DSC 152: Applied Statistical Data Analysis and Inference

Lecture #17 Part 1
Introduction to
Time Series Analyses

Recall. . .

Conditions for validity of inference in linear regression

- The relationship between X and Y, if there is one, is actually **L**inear
  - e.g. not quadratic, exponential, etc.
- **I**ndependence of observations
- **N**ormality of ϵ_i
  - Note that this can also be achieved, due to the central limit theorem, with a large sample size even if ϵ_i does not follow a normal distribution
- **E**qual variance across all values of X
  - Also known as homoskedasticity

Time Series Data

Time Series data violate the independence condition in a very particular way.

Simply put, time series data are data that are collected over time.

Examples:

- Stock market prices
- Temperature of a particular location over time
- Quiz scores within individual students over time

TSM is the stock symbol for Taiwan Semiconductor Manufacturing Company. The R package `quantmod` lets you grab pricing information for stocks:

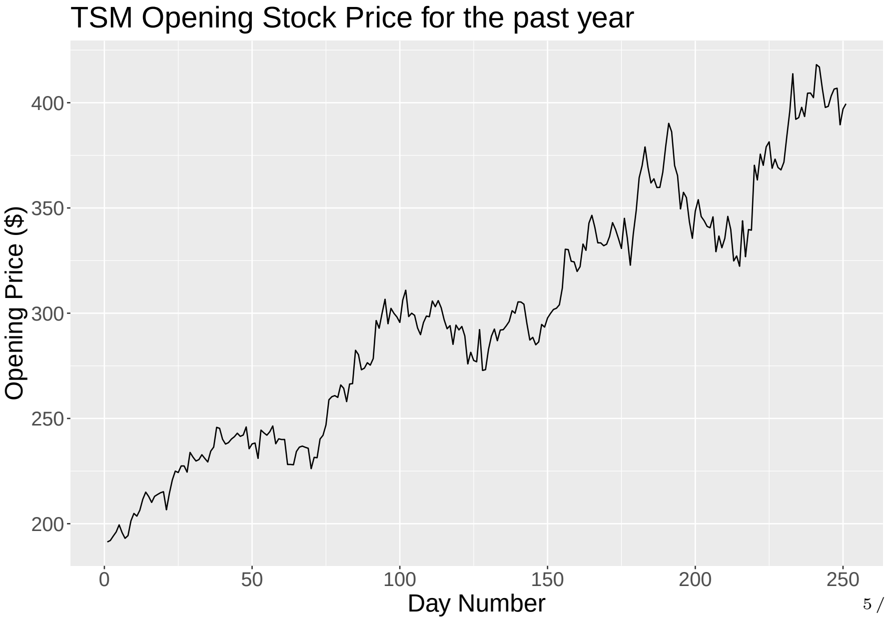
```
install.packages("quantmod")
library(quantmod)
getSymbols("TSM")
```

```
## [1] "TSM"
```

TSM.Open	TSM.High	TSM.Low	TSM.Close	TSM.Volume	TSM.Adjusted
191.34	196.83	191.34	196.19	11825000	193.7898
192.02	192.80	190.03	191.98	9403100	189.6313
194.10	198.31	193.70	197.68	12228200	195.2616
196.09	198.07	195.48	196.14	17787600	193.7404
199.46	199.80	196.08	197.15	14609900	194.7381

