Technical Analysis of Capital Market Data in R - First Steps [*]

Prof. Dr. Michael Feucht [†]

April 25th, 2018

Abstract

To understand the classical textbook models of Modern Portfolio Theory and critically reflect their assumptions, we need convenient ways to download real market data and analyze it statistically. There are data providers who offer an automated programming interface (API) for downloading and automatically updating price data as an irregular time series into the free software R. In this tutorial, we show the first steps to visualize and analyze capital market data from the free sources of the data provider Quandl.

Keywords:
Modern Portfolio Theory, Markowitz, Mandelbrot, stock price, continuous return, technical analysis, time series, random walk, normal distribution, Quandl, R

Contents

1 Introduction

When students start studying the model world of Modern Portfolio Theory, the assumptions of the models are usually not discussed in depth. They are informed that stock prices follow a random walk and that this corresponds to returns being normally distributed. Most often it is not even mentioned that they have to think about continuous (i.e. logarithmic) returns. The joint behavior of risky assets is typically modeled with a multivariate normal distribution so that expected returns, standard deviations of returns (which are called volatility of the asset), and the bilateral covariances or correlations between different assets suffice to fully describe the stochastic behavior of the market. In textbook examples, the $\mu's$, $\sigma's$ and $\rho'_{ij}s$ are most often given without being questioned.

[*] DOI: 10.23779/0003

[†] Faculty of Business, Augsburg University of Applied Sciences, An der Hochschule 1, 86161 Augsburg, michael.feucht@hs-augsburg.de

Although the classical μ-σ-world of Harry Markowitz (originally developed in [Mar52], the Capital Asset Pricing Model (CAPM) - nicely summarized in [Wil90], William F. Sharpe's nobel lecture - , or the Black-Scholes-Merton formula for pricing options (as developed in [BS73]) are important for understanding and studying market behavior in general, it is also necessary to understand that models are always a **simplified** image of reality. A restrictive model assumption does not necessarily invalidate the result of an analysis, but sometimes the behavior of real markets deviates considerably from textbook modeling. Benoît Mandelbrot, who was the first to define various degrees of randomness, stated that the classical models (under normal distribution assumptions) describe mild random behavior whereas reality seems to be wildly random [Man97]. A popular, but nevertheless fascinating description of Mandelbrot's ideas can be found in [MH05]. For a detailed overview on Modern Portfolio Theory (including some alternatives to the assumption on normally distributed returns), see [Fra13].

To experience the wild randomness of price behavior in real capital markets and to understand, what the deficits of the traditional models are - in particular in which situations models will let us down or we need more sophisticated assumptions - , students need to analyze real market data. This paper gives a brief overview on how to get started with technical capital market analysis in R. We will download stock price data from the data provider Quandl, visualize the prices, calculate daily continuous returns and then perform a first basic check for normal distribution.

2 Import of Market Data as Time Series

For technically analyzing capital market data, in particular for empirically testing Modern Portfolio Theory and its assumptions, it is sufficient to look at one price for the respective financial instrument per day, typically a closing or adjusted closing price. We will not need intraday data (or - even beyond prices - order book information), which is only necessary for trading simulations or automated trading (algorithmic trading/high-frequency trading). As the tool for analysis, we will use the free software R. All examples were developed on R version 3.4.2 with RStudio Version 1.1.383.

The data base we will download our market data from is Quandl (`http://www.quandl.com`). Quandl is a data platform which with its free account still provides a wide range of data sets with a nice and easy to handle API (automated programming interface) for R. To make use of this, we will have to install the package `Quandl` in R. For the further data analysis, we will need the library `zoo`. `zoo` objects are in particular useful, when dealing with irregular time series. With trading data, this is usually the case, as we have non-trading dates, which we would like to ignore (Saturdays, Sundays and bank holidays). So the first step in our R script should be to import the necessary libraries (make sure that the corresponding packages were installed in RStudio before this step):

```r
library("Quandl")
library("zoo")
```

To access Quandl data, we have to set our (individual) API key. To have an API key, you have to register with the Quandl website. For our purposes, it is absolutely sufficient to have the free Quandl account. With your API key, the corresponding line in R should read

```r
Quandl.api_key("YourIndividualAPIKey")
```

Now you can access the Quandl database. For all further considerations, we read in the daily stock prices for BMW AG common stock (available since January 18th, 2000, but we will restrict ourselves to the last 10 years). The original source is Deutsche Börse (Frankfurt Stock Exchange).

```r
# Read in daily trading data from Quandl.  Import as object 'zoo'.
# Non-trading days are automatically omitted from time series.  For BMW AG,
# all trading data since January 18th, 2000 are available.  We will only
# analyze the last 10 years (since January 1st, 2008).  Original Source is
# Deutsche Börse Frankfurt.
```

```
bmw_quote <- Quandl("FSE/BMW_X", type = c("zoo"), start_date = "2008-01-01",
    force_irregular = FALSE)
```

