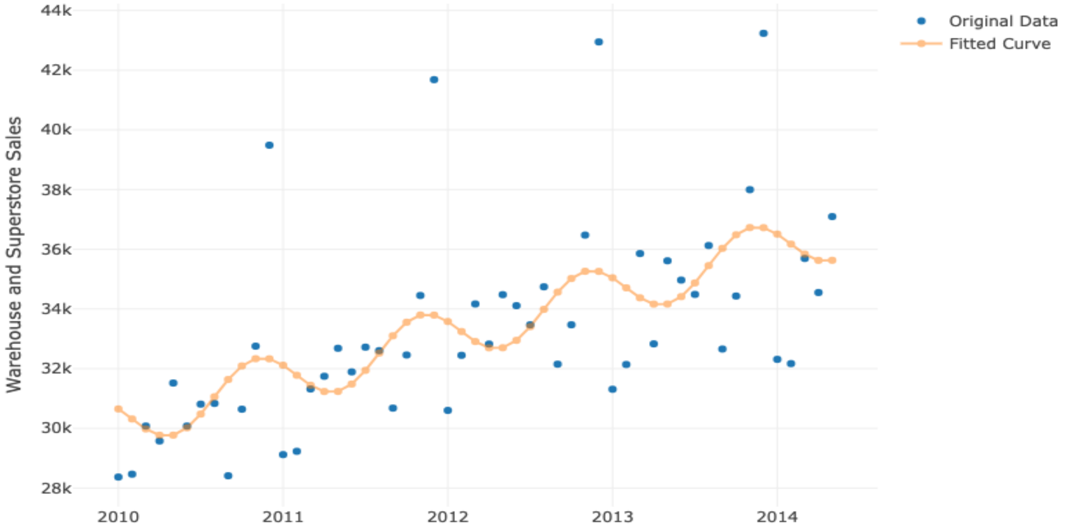
seasonality period = 12

max : m = 6

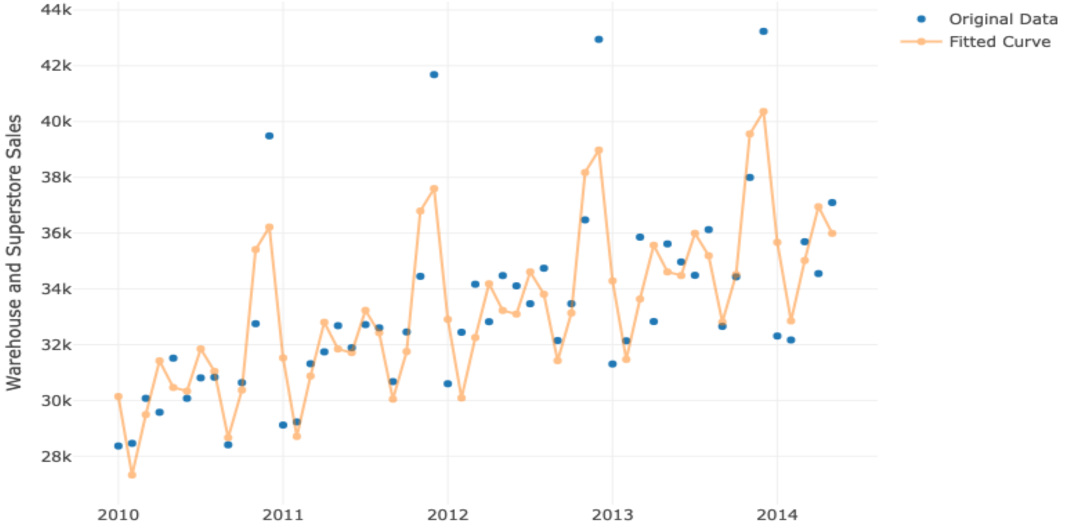


Frequency = 1 (R-sq = 0.321)

$m = 1$, 1 pair $\to \cos(2\pi t)$, $\sin(2\pi t)$.



