Unser Titel

- und Untertitel -

Project Report

Group: Nikos Bosse and Felix Süttmann

Georg-August-University Göttingen
Statistisches Praktikum

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Faculty of Business and Economics
Georg-August-University Göttingen
https://www.uni-goettingen.de/

Title:
Project Title

Theme:
Scientific Theme

Project Period:
Summer Semester 2018

Project Group:
6

Participant(s):
Nikos Bosse
Felix Süttmann

Supervisor(s):
Paul Wiemann
Peter Pütz
Stanislaus Stadlmann

Copies: 1

Page Numbers: 26

Date of Completion:
September 1, 2019

Abstract:

XXXXXX.

Contents

1. Introduction

1.1 Our Goal / Scope of this report

General Idea:
1. Predict Stocks using • Time Series • Machine Learning and Sentiments from Text 2. Use Predictions to Implement Trading Strategies 3. Evaluate and Compare to Other Trading Strategies

2. The Data

2.1 Stock Selection

The stock data comprises 10 selected companies from the NASDAQ stock index. The stocks were determined as those are the stocks we have Ravenpack data and analyst reports about.

2.2 Ravenpack Sentiment Data

2.3 Analyst Reports

2.4 Stock Data

2.4.1 Overview over the Stock Data

The stock data were downloaded from Yahoo Financial Data Base. Table 2.1 provides a small overview over the raw data. Figure 2.1 shows the Closing Prices of the selected assets. The closing prices have been provided adjusted for dividends by Yahoo. Time series analysis is easiest with data that are at least weakly stationary. Weak stationarity implies that the mean of the time series is constant over time and that the covariance between two observations y_t and y_{t+h} depends only on h, not on t (see [Shumway and Stoffer 2011]).
From figure 2.1 it can be clearly seen that most of the stocks exhibit a strong trend. Also the variance of most stocks increases steadily with time over the observed period. This increase in variance is illustrated in figure 2.2 where the standard deviation of the time series is shown. The data are therefore clearly not stationary. We can also formally test for stationarity using the augmented Dickey-Fuller test [REFERENCE]. This test, applied to all time series in no case is able to reject the null hypothesis of a unit root (implying non-stationarity) at any reasonable confidence level.

Stock Data

Date	Open	High	Low	Close	Adj Close	Volume	ticker
2012-01-03	83.76	84.44	83.36	83.49	68.41	3380100	MMM
2012-01-04	83.13	84.26	83.11	84.18	68.98	3007400	MMM
2012-01-05	83.53	83.87	82.70	83.80	68.67	3116400	MMM
...	...	...	...	...	...	...	...
2015-11-27	29.11	29.21	29.03	29.19	25.97	34469600	GE
2015-11-30	29.16	29.28	28.79	28.79	25.61	82905200	GE
2015-12-01	28.84	29.09	28.72	29.01	25.80	56414600	GE
...	...	...	...	...	...	...	...
2017-12-27	108.42	108.55	107.46	107.64	105.31	5624000	DIS
2017-12-28	108.00	108.05	107.06	107.77	105.43	3477700	DIS
2017-12-29	108.05	108.34	107.51	107.51	105.18	4538400	DIS

Table 2.1

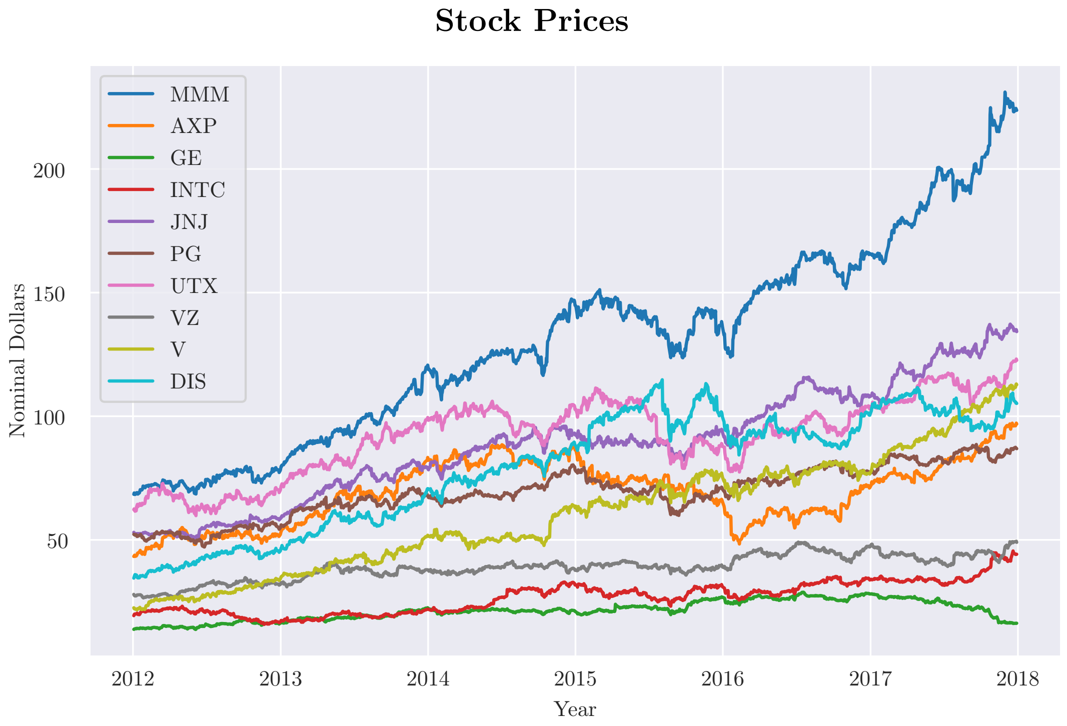

Figure 2.1: Time series of the adjusted closing prices of all 10 stocks looked at in this paper

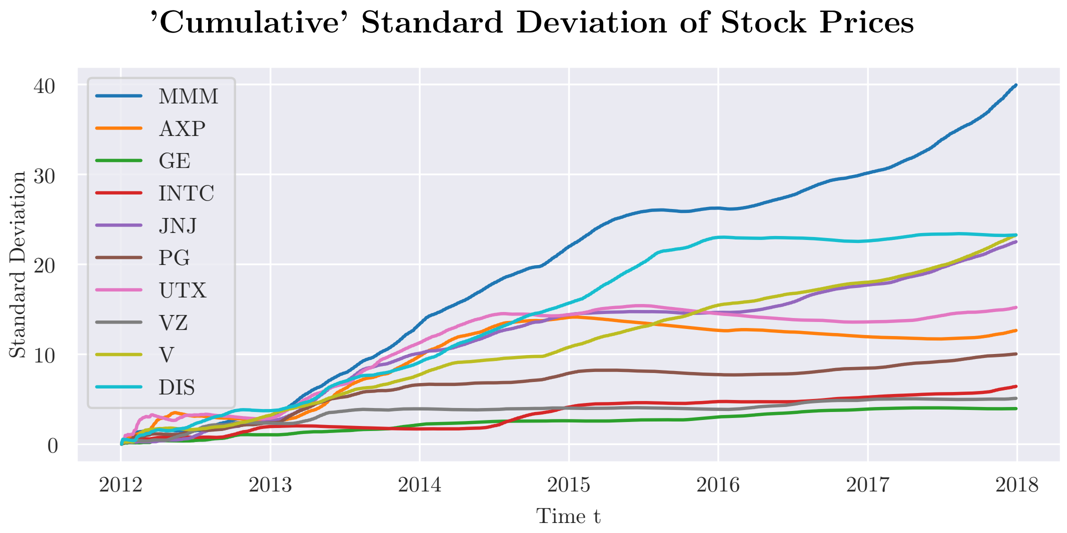

Figure 2.2: Standard deviation for the time series of stock prices. The value of the graph at point t is calculated as the standard deviation of all recorded values of the respective stocks up to that point t.

2.4.2 Data Transformation

In order to obtain weakly stationary time series the data needs to be transformed. There are different ways to proceed that are often equivalent or very similar to each other. Economists usually work with either returns or log-returns, albeit the nomenclature may be a bit confusing. (Daily) returns can be calculated as

$$r_t^{(1)} = \frac{r_t}{r_{t-1}} \qquad \text{or as} \qquad r_t^{(2)} = \frac{r_t - r_{t-1}}{r_{t-1}} = r_t^{(1)} - 1$$

Usually $r_t^{(2)}$ is used and is called either returns or log-returns, even though no logging takes place. For increased conceptual clarity, $r_t^{(1)}$ will be called returns and $log(r_t^{(1)})$ will be called log-returns. $r_t^{(2)}$ will not be explicitly used. Log-returns are computationally convenient and numerically stable. For very small values they are also very close to the returns $r_t^{(2)}$ often used in economic literature as $\log(r_t^{(1)}) \approx r_t^{(1)} - 1 = r_t^{(2)}$ for values of $r_t^{(1)}$ close to 1. Using returns or log-returns instead of stock prices can make the time series stationary. Figure **??** illustrates that the trends in the time series have vanished after looking at log-returns. The data visually now looks like white noise.

