

2. Мультиколлинеарность

```
data <- read.csv(file='../\\dataset.csv')
data$date <- as.Date(data$date)
data$year <- as.numeric(format(data$date, format='%Y'))
```

```
apl <- subset(data, data$Name=='AAPL')
int <- subset(data, data$Name=='INTC')
nvda <- subset(data, data$Name=='NVDA')
pfe <- subset(data, data$Name=='PFE')
msft <- subset(data, data$Name=='MSFT')
ndq <- subset(data, data$Name=='NDAQ')

companies <- data.frame(Date=apl$date, Apple=apl$open,
                        Intel=int$open, NVIDIA=nvda$open, Pfizer=pfe$open,
                        Microsoft=msft$open, Nasdaq=ndq$open)
head(data.matrix(companies))
```

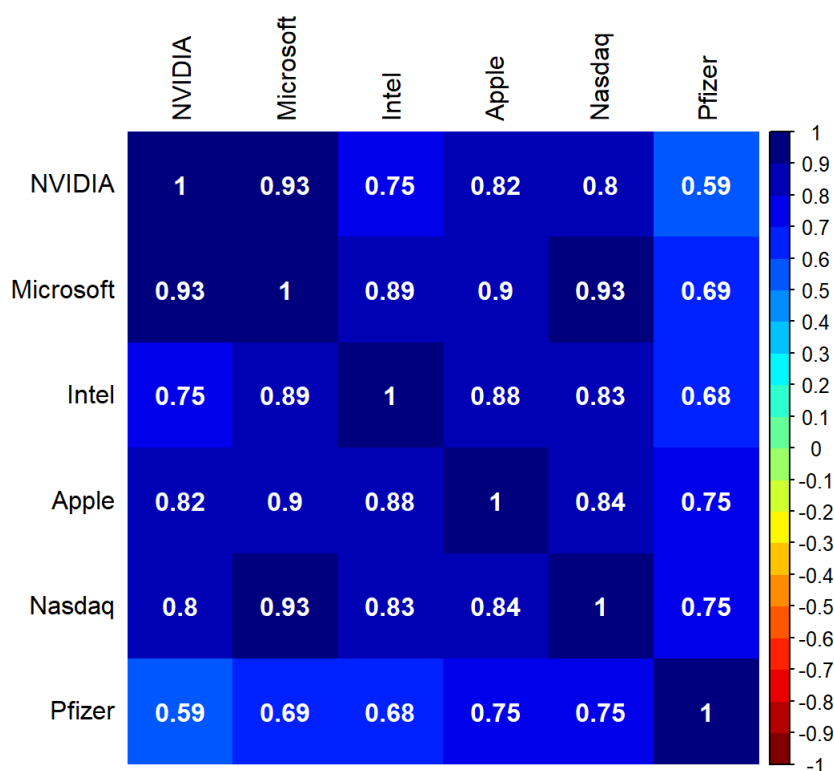
```
##      Date   Apple Intel NVIDIA Pfizer Microsoft Nasdaq
## [1,] 15744 67.7142 20.90 12.370 26.96      27.35 29.40
## [2,] 15747 68.0714 21.00 12.470 26.89      27.65 29.49
## [3,] 15748 68.5014 21.03 12.395 27.14      27.88 31.33
## [4,] 15749 66.7442 21.22 12.440 27.01      27.93 30.44
## [5,] 15750 66.3599 21.13 12.200 26.91      27.92 31.51
## [6,] 15751 66.9785 21.23 12.700 27.15      28.04 30.56
```

2.1 Корреляционная матрица

```
library(corrplot)
```

```
## corrplot 0.90 loaded
```

```
col4 <- colorRampPalette(c("#7F0000", "red", "#FF7F00", "yellow",
                          "#7FFF7F", "cyan", "#007FFF", "blue", "#00007F"))
corrplot(cor(companies[-1]), method = "shade", col = col4(20), cl.length = 21,
         order = "AOE", tl.col = "black", addCoef.col = "white")
```



