STAT 436 / 536 - Lecture 16: Key

Modeling Non-Stationary Time Series

- Many time series models are non-stationary. Recall a time series is stationary if the mean and variance are constant in time and the autocorrelation only depends on the lag between two time points.

- One of our current strategies for non-stationary time series models is to include a regression component for trends or seasonal cycles.

- When using a regression model, the interest shifts to the residuals. In particular, the residuals should satisfy stationarity.

- We also have seen that differencing a random walk results in a stationary series. A random walk can be written as
$$x_t = x_{t-1} + w_t$$
and then the differenced series
$$\nabla x_t = x_t - x_{t-1} = w_t$$
just results in white noice, and hence, is stationary.

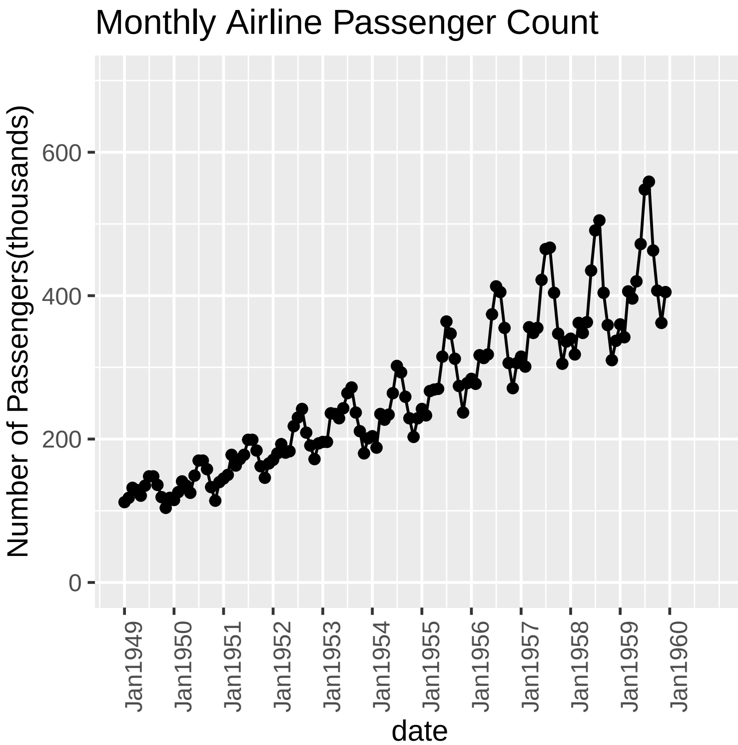
So why is stationarity important?

- Stationarity is a particular kind of dependence structure, one that enables easier modeling of the sequence of random variables. Given a stationary sequence, the ARMA suite of tools can be used to model and explain the dependence structure.

- When stationarity is not present, differencing (as in the random walk), is used to try to obtain a resulting series that is stationary.

Diagnosing a non-stationary series

- Sometimes a non-stationary series can be diagnosed visually:



In fact, this is often what we use for residual diagnostics.

- The characteristic equation can be used to determine if a series is stationary, assuming the parameter values are known.

$$
\begin{aligned}
x_t &= x_{t-1} - \frac{1}{4}x_{t-2} + w_t \\
\left(\frac{1}{4}B^2 - B + 1\right)x_t &= w_t
\end{aligned}
$$

```
polyroot(c(1, -1, .25))
```

```
## [1] 2+0i 2-0i
```

stationary

$$
\begin{aligned}
x_t &= x_{t-1} + w_t \\
(B-1)\,x_t &= w_t
\end{aligned}
$$

```
polyroot(c(-1, 1))
```

```
## [1] 1+0i
```

non-stationary

$$
\begin{aligned}
x_t &= \frac{1}{2}x_{t-1} + \frac{1}{2}x_{t-2} + w_t \\
\left(-\frac{1}{2}B^2 - \frac{1}{2}B + 1\right)x_t &= w_t
\end{aligned}
$$

```r
polyroot(c(1, -.5, -.5))
```

```
## [1]  1-0i -2+0i
```

non-stationary

- One last way to diagnose stationarity is to use a unit root test. This is closely related to the idea of a random walk as a unit root corresponds to the solution of the polynomial equation of an AR 1 model.

