W2453 – Angewandte Zeitreihenanalyse und Wirtschaftsprognose

Wintersemester 2019/2020

Projektbericht

Dozent: Prof. Yuanhua Feng

Beginn des Projekts: 10.12.2019

Ende des Projekts: 20.01.2020, 16:00

Aufgabe No.	1	2	3	Total
Punkte				

Gruppennummer: 22

Eingereicht von: Cem Öztürk

Mitglieder der Gruppe: Cem Ötztürk (7000548)

Hazar Erdogan (7012327)

Erika Just (7161492)

Lynn Krell (7152605)

Carla Marie Henke (7164541)

Paderborn, 22.01.2020

Aufgabe 1. Trendschätzung und Anpassung von ARMA-Prozessen

Aufgabe 1a:

Bei der ***nichtparametrischen Glättung*** geht es darum, nichtparametrisch und nicht normalverteilte Daten zu schätzen. Hierbei werden die Datenpunkte an eine Regressionsgerade angepasst. Die durch nicht zufälliges Verhalten entstandenen Daten werden von den zufälligen Schwankungen getrennt, entfernt oder reduziert.

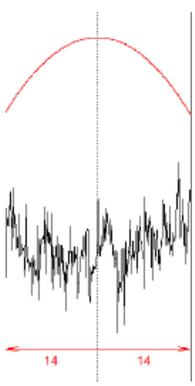
Die ***lokale lineare Schätzung***, auch lokale polynomiale Regression genannt, impliziert, dass man zu jedem der Beobachtungwerte einen Schätzwert durch Minimierung der lokal gewichteten Summe der Residuenquadrate berechnet. Dabei werden die Abweichungen vom Mittelwert, die Residuen, jedes Beobachtungszeitpunktes anhand der nachfolgenden Formel verringert:

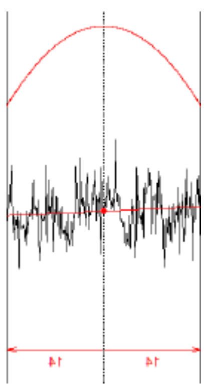
Mit h als Brandbreite, p als Grad des Polynoms und W als Gewichtungsfunktion auf dem Intervall . Sei außerdem v die Ordnung der zu schätzenden Ableitung p-v ungerade.

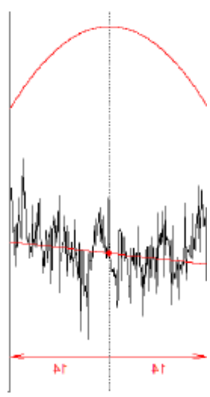
$\hat{m}^{(\nu)}(x) = \nu!\,\hat{b}_{\nu}(x) \rightarrow \hat{m}^{(0)}(x) = \hat{b}_0(x)$ ➔ für den Trend.

Besonders wichtig hierbei ist die gewählte Bandbreite, da diese den meisten Einfluss auf das Ergebnis hat.

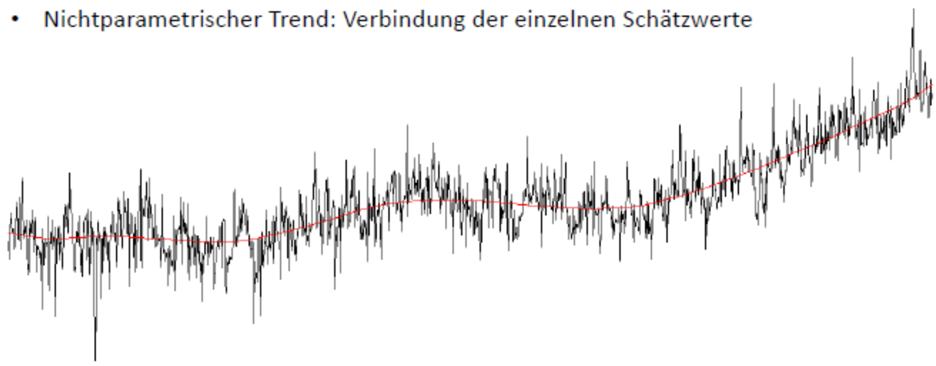
In der ersten Grafik ist ein Ausschnitt einer Zeitreihe mit der gewählten Bandbreite zu sehen. Man betrachtet nur die Beobachtungen im gewählten Intervall (im Beispiel $\pm$ 14). Die vertikal gestrichelte Linie zeigt den zu betrachtenden Zeitpunkt. Durch die lokal polynomiale Regression im betrachteten Intervall berechnet man die rote Gerade. Der Schätzwert liegt im Schnittpunkt der roten und vertikal gestrichelten Geraden (Grafik 2). Nur dieser Schnittpunkt ist relevant, der Rest der roten Gerade ist irrelevant. Dieses Verfahren wird für alle weiteren Zeitpunkte der Zeitreihe wiederholt. Die dritte Grafik zeigt einen weiteren Beobachtungspunkt, zu dem ebenfalls der Schätzwert berechnet wurde. Auch dieses Mal betrachtet man nur die Beobachtungen vom gewählten Beobachtungszeitpunkt ausgehend herum, die innerhalb der gewählten Bandbreite liegen, um den Schätzwert zu berechnen. Durch die Verbindung aller Schätzwerte miteinander, wird der nichtparametrische Trend sichtbar (Grafik 4).







1.. 2.. 3..



4.

Es gibt zwei Modelle, die einen nichtparametrischen Trend bestimmen: das additive und das multiplikative Modell.

