universität innsbruck

Methods of Empirical Finance

Seminar (UE)

Christoph Huber
University of Innsbruck

Master in Banking and Finance
Winter term 2019/20 (this version: 2019-12-03)

Regression Analysis

Methods of Empirical Finance

Regression Analysis

A simple (univariate) linear regression model

$$y_i = \alpha + \beta\, x_i + \varepsilon_i$$

with $i \in [1, 2, \ldots, n]$ referring to the ith observation

$\alpha\ldots$ intercept (often denoted β_0)

$\beta\ldots$ slope coefficient

$\varepsilon\ldots$ a random disturbance term

Regression Analysis

Example: Download stock price data

```r
library(tidyverse)
library(tidyquant)

sp500 <- tq_get("^GSPC")  # download S&P500 prices
nflx <- tq_get("NFLX")      # download Netflix (NFLX) prices
msft <- tq_get("MSFT")    # download Microsoft (MSFT) prices
```

Stata:

`getsymbols` command (install using `ssc install getsymbols`)

Regression Analysis

Example: MSFT prices

Show 8 entries Search:

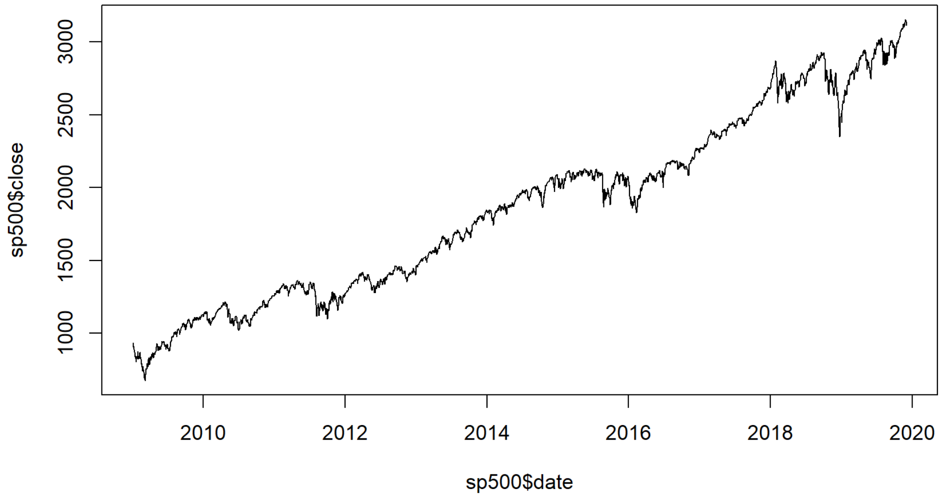
	date	open	high	low	close	volume	adjusted
1	2009-01-02	19.530001	20.4	19.370001	20.33	50084000	15.635055
2	2009-01-05	20.200001	20.67	20.059999	20.52	61475200	15.781175
3	2009-01-06	20.75	21	20.610001	20.76	58083400	15.96575
4	2009-01-07	20.190001	20.290001	19.48	19.51	72709900	15.004425
5	2009-01-08	19.629999	20.190001	19.549999	20.120001	70255400	15.473547
6	2009-01-09	20.17	20.299999	19.41	19.52	49815300	15.012109
7	2009-01-12	19.709999	19.790001	19.299999	19.469999	52163500	14.97366
8	2009-01-13	19.52	19.99	19.52	19.82	65843500	15.242832