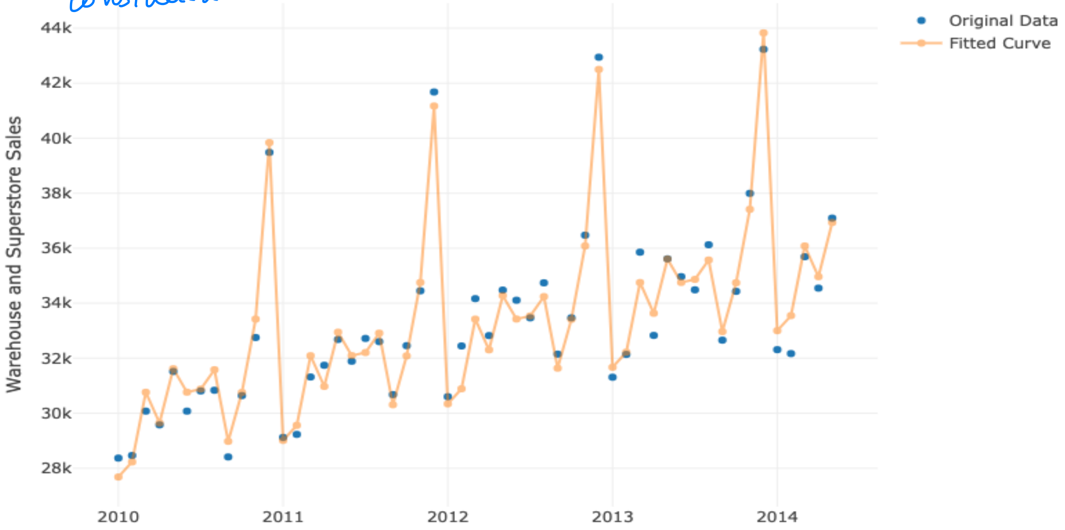
Frequency = 3 (R-sq = 0.691)



Frequency = 6 (R-sq = 0.964)

of ...ies considered.

m = 6.



Comparison Metrics - ANOVA Identity

X_t

- In a regression setup, the ANOVA identity can be written as:

$$\text{Total variation in } X_t = \text{Variation due to Model} + \text{Variation due to Error}$$

SST

$$\text{TSS} = \text{SSR} + \text{SSE}$$

- If a model fits the data well:
  - Most of the variation should be because of the model: (relatively) high SSR
  - Little variation explained by other factors not accounted by the model (error): (relatively) low SSE
- R^2 can be used to assess the proportion of variation explained by the model.
- An adjusted R^2 measure is similar, but penalizes models with more parameters.

$$TSS = SSR + SSE$$

$$R^2 = \frac{SSR}{TSS} = 1 - \frac{SSE}{TSS}.$$

$$\text{Adj } R^2 = \frac{SSR / df_{reg}}{TSS / df_T}.$$

Wholesale and Superstore Sales - lm() summary

R^2

Number of frequencies	RSE	Multiple R-squared	Adjusted R-squared
1	2759.573	0.360	0.321
2	2359.409	0.552	0.504
3	1863.179	0.732	0.691
4	1451.423	0.845	0.812
5	897.923	0.943	0.928
6	635.959	0.972 ✓	0.964 ✓

RSE $\downarrow$. $R^2 \uparrow$, adj. R^2.

RSE: Residual standard error

Information Criteria: Beyond R^2

- The Akaike Information Criterion (AIC), Bias corrected AIC (AICc), and Bayesian Information Criterion (BIC) may be used.

joint prob. dist
based on the
parameter

- These are likelihood-based model selection criteria.
- They are smaller for better fitting models, but they also penalize models with more parameters.
- One model selection strategy is to pick the model with the smallest value of the chosen criterion.

Information Criteria

For a model with k estimated parameters and ~~log~~-likelihood given by L:

- Akaike Information Criterion (AIC)

$$AIC = 2k - 2ln(L)$$

- Bias corrected AIC (AICc)

$$AICc = AIC + \boxed{\frac{2k(k+1)}{n-k-1}} \to 0$$

$$\lim_{n \to \infty} AICc$$

$$AICc \simeq AIC.$$

- Bayesian Information Criterion (BIC)

prior dist

$$BIC = k\ ln(n) - 2ln(L)$$

Information Criteria

prediction. cons. estimation.

AIC. BIC.

- AICc is better than AIC for smaller sample sizes.
- BIC has harsher penalty than AIC for additional parameters and also accounts for sample size.
- These criteria should mostly agree for the “best” model. If they don’t, report it.
- AIC and BIC are intended for different purposes and should be preferred accordingly.
- It only makes sense to compare models via these criteria if the response variable is exactly the same for both models.

