

2. Theory

It is assumed that the reader of this paper already has basic knowledge of the mathematical principles of time series analysis. Therefore, this section will only briefly describe the mathematical models and processes.

2.1. Time-Series

Almost anything with a data point to a given timestamp can be named a time-series. The monthly gross domestic product, the weekly US gasoline production, or daily stock price movements. In this paper lies the focus of the analysis of financial time series. Due to trades often only take place during the week, there are gaps in the time series on the weekends, an exception would be the trading of cryptocurrencies like Bitcoin which are also tradeable at the weekends.

A series of data points with more or less equidistant time points t with the sample length of T , is called a time-series $x_t, t = 1, \dots, T$ [eco1]. The analysis of a time-series x_t involves creating a reasonable model that can be utilized to perform forecast predictions.

2.1.1. Stationarity

In order to fit a suitable model with a given time series x_t , the assumptions of stationarity must be met. In this practical application, only the following weak-stationarity properties are required.

$$E[x_t] = \mu \quad (1)$$

$$Var(x_t) = \sigma_x^2 \quad (2)$$

$$Cov(x_t, x_{t-k}) = R(k) \quad (3)$$

Many financial time-series are subject to shift, trends or changing volatility. In figure ?? are the stock prices of Alphabet Inc Class A (Google) visualized. This time-series shows a clear upwards drift and towards the end the volatility increases.

```
## 'getSymbols' currently uses auto.assign=TRUE by default, but will
## use auto.assign=FALSE in 0.5-0. You will still be able to use
## 'loadSymbols' to automatically load data. getOption("getSymbols.env")
## and getOption("getSymbols.auto.assign") will still be checked for
## alternate defaults.
##
## This message is shown once per session and may be disabled by setting
## options("getSymbols.warning4.0"=FALSE). See ?getSymbols for details.
```

Adjusted Prices ~ Google

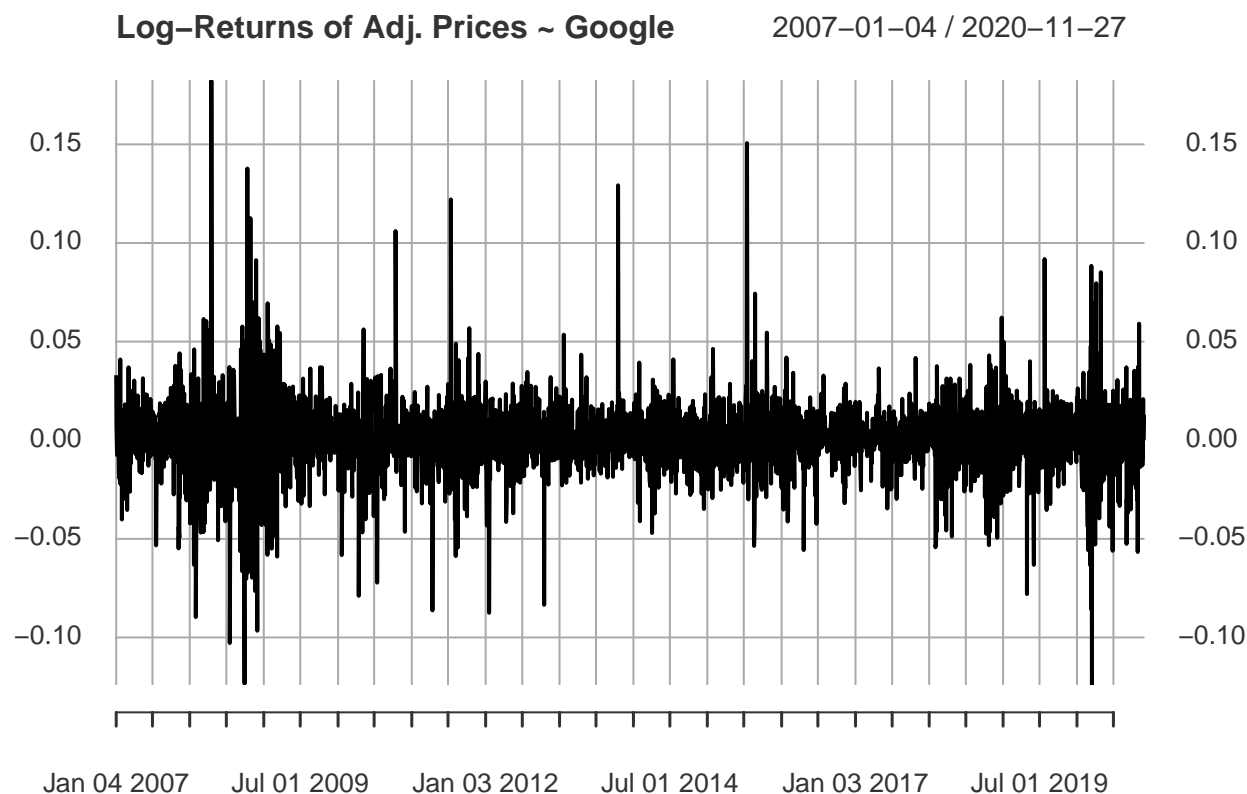
2007-01-03 / 2020-11-27



To improve the violated properties the first difference can be applied and additionally a logarithmic transformation can be performed [slide_eco3_1]. The log-returns transformation can only be performed to strict positive data.

$$\text{LogReturn} = \log(x_t) - \log(x_{t-1})$$

The result is the so-called log-returns.



Applying the transformation to the data causes the drift to disappear, but the series still contains stronger and weaker volatile phases. This effect often occurs in non-stationary financial data and is called volatility cluster. This special property is used for the modeling of forecast models, which will be discussed in chapter 2.2.