Showing 1 to 8 of 20 entries Previous 1 2 3 Next

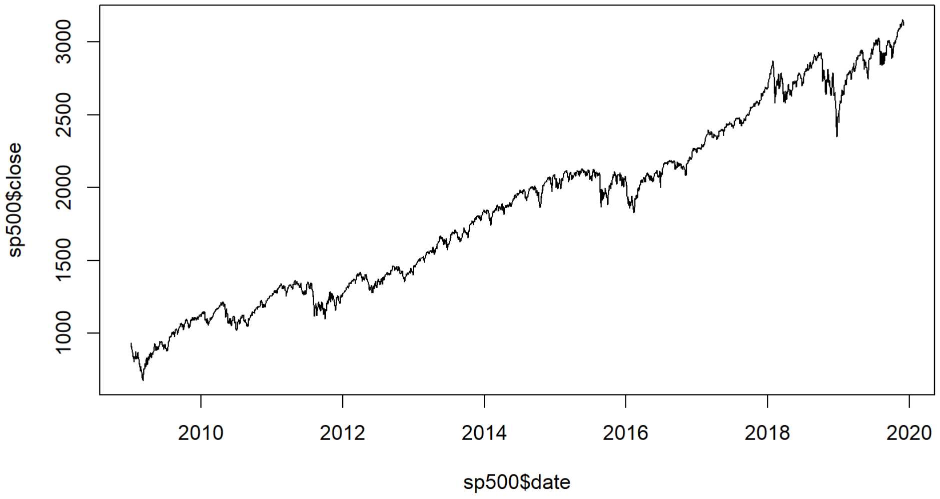
Regression Analysis

Example: Plot price development

```
plot(sp500$date, sp500$close, type = "l")
```




```
plot(sp500$date, sp500$close, type = "l")
```

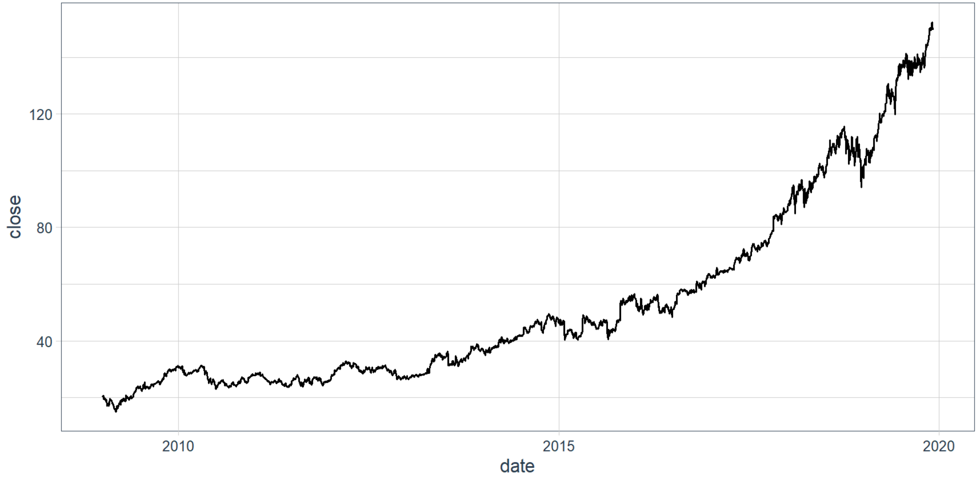


Regression Analysis

Example: Plot price development (`ggplot2`)

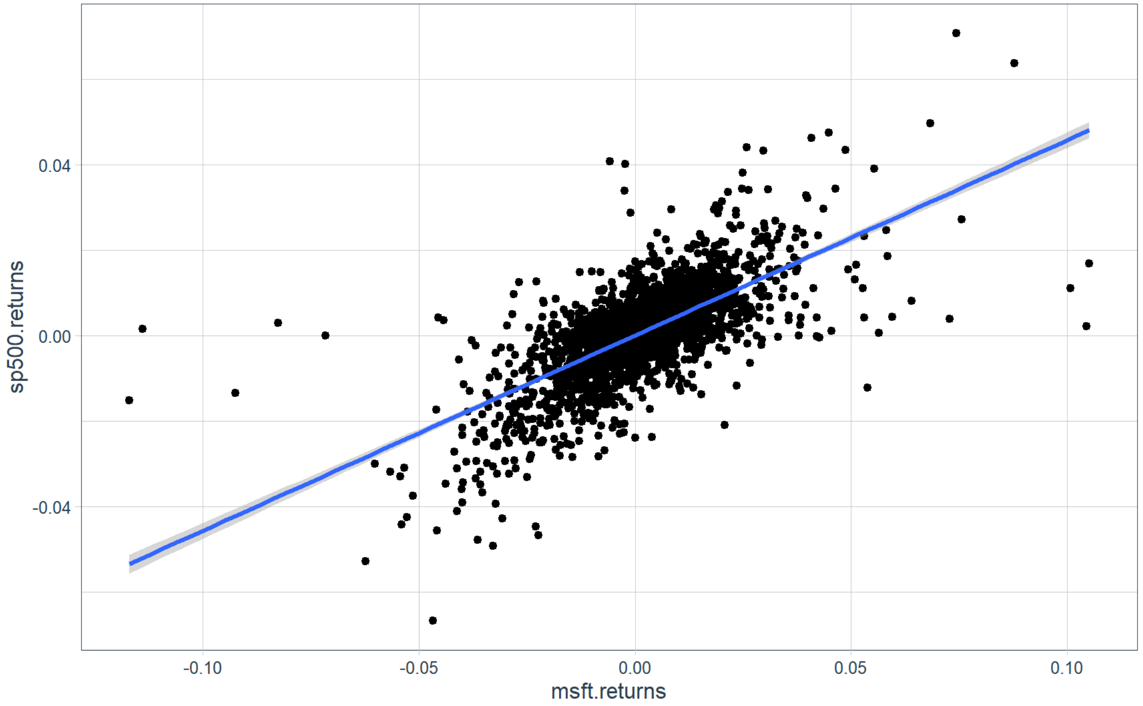
```
sp500 %>%
  ggplot(aes(x = date, y = close)) +
  geom_line() +
  theme_tq()
```


```
msft %>%
  ggplot(aes(x = date, y = close)) +
  geom_line() +
  theme_tq()
```