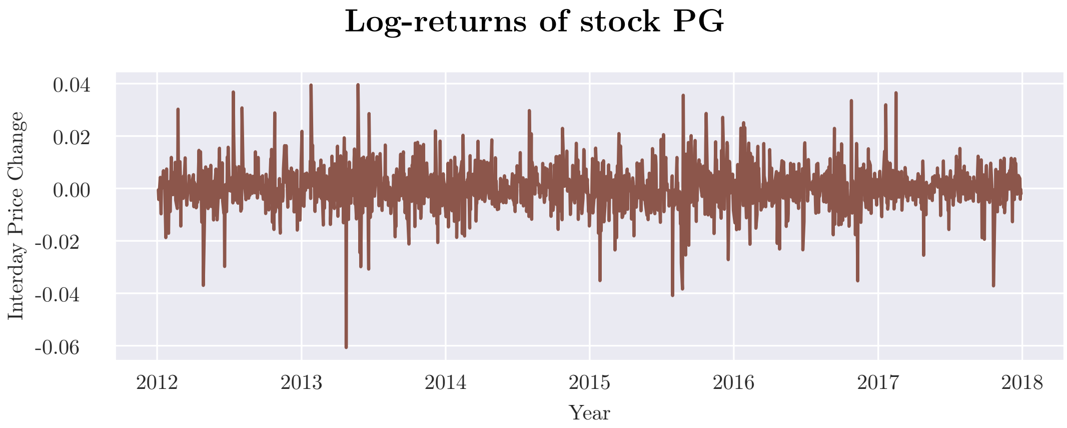

Figure 2.3: Log-returns (or equivalently, first differences of the log of adjusted closing prices) of the PG. Visually the data looks similar to white noise. The entire data can be seen in the appendix in figure A.2

That this is indeed a suitable transformation can be more intuitively understood by looking at a different route of transforming the data: taking the log of the stock prices and then using the first difference of the logged time series. Logging the time series transforms an exponential trend in the time series into a linear one and also serves to stabilize the variance. However, the trend does not vanish and after the log-transformation, the log value of a stock at time t is still mostly determined by its log-value at time t-1. To remove this effect, the time series needs to be differenced. To put this into a clearer perspective we can look at the autocorrelation and partial autocorrelation function of the series.
The autocorrelation at lag j is the correlation between an observation at time t with the observation at t-j. As the series is assumed to be stationary, the autocorrelation function (ACF) does not depend on t, but only on the number of periods that lie between one observation y_t and another y_{t+h}

$$\text{ACF}(h) = corr(y_t, y_{t+h}) \tag{2.1}$$

Partial autocorrelation between an observation y_t and another observation y_{t+1} is the correlation between y_t and y_{t+h} that is not already explained by a linear dependence on the observations in between y_t and y_{t+h}. Formally this can be defined as

$$\text{PACF}(h) = corr(y_t - \hat{y}_t, y_{t+h} - \hat{y}_{t+h}) \tag{2.2}$$

where $\hat{y}_{t+h} = \beta_1 y_{t+h-1} + \beta_2 y_{t+h-2} + ... + \beta_{h-1} y_{t+1}$

and $\hat{y}_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + ... + \beta_{h-1} y_{t+h-1}$

are the linear combinations $\{y_{t+1}, ..., y_{t+h-1}\}$ that minimize the mean squared error of a regression of y_{t+h}, and y_t respectively, on $\{y_{t+1}, ..., y_{t+h-1}\}$. Both $y_t - \hat{y}_t$ and $y_{t+h} - \hat{y}_{t+h}$ are uncorrelated with $\{y_{t+1}, ..., y_{t+h-1}\}$. For now, however, it suffices to know that an ACF which is very slowly decaying to zero is an indicator that differencing may be appropriate (see Shumway and Stoffer 2011, p. 145) to make the series stationary. A large partial autocorrelation at lag 1, as shown in figure A.1 also supports the conjecture that the dependence of the current on the previous value can be eliminated through differencing. After differencing we arrive again at the log-returns as $\log r_t^{(1)} = \log \frac{y_t}{y_{t-1}} = \log y_t - \log y_{t-1}$. We can see now that the the partial autocorrelation at lag 1 has vanished after differecing and that the autocorrelation has also dropped to insignificance as is illustrated in figure 2.5. We can also see that we have not induced any negative autocorrelation. The data therefore is not overdifferenced. The means of our time series is very close to zero (as shown in table 2.2). Overall this pattern strongly suggests the time series are stationary now. However we can also perform a formal test whether our data is stationary or not. The augmented Dickey-Fuller test (ADF). P-values of the ADF for all log-returns are smaller than 10^{-12} even after correcting for multiple testing so we can safely assume the time series are stationary.

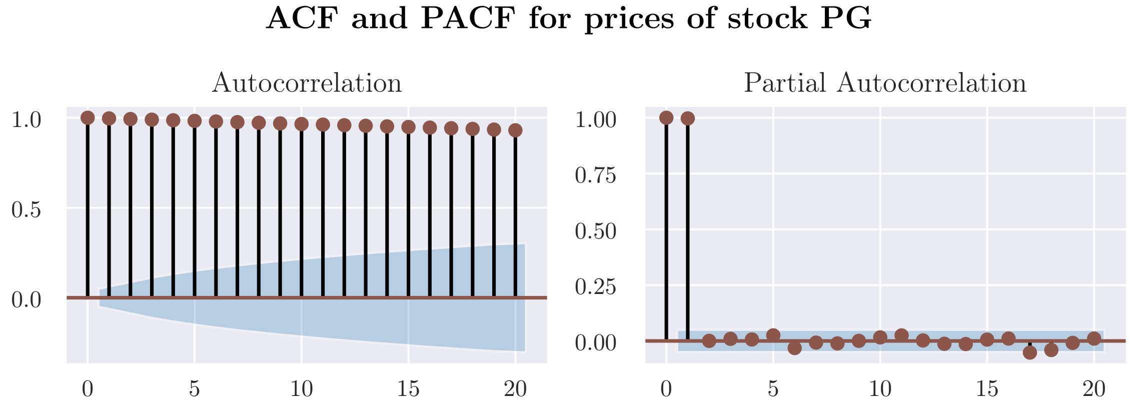

Figure 2.4: Autocorrelation function (ACF) and partial autocorrelation function (PACF) for log-returns of PG. (For convenience, only one stock is shown. ACF and PACF for other stocks can be seen in the appendix in figure A.1)

Means of the log-returns for all Stocks

	MMM	AXP	GE	INTC	JNJ	PG	UTX	VZ	V	DIS
mean	0.000786	0.000534	0.000110	0.000548	0.000616	0.000337	0.000450	0.000372	0.001068	0.000739

Table 2.2

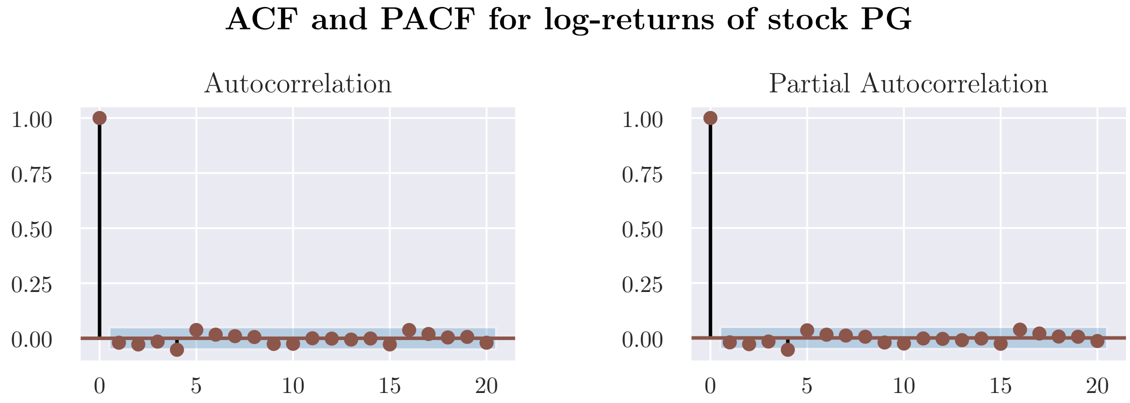

Figure 2.5: Autocorrelation and partial autocorrelation for the log-returns of stock PG (For convenience, only one stock is shown. ACF and PACF for other stocks can be seen in the appendix in figure A.3).

The data is, however, not quite normally distributed. While the mean of the log returns is close to zero, we can see by looking at figure 2.6 that the distribution has fat tails: extrem events appear more often than would be expected if the data was normally distributed.

