STA 371G: Statistics and Modeling

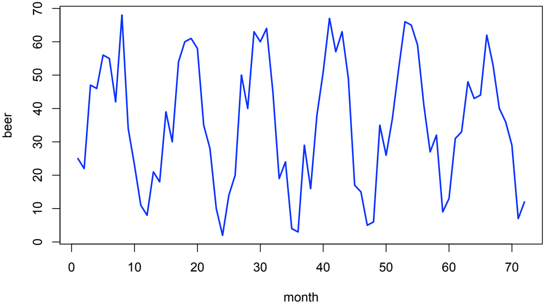
Time Series: Modeling Seasonality

Mingyuan Zhou
McCombs School of Business
The University of Texas at Austin

http://mingyuanzhou.github.io/STA371G

The Seasonal Model

- Many time-series data exhibit some sort of seasonality
- The simplest solution is to add a set of dummy variables to deal with the "seasonal effects"



Y_t = monthly U.S. beer production (in millions of barrels).

The Seasonal Model

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.959010553
R Square	0.919701241
Adjusted R Square	0.904979802
Standard Error	0.588667988
Observations	72

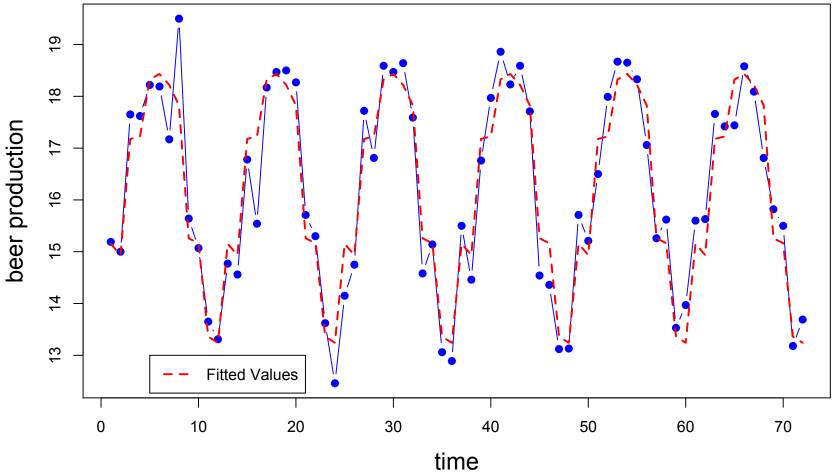
ANOVA

	df	SS	MS	F	Significance F
Regression	11	238.138728	21.649	62.47359609	1.20595E-28
Residual	60	20.7918	0.34653		
Total	71	258.930528			

	Coefficients	Standard Erro	t Stat	P-value	Lower 95%	Upper 95%
Intercept	13.24166667	0.2403227	55.0995	4.32368E-53	12.7609497	13.72238
X Variable 1	1.911666667	0.33986762	5.62474	5.15088E-07	1.23183021	2.591503
X Variable 2	1.693333333	0.33986762	4.98233	5.64079E-06	1.013496877	2.37317
X Variable 3	3.936666667	0.33986762	11.5829	6.13313E-17	3.25683021	4.616503
X Variable 4	3.983333333	0.33986762	11.7202	3.74305E-17	3.303496877	4.66317
X Variable 5	5.083333333	0.33986762	14.9568	6.59589E-22	4.403496877	5.76317
X Variable 6	5.19	0.33986762	15.2707	2.44866E-22	4.510163543	5.869836
X Variable 7	4.978333333	0.33986762	14.6479	1.77048E-21	4.298496877	5.65817
X Variable 8	4.581666667	0.33986762	13.4807	8.22861E-20	3.90183021	5.261503
X Variable 9	2.016666667	0.33986762	5.93368	1.58522E-07	1.33683021	2.696503
X Variable 10	1.923333333	0.33986762	5.65907	4.52211E-07	1.243496877	2.60317
X Variable 11	0.118333333	0.33986762	0.34817	0.728927584	-0.561503123	0.79817

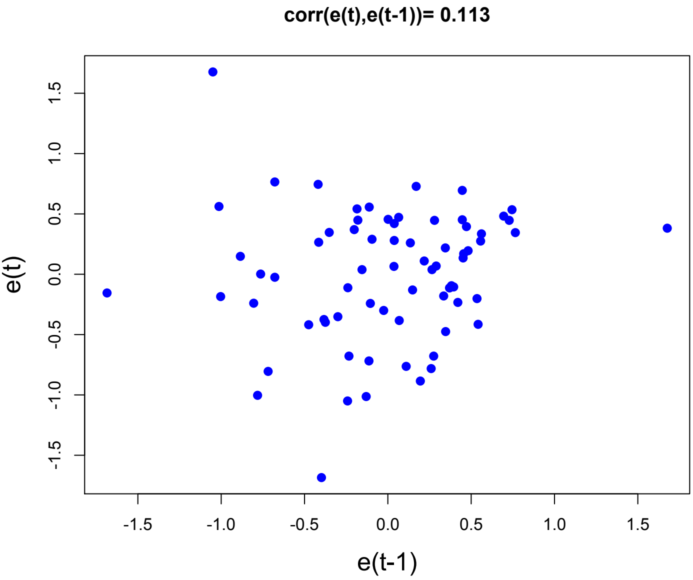
Let's look at the Excel file...

