THE UNIVERSITY OF TEXAS AT AUSTIN
McCOMBS SCHOOL OF BUSINESS

Time Series: Autocorrelation

Lecture 25

STA 371G

Predicting oil prices

Date	Oil price ($)
1/1/2013	112.98
1/1/2014	107.94
1/1/2015	55.38
1/1/2016	36.85
1/1/2017	55.05
1/1/2018	?

Predicting oil prices

Date	Oil price ($)
1/1/2013	112.98
1/1/2014	107.94
1/1/2015	55.38
1/1/2016	36.85
1/1/2017	55.05
1/1/2018	?

- What's the best prediction of the price of oil on January 1?

Predicting oil prices

Date	Oil price ($)
1/1/2013	112.98
1/1/2014	107.94
1/1/2015	55.38
1/1/2016	36.85
1/1/2017	55.05
1/1/2018	?

- What's the best prediction of the price of oil on January 1?
- Does next year's price depend on this year's?

Time series

In a time series, data are not necessarily independent. (Often it is not!)

Time series

In a time series, data are not necessarily independent. (Often it is not!)

Time series:

- A sequence of measurements of the same variable collected over time.
- The measurements are made at regular time intervals (most commonly daily, weekly, monthly, quarterly, or yearly).
- The variances are not necessarily constant over time either.

Some examples

- S&P 500 index (or any stock price)
- iPhone sales worldwide
- U.S. unemployment rate
- U.S. inflation rate
- Crime rate in Austin

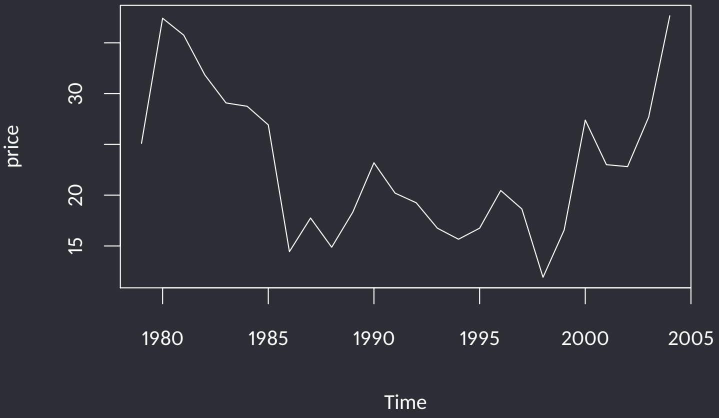
Some examples

- S&P 500 index (or any stock price)
- iPhone sales worldwide
- U.S. unemployment rate
- U.S. inflation rate
- Crime rate in Austin

Any of these could be measured at weekly, monthly, yearly etc. intervals; and each would be a different time series.

Oil Prices 1979-2004

```
# Convert the data into a time series object
price <- ts(oil$price, start=1979, frequency=1)
# Frequency: # of data points per year
plot(price)
```



Oil Prices 1979-2004

We argued that oil prices are not independent year-over-year.

Oil Prices 1979-2004

We argued that oil prices are not independent year-over-year. To predict the oil price in a given year, can we use the previous year's price?

Oil Prices 1979-2004

We argued that oil prices are not independent year-over-year. To predict the oil price in a given year, can we use the previous year's price?

y_t: The oil price at the end of the year t

Oil Prices 1979-2004

We argued that oil prices are not independent year-over-year. To predict the oil price in a given year, can we use the previous year's price?

y_t: The oil price at the end of the year t

t	y_t	y_{t-1}
...	...	...
1999	16.56	11.91
2000	27.39	16.56
2001	23	27.39
2002	22.81	23
...	...	...

Oil Prices 1979-2004

We argued that oil prices are not independent year-over-year. To predict the oil price in a given year, can we use the previous year's price?

y_t: The oil price at the end of the year t

t	y_t	y_{t-1}
…	…	…
1999	16.56	11.91
2000	27.39	16.56
2001	23	27.39
2002	22.81	23
…	…	…

y_{t-1} column is obtained by shifting y_t by 1.

Oil Prices 1979-2004

We argued that oil prices are not independent year-over-year. To predict the oil price in a given year, can we use the previous year's price?

y_t: The oil price at the end of the year t

t	y_t	y_{t-1}
…	…	…
1999	16.56	11.91
2000	27.39	16.56
2001	23	27.39
2002	22.81	23
…	…	…

y_{t-1} column is obtained by shifting y_t by 1.