Time Series Data

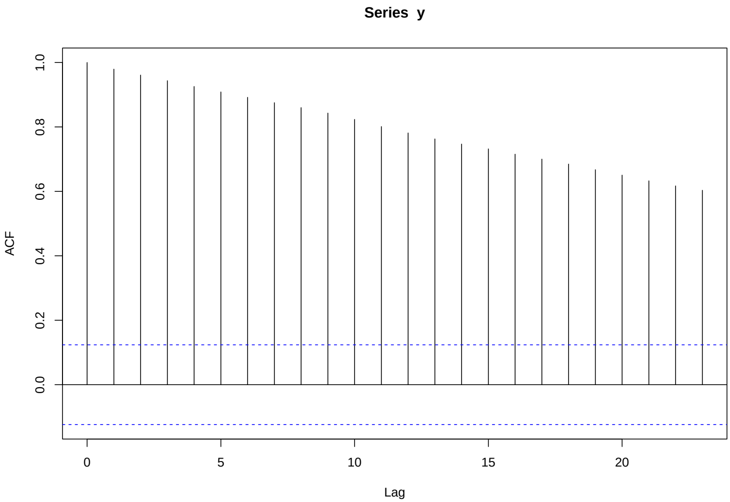
Example #1: TSM Stock Price



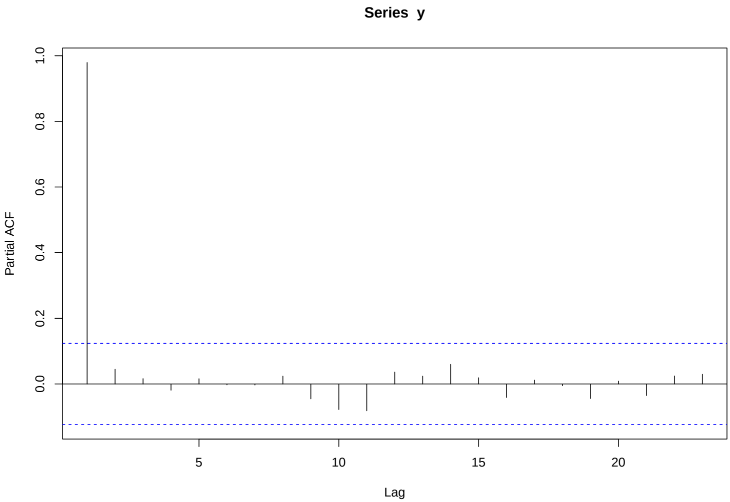
To assess autocorrelation, the standard approach is to look at correlograms of two things:

- autocorrelation function (ACF)
  - This shows the correlation between x_t and x_{t+k} for any k, where k is the number of steps in the future. For example, for $k = 1$, this would just be looking at the correlation between subsequent days
- partial autocorrelation function (PACF)
  - Even if the data generating process only has e.g. one-step correlations, it will show up as correlations between steps that are further apart as well due to carry-over. The PACF controls for these intermediate steps in an attempt to isolate only the correlations that are actually driving any relationship.

```
acf(y)
```



```r
pacf(y)
```



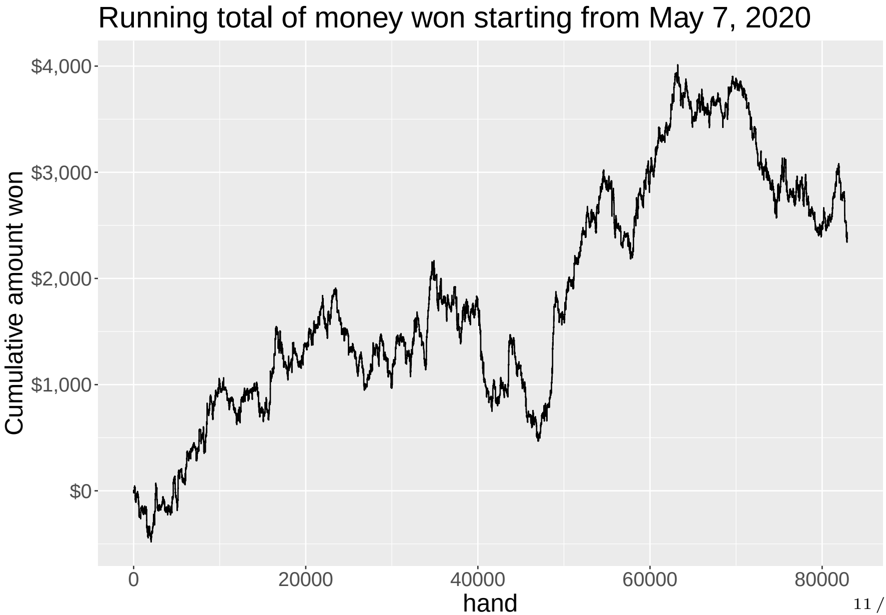
So, let's try an AR(1) model

```
library(forecast)
t <- 1:length(y)
stock_model <- Arima(y, order=c(1,0,0), xreg=t)
stock_model
```

```
## Series: y
## Regression with ARIMA(1,0,0) errors
##
## Coefficients:
##          ar1  intercept    xreg
##       0.8860   198.3573  0.7880
## s.e.  0.0285     7.3983  0.0501
##
## sigma^2 = 52.02:  log likelihood = -851.35
## AIC=1710.7   AICc=1710.87   BIC=1724.81
```

```
library(lmtest)
coeftest(stock_model)
```

```
##
## z test of coefficients:
##
##             Estimate Std. Error z value  Pr(>|z|)
## ar1          0.885996   0.028485  31.104 < 2.2e-16 ***
## intercept 198.357307   7.398347  26.811 < 2.2e-16 ***
## xreg         0.787985   0.050125  15.720 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
```


Running total of money won starting from May 7, 2020
Cumulative amount won
$4,000
$3,000
$2,000
$1,000
$0
0
20000
40000
60000
80000
hand

Your Turn #1

Example #2: Poker data from Lecture #5

- Load the data into R (`PokerHands1NL.csv` file on course website)

- Using the `won_cumul` column in that dataframe, recreate the graph on the previous slide

- Create ACF and PACF plots with the `won_cumul` column

- Run the appropriate Arima model and find the p-value for whether there is a significant directional trend, adjusting for any autocorrelation.