? Bei einem ***additiven Komponentenmodell*** werden die Effekte einzelner Faktoren zur Datenmodellierung subtrahiert und addiert.

Dieses Modell mit nichtparametrischem Trend ist definiert mit:

mit als beobachtete Zeitreihe, als nichtparametrischer Trend in , als auf dem Intervall skalierte Zeit, als Saisonkomponente und als trendbereinigter, stationärer Rest mit . Wir vernachlässigen in unserer Betrachtung die Saison .

Daraus folgt, dass die Zeitreihenwerte durch drei Komponenten, die sich in ihrer Wirkung additiv überlagern, beschrieben werden können: die glatte Komponente bzw. Trendkomponente , die Saisonkomponente und die residuale Komponente .

Man entscheidet sich für das additive Modell, wenn die Werte im Allgemeinen über den Zeitverlauf eine geringe Streuung bzw. wenig Schwankungen aufweisen. Hinzuzufügen ist, dass das Modell bei niederer als intervallskalierten Daten keine Anwendung findet, da sich dort ausschließlich die Häufigkeiten jedoch keine Mittelwerte für einzelne Gruppen angeben lassen.

?Das ***multiplikative Komponentenmodell*** mit nichtparametrischer Skalierungsfunktion ist folgendermaßen definiert: .

entspricht der beobachteten Zeitreihe, ist die nichtparametrische Skalierungsfunktion, ist die saisonale Komponente und ist der Rest mit als stationärer Prozess mit . Aus dem multiplikativen Modell folgt, die multiplikative

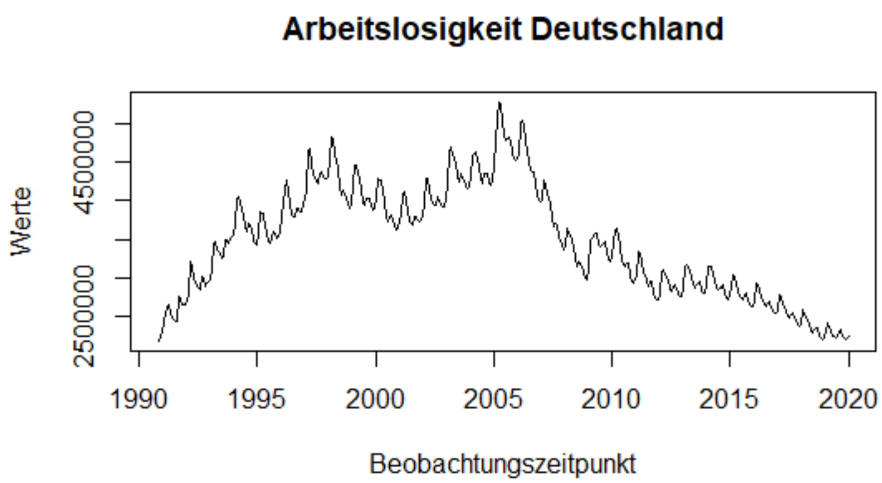
Verknüpfung von glatter Komponente und Saisonkomponente . Definiere . Durch Logarithmieren erhalten wir folglich das additive Modell:

,   (wir vernachlässigen und)

welches bekannt ist und mit dem smoots-Paket geschätzt werden kann.

Der Trend einer nichtstationären Zeitreihe kann durch das smoots-Paket in R geschätzt werden. Die wichtigste Aufgabe des Pakets ist, die optimale Bandbreite für Zeitreihen zu bestimmen, da diese den meisten Einfluss auf das Ergebnis hat. Durch die zentrale Funktion msmooth wird eine lokale polynomische Regression des Trends auf die Basis einer optimalen Bandbreite durchgeführt, die durch einen iterativen Plug-In-Algorithmus erhalten wird. Nach der Bestimmung der optimalen Bandbreite, berechnet das Programm durch die Funktion dsmooth die Ableitung des Trends.

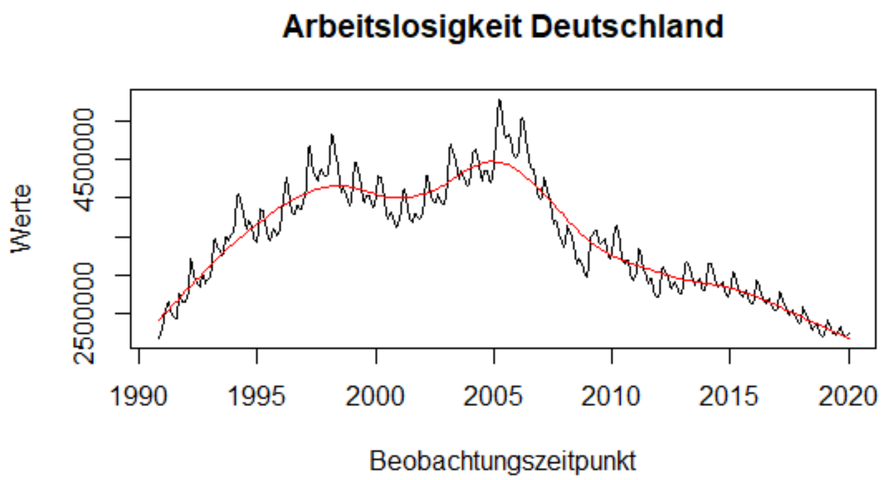
Aufgabe 1b:



Der Plot stellt eine Zeitreihe mit 352 Beobachtungen dar, welche die Arbeitslosigkeit in Deutschland im Zeitraum von September 1990 bis Dezember 2019 zeigt. Bei den Daten handelt es sich um absolute Angaben und die Periodizität ist monatlich. Dabei sind die angegebenen Werte Ursprungswerte, d.h. die Daten sind unbereinigt. Außerdem wurde der Arbeitsmarkt nicht in Segmente unterteilt. Den Datensatz haben wir bei der deutschen Bundesbank gefunden. Allgemein lässt sich sagen, dass die Arbeitslosigkeit in Deutschland starken Schwankungen unterliegt. Eine besonders hohe Arbeitslosigkeit ist im Jahr 2005 erkennbar. Dies ist zurückzuführen auf die Hart IV Reform wodurch ab diesem Zeitpunkt auch erwerbsfähige Sozialhilfeempfänger in der Statistik mitberücksichtigt werden. Seit 2005 sinkt die Arbeitslosigkeit stetig.

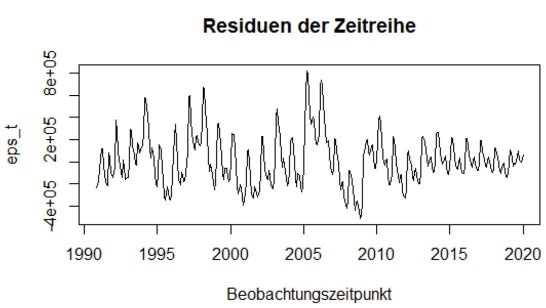
Aufgabe 1c:

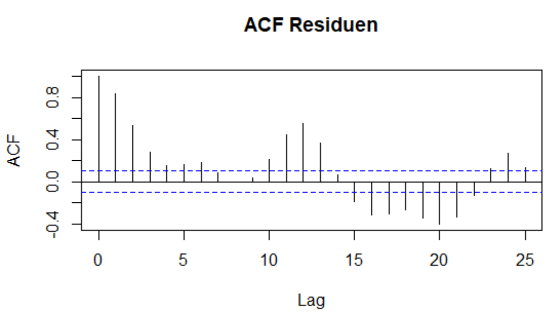
Originale Zeitreihe mit Trend:



Die Zeitreihe wurde in diesem Aufgabenteil durch das multiplikative Modell stationarisiert. Da wesentliche Charakteristika eines additiven Modells periodisch wiederkehrende gleichbleibende Schwankungen der beobachteten Zeitreihenwerte um die glatte Komponente sind, lässt sich das additive Modell im vorliegenden Fall ausschließen. Unsere Zeitreihe ist von großen Schwankungen und somit einer breiten Streuung gekennzeichnet, weshalb wir uns für das multiplikative entschieden haben.

Trendbereinigte Zeitreihe: Die Werte der trendbereinigten und damit stationären Zeitreihe bilden die Residuen aus der Schätzung der Trendparameter.





Aufgabe 1d:

Unter Berücksichtigung des kleinsten BIC-Wertes, 9041.168, wurde ein ARMA(3,3) – Modell ausgewählt.

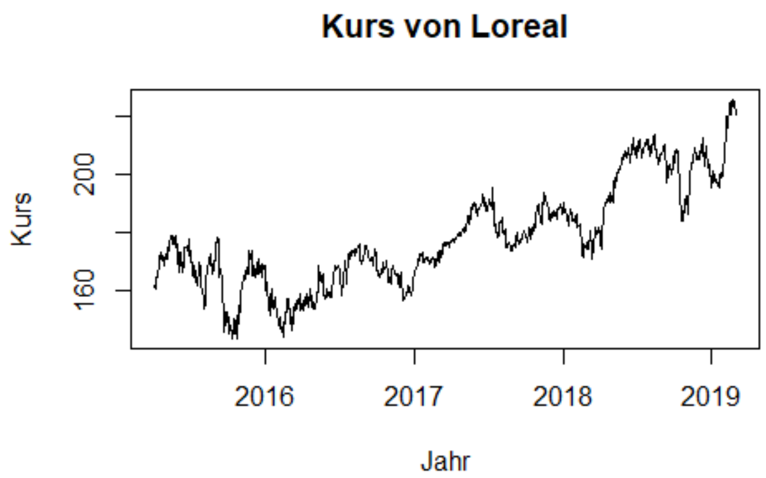
	q = 0	q = 1	q = 2	q = 3
p = 0	9683.980	9356.193	9192.113	9131.853
p = 1	9268.082	9151.883	9146.645	9121.470
p = 2	9153.816	9143.182	9115.346	9088.208
p = 3	9142.382	9148.975	9117.411	9041.168

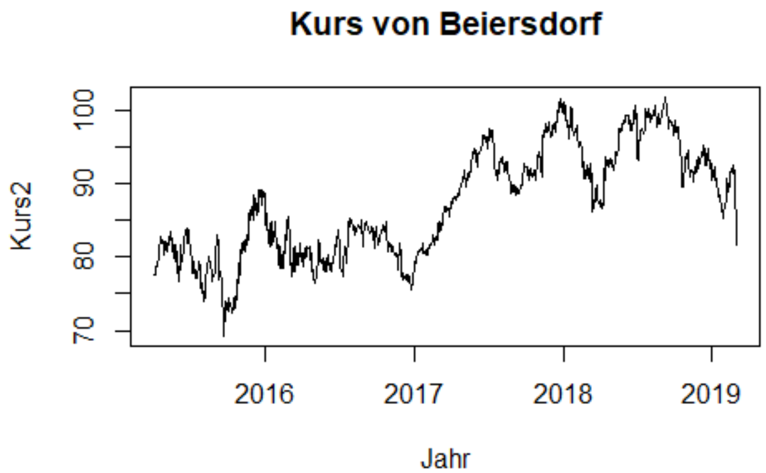
Das geschätzte ARMA(3,3) – Modell:

Yt	ar1	ar2	ar3	ma1	ma2	ma3
	1.7147	-1.6658	0.7202	-0.4208	0.7482	0.2833
s.e.	0.0561	0.0630	0.0519	0.0942	0.0644	0.0937

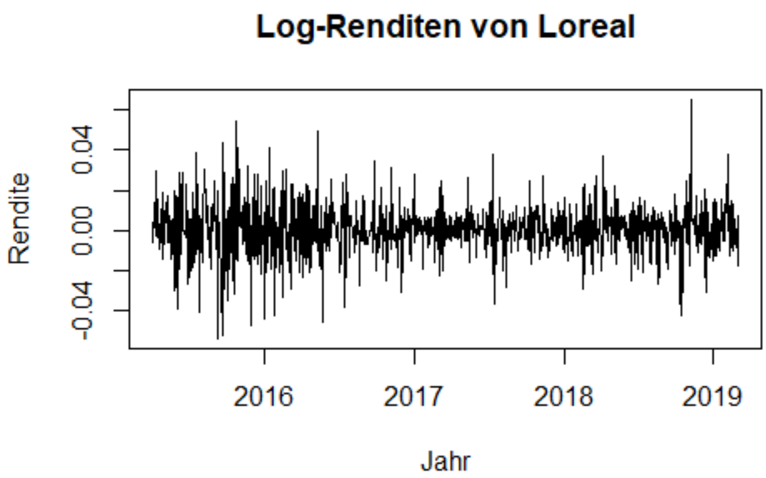
Aufgabe 2. Finanzmarktanalyse

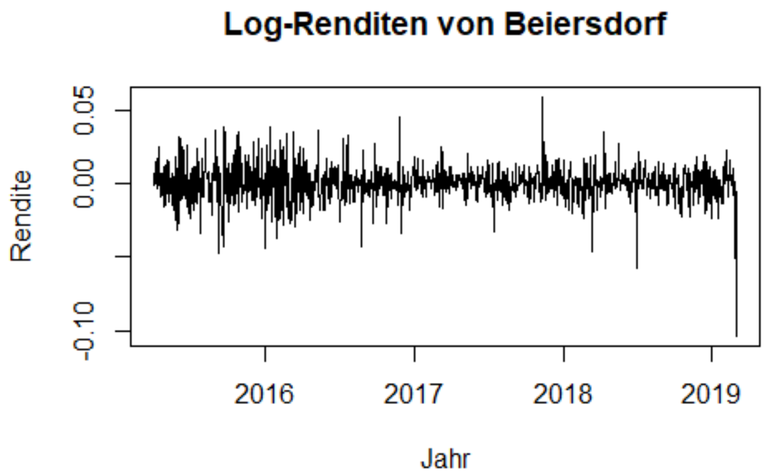
Aufgabe 2a:





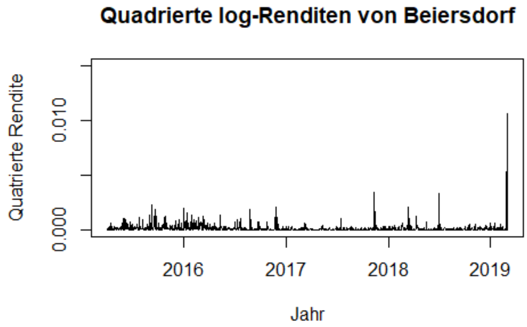
Zu sehen sind die Aktienkurse der Unternehmen Loreal und Beiersdorf im Zeitraum von März 2015 bis Februar 2019. Hierbei handelt es um Finanzzeitreihen und sie erfüllen viele charakteristischen Eigenschaften von Finanzzeitreihen. Die Beobachtungen beider Zeitreihen sind positiv und nicht normalverteilt. Zudem sind die Anzahl sowie die Variation der Beobachtungen sehr groß. In beiden Grafiken sind „Drifts“ erkennbar, langfristig steigende oder abfallende Tendenzen, beispielsweise zu sehen bei dem Kurs von Beiersdorf zwischen 2017 und Mitte 2018. Beide Finanzzeitreihen sind nicht stationär, aber integriert. Es ist klar zu erkennen, dass die Mittelwertstationarität und Varianzstationarität nicht erfüllt sind. Damit Finanzzeitreihen stationär werden, müsste man sie d-mal differenzieren. Genau dann heißt der Prozess integriert der Ordnung d.