The lag between y_t and y_{t-1} is one time-step.

Compute one-lag time series

```
# Create lag 1 time series.
priceL1 <- lag(price, k=-1)
# Put them together
price_all <- cbind(price=price, priceL1=priceL1)
price_all[1:5,]

     price priceL1
[1,] 25.10      NA
[2,] 37.42   25.10
[3,] 35.75   37.42
[4,] 31.83   35.75
[5,] 29.08   31.83
```

Compute one-lag time series

```
# Create lag 1 time series.
priceL1 <- lag(price, k=-1)
# Put them together
price_all <- cbind(price=price, priceL1=priceL1)
price_all[1:5,]
```

```
     price priceL1
[1,] 25.10      NA
[2,] 37.42   25.10
[3,] 35.75   37.42
[4,] 31.83   35.75
[5,] 29.08   31.83
```

`priceL1` in the first row is NA because we did not have data from 1978 to put under y_{t-1} column of 1979.

Linear regression model

The simple linear regression model is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

Linear regression model

The simple linear regression model is:

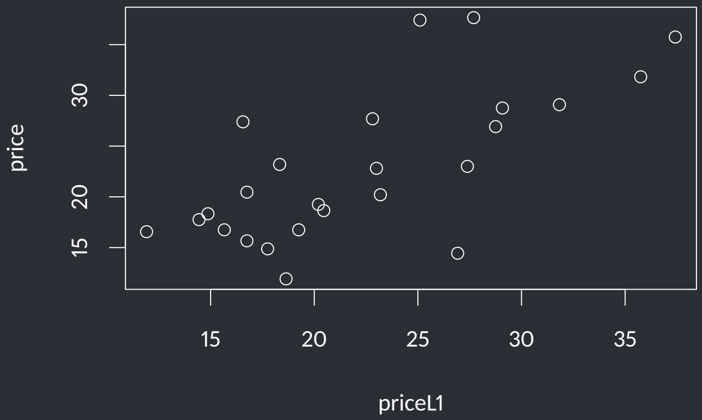
$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

Note that we obtained our predictor from the response itself, lagged 1 time step!

When we use such a model, we expect to see a linear relation between the predictor and the response. Let’s see if there is such a relation!

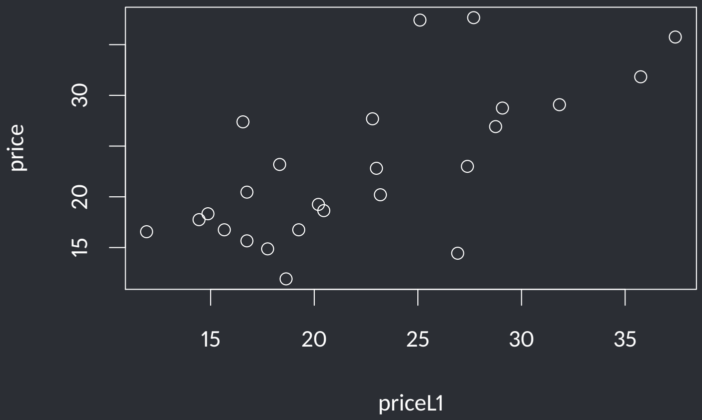
When we use such a model, we expect to see a linear relation between the predictor and the response. Let's see if there is such a relation!

```
plot(price ~ priceL1, xy.labels=F, xy.lines=F)
```



When we use such a model, we expect to see a linear relation between the predictor and the response. Let's see if there is such a relation!

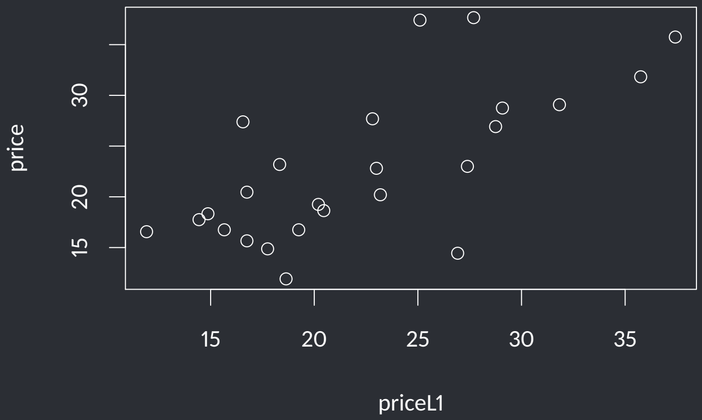
```
plot(price ~ priceL1, xy.labels=F, xy.lines=F)
```



The oil prices seem to be correlated with its first lag!