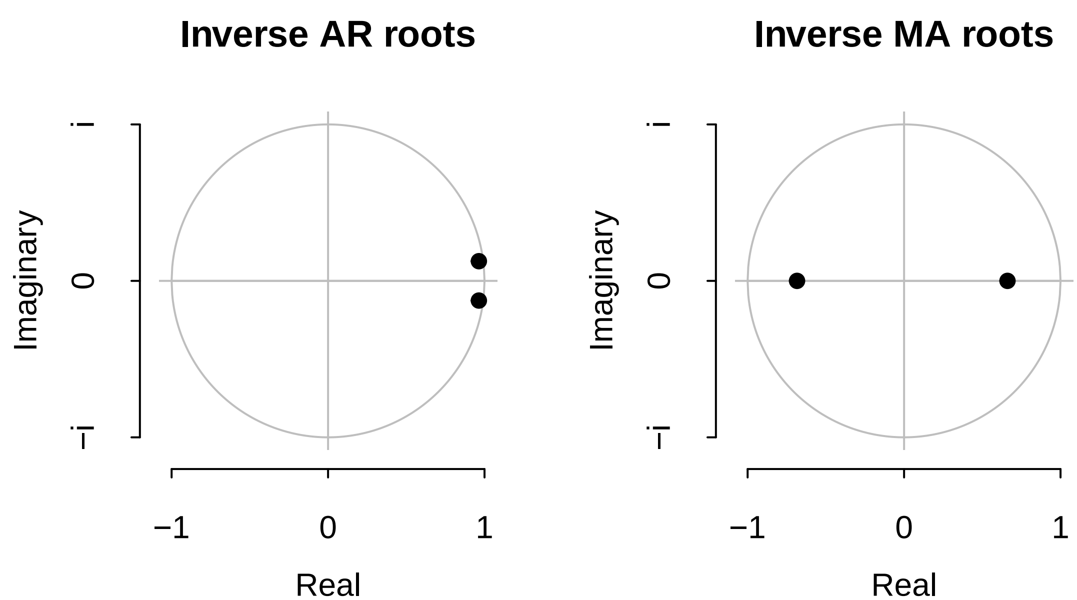
- The `arima.sim()` function will return an error if you try to simulate a non-stationary model.

```r
arima.sim(n = 100, list(ar = c(1)))
```

```
## Error in arima.sim(n = 100, list(ar = c(1))): 'ar' part of model is not stationary
```

- The `Arima` and `auto.arima` can be used to assess for stationary with the a fitted model; however, there are restrictions in the model that generally result in non-stationary models being fitted (with the integrated piece).

```r
library(forecast)
#autoplot(WWWusage)
plot(auto.arima(WWWusage, stationary = T))
```



- Furthermore, the `auto.arima` package will also select models that "integrate" the data to create a differenced series that will be stationary.

```
auto.arima(WWWusage)
```

```
## Series: WWWusage
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1     ma1
##       0.6504  0.5256
## s.e.  0.0842  0.0896
##
## sigma^2 estimated as 9.995:  log likelihood=-254.15
## AIC=514.3   AICc=514.55   BIC=522.08
```