In the following examples, we will only work with a section of the time series, as it often makes no sense to look long into the past. The further a value lies in the past, the smaller its influence on a future value will be.

2.1.2. Autocorrelation

The autocorrelation function (ACF) reveals how the correlation between any two data points of the time series changes as their separation changes [acf]. More precisely, acf measures the dependence between x_t and $x_{t \pm k}$ at lag k . The partial autocorrelation (PACF) measures the dependency between x_t and x_{t-k} at lag k [eco1]. For stationary time series, ACF can be used to identify the model order of a MA-process, PACF for AR-processes.

In the following figure ?? are ACF and PACF of the non-stationary adjusted Google stock visualized. Both graphics show the typical pattern of a non-stationary time series. The plot above shows the dependence structure of the time series. This means that it takes a long time until the series changes. Often a large

value is followed by another large value, which indicates a strong trend. This property of the series can be seen in figure ?? as the long upward drift. The plot below indicates a significant partial autocorrelation at lag $k = 1$.

In the following section 2.2. the characteristics of the autocorrelation function can be used for the verification of ARIMA and ARCH-processes.

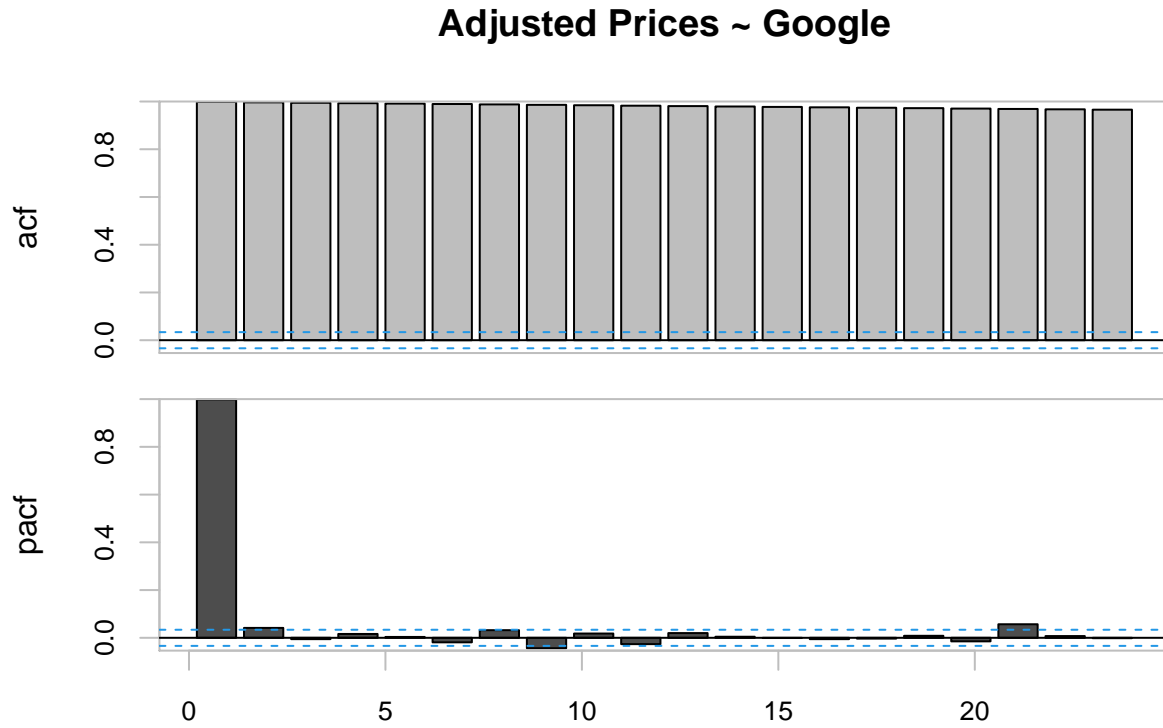


Figure 1: Acf and Pacf of the Adjusted Prizes of Google.

2.2. Models

The following processes are used to determine certain properties and characteristics of a time series so that they are transformed into a model. The goal is to fit the time series as well as possible in order to create reliable forecasts.

2.2.1. ARIMA

An ARIMA(p, d, q) process is defined as follows.

$$x_t = c + a_1x_{t-1} + \dots + a_px_{t-p} + \epsilon_t + b_1\epsilon_{t-1} + \dots + b_q\epsilon_{t-q} \quad (4)$$

- p and q are the AR- and MA-model orders
- a and b are the AR- and MA-model parameters
- d is the differential parameter
- ϵ_t is a white noise sequence
- x_t is the given data x_1, \dots, x_T

The mean of an ARIMA-process can be computed as:

$$\mu = \frac{c}{1 - a_1 - \dots - a_p}$$

ARIMA processes can be divided into 4 different models. Choosing a model that best represents the time series is a difficult task. The goal is to find the best possible model with as few parameters as possible.

The previously introduced ACF and PACF can help to determine the orders of simple models. Provided that the time series is stationary, the model orders can be determined directly. For an AR(p)-process (ARIMA($p, 0, 0$)), the ACF plot will gradually decrease and simultaneously the PACF should have a sharp drop after p significant lags. For an MA(q)-process (ARIMA($0, 0, q$)) the opposite is true, the ACF should show a sharp drop after a certain q number of lags while PACF should show a gradual decreasing trend. If both ACF and PACF show a gradual decreasing pattern, then the ARIMA($p, 0, q$)-process should be considered for modeling [arima]. If the time series is not stationary, differentiation can be considered (ARIMA(p, d, q)).