Regression Analysis

Example: Correlation of Returns



Regression Analysis

Example: Estimating portfolio characteristics using OLS

CAPM:

$$E(R_i) - R_f = \beta_i(E(R_M) - R_f)$$

This implies the following linear model:

$$E(R_i) = \alpha_i + \beta_i E(R_M),$$

where $\alpha_i \neq 0$ unless $\beta_i = 1$.

We can then express the single index model in the form

$$R_{it} = \alpha_i + \beta_i X_t + \varepsilon_{it}, \qquad \varepsilon_{it} \sim i.i.d.\,(0, \sigma_i^2).$$

Regression Analysis

Example: Estimating portfolio characteristics using OLS

```
msft_weekly <- msft %>%
  tq_transmute(
    adjusted, periodReturn, period ="weekly", col_rename = "msft.weekly"
  )
sp500_weekly <- sp500 %>%
  tq_transmute(
    adjusted, periodReturn, period="weekly", col_rename = "sp500.weekly"
  )

weekly_combined <- left_join(msft_weekly, sp500_weekly, by = "date")
```

Regression Analysis

Example: Estimating portfolio characteristics using OLS

```
msft_ols <- lm(msft.weekly ~ sp500.weekly, data = weekly_combined)
summary(msft_ols)
```

```
##
## Call:
## lm(formula = msft.weekly ~ sp500.weekly, data = weekly_combined)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.128693 -0.010955 -0.000378  0.011463  0.131208
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.002231   0.001025   2.176     0.03 *
## sp500.weekly 0.953651   0.047945  19.890   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02435 on 569 degrees of freedom
## Multiple R-squared:  0.4101,    Adjusted R-squared:  0.4091
## F-statistic: 395.6 on 1 and 569 DF,  p-value: < 2.2e-16
```

Regression Analysis

Example: Estimating portfolio characteristics using OLS

```
weekly_combined %>%
  tq_performance(Ra = msft.weekly, Rb = sp500.weekly, performance_fun = table.CAPM)
```

```
## # A tibble: 1 x 12
##   ActivePremium  Alpha AnnualizedAlpha  Beta `Beta-` `Beta+` Correlation
##           <dbl>  <dbl>           <dbl> <dbl>   <dbl>   <dbl>       <dbl>
## 1         0.112 0.0022           0.123 0.954    1.05   0.826       0.640
## # ... with 5 more variables: `Correlationp-value` <dbl>,
## #   InformationRatio <dbl>, `R-squared` <dbl>, TrackingError <dbl>,
## #   TreynorRatio <dbl>
```

Regression Analysis

A multivariate regression model

$$y_i = \beta_1\, x_{i,1} + \beta_2\, x_{i,2} + \ldots + \beta_k\, x_{i,k} + \varepsilon_i, \quad i = 1, \ldots n$$

