

Zeitreihenanalyse des DAX 40

In der Analyse des DAX-40, die Daten aus dem Zeitraum vom 1. Januar 2016 bis zum 10. Dezember 2023 umfasst, wird eine detaillierte Zeitreihenanalyse durchgeführt. Zu Beginn erfolgt eine sorgfältige Datenbereinigung, bei der die täglichen Schlusskurse des DAX 40 gesammelt und auf fehlende oder fehlerhafte Werte überprüft werden, um eine verlässliche Auswertungsgrundlage sicherzustellen.

Einen wichtigen ersten Schritt der Auswertung stellt die *deskriptive Analyse* dar. Dazu werden relevante Lage- und Streuungsmaße ermittelt, die wichtige Einblicke in die Zeitreihendaten geben und dabei helfen, die Beurteilung der Volatilität des DAX-Kurses zu bestimmen.

Ein zusätzlicher Schritt befasst sich mit der Testung auf *Stationarität*. Hierbei wird die Zeitreihe auf bestehende systematische Trends geprüft, wofür Methoden wie der Dickey-Fuller-Test und der KPSS-Test zum Einsatz kommen. Diese Tests sind unerlässlich, um die Eignung der Daten für weitere Analysen zu bestimmen und die Genauigkeit der Modellierung zu sichern.

Abschließend werden fortgeschrittene *Zeitreihenmodelle* wie ARIMA, SARIMA und GARCH angewendet, um die Zeitreihe zu modellieren und Prognosen für künftige Kursentwicklungen zu ermöglichen. Diese Modelle erweisen sich als effektiv, um Muster in Zeitreihendaten zu identifizieren, die genutzt werden können, um fundierte datengetriebene Vorhersagen zu treffen. Mithilfe der Kombination dieser verschiedenen Modelle wird das Ziel verfolgt, ein umfassendes Bild der Marktdynamik und potenzieller Risiken im analysierten Zeitraum des DAX 40 zu erstellen.

4.1 | Deskriptive Analyse

Im ersten Teil der deskriptiven Analyse der Kursdaten des DAX 40 werden Minimum, Maximum, Mittelwert, Median, Varianz und Standardabweichung betrachtet. Sie bieten einen Überblick über die Marktleistung der führenden deutschen Großunternehmen im Zeitraum zwischen 2016 und 2023. Der Minimalwert der Zeitreihe betrug am 18.03.2020 8.442 Punkte, während der Maximalwert am 08.12.2023 bei 16.759 Punkten auszumachen war. Damit ergibt sich eine Spannweite von 8.317 Punkten, was eine hohe Schwankungsbreite indiziert. Der Mittelwert des Schlusskurses des DAX-40 im besagten Zeitraum liegt bei 12.994 Punkten. Dieser Wert ist ein Indikator für die zentrale Tendenz des Marktes. Der Median, der den mittleren Wert der Schlusskurse darstellt liegt bei 12.806 Punkten. Die Ähnlichkeit von Median und Mittelwert deutet darauf hin, dass die Kurswerte recht gleichmäßig um dieses Niveau verteilt sind. Die geringe Differenz zwischen beiden Lagemaßen illustrieren demnach, dass die Daten keine starken Asymmetrien um den Schwerpunkt der Kursdaten aufweisen.

Tabelle 4.1: Deskriptive Kennzahlen

Mean	Median	Variance	SD	Minimum	Maximum
12994	12806	3361888	1834	8442	16759

Die Analyse der gängigen Streuungsmaße, hier der Varianz und Standardabweichung, erweist sich als besonders aufschlussreich für die Bewertung der Marktdynamik des DAX 40. Eine erhebliche Varianz von 3.361.888 signalisiert eine substantielle Streuung der täglichen Schlusskurse um den ermittelten Mittelwert, was auf eine ausgeprägte Volatilität während des untersuchten Zeitraums hindeutet. Dies wird weiter untermauert durch die Standardabweichung, die mit einem Wert von 1.834 eine deutliche Abweichung der einzelnen Schlusskurse vom Mittelwert anzeigt. Dieser relativ hohe Wert der Standardabweichung ist ein Indikator für eine erhöhte Volatilität des Marktes, die wiederum ein gesteigertes Risiko im Zusammenhang mit Anlageentscheidungen impliziert. Diese Erkenntnisse sind essenziell für eine tiefgehende Risikoanalyse und strategische Entscheidungsfindung im Kontext des Börsenhandels.