The application of an analysis method to Google prices finds an $ARIMA(1,1,0)$ as the optimal model. This makes sense if you look back at figure ?? . Long dependency structures in the ACF plot indicating an $AR(p)$ process and at the same time after $lag=1=p$ the PACF has a strong drop. The differential operator $d=1$ transforms the non-stationary series into a stationary one.

To convince yourself of the quality of the model, you can use the Ljung-Box statistics shown in figure ?? . For the lags where the p-values are above the 5% line, the forecasts are reliable. Here the values from the 6 lag are below the significant line, but since one only wants to make short-term forecasts (for example 1 day into the future), the lag is sufficient up to 5. If you want to forecast further than 5 days into the future, you might have to adjust the ARIMA model.

In figure ?? you can see the prediction of the model. The whole representation is shifted by one day so that one can compare the model with a true value. The **green dot** is the actual value of the time series. The **red dots** indicate the upper and lower 95% interval limits respectively. These indicate that a future value will be within this band. The **blue dot** is the point forecast predicted by the model.

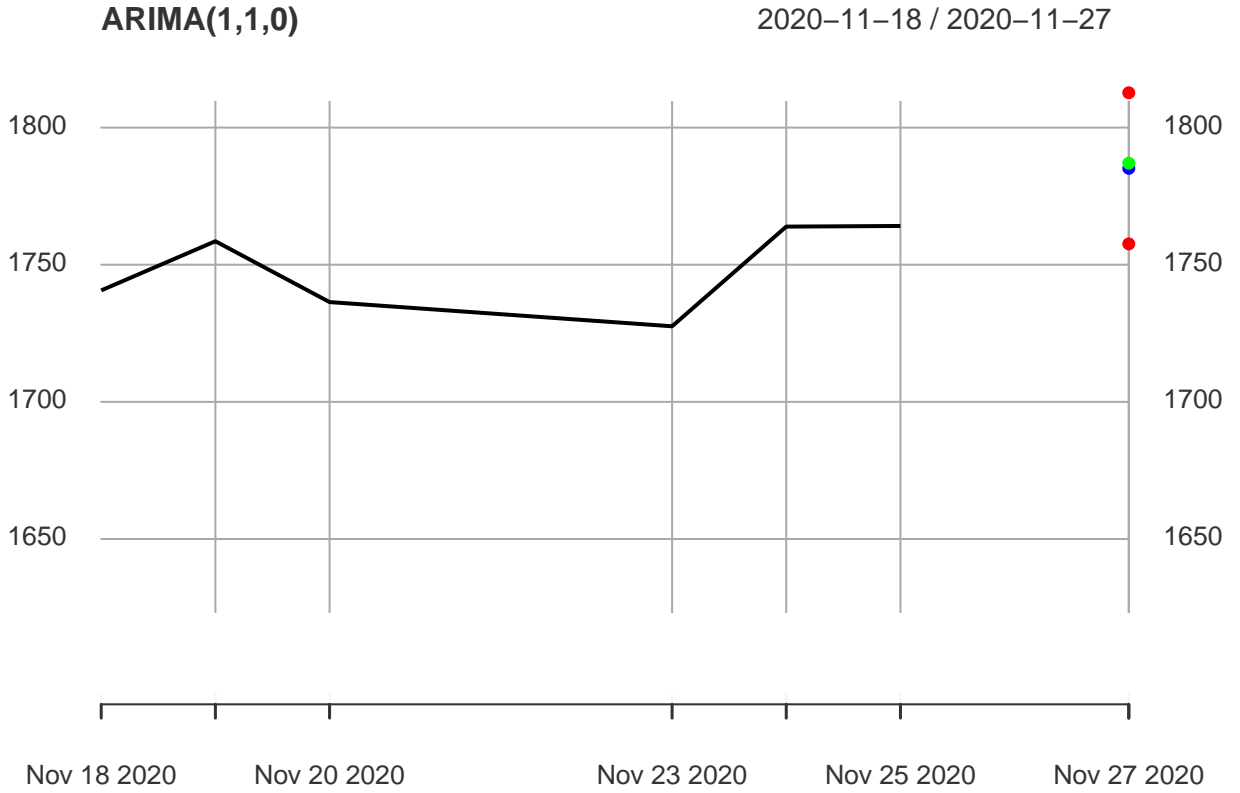


Figure 2: ARIMA-Forecast.

2.2.2. GARCH

The volatility clustering mentioned in section 2.1.1 can be handled with an generalized auto-regressive conditional heteroscedastic process.

$$\begin{aligned}
 \epsilon_t &= \log(x_t) - \log(x_{t-1}) \\
 \epsilon_t &= \sigma_t u_t \\
 \sigma_t^2 &= c\sigma^2 + \sum_{j=1}^n \alpha_j \sigma_{t-j}^2 + \sum_{k=1}^m \beta_k \epsilon_{t-k}^2
 \end{aligned} \tag{5}$$

with

- x_t is the original data (often non-stationary)
- ϵ_t is the stationary log-return
- u_t is independent and identically distributed (iid) and a standardized random variable
- σ^2 is the unconditional variance of the process ϵ_t .
- σ_t^2 is the conditional variance of the process ϵ_t .

With a GARCH(n,m)-process it is possible to model the volaclusters of a time series. The GARCH($1,1$) model has become widely used in financial time series modeling and is implemented in most statistics and

econometric software packages. Those models are favored over other stochastic volatility models by many economists due to their relatively simple implementation [Garch11].

Table 1: Coefficients GARCH(1,1).

	Estimate	Std. Error	p-Value
ω	1.091128e-05	1.847293e-06	3.491711e-09
α_1	7.387656e-02	1.082633e-02	8.867351e-12
β_1	8.948563e-01	1.380181e-02	0.000000e+00

For an optimal model some conditions must be fulfilled. Suppose you want to model the Google time series with a GARCH(1,1).