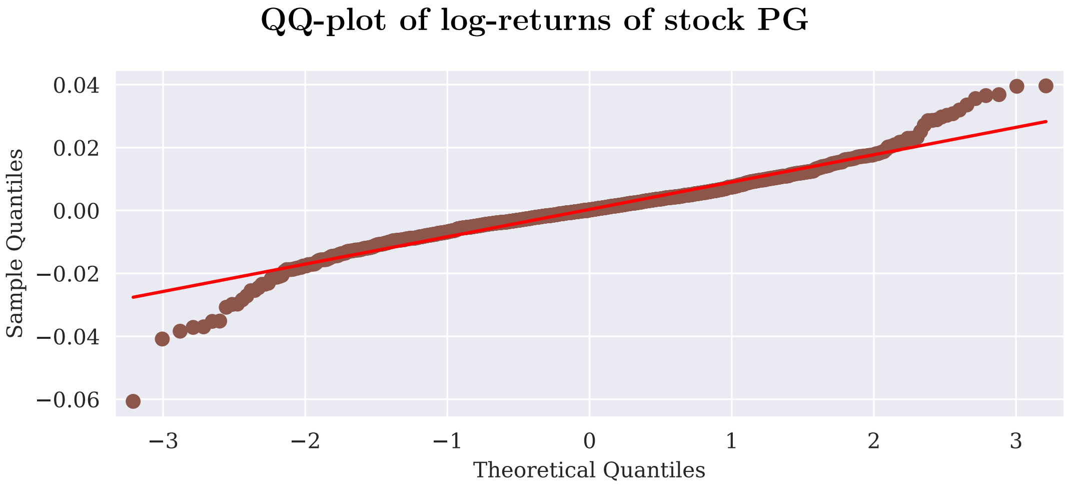

Figure 2.6: QQ-Plot for log-returns of stock PG. QQ-plots for the other stocks can be seen in the Appendix in figure A.4.

The data are also not homoscedastic. While stationarity implies that the conditional variance is constant over time, the variance of the time series fluctuates conditional on past observations [Beschreibung nachgucken, QUELLE!]. This conditional heteroskedasticity is quite common in financial data. (SOURCE). The pattern can be observed in figure 2.7. Figure 2.8 shows the ACF and PACF of the squared residuals of some selected stocks. The patterns indicate that at least some of the volatility can be modeled using time series approaches.

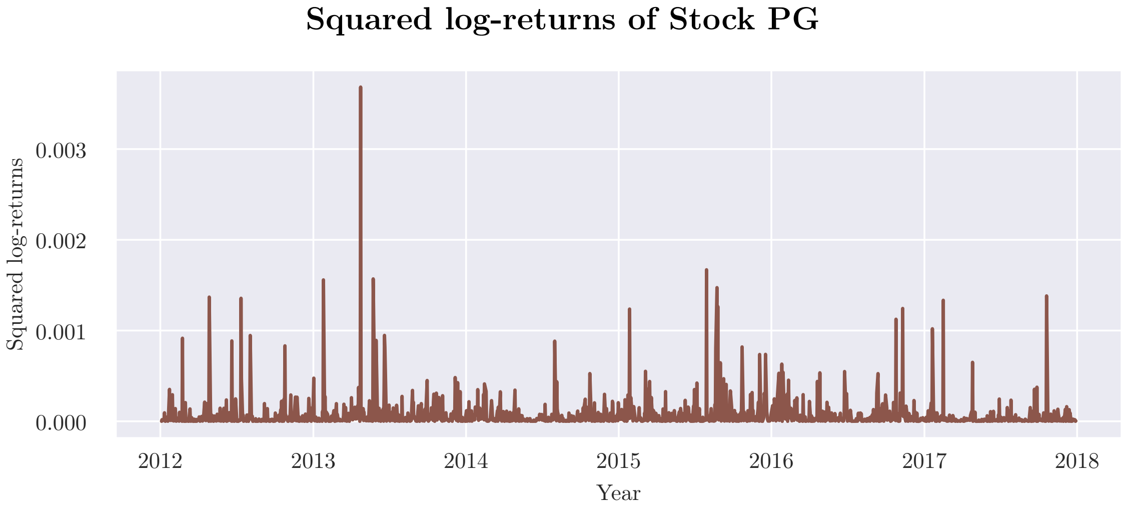

Figure 2.7: Plot of squared log returns of stock PG. This serves as an approximation of the variance of the log returns, as $Var(x) = E[(x - E(x))^2$ and the mean of the log returns is close to zero. The pattern looks similar for all stocks, therefore only one is shown.

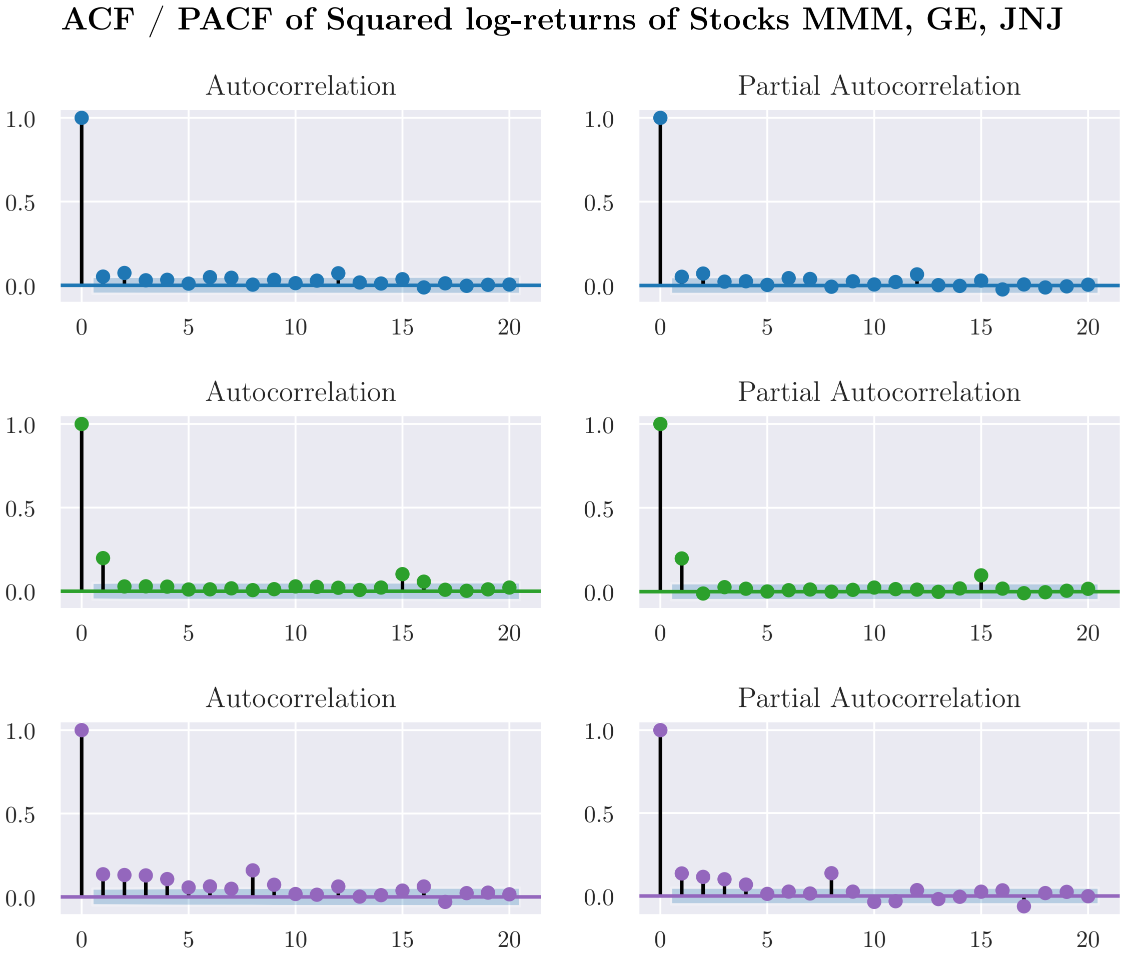

Figure 2.8: ACF and PACF of squared residuals of stocks MMM, GE and JNJ. We see that some of the squared log-returns exhibit indeed autocorrelation, while others do less so. Strong autocorrelation implies that there is information about the future in the time series that can be modeled.

3. Predicting Stocks

3.1 Predictions Using Machine Learning

3.2 Predictions Using Time Series

3.2.1 Idea/Process and Evaluation

irgendwas in der Richtung: wir benutzen Time Series Modelle, machen Predictions und gucken uns am Ende dann den Mean Squared Error an. Sinnvollerweise immer die ersten 10 Perioden verwerfen, um den MSE vergleichbar zu machen zwischen allen Gruppen, auch denen, bei denen die ersten paar Perioden nicht definiert sind. In-Sample vs. Out of Sample Prediction?