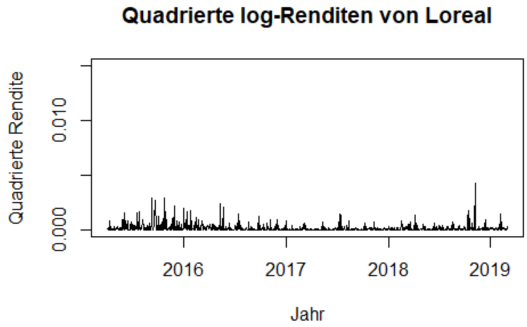


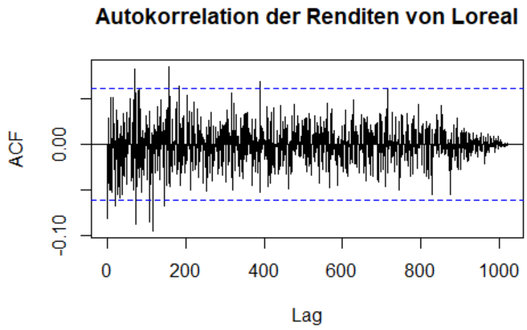


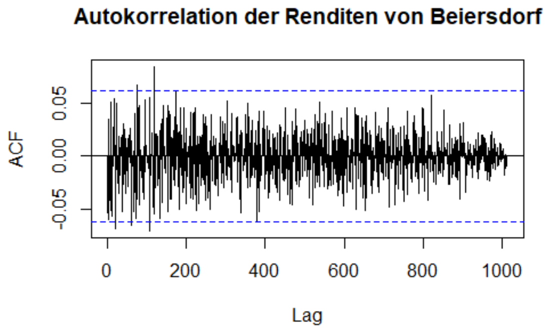
Zu sehen sind die Log-Renditen der Unternehmen Loreal und Beiersdorf. Eine annähernde Stationarität ist bei beiden Renditen erkennbar. Die Mittelewertstationarität ist bei den zwei Unternehmen erfüllt. Jedoch wird die Varianzstationarität sowie die Kovarianzstationarität in beiden Fällen verletzt. Dies ist an der abwechselnd steigenden und abnehmenden Streuung zu erkennen. Da die Kovarianzstationarität die Varianzstationarität impliziert und diese nicht gegeben ist, kann der Prozess nicht kovarianzstationär sein.

<ins>Aufgabe 2b:</ins>

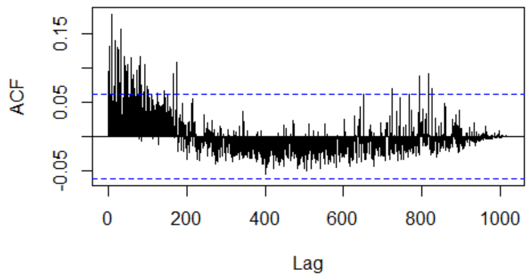




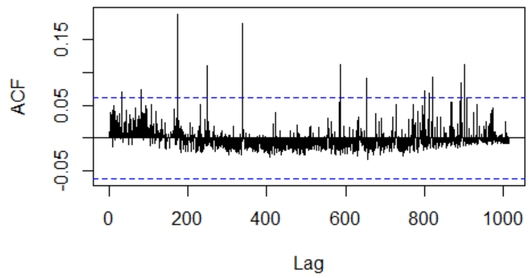




Autokorrelation der quadrierten Rendite von Loreal



Autokorrelation der quadrierten Rendite von Beiersdorf

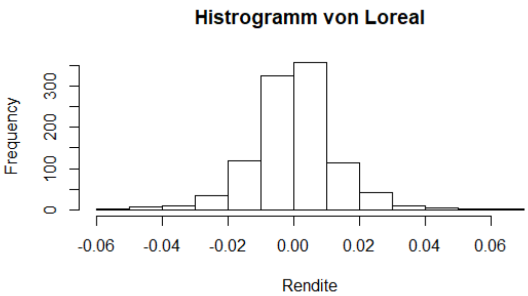


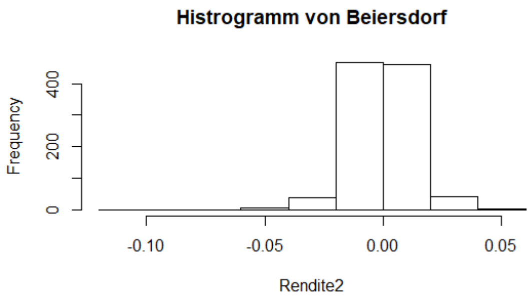
Der Plot der Autokorrelationsfunktion heißt Korrelogramm, d.h. das Korrelogramm ist die graphische Darstellung dieser. Es zeigt die geschätzten ACFs und wird zusammen mit 2 Grenzen $\pm$ 2/ ausgegeben. Diese Grenzen werden auch als Bartlett-Grenzen bezeichnet. Das Korrelogramm enthält wesentliche Informationen über zeitliche Abhängigkeiten, was bedeutet, dass sich Zusammenhänge zwischen den beobachteten Ergebnissen zu verschiedenen Beobachtungszeitpunkten feststellen lassen. Eine signifikante Autokorrelation liegt vor, wenn der Anteil der Werte der ACF außerhalb der Grenzen größer als 5% ist. Die Autokorrelationsfunktion der einfachen Renditen befindet sich überwiegend bei beiden Unternehmen innerhalb der beiden blau gekennzeichneten Grenzen. Lediglich ein paar ACF zwischen Lag 0 und 200 liegen außerhalb der Grenzen. Dies könnte darauf hinweisen, dass die Daten miteinander korreliert sind. Der größte Teil der ACF befindet sich allerdings innerhalb der Grenzen, welches es wahrscheinlicher macht, dass die Beobachtungen nicht signifikant miteinander korreliert sind.