In table 1 the estimated coefficients of the process can be seen. The p-values are all lower than 0.05 and thus indicate that they are essential for the model. (Note: $\omega = c\sigma^2$)

```
## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
## Consider formula(paste(x, collapse = " ")) instead.
```

The following parameter restrictions are also examined:

$$c + \sum_{j=1}^n \alpha_j + \sum_{k=1}^m \beta_k = 1 \quad (6)$$

with

$$c > 0, \alpha_k \geq 0, j = 1, \dots, n, \beta_k \geq 0, 1, \dots, m$$

To satisfy formula 6, c needs to be determined from ω . First calculate the unconditional variance.

$$\sigma^2 = \frac{\omega}{1 - \alpha_1 - \beta_1}$$

Calculate c with:

$$c = \frac{\omega}{\sigma^2}$$

and then check for the restriction in 6.

For the coefficients of GARCH(1,1) the restrictions are fulfilled. You can see that $c = 1 - \alpha_1 - \beta_1$. So this restriction can be determined easily with:

$$\sum_{j=1}^n \alpha_j + \sum_{k=1}^m \beta_k < 1 \quad (7)$$

If the parameter restrictions are not fulfilled, complications may arise, the forecast of the conditional variance $\hat{\sigma}_t^2$ may diverge to the unconditional variance σ^2 of the process.

Furthermore, the Ljung-Box statistics are important for the standardized residuals. Looking back at formula 5 standardized residuals u_t are proportional to the conditional volatilities σ_t , which should lead to the log returns ϵ_t . The conditional volatilities map the volacluster in the time series. To achieve the best possible model, one does not want to find these volacluster effects in the standardized residuals, but only in the conditional volatilities. The Ljung-Box statistics check this property. In figure ?? Ljung-Box statistics of the \hat{u}_t and the \hat{u}_t^2 are shown.

The plot shows reliable statistics for forecasts from the Google-GARCH(1,1) model. For all lags up to 20, the p-values are above the 5% line and thus hypothesis tests are discarded. However, if for a given lag=k the p-values would fall below the 5% line, then forecasts would only be reliable up to a forecast horizon k.

If one wants to improve the the standardized residuals \hat{u}_t (if autocorrelations exists in the \hat{u}_t), an ARMA part would have to be added to the existing model (see 2.2.3). This can again be optimized with different model orders. If you want to improve the squared standardized residuals \hat{u}_t^2 (if volaclustering exists within the \hat{u}_t^2), then you should modify the GARCH model order.

Now an optimal model has been found and a forecast can be made. Since a GARCH(n,m) process is white noise sequence the expected value $E[\epsilon_{T+h}|\epsilon_T, \dots, \epsilon_1] = \mu = 0$ can be assumed (if the mean value in the fit object was also estimated and is significant, then the expected value is the estimated μ).

Calculating the forecast variance is a recursive process. With increasing model order the calculation becomes more and more difficult. For this work the rather simple calculation for a GARCH(1,1) model is sufficient. One receives:

$$\hat{\sigma}_{T+h}^2 = \omega + (\alpha_1 + \beta_1)^2 \hat{\sigma}_{T+h-1}^2$$

If the parameter restriction from formula 7 is true, the forecast variance converges with the increasing forecast horizon to the unconditional variance of the process.

$$\hat{\sigma}_{T+h}^2 = \frac{\omega}{1 - \alpha_1 - \beta_1} = \sigma^2$$

The 95% forecast interval is calculated as follows:

$$E[\epsilon_{T+h} | \epsilon_T, \dots, \epsilon_1] \pm 1.96 \sqrt{\sigma_{T+h}^2}$$

Figure ?? shows the GARCH(1,1) forecast for 20 days. The blue line represents the expected value of the time series. Reds are the two 95%-interval limits and green shows the actual values of the series.