4.2 | Visualisierung der Zeitreihendaten

In der grafischen Darstellung der Entwicklung des DAX 40 werden die tatsächlichen Schlusskurse als schwarze Linie präsentiert, während die Trendlinie, die mittels eines gleitenden Durchschnitts berechnet wurde, in Blau dargestellt ist. Diese Trendlinie wird

genutzt, um die grundlegende Kursentwicklung der Daten in einer simplen Form, bereinigt von tagesbedingten Schwankungen, darzustellen. Dies wird möglich, indem kurzfristige Schwankungen über Mittelwertbildungen glättet werden. Dieses Vorgehen ist besonders hilfreich bei der Identifikation zugrunde liegender Muster in volatilen Finanzdaten, bei denen tägliche Preisbewegungen stark variieren können. Aus der nachfolgenden Grafik wird ersichtlich, dass bis etwa 2018 ein Aufwärtstrend im DAX 40 zu beobachten ist, was auf eine positive Marktstimmung und wirtschaftliche Wachstumspotentiale zurückgeführt werden kann. Im Jahr 2020, zur Zeit der COVID-Pandemie und damit einhergehender wirtschaftspolitischer Eingriffe sowie globaler ökonomischer Effekte ist ein deutlicher Rückgang in den Kursen erkennbar, der wirtschaftliche Krisen und damit verbundene negative Erwartungen und Unsicherheiten von Investoren repräsentiert. Dem Einbruch folgt jedoch eine starke Erholung, was auf eine schnelle Reaktion der Marktkräfte zur Wiederherstellung der früheren Werte hindeutet.

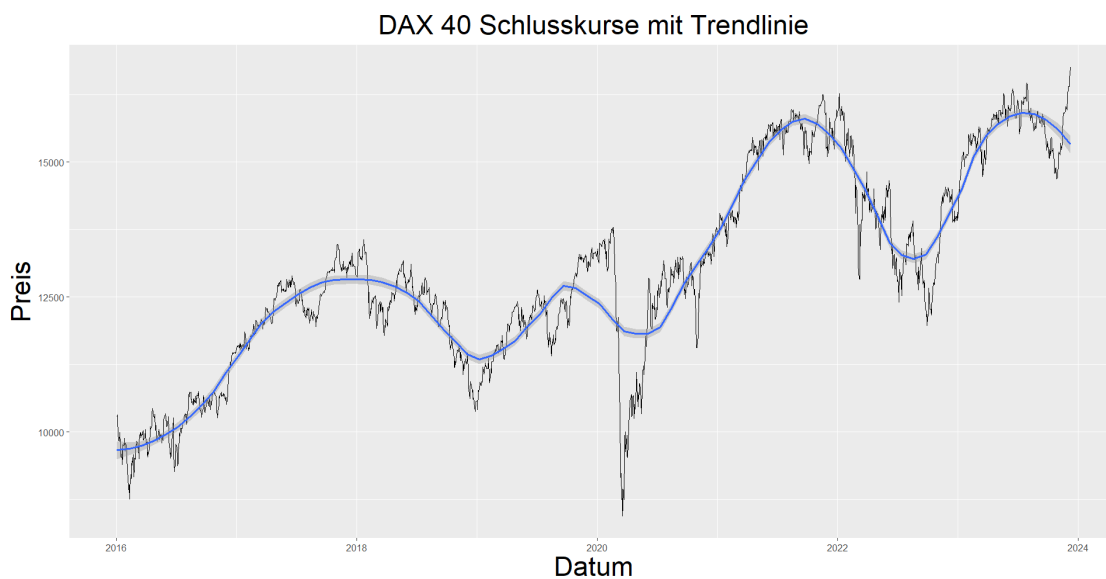


Abbildung 4.1: Beschreibung der Grafik

In einer erweiterten visuellen Aufbereitung wird eine zweigeteilte Darstellung der DAX 40 Schlusskurse offeriert. Im oberen Segment der Grafik wird ein Liniendiagramm abgebildet, welches die Schlusskurse des DAX 40 über einen mehrjährigen Zeitraum bis einschließlich 2023 abbildet. Diese graphische Repräsentation illustriert die oben bereits elaborierten Trends, inklusive eines signifikanten negativen Trends im Jahr 2020, welcher als Marktreaktion auf die COVID-19-Pandemie interpretiert werden kann, gefolgt von einer beachtlichen Marktrevitalisierung. Im unteren Segment der Darstellung sind Boxplots implementiert, die die monatliche Distribution der Schlusskurse innerhalb desselben

zeitlichen Rahmens illustrieren. Jeder Boxplot zeigt den Median, das obere und untere Quartil sowie die sogenannten „Whisker“, welche den Bereich der Mehrheit der Datenpunkte umschließen. Die vertikale Breite der Boxen und Whisker dient als Indikator für die Volatilität der Kurswerte in den jeweiligen Monatserhebungen. Eine sequenzielle Anordnung dieser Boxplots ermöglicht eine Analyse der zeitlichen Variationen in der Verteilung der Schlusskurse auf Monatsebene, einschließlich der Fluktuationen in der Volatilität sowie der Verschiebungen im Medianwert.