This gives you zoo objects `bmw_quote` with several dimensions: `bmw_quote` includes (in this order) opening, highest, lowest, closing price of the day, change (daily return) - but this dimension does not contain reasonable data, so we will calculate returns ourselves-, traded volume and turnover. The additional fields ”last price of the day”, ”daily traded units” and ”daily turnover” are not filled for the data set `FSE/BMW_X`.

It would be possible to restrict the time frame of the time series with an end date by specifying `end_date= "yyyy-mm-dd"`. Without this parameter, the code sequence has the convenient property of updating to the latest data set available, whenever it is rerun.

To simplify the further analysis, we can assign an individual variable for the closing price of the day (in Quandl, you can see that the closing price is the fourth dimension of the dataset):

```
bmw_close <- bmw_quote[, 4]
```

3 Graphing Market Prices in R

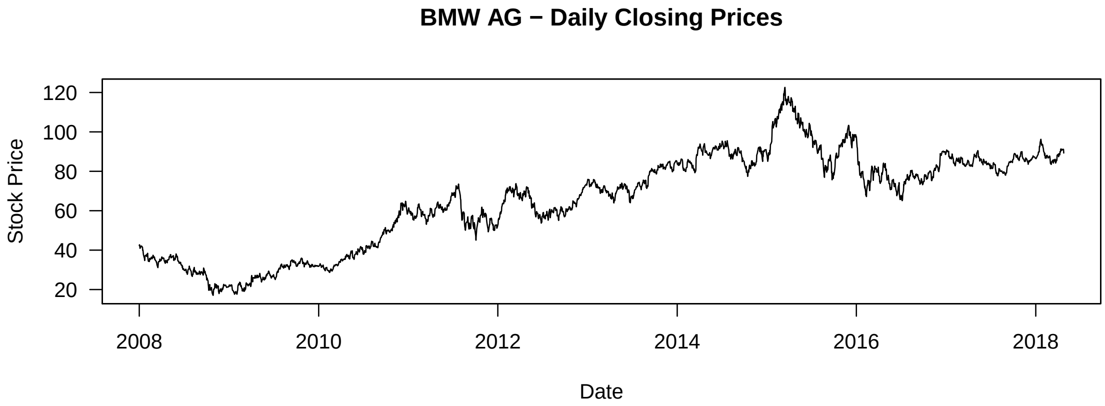
Before we will start with statistical analysis of market data, it is helpful to get a first visual impression of the time series. Graphical analysis is the starting point of any time series statistics (and gives us the possibility to display our findings in the end). This brief section will give you a first insight into the graphical possibilities of R.

3.1 Absolute Prices

The easiest way to display a time series of prices graphically is just a line plot of the absolute amounts. Let us start with a plot of our BMW stock closing prices:

```
# Parameters: las=1 makes sure that axis labels are always horizontal; lty=1
# is line type 'solid; and mai specifies the margins (in inches)
par(las = 1, lty = 1, mai = c(1, 1, 1, 1))

plot(bmw_close, main = "BMW AG - Daily Closing Prices", xlab = "Date", ylab = "Stock Price",
    col = "black")
```



The result is the typical zig-zag-line visualization of stock prices as a time series.