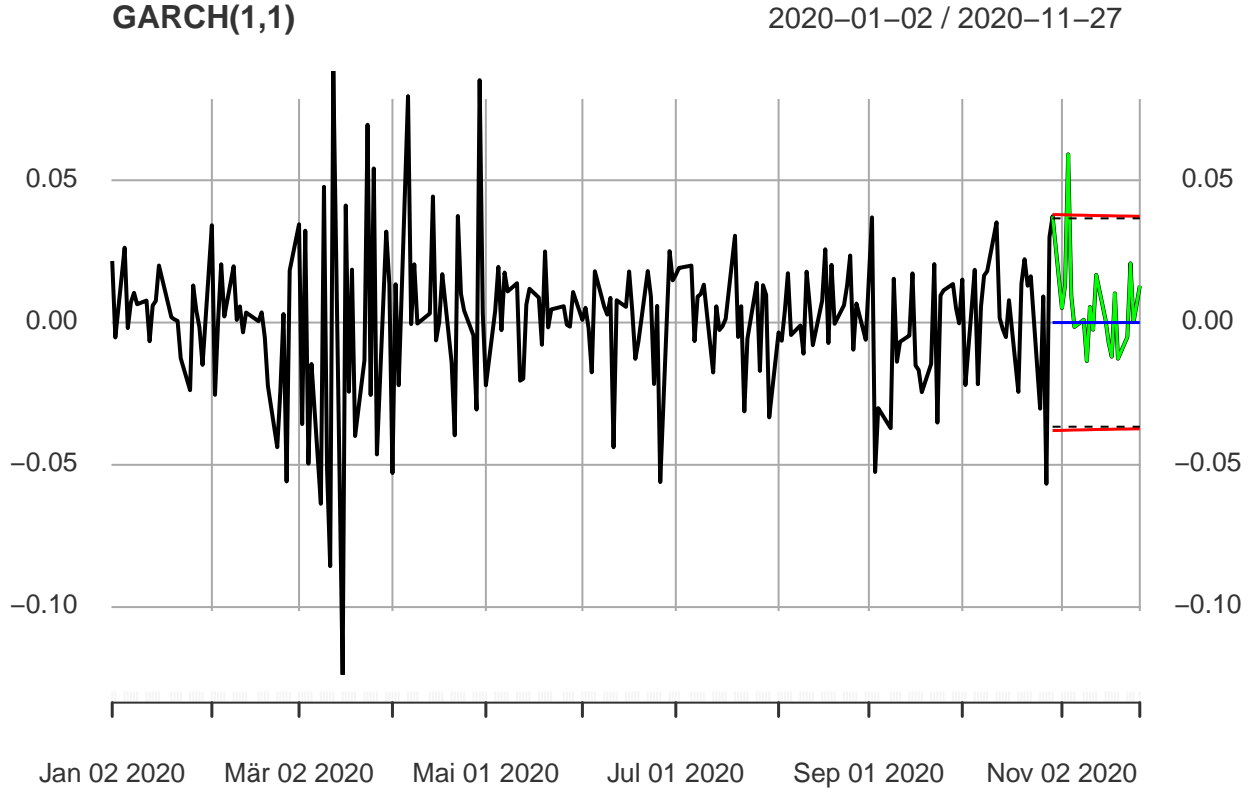


Figure 3: GARCH-Forecast.

You can see how the red lines are approaching the black dotted lines ($0 \pm 1.96 * \sqrt{\sigma^2}$) as the forecast horizon increases. This means that the forecast variance converges to the unconditional variance and thus there is another proof for the satisfaction of the parameter restrictions.

2.2.3. ARMA-GARCH

If you want to adopt a GARCH model and you discover non-vanishing autocorrelations in the standardized residuals \hat{u}_t , the model can be extended with an ARMA part.

$$y_t = \mu + a_1 y_{t-1} + \dots + a_p y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + \dots + b_q \epsilon_{t-q} \quad (8)$$

$$\begin{aligned} \epsilon_t &= \sigma_t u_t \\ \sigma_t^2 &= c\sigma^2 + \sum_{j=1}^n \alpha_j \sigma_{t-j}^2 + \sum_{k=1}^m \beta_k \epsilon_{t-k}^2 \end{aligned} \quad (9)$$

This model contains now 4 different model orders: p, q, n and m . Experience shows that for financial time series, a model order of $n=m=1$ and $p, q \leq 1$ is often sufficient to fit the data to an ARMA-GARCH model.

The equation 8 is called the *mean-equation* and describes the conditional mean of the process, which is used to obtain optimal point forecast. The equation 9 defines the variance of the process and is called *variance-equation*, which is used for the forecast error variance [eco2].

2.3. Moving Average Filters

Moving average filters are basically used to identify trends and smooth out price fluctuations. As a commonly used tool, moving average filters are very simple in their usage, historical data from a timeframe L gets summarized and divided by the length of the filter (L). Depending on the length of the filter, the application to the timeseries the MA gets shifted, longer filters have a higher shift then shorter ones we visualize this behavior in ???. Many different indicators are built upon the Moving Average principle, mostly they are used in combinations of different lengths to create signals. In the following section, we introduce some of the most popular indicators based on the MA principle.

The actual challenge in using Moving average filters, is to figure out which length of the filter brings the most useful information.

2.3.1. Equally-weighted Moving Average or SMA

SMA stands for Simple Moving Average, depending on the length of the filter(L), L observations since the last noted observation will be considered. The observations are getting summarized and divided by the filterlength L . As stated in the name, all past observations are weighted equally. For every timestep, a new observation is considered and the last one eliminated [SMA]. SMA's are very easily customized by changing the length of the filter.

EqMA

$$y_t = \frac{1}{L} \sum_{k=0}^{L-1} x_{t-k} \quad (10)$$

- L = filterlength
- x = original series price e.g.

2.3.1.1 Momentum Momentum is an indicator whether a market is bullish or bearish, it measures the “speed” of the trend direction in the market. for a timespan k the last price p_k , k timesteps ago is subtracted from the last price. This is equivalent to applying and EqMA on the returns of a series.

Momentum

$$y_t = p_t - p_{t-K} \quad (11)$$

- p_t = prices of the series
- K = Lag

2.3.2. Exponentially-weighted Moving Average

Since not all observations are having the same influence on the future value, we can apply a weight to past observations. One method will be exponentially weighted Moving average. So we chose an optimal parameter to give past observations weights decreasing by α^k . In comparison to the SMA 2.3.1. [EMA], a EMA from the same length L , reacts faster than to price changes.

A skillful trader chooses an optimal α to increase the performance of the measurement. Weights could also be given individually by adding a weight vector to the filter.

EMA

$$y_t = \frac{1}{\sum_{k=0}^m \alpha^k} \sum_{k=0}^m \alpha^k x_{t-k} \quad (12)$$

- $m = \text{filterlength}$
- $\alpha = \text{Parameter to weigh the observations}$

2.3.3. Moving Average Crossings

Moving average crossings are basically just different MA's with different lengths applied to a time-series. The points the filters then cross, will be used as a trading signal to go long, short or hold. The “death” and “golden” cross are very popular trading patterns [EMA]. If a shorter MA crosses the longer MA from above, its called a “golden cross” it is an indicator that the price will rise in the future and can be used to created the buy signal. In contrast stands the “death cross” vise versa, a shorter MA (popular $L = 50$) crosses a longer MA (popular $L = 200$) from below, signaling that further losses are in store.