The Seasonal Model



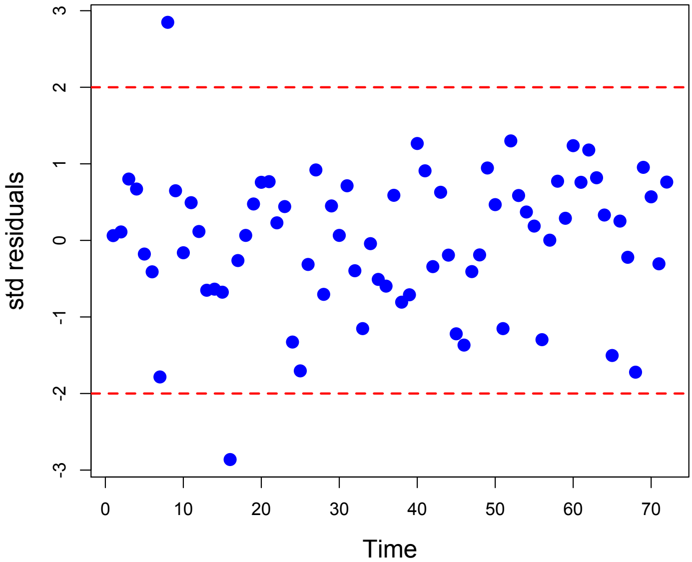
What would our future predictions look like?

The Seasonal Model



Okay... good enough.

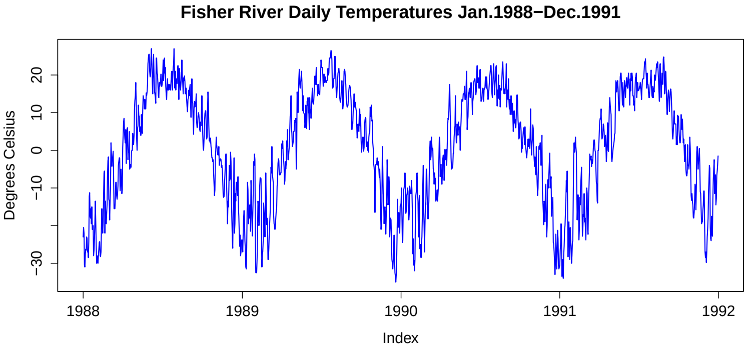
The Seasonal Model



Still, no obvious problems...

Fisher River Daily Temperatures

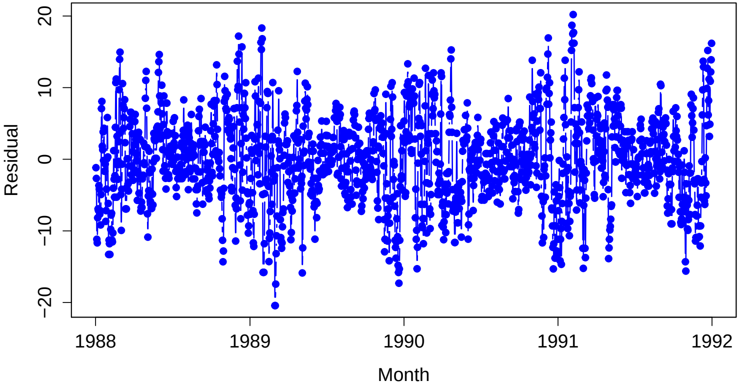
Fisher river daily temperatures Jan.1988-Dec.1991