Vector notation

$$y_i = \mathbf{x}_i'\, \beta + \varepsilon_i, \quad i = 1, \ldots n \qquad \rightarrow \mathbf{x}_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,k} \end{bmatrix} \text{ includes } k \geq 1 \text{ variable(s)}$$

Matrix notation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{with } \mathbf{X} = \begin{vmatrix} x_{1,1} & x_{1,2} & \ldots & x_{1,k} \\ x_{2,1} & x_{2,2} & \ldots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \ldots & x_{n,k} \end{vmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}, \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Regression Analysis

Assumptions of the linear regression model

1. the residuals have zero mean: $E(\varepsilon_i) = 0$
2. the variance of the residuals is constant and finite for all values of x_i: $\mathrm{Var}(\varepsilon_i) = \sigma^2 < \infty$
3. the residuals are linearly independent of each other: $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$
4. the residuals are linearly independent of the corresponding x variates: $\mathrm{Cov}(\varepsilon_i, x_i) = 0$
5. the residuals are normally distributed with zero mean and constant variance σ^2: $\varepsilon_i \sim N(0, \sigma^2)$

Regression Analysis

Example

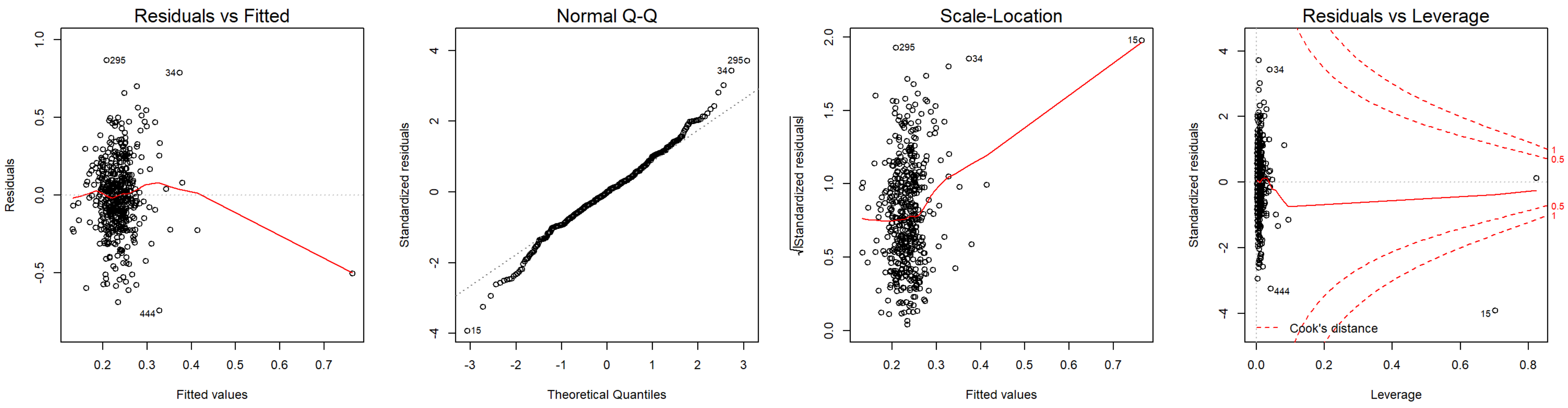
```r
data$PCT_CHANGE <- data$CHG_NET_YTD/(data$PX_LAST-data$CHG_NET_YTD)

reg <- lm(PCT_CHANGE ~ PE_RATIO + PCT_WOMEN_ON_BOARD + BOARD_AVERAGE_TENURE +
            BOARD_AVERAGE_AGE + AVERAGE_BOD_TOTAL_COMPENSATION,
          data = data)
summary(reg)
```

```
##
## Call:
## lm(formula = PCT_CHANGE ~ PE_RATIO + PCT_WOMEN_ON_BOARD + BOARD_AVERAGE_TENURE +
##     BOARD_AVERAGE_AGE + AVERAGE_BOD_TOTAL_COMPENSATION, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.74114 -0.14035  0.00104  0.13755  0.86784
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    3.615e-01  2.208e-01   1.637  0.10229
## PE_RATIO                       8.474e-04  2.862e-04   2.960  0.00323 **
## PCT_WOMEN_ON_BOARD            -2.845e-03  1.231e-03  -2.310  0.02130 *
## BOARD_AVERAGE_TENURE          -3.332e-04  3.675e-03  -0.091  0.92779
## BOARD_AVERAGE_AGE             -1.205e-03  3.601e-03  -0.335  0.73796
## AVERAGE_BOD_TOTAL_COMPENSATION -1.859e-09  3.542e-08  -0.053  0.95815
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2338 on 469 degrees of freedom
##   (30 observations deleted due to missingness)
## Multiple R-squared:  0.02927,    Adjusted R-squared:  0.01892
## F-statistic: 2.829 on 5 and 469 DF,  p-value: 0.01575
```