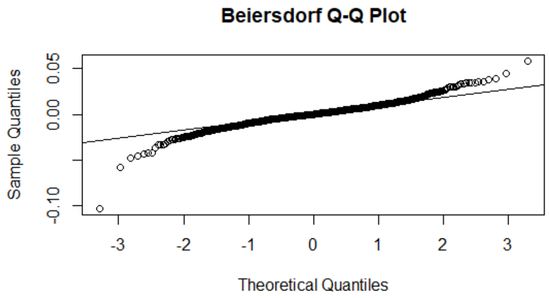
Bei der Autokorrelationsfunktion der quadrierten Renditen ist zu erkennen, dass im Gegensatz zu den einfachen Renditen, die ACF deutlich öfter die Grenzen überschreitet. Bei der Autokorrelationsfunktion der quadrierten Renditen von Loreal liegt der Anteil der Werte außerhalb der Grenzen erkennbar über 5%, was auf signifikante Autokorrelation schließen lässt. Hingegen bei Beiersdorf liegen nur vereinzelte Beobachtungen außerhalb der Grenzen, weshalb hierbei womöglich keine signifikante Korrelation vorliegt.

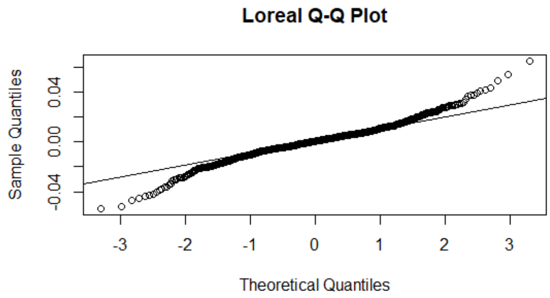
Aufgabe 2c:

	Loreal	Beiersdorf	Normalverteilung
Mittelwert	0.0003072538	0.0004911459	0
Varianz	0.0001607252	0.0001492199	1
Schiefe	-0.01281134	-0.6142513	0
Wölbung	2.548208	6.803424	0







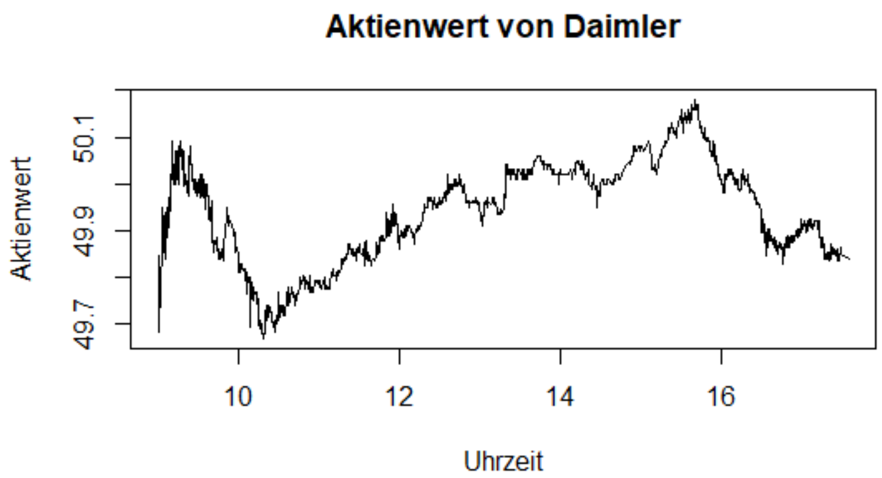


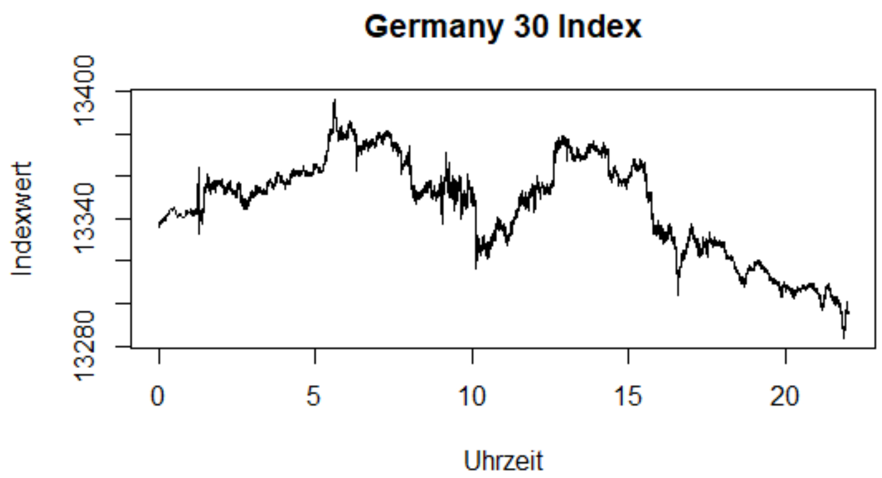
In diesem Aufgabenteil wird für die Renditen der Gesamtreihen die Werte des Mittelwertes, der Varianz, der Schiefe und Wölbung angegeben sowie das Histogramm dieser gezeigt. Außerdem wird numerisch und graphisch geprüft, ob die Renditenreihen normalverteilt sind. Ein übliches Merkmal von Aktienrenditen ist nämlich der Aspekt, dass diese nicht normalverteilt sind.