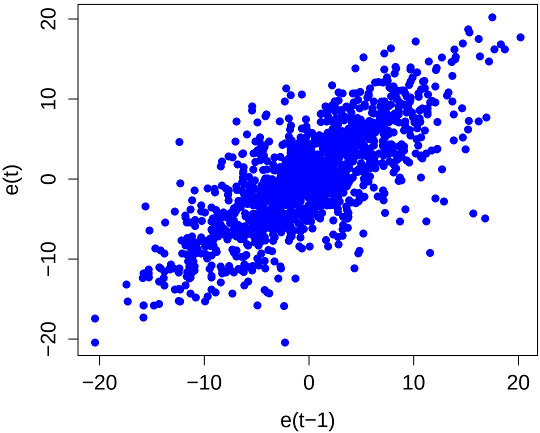
Fisher River Daily Temperatures

Residual plot for a seasonal model

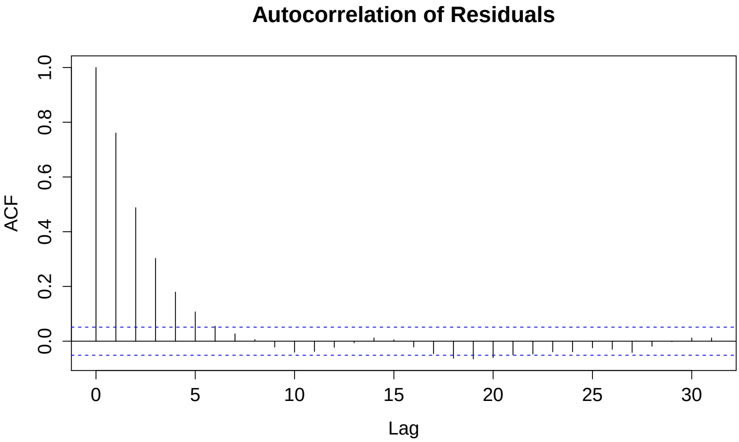
$$Y_t = \beta_0 + \beta_1 * Jan + \cdots + \beta_{11} * Nov + \epsilon_t$$