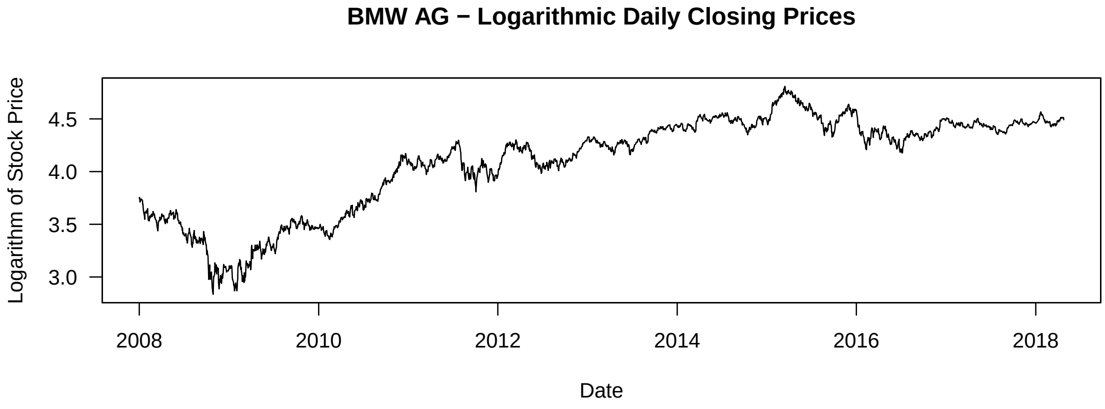
3.2 Logarithmic Prices

Although the actual price movement may also be useful, it is in many cases helpful to look at logarithmic prices. Exponential development - which is the result of cumulated growth over time - is translated into a linear trend of the logarithmic prices and shows as a straight line trend when you graph the logarithmic prices. In R, you can transform the time series `bmw_close` of BMW stock prices into the time series `bmw_close_log` of logarithmic prices with just one single R command line: (*log* is the natural logarithm)

```
bmw_close_log <- log(bmw_close)
```

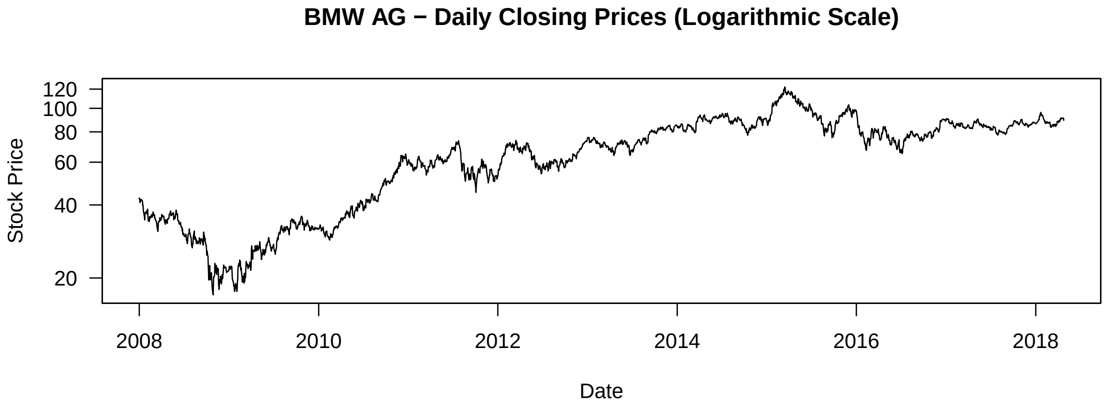
Again we plot these logarithmic prices:

```
par(las = 1, lty = 1, mai = c(1, 1, 1, 1))
plot(bmw_close_log, main = "BMW AG - Logarithmic Daily Closing Prices", xlab = "Date",
    ylab = "Logarithm of Stock Price", col = "black")
```



At least in some segments, we can see the linear trend. But this method has the disadvantage that now the y-axis labels do not show the original stock prices anymore, but rather the logarithm of the price (which does not have a reasonable economic interpretation). To overcome this disadvantage, R allows us to directly plot the actual prices on a logarithmic scale:

```
par(las = 1, lty = 1, mai = c(1, 1, 1, 1))
plot(bmw_close, main = "BMW AG - Daily Closing Prices (Logarithmic Scale)",
    xlab = "Date", ylab = "Stock Price", log = "y", col = "black")
```



Remark that here the y-axes labels still show the original prices whereas in the previous graph the y-axes was labeled with the logarithm of the price.

4 Returns

4.1 Two Ways to Define Daily Returns

In general, there are two way to calculate the return from a price change: either as a discrete return r^d (this is how we intuitively think of return) or as a continuous logarithmic return r^c (this is how we think of return in the field of finance).

If S_0 denotes the actual spot price ("now") and S_1 is the future price after one unit of time (which may be one day, one month, one year, but also one hour, one minute or even one second), then the formulas for return after one unit of time are

$$r^d = \frac{S_1 - S_0}{S_0}$$

and

$$r^c = \ln\left(\frac{S_1}{S_0}\right) = \ln S_1 - \ln S_0$$

The two definitions correspond to different growth models:

$$S_1 = S_0\left(1 + r_d\right)$$

characterizes a discrete growth process, whereas the exponential equation

$$S_1 = S_0 e^{r_c}$$

describes continuously compounded returns.