Wholesale and Superstore Data

smallest

Number of frequencies	AIC	AICc	BIC
1	996.07	997.35	1005.92
2	981.25	983.74	995.05
3	957.92	962.11	975.65
4	933.04	939.48	954.71
5	883.61	892.94	909.22
6	847.74	858.79	875.32

Timeline for Analyzing Models

- Explored decomposition of time series
  - Additive Model: $X_t = T_t + S_t + C_t + I_t$
  - Multiplicative Model: $X_t = T_t \times S_t \times C_t \times I_t$

$$\log X_t = \log T_t + \log S_t + \log C_t + \log I_t .$$

- In order to de-trend or de-seasonalize the series, we need to estimate these components.
- Estimation can be done together or separately using different techniques.

What have we done so far?

exponential smoothing - weights.

Holt-Winter's smoothing.

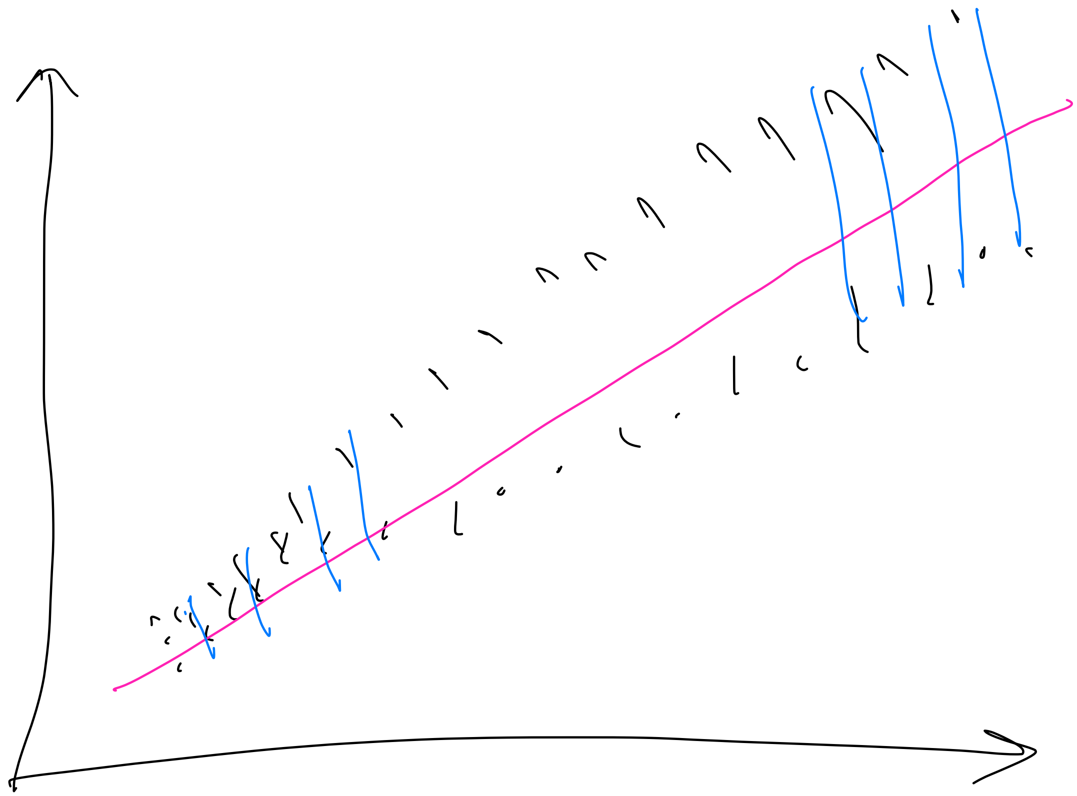
- Estimate trend using two approaches:
  - moving average smoothing / any smoothing
  - linear regression
- Estimate seasonality using two approaches:
  - calculate the mean values of the de-trended series for each specific period (for example, for quarterly data - to estimate the seasonal effect of quarter 1 - average the de-trended values for all quarter 1 values in the series, etc.)
  - along with trend in a regression setup
    - using dummy variables
    - using harmonic functions

Additive and Multiplicative Model with Regression

- We will examine the AirPassengers data again with respect to an Additive and Multiplicative model but in the regression framework.
- This framework allows us to handle trend and seasonality together.
- Let's consider three models:
  - Model 1: $X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t$ dummy variably.
  - Model 2: $X_t = \underline{\beta_0} + \underline{\overbrace{\beta_1 t + \beta_2 t^2}^{T}} + \underline{\gamma_2 d_{2,t} + \cdots + \gamma_{12} d_{12,t}} + \epsilon_t$ Additive.
  - Model ~~2~~ 3: $log(X_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \gamma_2 d_{2,t} + \cdots + \gamma_{12} d_{12,t} + \epsilon_t$

multiplicative model.

stabilize variance.