Fisher River Daily Temperatures

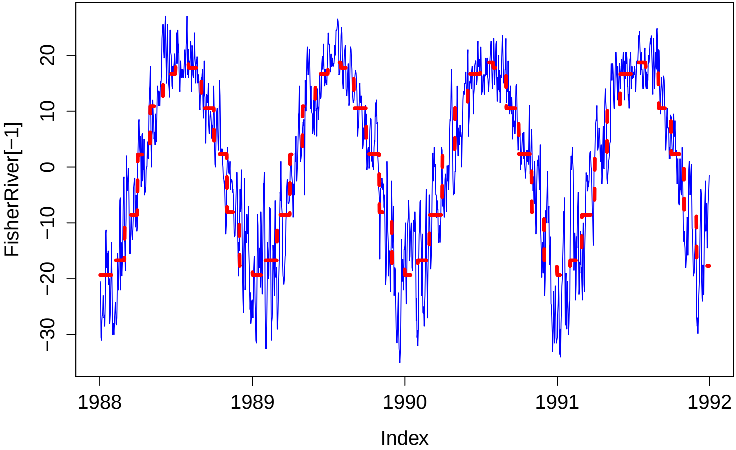


Fisher River Daily Temperatures



Fisher River Daily Temperatures

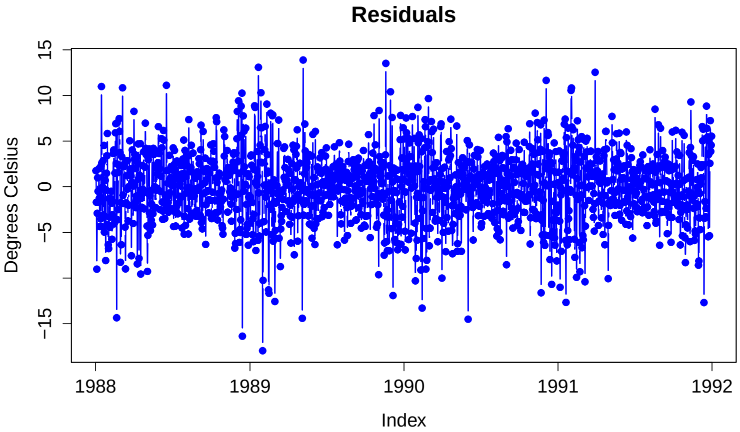
Fitted results



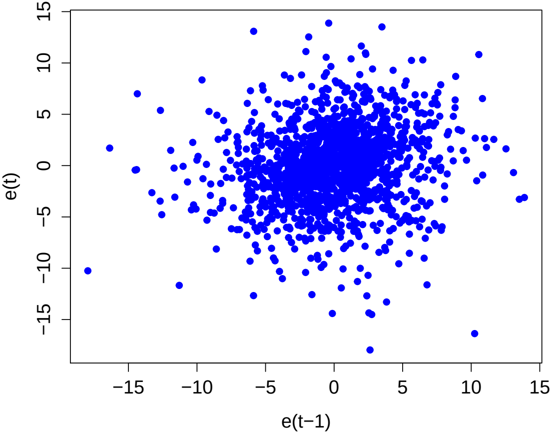
Fisher River Daily Temperatures

Residual plot for a seasonal & autoregressive model

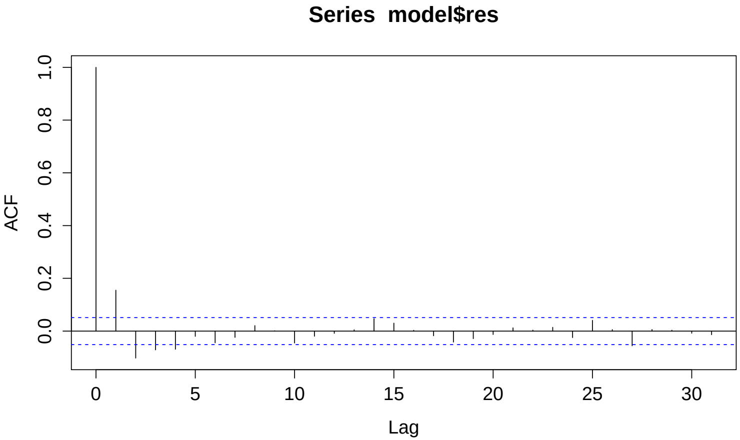
$$Y_t = \beta_0 + \beta_1 * Jan + \cdots + \beta_{11} * Nov + \beta_{12} Y_{t-1} + \epsilon_t$$



Fisher River Daily Temperatures

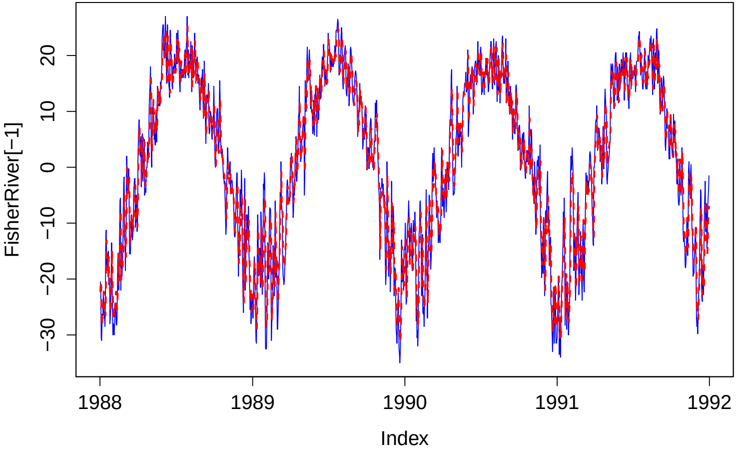


Fisher River Daily Temperatures



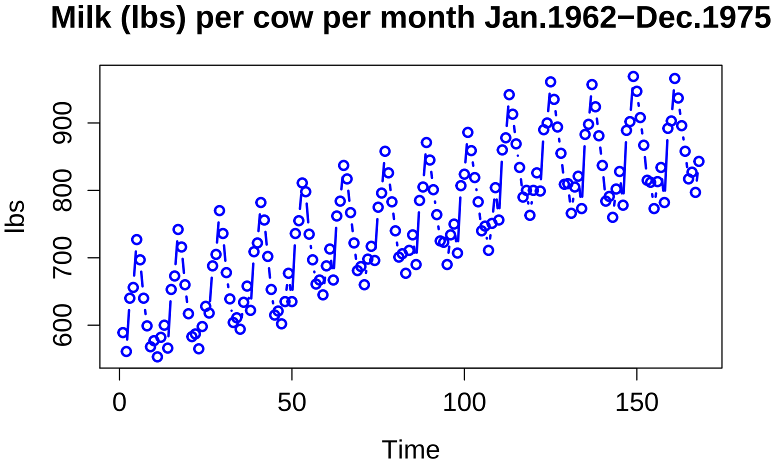
Fisher River Daily Temperatures

Fitted results



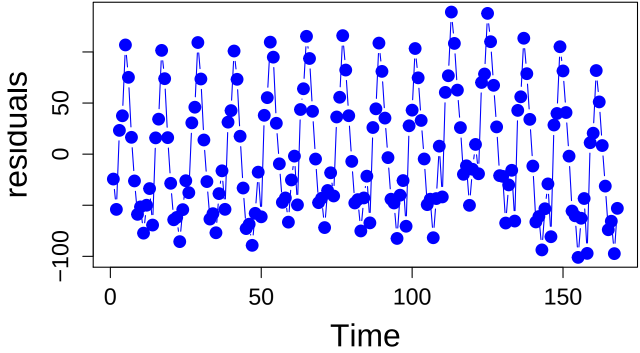
Milk Production Jan 1962 - Dec 1975

Milk production in lbs per cow per month Jan 1962 - Dec 1975



Milk Production Jan 1962 - Dec 1975

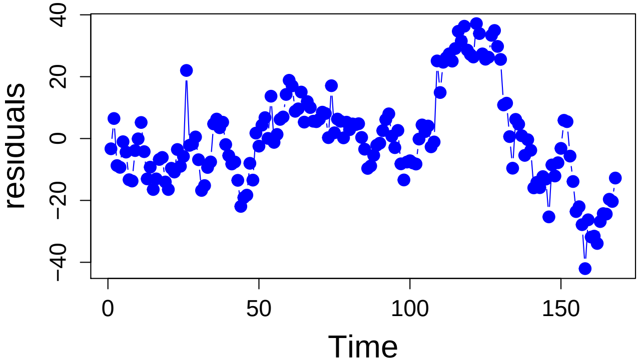
Residual plot for a trend model $Y_t = \beta_0 + \beta_1 t + \epsilon_t$