Т.к. Pfizer - компания из другого сектора экономики, по сравнению с остальными компаниями, то логично, что коэффициенты корреляции с данной акцией ниже

Корреляция между Microsoft и Nasdaq велика. Это объясняется тем, что Microsoft была одной из первых компаний, опубликовавших на Nasdaq, и послужила своеобразной рекламой биржи

Так же заметно, что все значения достаточно велики (>0.5). Это подтверждает факт экономической связи всего рынка (экономические показатели растут и падают вместе с макроэкономикой на больших промежутках времени)

2.2 VIF

```
library(car)
```

```
## Загрузка требуемого пакета: carData
```

```
vif(lm(Nasdaq~., data=companies[-1]))
```

```
##      Apple      Intel      NVIDIA      Pfizer Microsoft
## 7.052822 7.511770 10.230098 2.343868 23.298384
```

Здесь опять же заметен малый вклад Pfizer (т.к. Nasdaq - биржа, специализирующаяся на акциях высокотехнологических компаний) и большой вклад Microsoft

3. Форма связи между категориальными переменными

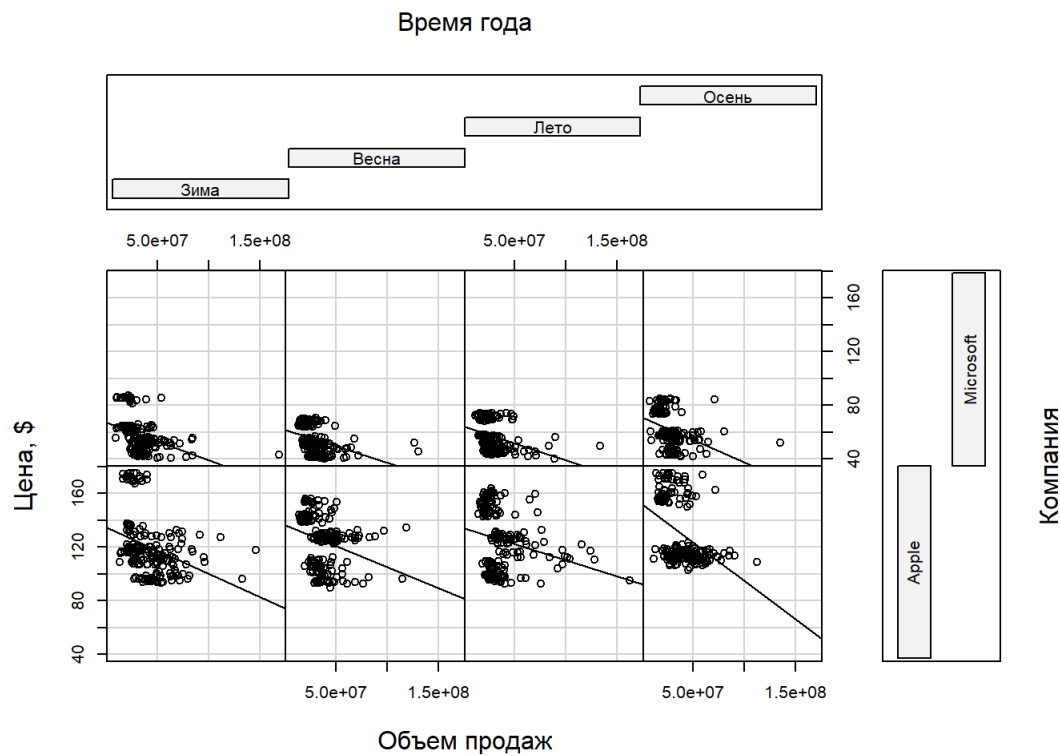
Рассмотрим связь цены акции и объема продаж, разбитых по категориям “время года” и “компания” за 2015-2017 годы Для Apple и Microsoft

```
season <- function(x){ #номер времени года по номеру месяца
  if (x %in% c(12,1,2))
    return(1)
  if (x %in% 3:5)
    return(2)
  if (x %in% 6:8)
    return(3)
  if (x %in% 9:11)
    return(4)
}
```

```
comp <- subset(data, (data$Name %in% c("AAPL", "MSFT")) & (data$year %in% 2015:2017))
comp$season <- sapply(as.numeric(format(comp$date, format='%m')), FUN = season)
```

```
f_season <- factor(comp$season, level=1:4,
  labels=c("Зима", "Весна", "Лето", "Осень"))
f_comp <- factor(comp$Name, levels=c("AAPL", "MSFT"),
  labels=c("Apple", "Microsoft"))
```

```
coplot(comp$open ~ comp$volume | f_season * f_comp,
  ylab = c("Цена, $", "Компания"),
  xlab = c("Объем продаж", "Время года"),
  panel = function(x, y, ...) {
    tmp <- lm(y ~ x)
    abline(tmp)
    points(x, y) })
```