In R, it is extremely easy to calculate with irregular and lagged data in time series. This is a sharp contrast to other statistical tools or spreaqdsheet calculations where it is often tedious and more complicated to work with time indexed data, especially when the time series is irregular - as it is always the case with market data that only exists for trading days. To calculate daily returns of the BMW closing prices time series, we just have to do the following:

```
# Discrete return
bmw_return_daily_d <- diff(bmw_close, lag = 1)/lag(bmw_close, 1)

# Continuous return
bmw_return_daily_c <- diff(log(bmw_close), lag = 1)
```

Both calculations give us a time series of daily returns. As a consequence of the different formulas, the values differ slightly. For further analysis, we will only use continuous returns. The next section 4.2 will explain briefly why this is reasonable.

4.2 Why We Should Always Use Continuous Returns

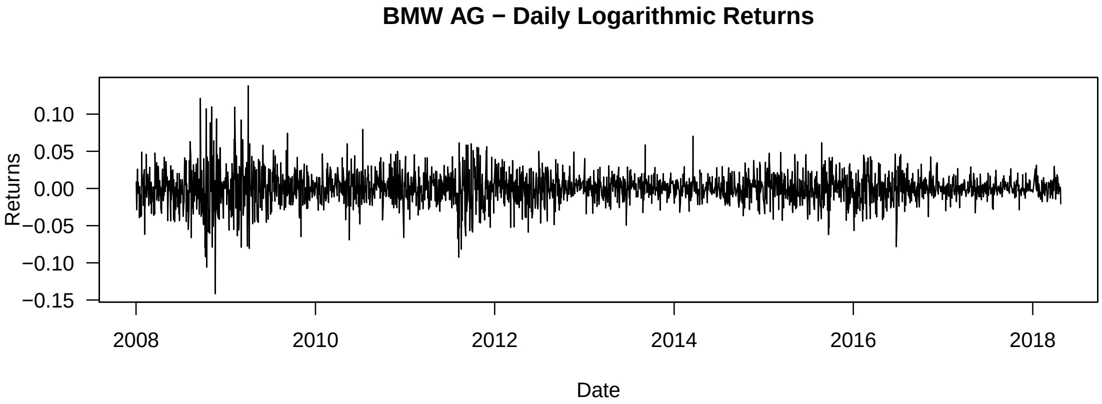
Before we continue, it is important to understand that the discrete return (interpreted as the observations of i.i.d. - independent and identically distributed - random variables) can never have the desired property of following a Gaussian normal distribution, as the normal distribution takes all values from $-\infty$ to $+\infty$ whereas r_d is limited by $-100\% = -1$. For small values, the difference between r_c and r_d is only small. Nevertheless, to be theoretically consistent, when analyzing stock prices or other capital market data, we always use continuous returns. This allows us to assume underlying probability distributions which are defined for all values on the real axis.

4.3 A Visual Analysis of Continuous Returns

Now we can visualize the daily continuous returns as a time series with

```
# Parameters: las=1 makes sure that axis labels are always horizontal; lty=1
# is line type 'solid; and mai specifies the margins (in inches)
par(las = 1, lty = 1, mai = c(1, 1, 1, 1))

plot(bmw_return_daily_c, main = "BMW AG - Daily Logarithmic Returns", xlab = "Date",
    ylab = "Returns", col = "black")
```



A histogram of the observed returns shows, that their distribution resembles the density of a normal distribution function at least a little bit:

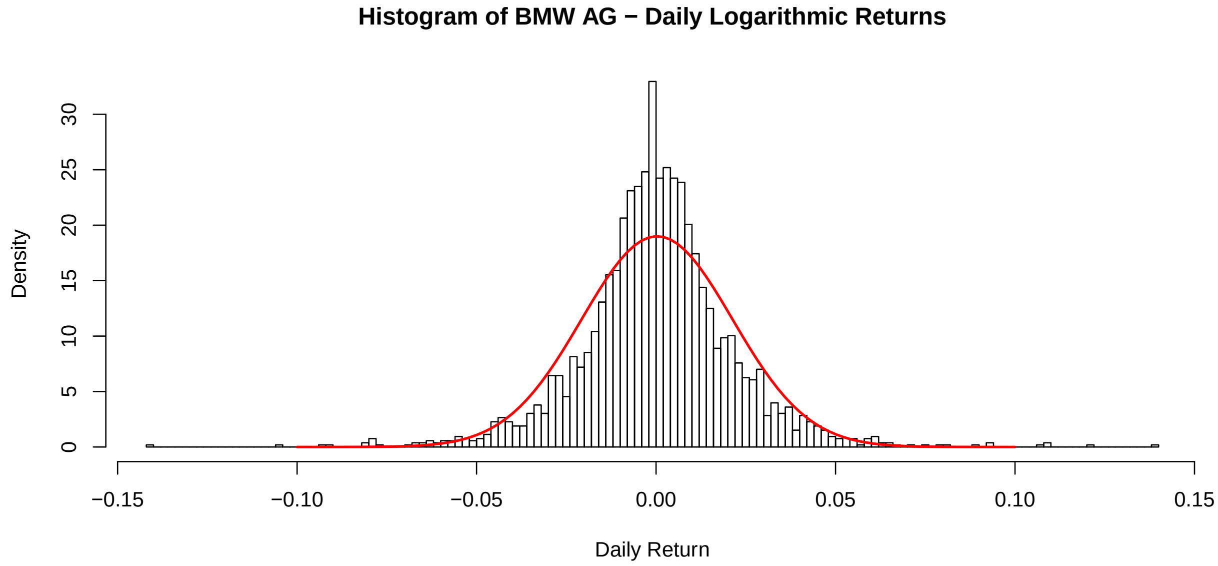
```
# freq=FALSE: probability density (instead of frequencies) is plotted

hist(bmw_return_daily_c, main = "Histogram of BMW AG - Daily Logarithmic Returns",
    breaks = 100, freq = FALSE, xlab = "Daily Return", plot = TRUE)

# For comparison: plot density function of a normal distribution.  Sample
# mean bmw_mu and sample volatility bmw_volatility have to be calculated
# beforehand

bmw_mu <- mean(bmw_return_daily_c)
bmw_volatility <- sd(bmw_return_daily_c)

normdist <- function(x) {
    dnorm(x, mean = bmw_mu, sd = bmw_volatility)
}
curve(normdist, from = -0.1, to = +0.1, add = TRUE, col = "red", lwd = 2)
```



But although the the histogram has the bell shape of a normal distribution, we can observe two important deviations:

1. Events which are close to the mean appear more often than the normal distribution would suggest. So in comparison to the density of a normal distribution, the density of observed returns is more peaked in the middle.
2. Extreme events (i.e. excessively large positive or negative returns) also appear more frequently than expected. This is called the "fat tails" or "heavy tails" of the distribution.

Both effects are the result of the leptocurtic property of the distribution. It can also be said that the distribution shows excess curtosis. Furthermore, asset returns are usually not symmetric (as a normally distributed random variable would be), but rather skewed to the left or to the right. Theoretically, this could be overcome by using so-called stable Paretian distribution functions instead of the normal distribution. While this approach solves some of the problems, it on the other hand invalidates the classical Markowitz portfolio optimization based upon return vector and covariance matrix.
Furthermore, it is quite obvious that the assumption of all returns being i.i.d. (independent and identically distributed) is far from being realistic. Even if we accept the notion of every single return being the result of a stochastic experiment with a certain expectation and standard deviation, these parameters will obviously be time dependent (i.e. μ_t and σ_t). There exist many approaches for time-dependent modeling of volatilities which may also be analyzed with R but this will have to be the focus of another working paper.

5 Conclusion

This short paper gives an introduction to technical capital market analysis in R. With some easy to understand lines of code, we downloaded a time series of BMW AG stock prices for the last 10 years from data provider Quandl into R. Readers will be able to replicate this procedure with any risky asset they like. We calculated logarithmic returns as a time series, and then - with one line of code each - static empirical estimators for expected return (μ) and volatility (standard deviation of returns, σ) for BMW AG. We compared the real distribution of returns over the time horizon of 10 years with the density function of the corresponding normal distribution. It turned out that, although the shape at a first glance seemed to be similar to a normal distribution, our histogram showed certain properties which were not in line with the assumption of a normal distribution.

References

[BS73] Fischer Black and Myron Scholes.
The Pricing of Options and Corporate Liabilities.
Journal of Political Economy, 81(3):637–654, 1973.

[Fra13] Jack Clark Francis.
Modern portfolio theory.
Wiley finance series. J. Wiley & Sons, Hoboken and N.J, 2013.

[Man97] Benoît B. Mandelbrot.
Fractals and scaling in finance: Discontinuity, concentration, risk, volume selected works of Benoit B. Mandelbrot, reprinted, translated or new with annotations and guest contributions ; Vol. E of *Selecta*.
Springer, New York, NY, 1997.

[Mar52] Harry Markowitz.
Portfolio Selection.
The Journal of Finance, 7(1):77–91, 1952.

[MH05] Benoît B. Mandelbrot and Richard L. Hudson.
The (mis)behaviour of markets: A fractal view of risk, ruin, and reward.
Profile Books, London, [nachdr.] edition, 2005.

[Wil90] William F. Sharpe.
Capital Asset Prices with and without Negative Holdings: Nobel Lecture, December 7, 1990.
1990.