Milk Production Jan 1962 - Dec 1975

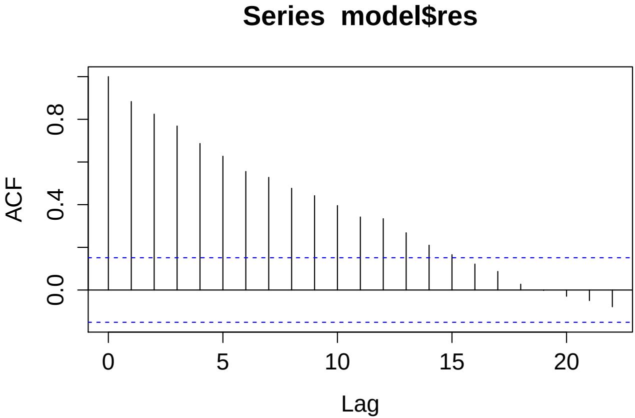
Residual plot for a trend & seasonal model

$Y_t = \beta_0 + \beta_1 t + \beta_2 * Jan + \cdots + \beta_{12} * Nov + \epsilon_t$



Milk Production Jan 1962 - Dec 1975

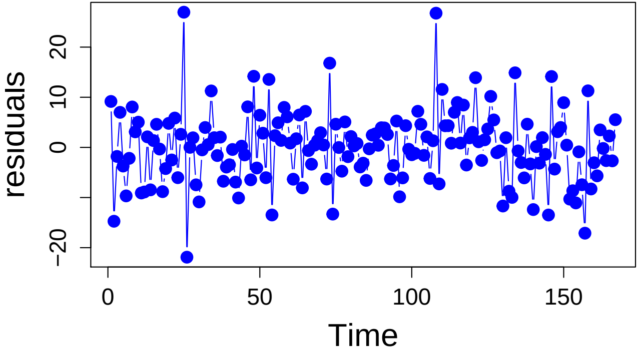
The residuals are still highly autocorrelated!



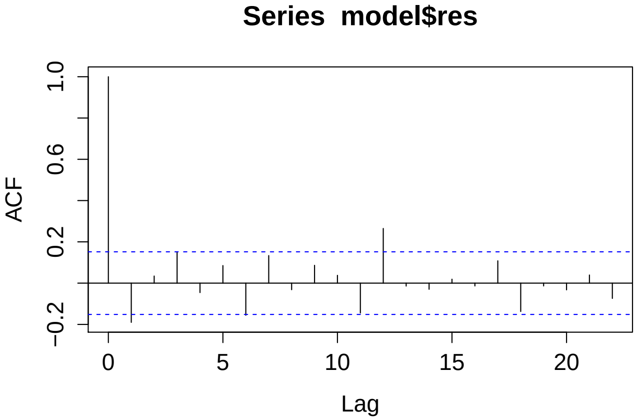
Milk Production Jan 1962 - Dec 1975

Residual plot for a trend & seasonal & autoregressive model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 * Jan + \cdots + \beta_{12} * Nov + \beta_{13} Y_{t-1} + \epsilon_t$$

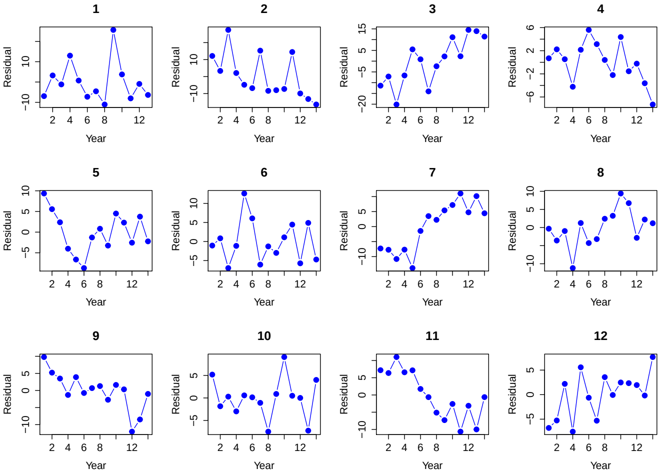


Milk Production Jan 1962 - Dec 1975



Can we further improve the model?