3.2.2 Theoretical Overview

Time series predictions in this paper will be made using Random Walks, autoregressive (AR), moving average (MA) and generalized autoregressive conditional heteroscedasticity (GARCH) models. The following will first give a short theoretical overview. Then different models will be applied to a training data set of two chosen stocks. Then predictions will be made for the other stocks using the above mentioned techniques.

Random Walks
Random walks serve as the baseline against which every prediction can be compared. Assuming a random walk as the underlying process implies that we know nothing about the future and can do no better than assuming tomorrow's stock price will on average be the same as today. Formally, a random walk follows

$$y_t = y_{t-1} + w_t \tag{3.1}$$

where y_t is the value of the time series at time t and w_t is a random realisation of a stationary white noise process with mean 0 and variance σ^2. We can expand equation 3.1 by allowing for a constant trend, a drift. A random walk with drift can be represented as

$$y_t = \delta + y_{t-1} + w_t \tag{3.2}$$

where δ is a drift parameter. Predictions for period t + 1 are therefore exactly the value at time t. As we have already eliminated the trend by transforming the data to log-returns we will not use the drift representation here. If we did our analysis on the original stock values then a drift would be appropriate. Note that the random walk (with or without drift) is not a stationary process.

Autoregressive Models
An autoregressive process of order p (AR(p)) implies that the current value of a time series can be described as a combination of the previous p values plus a random shock.

As those previous values intern depend on previous values, the current value is indirectly influenced by its entire past. Formally, an AR(p) process follows

$$y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + ... + \psi_p y_{t-p} + w_t \tag{3.3}$$

where y_t is stationary, $\psi_1, ..., \psi_p$ are constants and w_t is white noise. The mean of y_t is assumed to be zero. If the mean is μ instead of zero, equation 3.4 can be rewritten as

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + ... + \phi_p(y_{t-p} - \mu) + w_t \tag{3.4}$$

This can also be expressed as

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + w_t \tag{3.5}$$

with $\alpha = \mu(1 - \phi_1 - ... - \phi_p)$.

Moving Average Models

A moving average process of order q implies that the current value of a time series consists of the average of the previous q observations plus a random shock. As the mean of the time series μ is constant this average can also be simply expressed as an average of the past random shocks $\{w_{t-1}, ...w_{t-q}\}$. In constrast to the AR(p) process, the shocks affect the future directly (and not only indirectly through past values) and only affect the next q values. Formally, the MA(q) process can be expressed as

$$y_t = \mu + w_t + \theta w_{t-1} + ... + \theta w_{t-q} \tag{3.6}$$

where w_t represents white noise and $\theta_1, ..., \theta_q$ are parameters and q is the number of lags in the moving average.

Autoregressive Moving Average Models

Autoregressive Moving Average Models of order p and q (ARMA(p,q)) form a combination of the above described AR(p) and MA(q) models. Formally, an ARMA(p,q) process follows

$$y_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} \tag{3.7}$$

if the mean of y_t is μ, then the above results in

$$y_t = \alpha + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q} \tag{3.8}$$

with $\alpha = \mu(1 - \phi_1 - ... - \phi_p)$.

GARCH Models

The GARCH(p,q) model is specified as follows:

$$r_t = \sigma_t \epsilon_t \tag{3.9}$$

where ϵ_t is Gaussian white noise with ϵ_t $\mathcal{N}(0, 1)$ and

$$\sigma_t^2 = \alpha_0 + \underbrace{\alpha_1 r_{t-1}^2 + ... + \alpha_p r_{t-p}^2}_{\text{autoregressive part}} + \underbrace{\beta_1 \sigma_{t-1}^2 + ... + \beta_q \sigma_{t-q}^2}_{\text{moving average part}} \tag{3.10}$$

In equation 3.9 the returns r_t are modelled as white noise with mean zero and variance variance σ_t. When compared to a white Gaussian noise with constant variance this can produce a leptokurtic (fat-tailed) distribution similar to what we observed in the QQ-Plots in figure XXXXX. Equation 3.9 is called the mean model of the GARCH(p,q) process. This mean model can also be altered as needed. The GARCH model can then be specified in the following way:

$$r_t = x_t + y_t \tag{3.11}$$

where x_t can be any constant mean, regression or time series process and y_t is a GARCH process that satisfies equations 3.9 and 3.10. In a similar way, the distribution of ??? ϵ_t can be altered. In praxis, researchers often assume a t-distribution instead of a standard normal distribution. ??? is that truly the distribution of epsilon?

3.2.3 Approaching the Training Data

To get a better feeling about our data and to avoid overfitting we try to explore the two stocks INTC and V. We apply different time series models to the entire time series and check their model fit. Figure 3.1 shows the ACF and PACF for the log-returns of INTC and V. From looking at the plots we can presume that for V, AR and MA models of order one or two might be a reasonable try. For INTC it looks like there is very little information included as none of the lower order lags bears any significance.

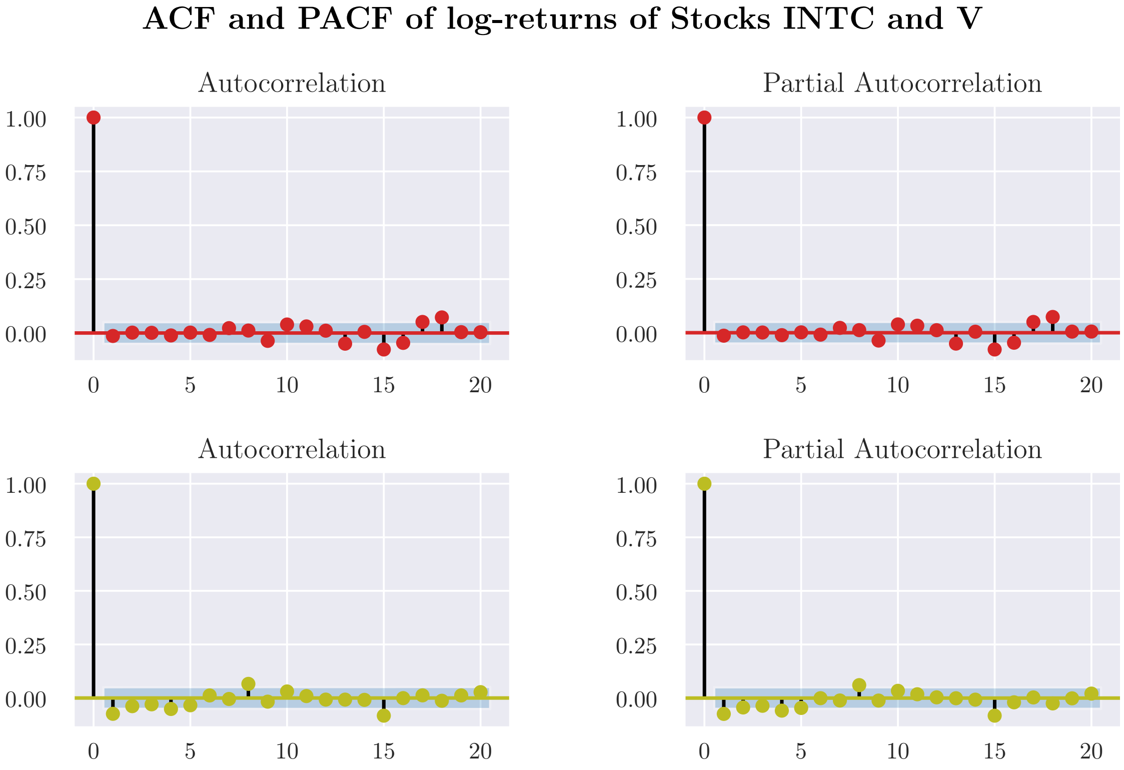

Figure 3.1

AR1 MA1 ARMA(1,1) ARMA(2,1) ARMA(1,2) V -8887.36 -8888.1 -8894.21 -8892.25 In 8701.9 8701.9 -8699.91 -8697.9

Fitting: Our Code fits an RW model, prediicts the next value, compares it to the real value for the MSE and then adds the true value to the time series used for predicting the next value. (we could have done this simpler, by just taking the values of the previous period as prediction for the current one.

Figure: Plot Real vs. Predicted values. (Real values? or FD of log-Values? Table: MSE

3.2.4 Summary of Prediction with Time Series Models

Table: MSE all Figure: All predicted values?