Moving Average Crossing

$$y_t = \frac{1}{L_1} \sum_{k=0}^{L_1-1} x_{t-k} - \frac{1}{L_2} \sum_{k=0}^{L_2-1} x_{t-k} \quad (13)$$

- $L_1 = \text{filterlength 1}$
- $L_2 = \text{filterlength 2}$
- $0 < \alpha < 1 = \text{Parameter to weigh the observations}$

An example of MA average crossings with 2 SMAs of different length is visualized in # ???. The prices are in green while the blue line represents an 50 day SMA and the red line a 250 (1 year) SMA. The crossing points of those two SMAs could now be used as trading signals. In this example these crossings wouldn't perform very well, therefore as mentioned earlier, finding the right length of the filter depends on each timeseries, their behavior and the preferences of the trader.

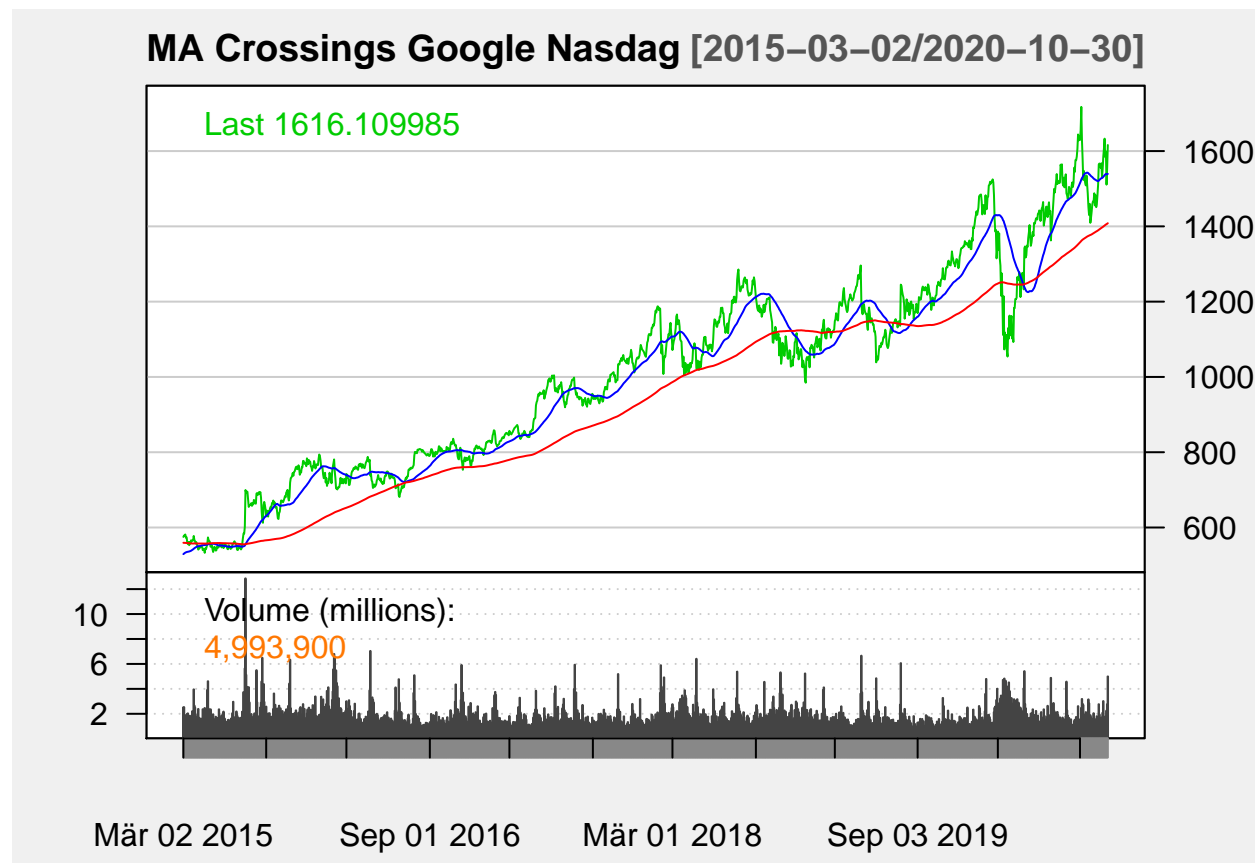


Figure 4: Moving Average Crossing

2.4. Relative Strength Index

The Relative strength index is a tool to measure momentum, the value indicates if positions are overvalued or undervalued. the scale goes from 0 to 100. [RSI].