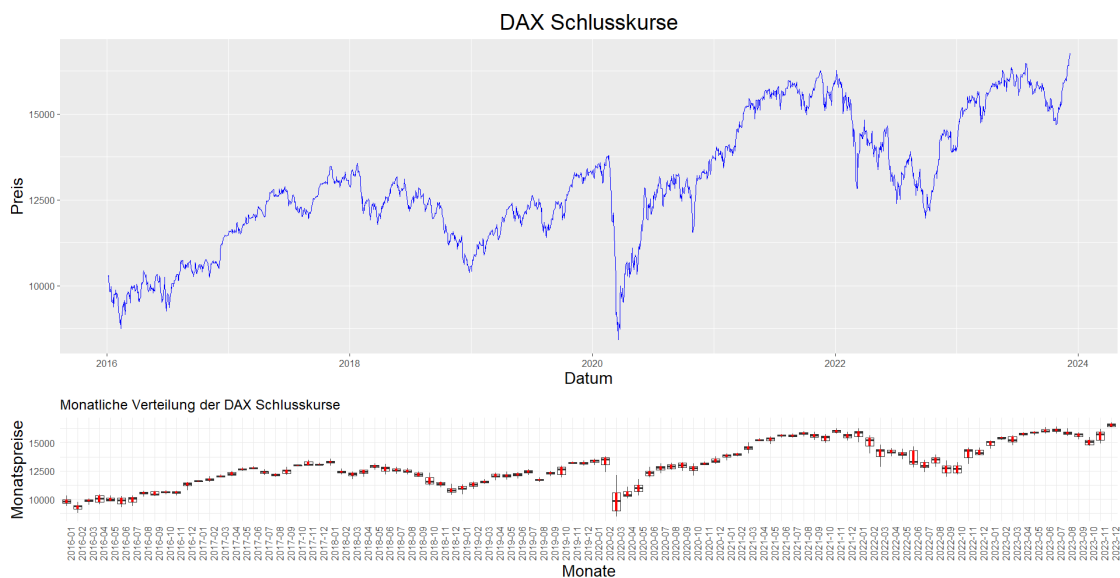


Abbildung 4.2: Beschreibung der Grafik

4.3 | ARIMA-Modellierung des DAX-40

Bei der Modellierung des DAX 40-Kurses wurde das optimale ARIMA-Modell mittels der `auto.arima()`-Funktion aus dem R-Paket `forecast` generiert. Diese Funktion setzt eine Kombination aus Einheitswurzeltests, der Minimierung des Akaike-Informationskriteriums (AIC) und der Maximum-Likelihood-Schätzung (MLE) ein, um ein ARIMA-Modell zu ermitteln. Im Rahmen des Hyndman-Khandakar-Algorithmus für automatische ARIMA-Modellierung wird der KPSS-Test verwendet, um die Anzahl der Differenzen (d) zu bestimmen. Anschließend werden die Parameter p (Ordnung des autoregressiven Teils) und q (Ordnung des Moving-Average-Teils) durch Minimierung des korrigierten Akaike-Informationskriteriums (AICc) festgelegt. Variationen des aktuellen Modells werden durch Änderung der Werte von p und/oder q um ± 1 vom aktuellen Modell sowie durch Einbeziehung oder Ausschluss der Konstanten c berücksichtigt. Der angestoßene Op-

timierungsprozess wird solange wiederholt, bis kein niedrigeres AIC mehr gefunden werden kann, wodurch eine optimale Balance zwischen Modellkomplexität und Anpassungsgüte gewährleistet wird.

Das Ergebnis der ARIMA-Modellierung der DAX 40 Daten ergibt ein optimales Modell in der Form eines **ARIMA(0,1,0)**-Modells. Dieses Modell zeichnet sich durch das Fehlen von autoregressiven (AR) und Moving-Average (MA) Termen aus, wie die Null-Werte des ersten und dritten Parameters anzeigen. Zudem nutzt es eine einmalige Differenzierung, wie sie durch den Wert des zweiten Parameters dargestellt wird.

Im Rahmen der iterativen Optimierung des ARIMA-Modells für die Zeitreihe spielen verschiedene statistische Kennwerte eine entscheidende Rolle, um die Modellgüte im Rahmen der iterativen Modelloptimierung zu evaluieren. Diese Werte sind demnach ein integraler Bestandteil des Modellierungsprozesses, um ein optimales Gleichgewicht zwischen Modellkomplexität und Anpassungsgenauigkeit zu finden.

Die Varianz der Fehler σ^2 mit einem Wert von 22.276, ist ein zentraler Indikator für die Präzision der Modellanpassung. Eine niedrigere Varianz signalisiert, dass die Residuen, also die Abweichungen der Modellvorhersagen von den tatsächlichen Daten, gering sind, was auf eine hohe Modellgenauigkeit hindeutet. Diese Information ist besonders wichtig bei der iterativen Feinabstimmung des Modells, um die beste Anpassung an die Daten zu erreichen.