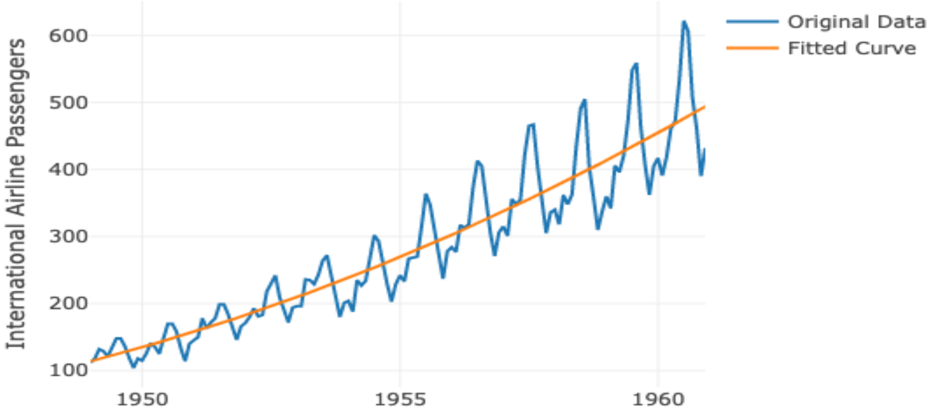
Example - AirPassengers

- Model 1 will help us ascertain the validity of a linear vs quadratic trend.
- Model 2 is an additive model - additive trend and seasonality.
- Model 3 is a multiplicative model.
    - Remember, a multiplicative model can be converted to an additive model by taking logarithms.
    - The response variable in Model 3 is $log(X_t)$.

Model 1 (R-sq = 86%)

$x_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t .$

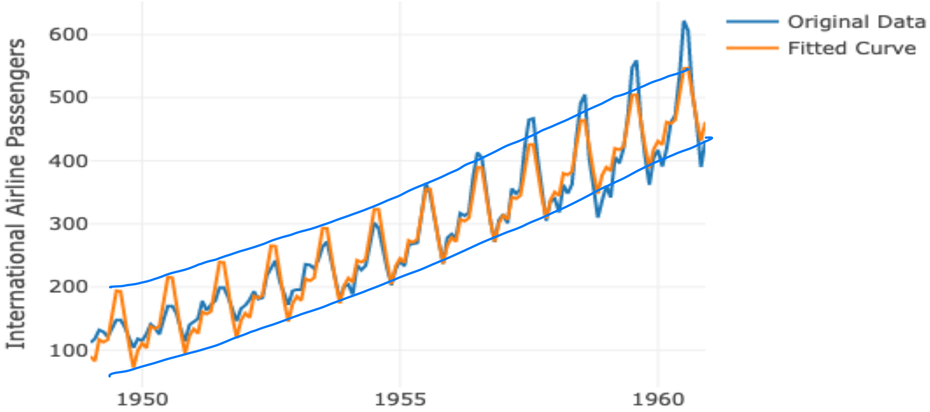
The model (and all terms) are significant. A quadratic trend model is appropriate for the data. The model doesn't capture any seasonality.



Model 2 (R-sq = 96%)

→ they should be included
→ non-significant terms.

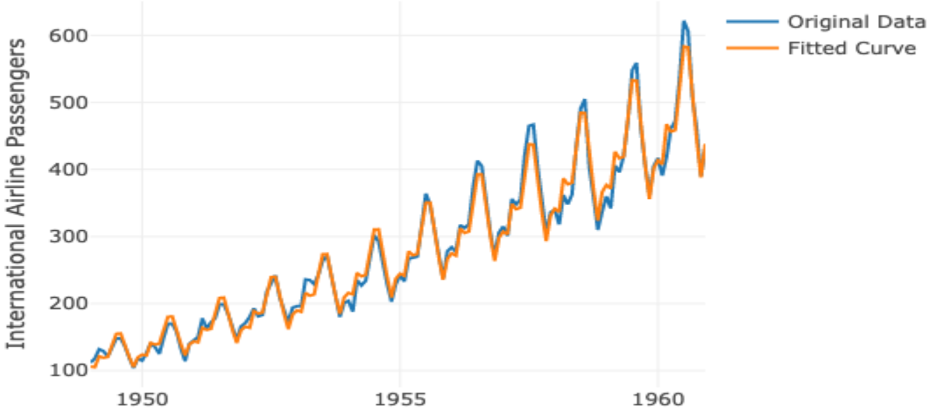
The model (and most terms) are significant. The seasonality doessn't change with trend indicating an additive model may not be appropriate.