3.3 Hybrid Prediction

The movement of time series data for financial data is influenced bei external effects and information. To leverage this we tried to add information from news sources to our trading strategies. For the hybrid prediction we predict sentiment scores on news sources. The original aim was to use financial news data to predict stock price movement and volatility for trading strategies. To achieve this, large amounts of text data would need to be preprocessed and analyzed regarding their connections to specific stocks, their topic and sentiment. The news data would need to be as precise as possible, because [. . .] mention that an effect on the stocks an only be measured up to 20min after the news appear. Other sources say that. . . . As we were not able to acquire access to a reliable and precise news sources, we tried to implement our approach on the available analyst reports regarding specific stocks. The problem with these reports is, that they are more an indicator of performance over the past month and a prediction about the future performance and don't cover sudden events. The reports also cluster around (meetings?) with long stretches of no or very few reports in between. This makes it unlikely that they are valuable for trading strategies. The goal was to identify the connection of specific articles to listed companies and compute a sentiment score for the article. There are many ways to calculate sentiment scores, like positive and negative, from text data. Many of these require a supervised approach . . . . Language is very context specific (. . .) making it unpractical to use other, labelled training data sets, than financial news data(. . .). With the use of intra day trading and news data one could have also used the movement or volatility of the period close after the news release to get a rough estimate of the impact certain news have. Such an approach was chosen by Robertson et al. (2007). Our data is only inter day and does not allow for a classification in that way.
A common approach for unsupervised
Analyst report data beschreiben...
To get reliable sentiment scores text data has to be preprocessed. The preprocessing was done using `R` (R Core Team, 2017). At first words where converted to lowercase and tokenized using the R package *tidytext* (Silge and Robinson, 2016). Next all the stop words where removed using the stop word library from the *tidytext* package, as well as a custom set. In the next step all links to websites, hyper-references, numbers and words with numbers are removed as well. The last step is lemmatizing the words using the *textstem* package (Rinker, 2018). Lemmatizing words means reducing them to their inflectional forms. Commonly stemming is also applied, because words sometimes have derivationally related forms. This was not done to have more flexibility for the later applied text analysis. Additionally we could have also used the term frequency–inverse document frequency (tf-idf) matrix (ZITIEREN) for further reductions in the number of words. The issue here would have been that highly informative words for the stock sentiment could have been removed.

3.3.1 ARMAX Predictions

ARMAX works like this: XXXXXX

Predictions Confidence Intervals
Predictors can be - Using weather forecasts –> ARMAX - Number of Tweets? - Sentiments from Machine Learning Algorithm - Predictions made by the Algorithm

3.3.2 Weighted Average of Predictions

a) of different time series models b) of time series and ML models

4. Trading Strategies

4.1 Trading Strategies - Introduction and Theory

4.1.1 Idea and Process

We compare different Trading Strategies. With the Goal of Predicting Stocks this is interesting in and of itself. For the purpose of the project it also serves as a baseline to compare our Trading Strategy against.

4.1.2 Mean Reversion Models and Momentum Trading - Theoretical Background

A vast body of scientific literature has tried to develop models that allow to understand and explain movements in stock markets. An even vaster community of traders has tried to implement these theories to do actual forecasting. While many of the theories are much more complex, two basic ideas can be summarized as "Momentum Based Trading" and "Mean Reversion Based Trading". The former theory hypothesizes that stocks that do well now will likely continue to do so in the future while the latter states that especially good or past performance is an exception and that stocks will eventually return to their average performance. A third strategy, Pairs Trading, will be detailed later.
In 1985 [Bondt and Thaler] were one of the early scholars to analyze mean reversion behaviour when they examined the hypothesis that markets tend to overreact. They looked at monthly returns of assets listed on the New York Stock Exchange in between 1926 and 1982 and constructed portfolios of winners and losers that were updated every three years. Winners and losers were those stocks that had performed the best / worst over the previous years. Their idea was that "if stock prices systematically overshoot, then their reversal should be predictable from past return data alone, with no use of any accounting data such as earnings. [...] Extreme movements in stock prices will be followed by subsequent price movements in the opposite direction." Indeed they observed that the portfolio of losers outperformed the market significantly. While they focus on a very long time horizon of three years, Jagadeesh(1991) suggests that this behaviour could also be observed in the shorter term: "These papers show that contrarian strategies that select stocks based on their returns in the previous week or month generate significant abnormal returns." While mean reversion behaviour has sometimes been observed in practice, it is not trivial to derive it from economic theory. Many economic theories like the famous model from Fama and French describe the return of the stocks of a company as the result of market properties like the baseline market return rate and inherent properties of that company, like for example its book-to-market-ratio [Fama and French 1993]. If such a relationship

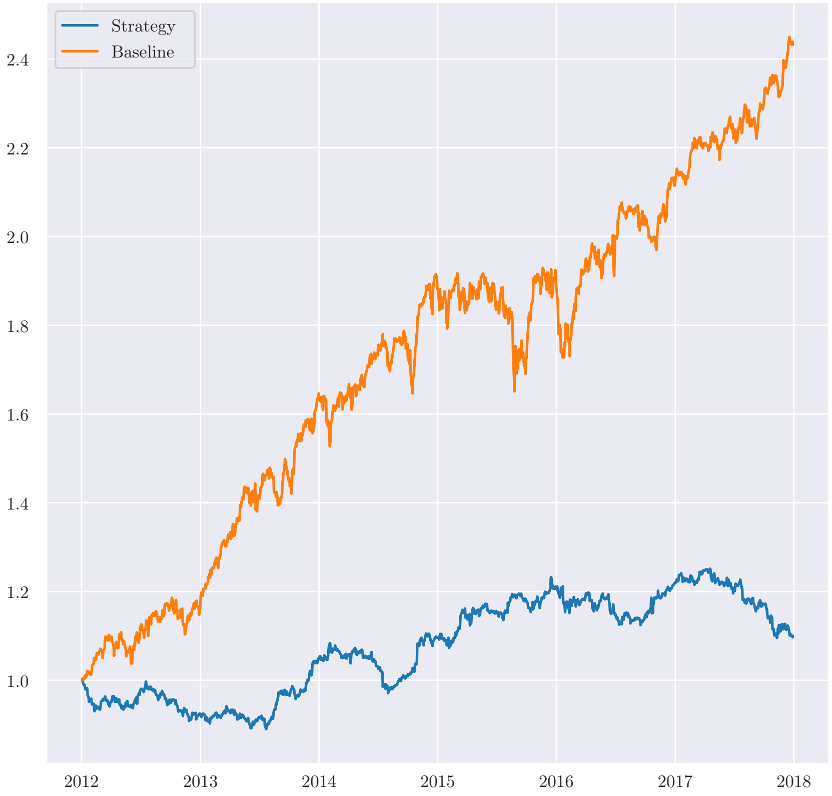

Figure 4.1: Caption

exists than this in turn implies that the actual observed stock movements should be random fluctuations around some much more slowly changing true return rate. However, it is unclear whether this mean reversion should happen in the form of a pendulum that swings forth and back or more in the form of coin tosses reverting to their equilibrium by flooding past observations with new random ones. Something about multiplicative connections?

On the opposite side of the spectrum of trading strategies lies the idea of momentum. [Japateesh and CX] in 1993 were one of the first to describe "that strategies which buy stocks that have performed well in the past and sell stocks that have performed poorly in the past generate significant positive returns over 3- to 12-month holding periods." [Japateesh and CX] looked at portfolios and saw mean reversion over 12 to 48 months. Thaler and Bondt observe mean reversion over the time frame of 36 months. Many scholares agree that mean reversion and momentum behaviour are not necessarily contradictions, but that the time frame determines in which way a stock will behave [Balvers and Wu].

PAIRS TRADING

4.2 Trading Strategies - Implementation

4.2.1 Mean Reversion Portfolio

While their analysis is focused on monthly returns over a much longer time frame, the basic idea that was replicated in our trading strategy was the same. While they found excessive returns of that strategy from 1932 to 1977 our implementation was unforturnately much less successful.

Implementation algorithm: rank portfolio everyday along their returns from the last day.