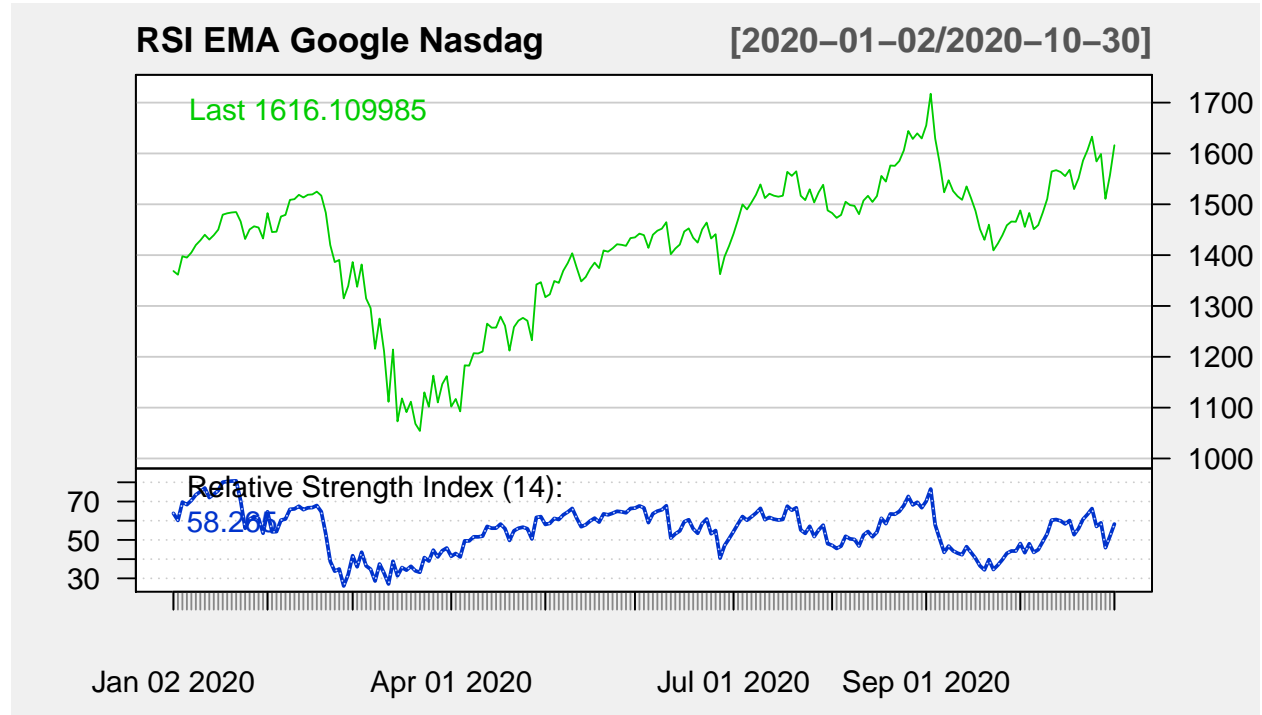
$$U_t = \begin{cases} 1, & \text{if } x_t \geq x_{t-1} \\ 0, & \text{otherwise} \end{cases} \quad D_t = \begin{cases} 1, & \text{if } x_t < x_{t-1} \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

- X_t =Original timeseries

We then apply SMA or EMA of length N to U_t and D_t which converts them in $up_t(N)$ and $down_t(N)$ The RSI is now computed by:

$$RSI_t(N) = 100 \frac{up_t(N)}{up_t(N) + down_t(N)} \quad (15)$$

The original developer J. Welles Wilder Jr. proposed a length of $N = 14$ in its work from 1978[slide_trading_indicators_univa]. Traditional trading signals based on the RSI are the upper 70 or lower 30 limit to buy or sell. Usually when the RSI is going over 70 it suppose that the asset is overvalued and in contrast when its under 30 then its undervalued.



2.5. Moving Average Convergence Divergence

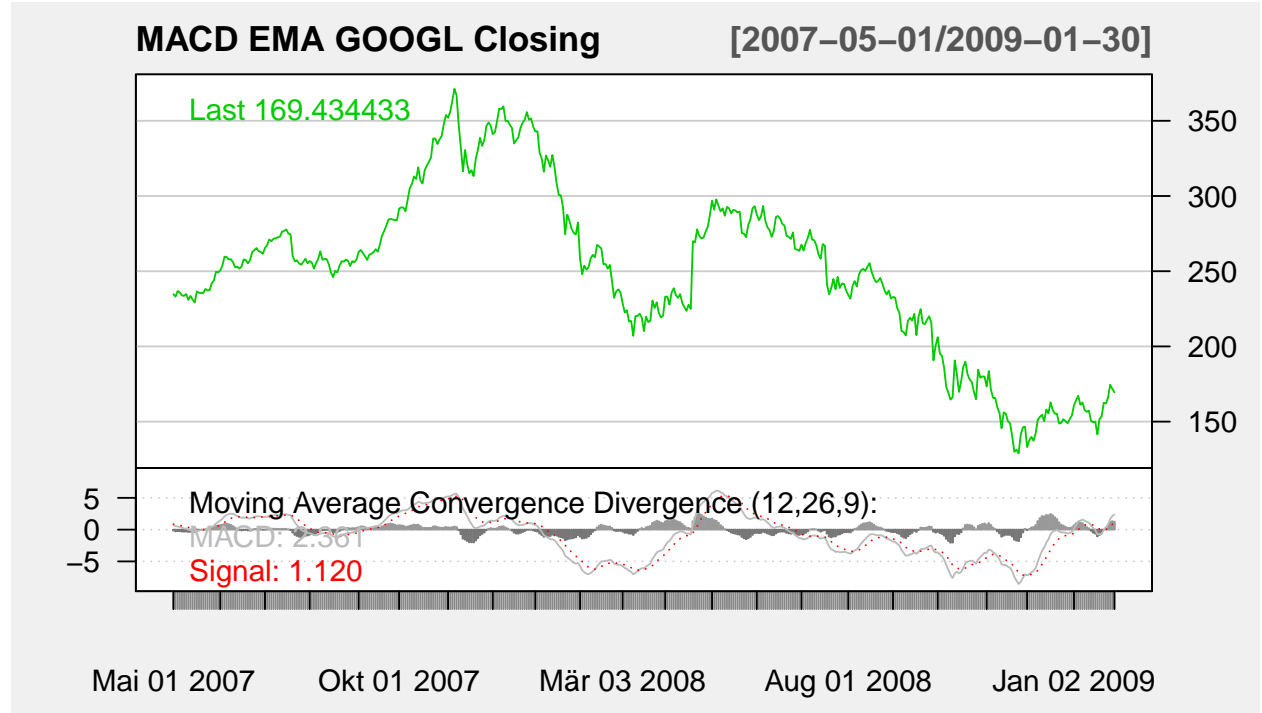
The MACD is also a commonly used filter. The basic principle is to Subtract a longer EMA with length L as in section 2.3.3 from a shorter EMA from length S then smooth the result with another EMA with length R . As a result with can use the crossing of the 2 generated curves for trading. As an alternative we could use SMA's instead of EMA's

$$macd_t = \frac{1}{\sum_{k=0}^{t-1} \alpha^k} \sum_{k=0}^{t-1} \alpha^k x_{t-k} - \frac{1}{\sum_{k=0}^{t-1} \beta^k} \sum_{k=0}^{t-1} \beta^k x_{t-k} \quad (16)$$

$$MACDsignal_t = \frac{1}{\sum_{k=0}^{t-1} \gamma^k} \sum_{k=0}^{t-1} \gamma^k macd_{t-k} \quad (17)$$