When we use such a model, we expect to see a linear relation between the predictor and the response. Let's see if there is such a relation!

```
plot(price ~ priceL1, xy.labels=F, xy.lines=F)
```



The oil prices seem to be correlated with its first lag! This is called autocorrelation.

```r
model <- lm(price ~ priceL1, data=price_all)
summary(model)
```

```


Call:
lm(formula = price ~ priceL1, data = price_all)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.9046  -2.9505  -0.8162   1.6303  12.4595 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.8724     3.8389   1.530 0.139722    
priceL1       0.7605     0.1642   4.632 0.000116 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.454 on 23 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4827,Adjusted R-squared:  0.4602 
F-statistic: 21.46 on 1 and 23 DF,  p-value: 0.0001164
```

```r
model <- lm(price ~ priceL1, data=price_all)
summary(model)
```

```


Call:
lm(formula = price ~ priceL1, data = price_all)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.9046  -2.9505  -0.8162   1.6303  12.4595 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.8724     3.8389   1.530 0.139722    
priceL1       0.7605     0.1642   4.632 0.000116 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.454 on 23 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.4827,Adjusted R-squared:  0.4602 
F-statistic: 21.46 on 1 and 23 DF,  p-value: 0.0001164
```

This is a first-order autoregressive, AR(1), model.

AR(2) model

Let's try to add one more lag.

```
# Create lag 2 time series.
priceL2 <- lag(price, k=-2)
# Put them together
price_all <- cbind(price=price, priceL1=priceL1, priceL2=priceL2)
price_all[1:5,]
```

```
     price priceL1 priceL2
[1,] 25.10      NA      NA
[2,] 37.42   25.10      NA
[3,] 35.75   37.42   25.10
[4,] 31.83   35.75   37.42
[5,] 29.08   31.83   35.75
```

AR(2) model

Let's try to add one more lag.

```
# Create lag 2 time series.
priceL2 <- lag(price, k=-2)
# Put them together
price_all <- cbind(price=price, priceL1=priceL1, priceL2=priceL2)
price_all[1:5,]
```

```
     price priceL1 priceL2
[1,] 25.10      NA      NA
[2,] 37.42   25.10      NA
[3,] 35.75   37.42   25.10
[4,] 31.83   35.75   37.42
[5,] 29.08   31.83   35.75
```

The model then becomes:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t$$

AR(2) model

```
model <- lm(price ~ priceL1 + priceL2, data=price_all)
summary(model)
```

```
Call:
lm(formula = price ~ priceL1 + priceL2, data = price_all)

Residuals:
     Min       1Q   Median       3Q      Max 
-10.6861  -3.0937   0.7269   2.3375  10.9071 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   7.1749     3.7363   1.920 0.068505 .  
priceL1       0.8427     0.2073   4.064 0.000557 ***
priceL2      -0.1646     0.2094  -0.786 0.440530    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.911 on 21 degrees of freedom
  (4 observations deleted due to missingness)
Multiple R-squared:  0.5411,Adjusted R-squared:  0.4974 
F-statistic: 12.38 on 2 and 21 DF,  p-value: 0.0002807
```

AR(2) model

- `priceL2` is not statistically significant, if we include `priceL1`.

AR(2) model

- `priceL2` is not statistically significant, if we include `priceL1`.
- This indicates that if we know last year's price, knowing the price from two years ago does not have a significant effect on our predictions!

AR(2) model

- `priceL2` is not statistically significant, if we include `priceL1`.
- This indicates that if we know last year's price, knowing the price from two years ago does not have a significant effect on our predictions!
- On its own, `priceL2` was significant, and still had a positive correlation with `price`... but that correlation was smaller than the correlation between `priceL1` and `price`.

Autocorrelation Function

The Autocorrelation Function (ACF) plots the correlation between the series and each of its lags, to help determine how many lags to include in our model.

```
acf(price)
```