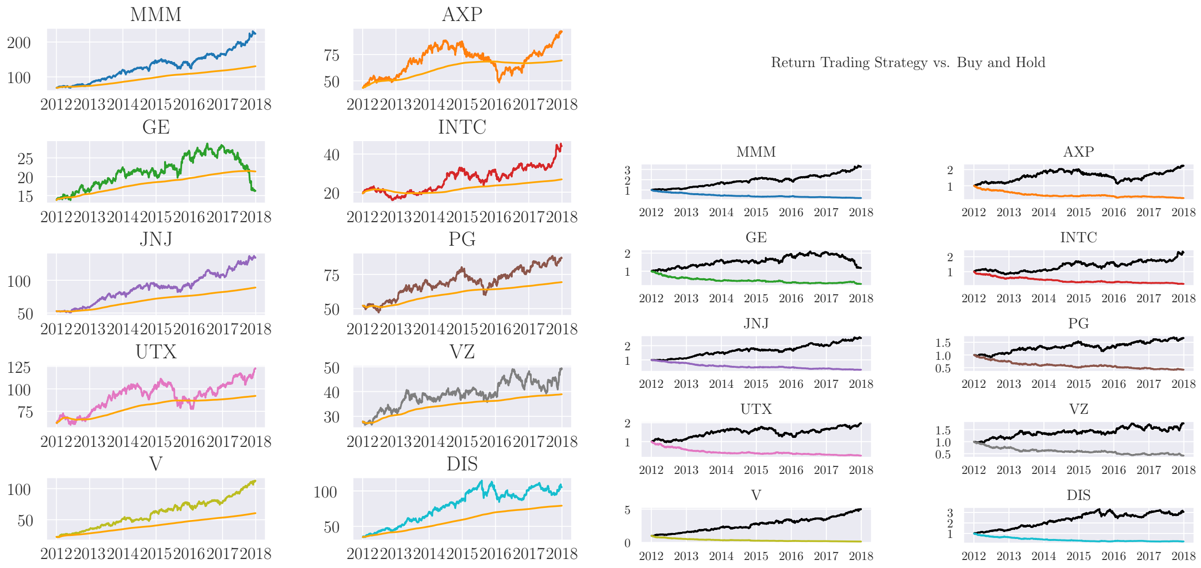

Figure 4.2: Caption

Then buy the two worst performers and sell the two best performers. Interestingly, reversing that does not produce better results. The time frame is obviously not right.

4.2.2 Mean Reversion - single stocks

Cumulative Mean - Stock prices

Idea: Plot: Cumulative Mean vs. Actual Time Series –> We see that the cumulative mean does not capture the time series well. Trend is always behind the current development, since we have a trend

Cumulative Mean - Returns

Comparison of 90d and 30d moving averages

Idea: If 30d averages is above 90d average, then sell. And vice versa

Mean Reversion Portfolio

Idea: Look at entire portfolio.

4.2.3 Momentum Based Trading

4.2.4 Pairs Trading

4.3 Trading Strategies Based on Our Predictions

4.4 Hybrid Trading Strategies

Take Mean of different predictions?

5. Conclusion

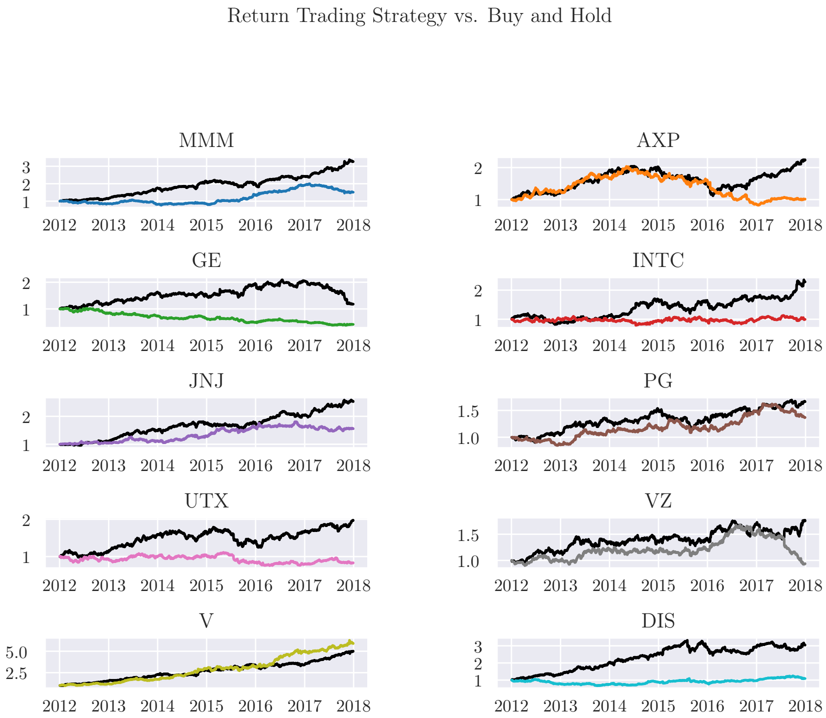

Figure 4.3: Mean Reversion based on past returns for single stocks

Bibliography

Calum S Robertson, Shlomo Geva, and Rodney C Wolff. News aware volatility forecasting: Is the content of news important? In *Proceedings of the sixth Australasian conference on Data mining and analytics-Volume 70*, pages 161–170. Australian Computer Society, Inc., 2007.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

Julia Silge and David Robinson. tidytext: Text mining and analysis using tidy data principles in r. *JOSS*, 1(3), 2016. doi: 10.21105/joss.00037. URL `http://dx.doi.org/10.21105/joss.00037`.

Tyler W. Rinker. *textstem: Tools for stemming and lemmatizing text*. Buffalo, New York, 2018. URL `http://github.com/trinker/textstem`. version 0.1.4.

A. Appendix A name

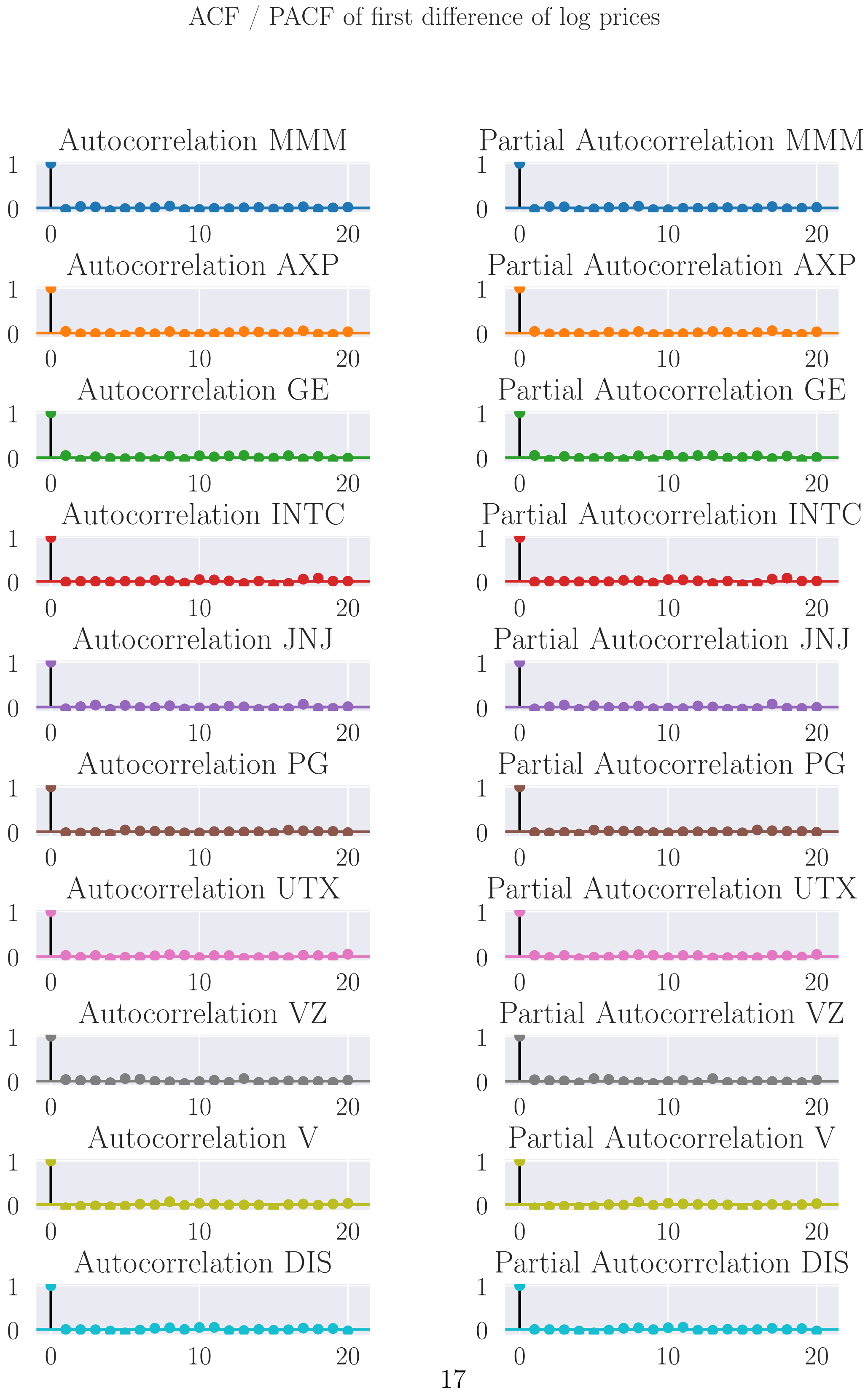

Figure A.3: Autocorrelation and partial autocorrelation for the first difference of log adjusted closing prices for all stocks

Results for an AR(1) process fit to the log-returns of V

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(1, 0)	**Log Likelihood**	4446.683
Method:	css-mle	**S.D. of innovations**	0.013
Date:	Fri, 30 Aug 2019	**AIC**	-8887.366
Time:	14:47:54	**BIC**	-8871.409
Sample:	0	**HQIC**	-8881.423