Можно заметить, что осенью у компании Apple наиболее сильная обратная зависимость, это связано с тем, что осенью Apple проводит презентации, где представляет новые продукты, что сподвигает рынок на совершение большого числа сделок

У Microsoft же нет таких различий между временами года

4. АКФ

Построим график АКФ изменения цены на акцию NASDAQ вокруг прямой непрерывного роста

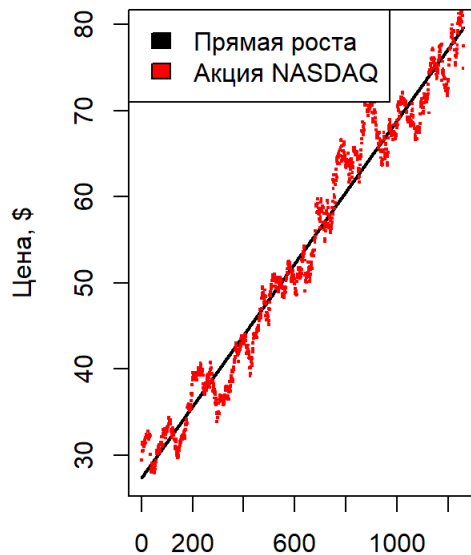
```
ndq$ind <- 1:nrow(ndq)
mean_line <- lm(open ~ ind, data=ndq)

par(mfrow=c(1,2))

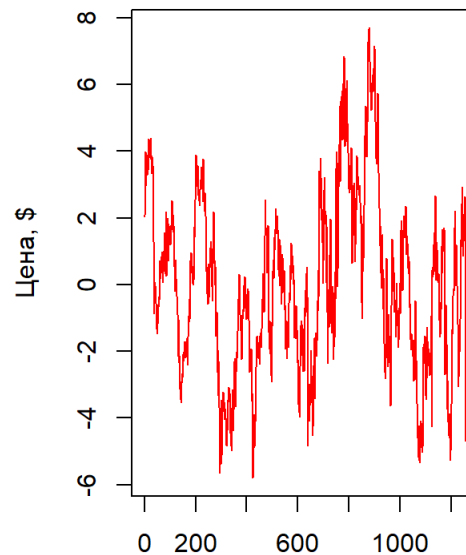
plot(predict(mean_line), xlab='', ylab='Цена, $', type='l', lwd=2, main='Исходная цена')
points(ndq$open, col='red', pch='.', cex=2)
legend('topleft', legend=c('Прямая роста', 'Акция NASDAQ'), fill=c('black', 'red'))

res <- ndq$open - predict(mean_line)
plot(res, main='Колебания вокруг среднего', xlab='', ylab='Цена, $', col='red', type='l')
```