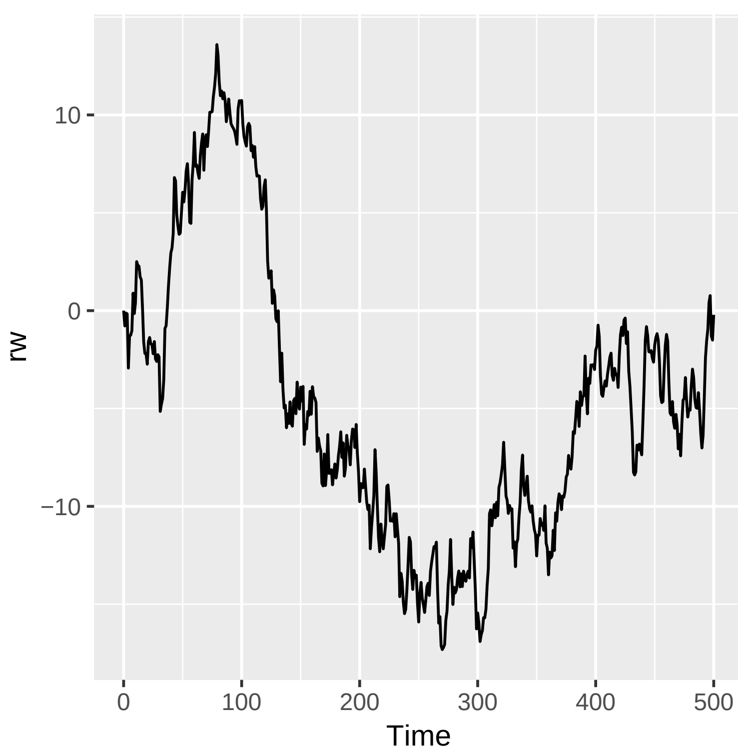
Integrated Model

- A model is 'integrated' with order d, denoted I(d), if the d^{th} difference of $\{x_t\}$ is white noise.

$$\begin{aligned} \nabla^d x_t &= w_t \\ (1-B)^d x_t &= w_t \end{aligned}$$

When $d = 1$ this is a random walk.

```
rw <- arima.sim(n=500, list( order = c(0,1,0)))
autoplot(rw)
```



```r
auto.arima(rw)
```

```
## Series: rw
## ARIMA(0,1,0)
##
## sigma^2 estimated as 1.018:  log likelihood=-714.03
## AIC=1430.06   AICc=1430.07   BIC=1434.27
```

- The integrated component can also be combined with ARMA models to form an ARIMA.

$$\theta_p(B)(1-B)^d x_t = \phi_q(B) w_t$$

  is an ARIMA(p, d, q) model.

- For instance,

$$\begin{aligned} x_t &= \alpha x_{t-1} + x_{t-1} - \alpha x_{t-2} + w_t + \beta w_{t-1} \\ x_t - \alpha x_{t-1} - x_{t-1} + \alpha x_{t-2} &= w_t + \beta w_{t-1} \\ (1-\alpha B)(1-B)^d x_t = (1+\beta B) w_t \end{aligned}$$

  is an ARIMA model of order (1,1,1) with an AR parameter of α and an MA parameter of β.

- Recall the taxi data set. Run the code below and discuss the results.

```r
taxi.rides <- read_csv('http://math.montana.edu/ahoegh/teaching/timeseries/data/taxi.csv')

taxirides.diff <- taxi.rides %>% arrange(year, month, day) %>% slice(-c(1:4)) %>%
  mutate(week.numb = rep(1:234, each = 7)) %>% group_by(week.numb) %>%
  summarize(total.rides = sum(n)) %>% select(total.rides) %>% pull() %>% diff()

auto.arima(taxirides.diff)
```

```
## Series: taxirides.diff
## ARIMA(3,0,1) with zero mean
##
## Coefficients:
##           ar1      ar2      ar3      ma1
##       -0.1069  -0.3432  -0.2949  -0.2583
## s.e.   0.1406   0.0650   0.0812   0.1419
##
## sigma^2 estimated as 2.644e+10:  log likelihood=-3124.72
## AIC=6259.45   AICc=6259.71   BIC=6276.7
```

```r
taxirides.summary <- taxi.rides %>% arrange(year, month, day) %>% slice(-c(1:4)) %>%
  mutate(week.numb = rep(1:234, each = 7)) %>% group_by(week.numb) %>%
  summarize(total.rides = sum(n)) %>% select(total.rides) %>% pull()
auto.arima(taxirides.summary)
```

```
## Series: taxirides.summary
## ARIMA(3,1,1)
##
## Coefficients:
##           ar1      ar2      ar3      ma1
##       -0.1069  -0.3432  -0.2949  -0.2583
## s.e.   0.1406   0.0650   0.0812   0.1419
##
## sigma^2 estimated as 2.644e+10:  log likelihood=-3124.72
## AIC=6259.45   AICc=6259.71   BIC=6276.7
```