Regression Analysis

Regression Diagnostics

```
par(mfrow=c(1,4))
plot(reg)
```



Regression Analysis

Heteroskedasticity

Breusch Pagan Test

```
library(lmtest, lib.loc = "I:/R/library")
bptest(reg)
```

```
##
##      studentized Breusch-Pagan test
##
## data:  reg
## BP = 33.823, df = 5, p-value = 2.583e-06
```

We can reject the null hypothesis that the variance of the residuals is constant $\rightarrow$ heteroskedasticity

Regression Analysis

Heteroskedasticity

Non-adjusted standard errors

```
coeftest(reg)
```

```
##
## t test of coefficients:
##
##                                  Estimate  Std. Error t value Pr(>|t|)
## (Intercept)                    3.6151e-01  2.2083e-01  1.6371  0.10229
## PE_RATIO                       8.4737e-04  2.8625e-04  2.9603  0.00323 **
## PCT_WOMEN_ON_BOARD            -2.8450e-03  1.2313e-03 -2.3105  0.02130 *
## BOARD_AVERAGE_TENURE          -3.3325e-04  3.6754e-03 -0.0907  0.92779
## BOARD_AVERAGE_AGE             -1.2055e-03  3.6011e-03 -0.3348  0.73796
## AVERAGE_BOD_TOTAL_COMPENSATION -1.8594e-09  3.5416e-08 -0.0525  0.95815
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression Analysis

Heteroskedasticity

Heteroskedasticity-consistent standard errors

```r
library(sandwich)
coeftest(reg, vcov. = vcovHC(reg, "HC0"))
```

```
##
## t test of coefficients:
##
##                                  Estimate  Std. Error t value Pr(>|t|)
## (Intercept)                    3.6151e-01  2.4987e-01  1.4468  0.14862
## PE_RATIO                       8.4737e-04  5.5162e-04  1.5361  0.12518
## PCT_WOMEN_ON_BOARD            -2.8450e-03  1.4126e-03 -2.0140  0.04458 *
## BOARD_AVERAGE_TENURE          -3.3325e-04  3.5304e-03 -0.0944  0.92484
## BOARD_AVERAGE_AGE             -1.2055e-03  3.8963e-03 -0.3094  0.75717
## AVERAGE_BOD_TOTAL_COMPENSATION -1.8594e-09  1.6788e-08 -0.1108  0.91186
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Regression Analysis

Multicollinearity

Variance inflation factors

```
library(car, lib.loc = "I:/R/library")
vif(reg)
```

```
##                        PE_RATIO             PCT_WOMEN_ON_BOARD
##                        1.012047                       1.027449
##            BOARD_AVERAGE_TENURE              BOARD_AVERAGE_AGE
##                        1.256336                       1.249016
## AVERAGE_BOD_TOTAL_COMPENSATION
##                        1.003105
```

Interpretation: The square root of the variance inflation factor indicates how much larger the standard error increases compared to if that variable had 0 correlation to other predictor variables in the model.

```
vif(reg)^(1/2)
```

```
##                        PE_RATIO             PCT_WOMEN_ON_BOARD
##                        1.006005                       1.013631
##            BOARD_AVERAGE_TENURE              BOARD_AVERAGE_AGE
##                        1.120864                       1.117594
## AVERAGE_BOD_TOTAL_COMPENSATION
##                        1.001551
```