- x_t = prices or log prices
- S = length of the $short_1$ EMA usually 12
- L = length of the $long_1$ EMA usually 26
- R = length of the “double smoothing ema” usually 9
- $\alpha = 1 - \frac{1}{S}$, $\beta = 1 - \frac{1}{L}$, $\gamma = 1 - \frac{1}{R}$

₁ short and long in the meaning $s < l$, not buy sell



2.6. Bollinger bands

Bollinger bands are a analysis tool founded by John Bollinger. It contains a moving average and an upper and lower band . The bands are defined by adding a constant K times a standard deviation σ_t to the *Moving Average* for the upper , and subtracting it for the lower band.

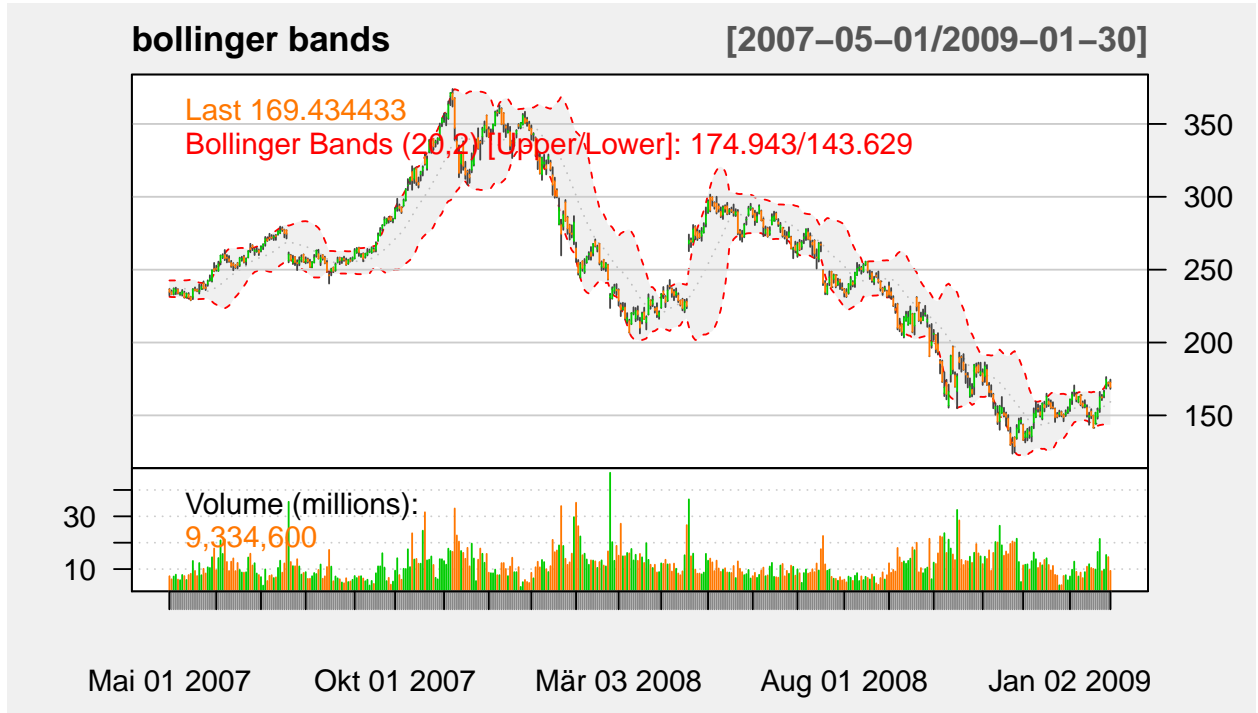
$$U_t = MA_t + K\sigma, L_t = MA_t - K\sigma \quad (18)$$

the variance from bollingers theory is calculated by:

$$\sigma_t^2 = \frac{1}{N} \sum_{k=0}^{N-1} (x_{t-k} - MA_t)^2 \quad (19)$$

The calculated σ_t could be problematic because its derived from the original series and increases with the level, its non stationary. Therefore an other method to calculate the standard deviation could be used. As done in section 2.2.2. σ_t could be provided by a GARCH, which would handle the increasing volatility.

- N = usually the filterlength and the length considered for σ are the same
- K = Constant usually equals 2
- σ_p = standard deviation of the series
- U_t = upper band
- L_t = lower band



2.7. Sharpe Ratio

Sharpe ratio is a very powerful and widely used ratio to measure performance. It describes return per risk.

$$\text{SharpeRatio} = \frac{R_p - R_f}{\sigma} \quad (20)$$

- R_p = Return of Portfolio
- R_f = Risk free Rate, mostly treasury bonds
- σ_p = standard deviation of portfolios excess return (risk)

2.8. Carry

Carry trades are trading strategies where usually money is borrowed at a lower interest rate, than the investment is giving in return. the risk of this strategy is based in the currency risk.