Model 3 (R-sq = 98.8%)

$\log x_t \cdot$

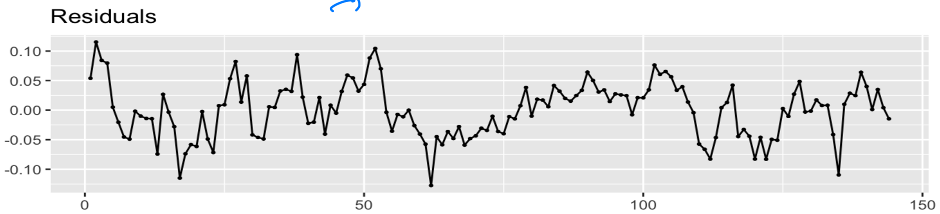
The model (and most terms) are significant. The fitted curve is plotted by exponentiating the fitted values from the model.



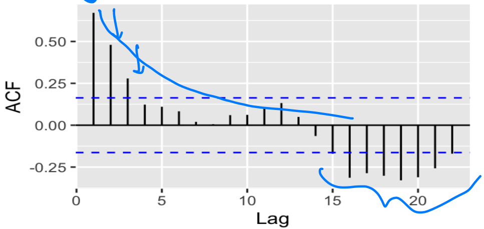
Analyzing Residuals

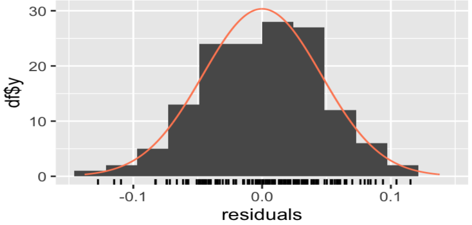
residual vs time .

ε_t



AR



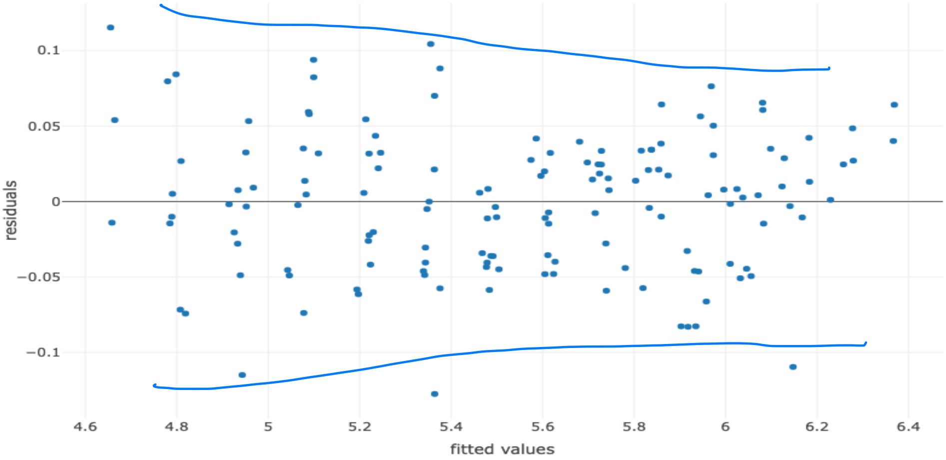


auto - correlated .

normality .

Analyzing Residuals

homoskedasticity.



Analyzing Residuals

histogram.

- Residuals are approximately normal centered at 0.

sample ACF.

- Residuals are autocorrelated. What does this mean?
  - A possible mis-specificatoon in the model.
  - More information may be extracted with a better model.
- No pattern in the plot of residuals vs. fitted values which suggests homoscedasticty.
- You may plot residuals vs. explanatory variables to check for patterns as well.

What's next?

$X_t = \mu_t + \epsilon_t$. $\rightarrow \epsilon_t$

$T_t + S_t$.

- Analyze the resulting irregular component (residuals/errors).
- Whichever method was used to decompose the series, the aim is to produce **stationary** residuals.
  - If residuals are white noise (along with stationary) then our job is done.
- Choose a model to fit the stationary residuals.
- Forecasting can be achieved by forecasting the residuals and combining with the forecasts of the trend and seasonal components.

AirPassengers Data - Residuals

unit root tests.

stationarity tests.