Die Log-Likelihood, die einen Wert von -12.952 aufweist, misst die Wahrscheinlichkeit der beobachteten Daten unter dem aktuellen Modell. Ein höherer Log-Likelihood-Wert (weniger negativ) ist ein Zeichen dafür, dass das Modell die Daten besser erklärt. Im Rahmen des iterativen Modelloptimierungsprozesses dient dieser Wert ebenfalls als Indikator, um Modellvariationen zu beurteilen, die eine bessere Anpassung an die Daten ermöglichen.

Das Akaike Information Criterion (AIC) mit einem Wert von 25.906 und der korrigierte AIC (AICc), der einen identischen Wert aufweist, sind wesentliche Maßstäbe zur Bewertung der Modellqualität. Sie bieten ein Kriterium, um verschiedene Modelle zu vergleichen, indem sie sowohl die Anpassung als auch die Anzahl der Modellparameter berücksichtigen. Im Rahmen der iterativen Modellfindung strebt man so nach dem Modell mit dem niedrigsten AIC, da ein niedrigerer Wert ein besseres Gleichgewicht zwischen Genauigkeit und Komplexität darstellt.

Ebenso wichtig ist der Bayesian Information Criterion (BIC) Wert von 25.912. Der BIC folgt einer ähnlichen Logik wie der AIC, legt jedoch mehr Gewicht auf die Modellkomplexität und bestraft Modelle mit einer größeren Anzahl von Parametern stärker. Dies ist besonders nützlich, um Überanpassung zu vermeiden und die Einfachheit des Modells im Optimierungsprozess zu bewahren.

Parameter	Wert
σ^2 (Varianz)	22.276
Log-Likelihood	-12.952
AIC	25.906
AICc	25.906
BIC	25.912

Tabelle 4.2: Kennwerte des Modells

In einem weiteren Schritt werden die Fehlermaße des Trainingssets als Indikatoren für die Genauigkeit und Zuverlässigkeit des implementierten Modells herangezogen.

Fehlermaß	Training Set
ME (mittlerer Fehler)	-3.2
RMSE (Root Mean Square Error)	149.2
MAE (Mean Absolute Error)	104.3
MPE (Mean Percentage Error)	-0.032
MAPE (Mean Absolute Percentage Error)	0.83
MASE (Mean Absolute Scaled Error)	0.06
ACF1 (Autocorrelation of Errors at Lag 1)	-0.015

Tabelle 4.3: Trainingsset Fehlermaße

Es wird eine leichte Tendenz zur Unterschätzung der beobachteten Werte deutlich, die durch den mittleren Fehler (ME) von -3.2 indiziert werden. Dies deutet auf eine minimale systematische Verzerrung im applizierten Modell hin. Der Root Mean Square Error (RMSE) mit einem Wert von 149.2 zeigt eine moderate Streuung der Fehler um den Mittelwert, was auf eine akzeptable Genauigkeit des Modells hinweist, auch wenn an dieser Stelle noch Optimierungspotenziale vorhanden sind.

Durch den Mean Absolute Error (MAE) von 104.2534 wird ersichtlich, dass die Vorhersagen des Modells im Durchschnitt um etwa 104 Einheiten von den tatsächlichen Werten abweichen, was ein nützliches Maß für die Größe der Fehler darstellt. Der Mean Percentage Error (MPE), der mit einem Wert von -0.03 nahe Null liegt, signalisiert, dass kein

signifikanter Bias im Modell vorliegt, was auf eine ausgewogene Vorhersagefähigkeit ohne systematische Über- oder Unterschätzungen hindeutet.

Von hoher Relevanz ist zudem der Mean Absolute Percentage Error (MAPE) von 0.83, der indiziert, dass die durchschnittlichen Vorhersagefehler weniger als 1% des tatsächlichen Wertes betragen, was eine hohe Vorhersagegenauigkeit impliziert. Der Mean Absolute Scaled Error (MASE) mit einem Wert von 0.062 zeigt an, dass das Modell im Vergleich zu einem naiven Benchmark eine effiziente Vorhersageleistung erbringt.

Abschließend wird durch die Autokorrelation des Fehlers bei Verzögerung 1 (ACF1) mit einem Wert von -0.015 deutlich, dass keine signifikante Autokorrelation der Fehler vorliegt, was auf eine adäquate Spezifikation des Modells hindeutet. Insgesamt wird durch diese Fehlermaße eine zufriedenstellende Anpassung des Modells an die Zeitreihe des DAX-40 gezeigt, die mit einer bedeutenden Vorhersagegenauigkeit und angemessener Modellspezifikation einhergeht.