Milk Production Jan 1962 - Dec 1975



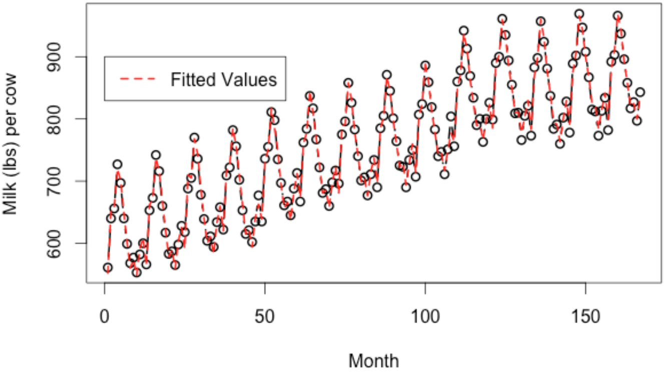
Can you guess what are these plots?

How shall we improve the model?

Milk Production Jan 1962 - Dec 1975

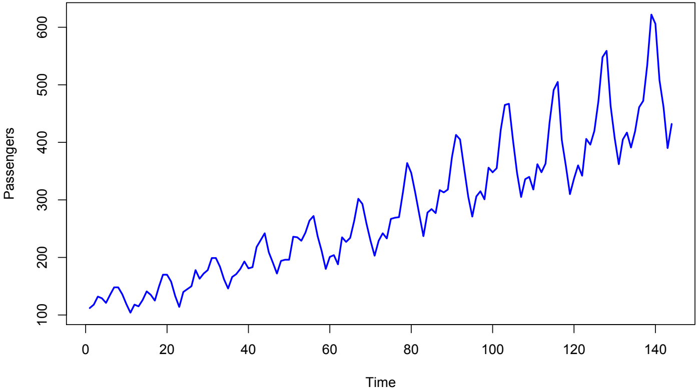
Fitted values for a trend & seasonal & autoregressive model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 * Jan + \cdots + \beta_{12} * Nov + \beta_{13} Y_{t-1} + \epsilon_t$$



Airline Data

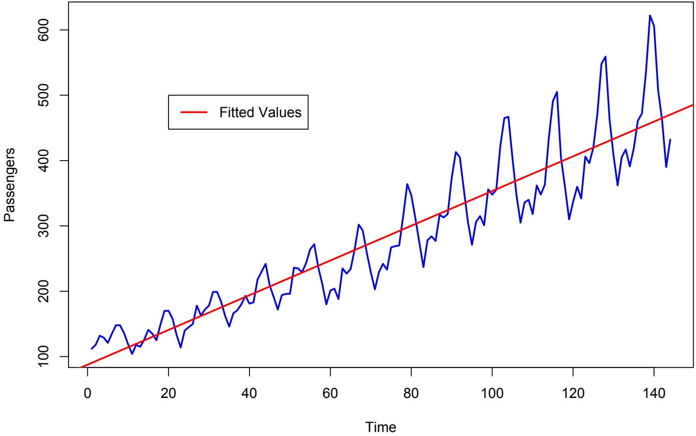
Monthly passengers in the U.S. airline industry (in 1,000 of passengers) from 1949 to 1960... we need to predict the number of passengers in the next couple of months.



Any ideas?

Airline Data

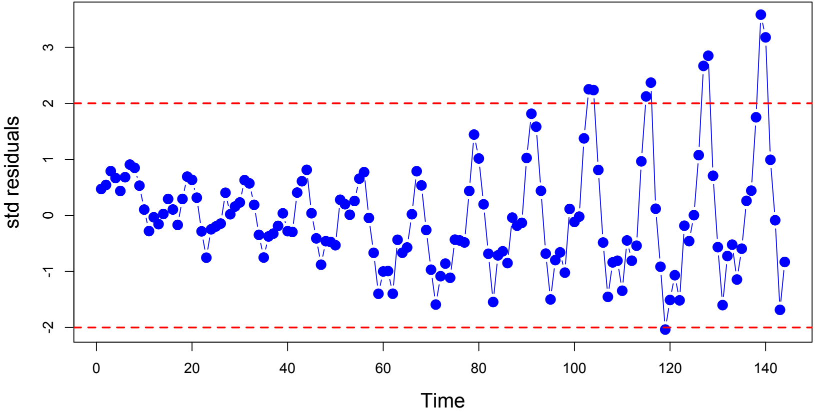
How about a "trend model"? $Y_t = \beta_0 + \beta_1 t + \epsilon_t$



What do you think?