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0011	0.000	3.503	0.000	0.000	0.002
ar.L1.log_returns	-0.0736	0.026	-2.868	0.004	-0.124	-0.023

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	-13.5846	+0.0000j	13.5846	0.5000

Table A.1

Results for an MA(1) process fit to the log-returns of V

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(0, 1)	**Log Likelihood**	4447.077
Method:	css-mle	**S.D. of innovations**	0.013
Date:	Fri, 30 Aug 2019	**AIC**	-8888.155
Time:	14:47:57	**BIC**	-8872.197
Sample:	0	**HQIC**	-8882.212

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0011	0.000	3.551	0.000	0.000	0.002
ma.L1.log_returns	-0.0809	0.027	-3.002	0.003	-0.134	-0.028

	Real	**Imaginary**	**Modulus**	**Frequency**
MA.1	12.3538	+0.0000j	12.3538	0.0000

Table A.2

Results for an ARMA(1,1) process fit to the log-returns of V

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(1, 1)	**Log Likelihood**	4451.108
Method:	css-mle	**S.D. of innovations**	0.013
Date:	Fri, 30 Aug 2019	**AIC**	-8894.215
Time:	14:48:01	**BIC**	-8872.938
Sample:	0	**HQIC**	-8886.291

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0011	0.000	4.194	0.000	0.001	0.002
ar.L1.log_returns	0.6156	0.116	5.307	0.000	0.388	0.843
ma.L1.log_returns	-0.6997	0.105	-6.681	0.000	-0.905	-0.494

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	1.6243	+0.0000j	1.6243	0.0000
MA.1	1.4293	+0.0000j	1.4293	0.0000

Table A.3

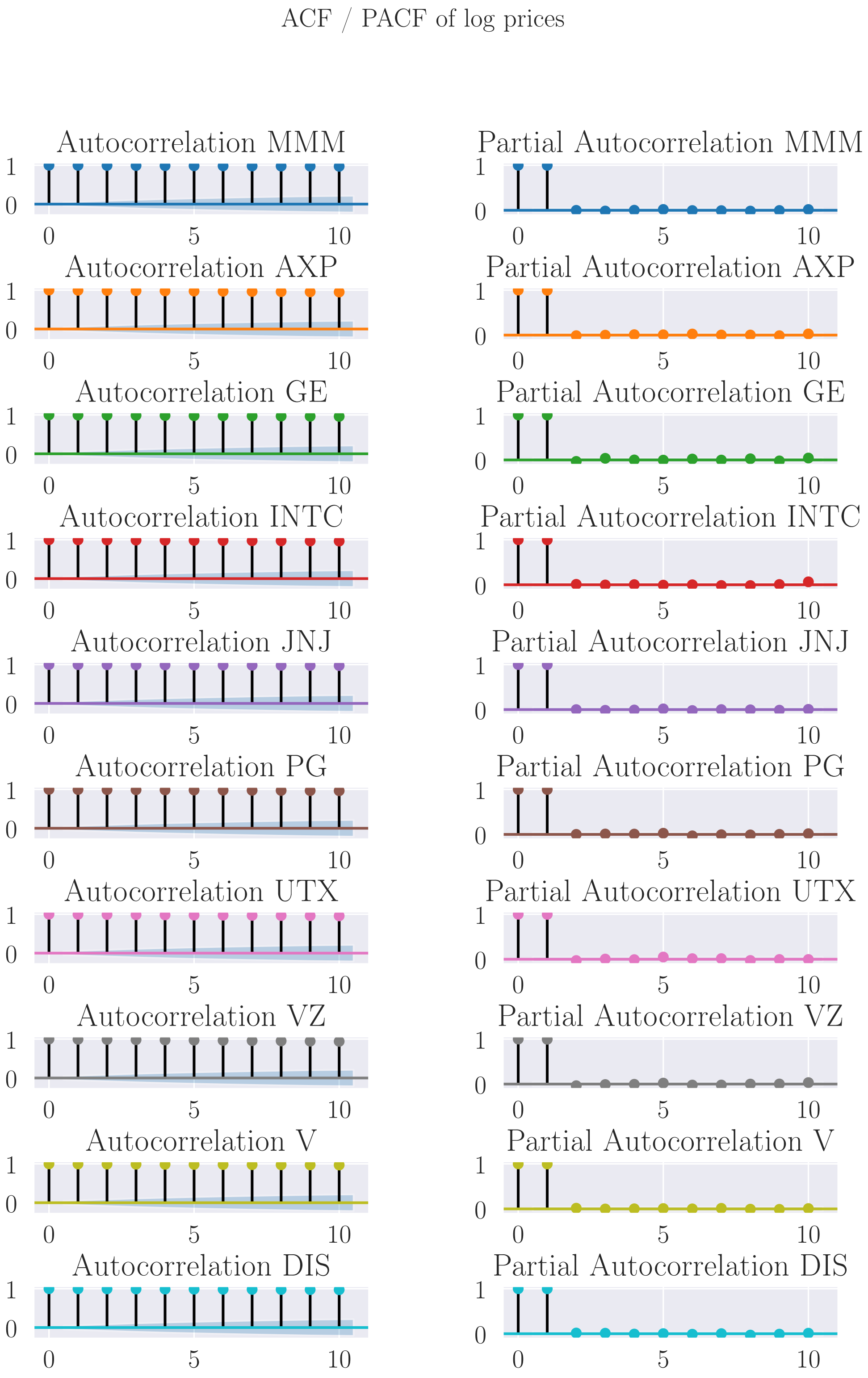

Figure A.1: Autocorrelation and partial autocorrelation for the log of the adjusted closing prices for all stocks