Исходная цена

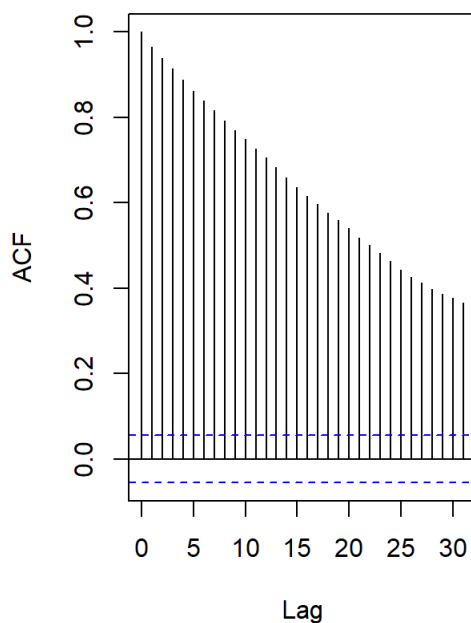


Колебания вокруг среднего

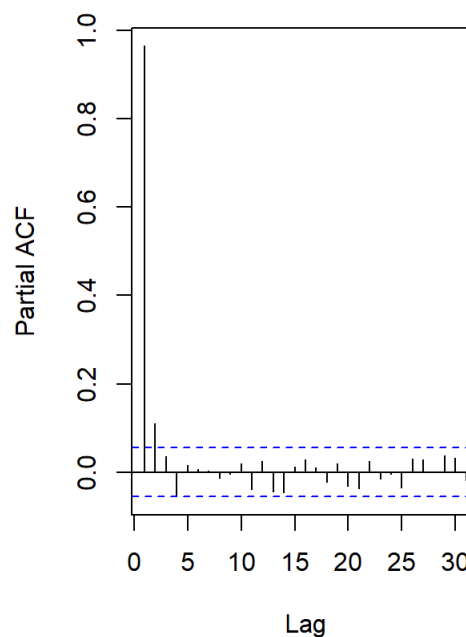


```
par(mfrow=c(1,2))
acf(res, main='АКФ')
pacf(res, main='Частичная АКФ')
```

АКФ



Частичная АКФ



5. Дисперсионный анализ

Рассмотрим зависимость оборота акций NASDAQ от оборота акций компаний, торгующихся на этой бирже

```
nasdaq_depend <- data.frame(Apple=(apl$open * apl$volume),
                             Intel=(int$open * int$volume),
                             NVIDIA=(nvda$open * nvda$volume),
                             Microsoft=(msft$open * msft$volume),
                             Nasdaq=(ndq$open * ndq$volume))
head(data.matrix(nasdaq_depend))
```

```
##           Apple      Intel      NVIDIA  Microsoft      Nasdaq
## [1,] 10710247755 588299763 94968646 911255669 25394426
## [2,] 8783213601 396818625 127288760 891644730 197802347
## [3,] 10400523927 597001995 86084081 1003424313 106050452
## [4,] 7924004579 451281942 265252808 1165114753 97732703
## [5,] 5893366579 745066409 269001887 911955818 178126692
## [6,] 6558844897 717760251 233000093 1392201086 84514627
```

```
lin <- lm(Nasdaq ~ ., data=nasdaq_depend)
data.matrix(anova(lin))
```

```
##           Df          Sum Sq      Mean Sq  F value      Pr(>F)
## Apple      1 9.044252e+15 9.044252e+15 11.45071 7.367343e-04
## Intel      1 2.010522e+16 2.010522e+16 25.45475 5.195592e-07
## NVIDIA     1 4.201294e+16 4.201294e+16 53.19159 5.345662e-13
## Microsoft  1 5.539720e+16 5.539720e+16 70.13709 1.463242e-16
## Residuals 1254 9.904616e+17 7.898418e+14      NA      NA
```

p-value мало, что говорит о статистической значимости всех компаний в цену акции NASDAQ

F value для Microsoft больше, чем для других компаний, что еще раз подтверждает мысль о большем вкладе Microsoft в акции NASDAQ