Der Mittelwert beider Renditen weicht minimal von dem Wert der Normalverteilung ab. Bei der Varianz ergibt sich ein größerer Unterschied im Vergleich zur Normalverteilung. Zudem können die Verteilungen der Renditen im Allgemeinen über ihre Form charakterisieren. Die Charakterisierung erfolgt über den Aspekt der Schiefe und Wölbung. Auf die Schiefe und Wölbung kann graphisch geschlossen werde, aber auch eine numerische Interpretation ist möglich. Für die Verteilungen beider Unternehme ergibt sich eine linksschiefe bzw. rechtssteile Verteilung mit einem Wert der Schiefe von -0.01281134 für Loreal und -0.6142513 für Beiersdorf. Die Werte der Wölbung von 2.548208 bei Loreal und 6.803424 bei Beiersdorf zeigen in beiden Fällen eine spitzere Verteilung als die Normaltverteilung, was man als leptokurtisch bezeichnet. Zudem sind Heavy Tails erkennbar, da die Verteilungen dickere Enden aufweisen und damit langsamer als exponentiell fällt.

Aufgabe 3. GARCH-Modelle

Aufgabe 3a:





In den Plots sind zwei Finanzmarkzeitreihen mit ultrahochfrequenten Daten/Tick-by-Tick Data über den 27.12.2019 zu sehen. Im linken Plot wird die Aktienpreisreihe des Unternehmens Daimler dargestellt, im rechten Plot ist der Germany 30 Index abgebildet.

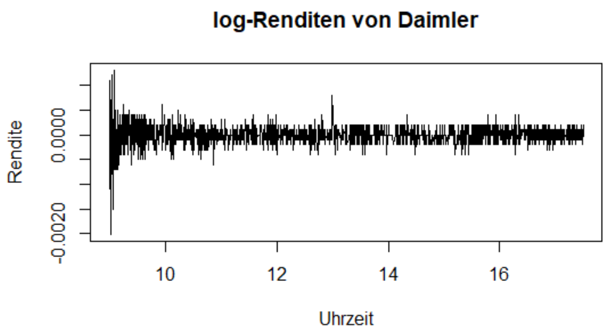
<u>Aufgabe 3b:</u>

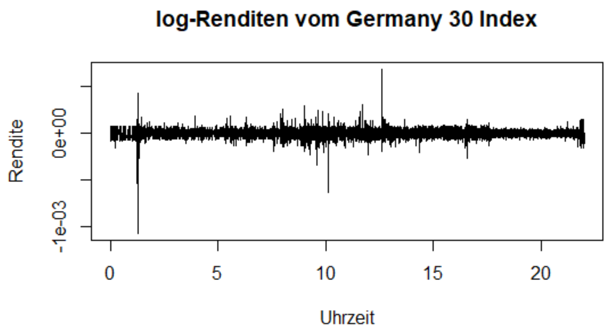
GARCH – Modelle werden genutzt um die Volatilität eines Aktienkurses, die Schwankungen einer Zeitreihe, zu modellieren. Aktienkurse folgen in der Regel keine zufällige oder gleichmäßige Volatilität, sondern Phasen geringer Kursschwankungen folgen wiederum Phasen intensiver Kursschwankungen. Aktienkurse oder Börsenindizes lassen über die Zeit oft Volatilitätscluster erkennen. Diese zu identifizieren und korrekt zu beschreiben ist das wesentliche Merkmal von GARCH – Modellen.

GARCH(1,1) – Modelle:

4.4956e-05	1.9751e-10	1.4368e-02	9.8669e-01

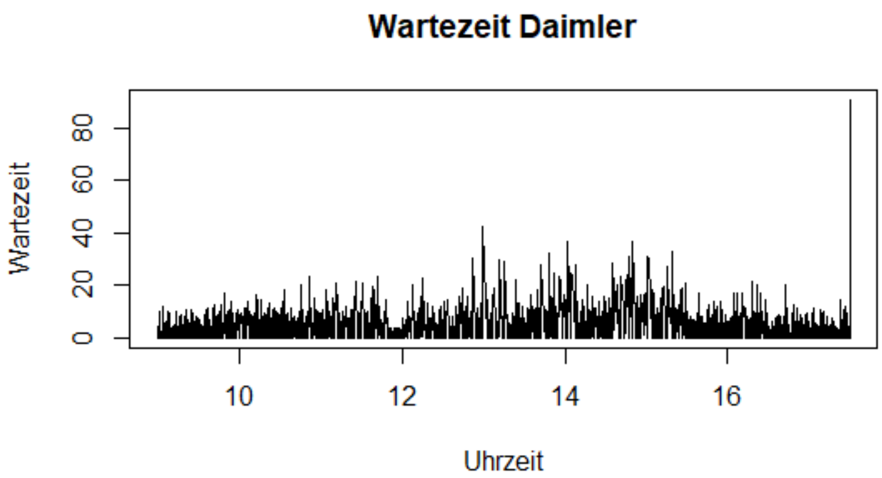
-4.8094e-06	1.7392e-08	1.3106e-02	9.8386e-01