Airline Data

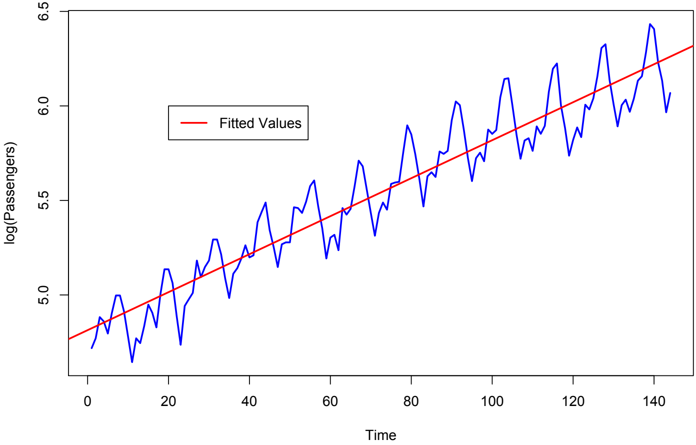
Let's look at the residuals...



Is there any obvious pattern here? YES!!

Airline Data

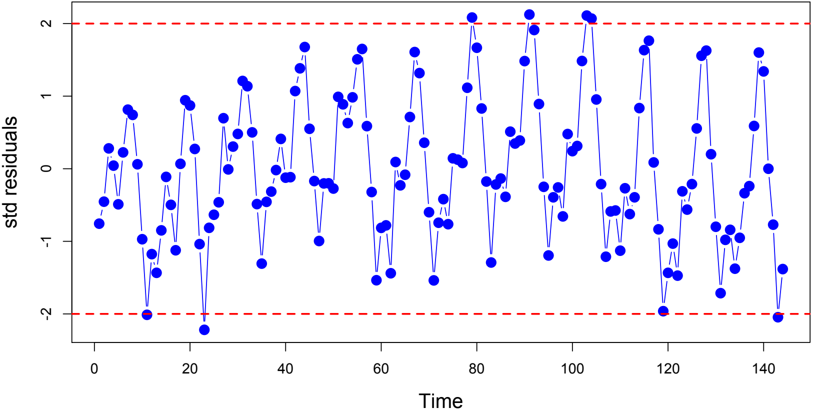
The variance of the residuals seems to be growing in time... Let's try taking the log. $\log(Y_t) = \beta_0 + \beta_1 t + \epsilon_t$



Any better?

Airline Data

Residuals...

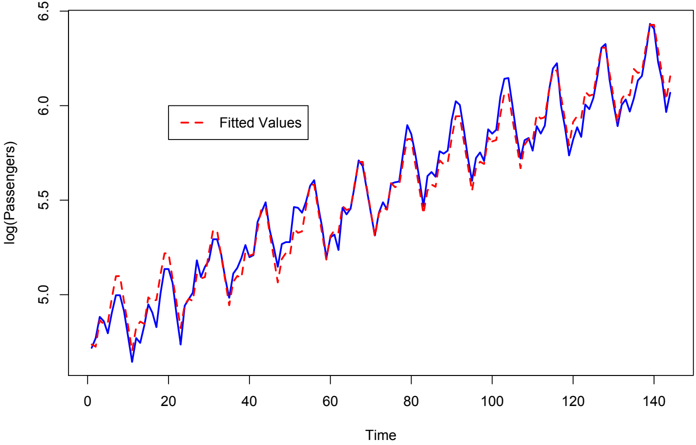


Still we can see some obvious temporal/seasonal pattern....

Airline Data

Okay, let's add dummy variables for months (only 11 dummies)...

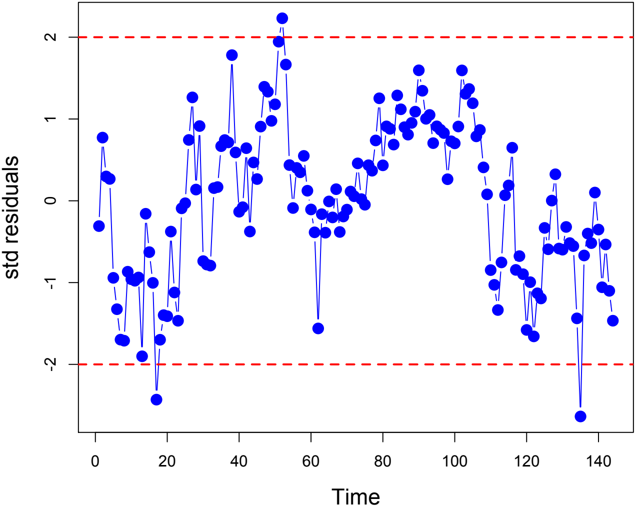
$\log(Y_t) = \beta_0 + \beta_1 t + \beta_2 Jan + ...\beta_{12} Nov + \epsilon_t$



Much better!!