- The `arima` function can also be used to fit a specific order of an ARIMA model.

- As we saw before with ARMA models, AR, ARI, IMA, and ARMA models are all special cases of the ARIMA framework.

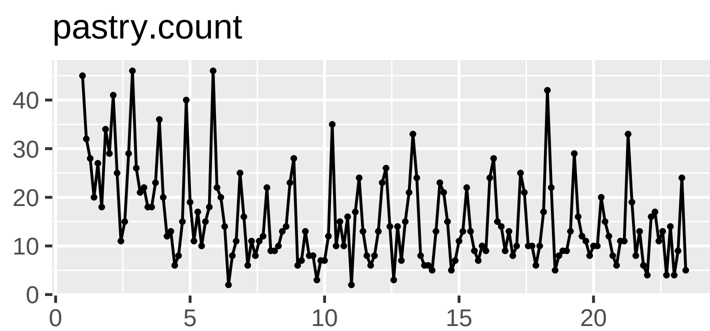
Seasonal Arima

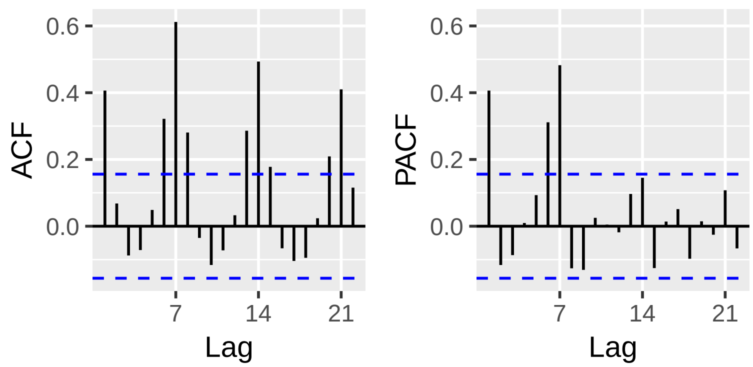
- ARIMA models can also have a seasonal component, where the lag corresponds to the seasonal frequency. For example, consider the following model for a time series with weekly seasonal frequency:

$$x_t = \alpha x_{t-7} + w_t$$

  then this model is a seasonal $ARIMA(0,0,0)(1,0,0)_7$.

```r
library(forecast)
bakery.sales <- read_csv('http://math.montana.edu/ahoegh/teaching/timeseries/data/BreadBasket.csv')
pastry.count <- bakery.sales %>% filter(Item %in% c('Pastry','Scandinavian','Medialuna','Muffin','Scone
ggtsdisplay(pastry.count)
```






```r
auto.arima(pastry.count)
```

```
## Series: pastry.count
## ARIMA(2,0,2)(1,1,1)[7] with drift
##
## Coefficients:
##           ar1      ar2     ma1     ma2     sar1     sma1    drift
##       -0.0402  -0.3802  0.3757  0.5374  -0.0920  -0.4636  -0.1164
## s.e.   0.2783   0.3361  0.2300  0.3482   0.1744   0.1618   0.0504
##
## sigma^2 estimated as 40.24:  log likelihood=-491.05
## AIC=998.11   AICc=999.12   BIC=1022.25
```