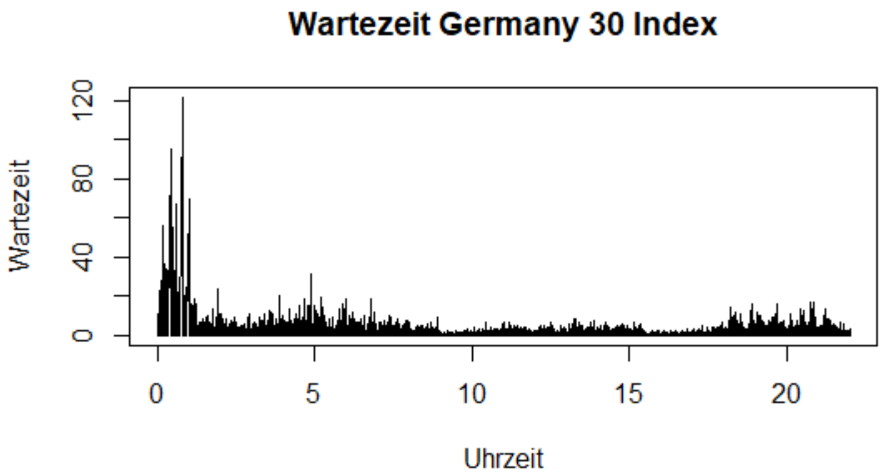




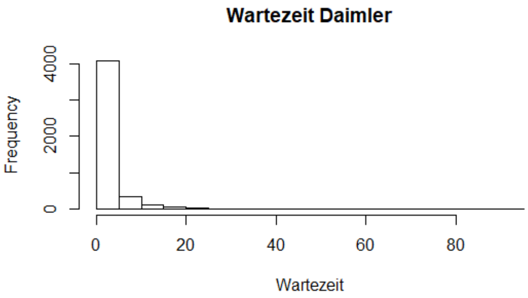
In den folgenden Grafiken sind die log-Renditen des Aktienkurses von Daimler sowie des Germany 30 Index zu sehen. Es ist erkennbar, dass die Varianzen und die bedingte Standardabweichung nicht konstant über die Zeit sind. Dies lässt darauf schließen, dass es in verschiedenen Zeiträumen zu einem starken Abfall oder Steigen der Kurse kam. In Zeiten von schneller Zunahme oder Abnahme der Kurse ist die bedingte Standardabweichung sehr groß, beispielsweise zu sehen in den log-Renditen von Daimler im Intervall von 0 bis 7. Allgemein gesagt, je größer die Standardabweichung ist, desto größer ist das Risiko und die Chancen. So kann die bedingte Standardabweichung im Risikomanagement genutzt werden, um zukünftige Prognosen für den Kurs vorherzusagen.

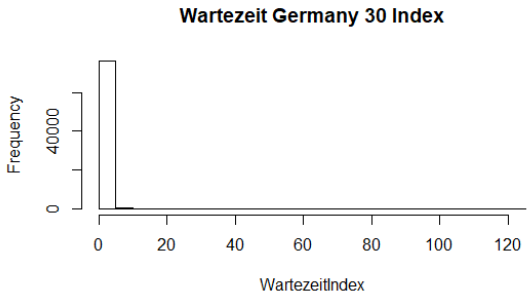
Aufgabe 3c:





Es lässt sich festhalten, dass der Aktienkurs von Daimler über den Verlauf von steigenden Wartezeiten gekennzeichnet ist. Dies bedeutet, dass sich der Wert der Aktie zu Beginn bis 13 Uhr häufig verändert hat und lässt somit darauf schließen, dass der Markt zu diesem Zeitpunkt deutlich aktiver war, als am Nachmittag. Bei dem Index sind die Wartezeiten von 0 bis 2 Uhr am längsten, was auf einen inaktiven Markt zu dieser Zeit zurückzuführen ist. Die sinkenden Wartezeiten im Zeitverlauf charakterisieren einen aktiver werdenden Markt. Bedingt durch die Öffnungszeiten der Börse, sind die Wartezeiten zwischen 9 und 17 Uhr am geringsten und somit der Markt am aktivsten. Im Vergleich zu den Wartezeiten von Daimler lassen sich Parallelen erkennen: beide Grafiken weisen zwischen 13 und 16 Uhr höhere Wartezeiten auf. Allgemein sind die Wartezeiten bei Daimler deutlich länger.





Die Histogramme zeigen die täglich durchschnittlichen Wartezeiten der Finanzmarktzeitreihen. Der Verlauf der Wartezeiten beider Zeitreihen ist sehr ähnlich. Der Großteil der Beobachtungen liegt bei Daimler sowie bei dem Index im Intervall von 0 bis 5. Jedoch ist auch hierbei zu erkennen, dass die Wartezeiten bei Daimler auf einem gering höheren Niveau sind als die von dem Index.

Quellenverzeichnis

Aufgabe 1:

http://www.reiter1.com/Glossar/Additives_Modell.html

https://books.google.de/books?id=hk3NpBcoJScC&pg=PA116&lpg=PA116&dq=charakteristika+multiplikatives+modell&source=bl&ots=8SzWlfZjwl&sig=ACfU3U3UQBTDBjrj6MMPGFiucdq4egJ04A&hl=de&s

a=X&ved=2ahUKEwjmlrne1pfnAhXOY1AKHSN0B2oQ6AEwAnoECAoQAQ#v=onepage&q=charakteristika%20multiplikatives%20modell&f=false

https://wirtschaftslexikon.gabler.de/definition/trend-50522

https://www.datadrivers.de/2018/05/07/zeitreihenanalysen/

https://www.bundesbank.de/dynamic/action/de/statistiken/zeitreihen-datenbanken/zeitreihen-datenbank/759778/759778?listId=www_s300_mb09_06c

Aufgabe 2:

https://support.minitab.com/de-de/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/store-descriptive-statistics/interpret-the-statistics/interpret-the-statistics/

Aufgabe 3:

https://www.statistik-nachhilfe.de/ratgeber/statistik/induktive-statistik/statistische-modellbildung-und-weitere-methoden/zeitreihenanalyse/garch-modelle

https://www.dukascopy.com/swiss/english/marketwatch/historical/

https://www.onvista.de/aktien/times+sales/Daimler-Aktie-DE0007100000