Airline Data

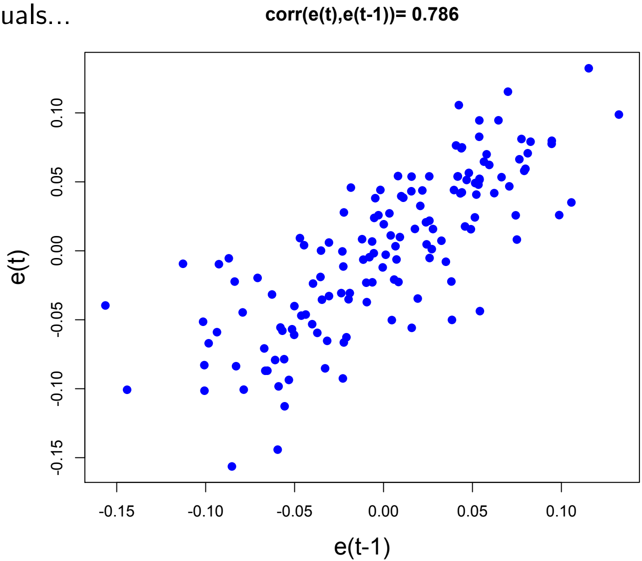
Residuals...



I am still not happy... it doesn't look normal iid to me...

Airline Data

Residuals...

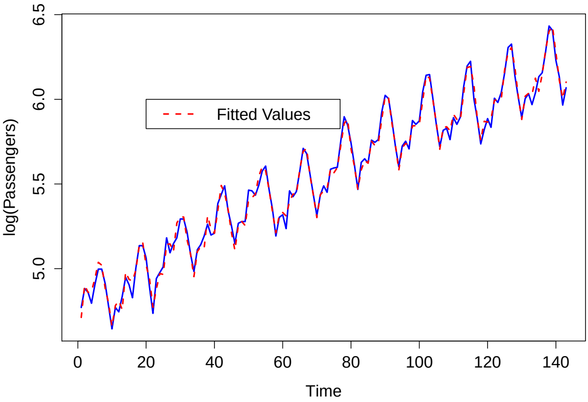


I was right! The residuals are dependent on time...

Airline Data

We have one more tool... let's add one legged term.

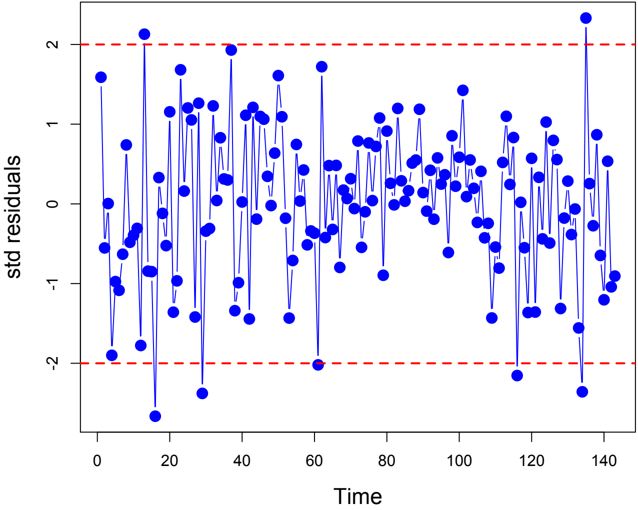
$$\log(Y_t) = \beta_0 + \beta_1 t + \beta_2 Jan + ... \beta_{12} Nov + \beta_{13} \log(Y_{t-1}) + \epsilon_t$$



Okay, good...

Airline Data

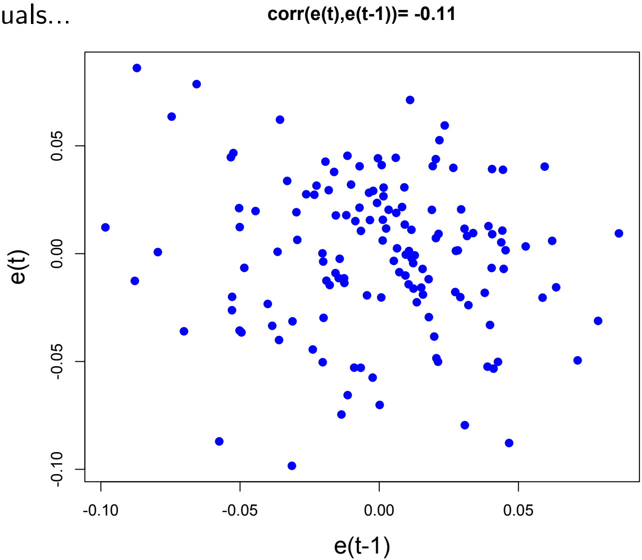
Residuals...



Much better!!

Airline Data

Residuals...



Much better indeed!!

Summary

Whenever working with time series data we need to look for dependencies over time.

We can deal with lots of types of dependencies by using regression models... our tools are:

- trends
- lags
- seasonal dummies