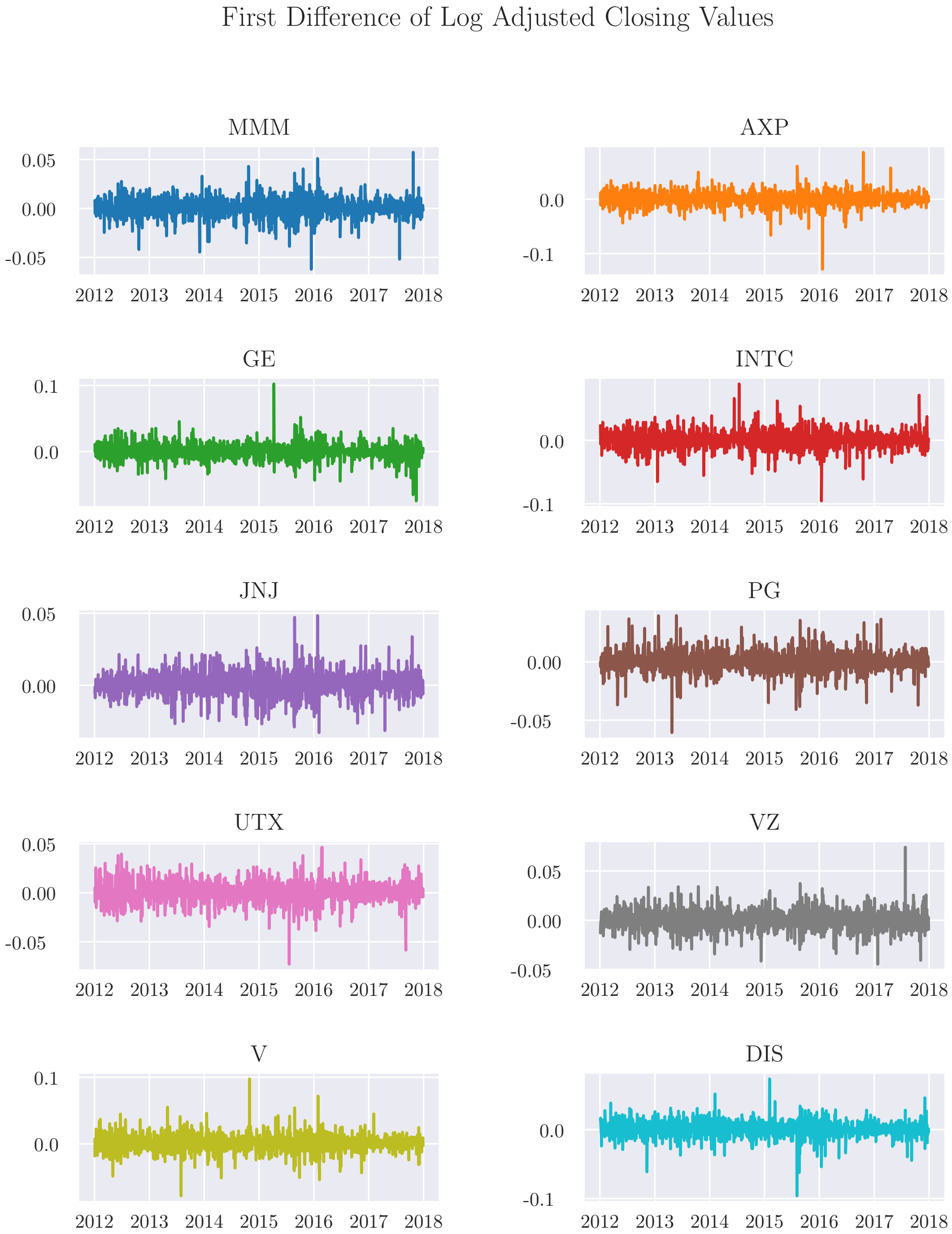

Figure A.2: First difference of log adjusted closing prices

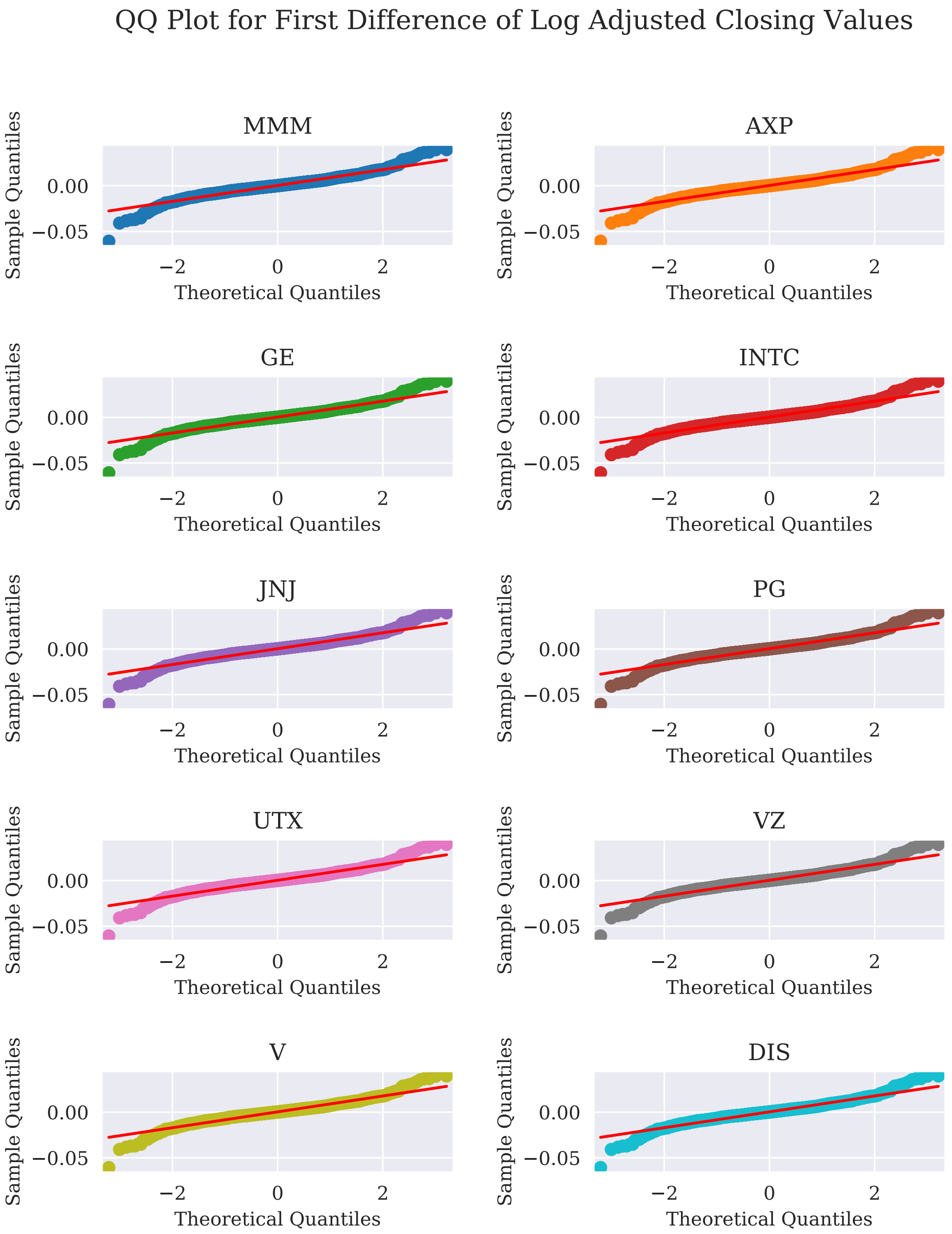

Figure A.4: QQ-Plots

Results for an ARMA(2,1) process fit to the log-returns of V

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(2, 1)	**Log Likelihood**	4451.124
Method:	css-mle	**S.D. of innovations**	0.013
Date:	Fri, 30 Aug 2019	**AIC**	-8892.248
Time:	14:48:02	**BIC**	-8865.652
Sample:	0	**HQIC**	-8882.343

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0011	0.000	4.189	0.000	0.001	0.002
ar.L1.log_returns	0.6033	0.135	4.467	0.000	0.339	0.868
ar.L2.log_returns	-0.0059	0.032	-0.182	0.855	-0.069	0.057
ma.L1.log_returns	-0.6850	0.133	-5.164	0.000	-0.945	-0.425

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	1.6851	+0.0000j	1.6851	0.0000
AR.2	101.4345	+0.0000j	101.4345	0.0000
MA.1	1.4597	+0.0000j	1.4597	0.0000

Table A.4

Results for an AR(1) process fit to the log-returns of INTC

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(1, 0)	**Log Likelihood**	4353.952
Method:	css-mle	**S.D. of innovations**	0.014
Date:	Fri, 30 Aug 2019	**AIC**	-8701.903
Time:	14:47:54	**BIC**	-8685.946
Sample:	0	**HQIC**	-8695.960

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0005	0.000	1.596	0.111	-0.000	0.001
ar.L1.log_returns	-0.0138	0.026	-0.534	0.593	-0.064	0.037

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	-72.7156	+0.0000j	72.7156	0.5000

Table A.5

Results for an MA(1) process fit to the log-returns of INTC

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(0, 1)	**Log Likelihood**	4353.951
Method:	css-mle	**S.D. of innovations**	0.014
Date:	Fri, 30 Aug 2019	**AIC**	-8701.902
Time:	14:47:57	**BIC**	-8685.945
Sample:	0	**HQIC**	-8695.959

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0005	0.000	1.596	0.111	-0.000	0.001
ma.L1.log_returns	-0.0137	0.026	-0.534	0.594	-0.064	0.037

	Real	**Imaginary**	**Modulus**	**Frequency**
MA.1	72.9645	+0.0000j	72.9645	0.0000

Table A.6

Results for an ARMA(1,1) process fit to the log-returns of INTC

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(1, 1)	**Log Likelihood**	4353.953
Method:	css-mle	**S.D. of innovations**	0.014
Date:	Fri, 30 Aug 2019	**AIC**	-8699.906
Time:	14:48:02	**BIC**	-8678.629
Sample:	0	**HQIC**	-8691.982

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0005	0.000	1.594	0.111	-0.000	0.001
ar.L1.log_returns	-0.1300	1.384	-0.094	0.925	-2.842	2.582
ma.L1.log_returns	0.1163	1.387	0.084	0.933	-2.603	2.835

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	-7.6907	+0.0000j	7.6907	0.5000
MA.1	-8.5960	+0.0000j	8.5960	0.5000

Table A.7

Results for an ARMA(2,1) process fit to the log-returns of INTC

Dep. Variable:	log_returns	**No. Observations:**	1509
Model:	ARMA(2, 1)	**Log Likelihood**	4353.952
Method:	css-mle	**S.D. of innovations**	0.014
Date:	Fri, 30 Aug 2019	**AIC**	-8697.904
Time:	14:48:03	**BIC**	-8671.308
Sample:	0	**HQIC**	-8687.999

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
const	0.0005	0.000	1.595	0.111	-0.000	0.001
ar.L1.log_returns	0.3549	nan	nan	nan	nan	nan
ar.L2.log_returns	0.0055	nan	nan	nan	nan	nan
ma.L1.log_returns	-0.3686	nan	nan	nan	nan	nan

	Real	**Imaginary**	**Modulus**	**Frequency**
AR.1	2.7036	+0.0000j	2.7036	0.0000
AR.2	-66.7356	+0.0000j	66.7356	0.5000
MA.1	2.7127	+0.0000j	2.7127	0.0000

Table A.8

Statutory Declaration

We declare that we have authored this thesis independently, that we have not used other than the declared sources / resources, and that we have explicitly marked all material which has been quoted either literally or by content from the used sources.

Georg-August-University Göttingen, September 1, 2019

Nikos Bosse
<nikos.bosse@stud.uni-goettingen.de>

Felix Süttmann
<felix.suettmann@stud.uni-goettingen.de>