- Are the residuals stationary?
  - There are hypothesis tests (ADF test, KPSS test, etc.) that can be done for stationarity. We will explore these later.
  - ADF test in R supports stationarity for residuals.
- Fit a model to extract additional information from the residuals.
  - A possible contender for the residuals is an AR(1) model given the characteristic decay in the model.

ar(fit_model3 $ residuals)

→ AR(1)

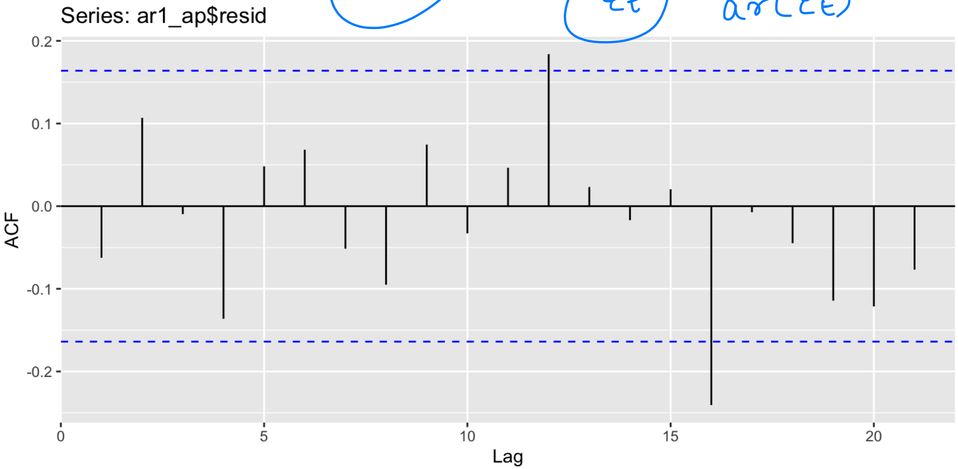
→ fit-ar $ residuals

AR(1) for Residuals

Reg. + AR(1) + w_t.

ε_t.

ar(ε_t)



What can we do get a stationary series?

$X_t = \hat{\mu_t} + \varepsilon_t$.

- De-trending and de-seasonalizing may work.
  - These are particularly useful when the key task is estimation.
- If estimation is not important, we may try differencing.
  - A special type of filtering, which is particularly useful for removing a trend, is simply to difference a given time series until it becomes stationary.

(a) $X_t = \beta_0 + \beta_1 t + w_t$.

(b) $y_t = X_t - X_{t-1}$ (differencing).

Differencing

∇ : nabla.

- Order 1 Differencing (useful for non-seasonal data)

first order diff. : $\underline{\underline{\nabla X_t}} = X_t - X_{t-1}$

- Occasionally second-order differencing is required

$$\underline{\nabla^2} X_t = \underline{\nabla}(\underline{\nabla X_t}) = \nabla X_t - \nabla X_{t-1} = X_t - 2X_{t-1} + X_{t-2}$$

- We may also use seasonal differencing.

$\nabla(x_t - x_{t-1})$

$$X_t = X_{t-1} + w_t$$

$$\operatorname{corr}(X_t, X_{t+1}) = \rho_x(t, t+1) = \frac{\gamma_x(t, t+1)}{\sqrt{\gamma_x(t,t)\gamma_x(t+1,t+1)}}$$

$$\begin{aligned}
X_t &= X_{t-1} + w_t \\
&= [X_{t-2} + w_{t-1}] + w_t \\
&= X_{t-3} + w_{t-2} + w_{t-1} + w_t \\
&= \dots = w_1 + w_2 + \dots + w_t = \sum_{i=1}^{t} w_i
\end{aligned}$$

$$X_{t+1} = \sum_{i=1}^{t+1} w_i$$

$$\gamma_x(t, t+1) = \operatorname{cov}(x_t, x_{t+1})$$

$$= \operatorname{cov}\left(\sum_{i=1}^{t} w_i, \sum_{i=1}^{t+1} w_i\right)$$

$$= \operatorname{cov}\left(\sum_{i=1}^{t} w_i, \sum_{i=1}^{t} w_i + w_{t+1}\right)$$

$$= \sum_{i=1}^{t} \operatorname{cov}(w_i, w_i) = t\sigma^2$$

$$\gamma_x(t, t) = t\sigma^2$$

$$\gamma_x(t+1, t+1) = (t+1)\sigma^2$$

$$\rho_x(t, t+1) = \frac{t\sigma^2}{\sqrt{t\sigma^2 \cdot (t+1)\sigma^2}} = \frac{t}{\sqrt{t(t+1)}}$$