

Zürcher Hochschule für Angewandte Wissenschaften
ZHAW SML
Master in Banking & Finance

Masterarbeit

*Wie können Schwankungen des Tradingvolumens von Privatpersonen in
den USA erklärt werden?*

Dannenmann Fabio

Eingereicht bei: Dr. Christoph Kley

12. Juni 2023

Management Summary

Das Handelsvolumen von Privatinvestoren erlebte in den letzten Jahren während und nach der Coronapandemie in den USA und weltweit neue Höchststände. Trotz grosser Medienabdeckung der erfahrenen Meme-Stock Hypes ist über die Einflüsse makroökonomischer Variablen auf das Handelsvolumen von Retailinvestoren wenig bekannt, während die individuellen Antriebe von Privatpersonen, an den Finanzmärkten teilzunehmen, bereits gut erforscht sind. Ein Grund hierfür liegt unter anderem in der schwierigen Abgrenzung der Handelsvolumina zwischen Institutionellen oder Privatinvestoren. Mittels der Charakteristiken des US-amerikanischen Systems zur Abwicklung von Retailaufträgen kann jedoch eine solche Abgrenzung vorgenommen werden.

Die vorliegende Arbeit beschreibt in einem ersten Schritt, wie diese Abgrenzung vorgenommen werden kann. In einem zweiten, empirischen Teil werden die monatlich gehandelten Volumina von 2009 bis 2022 der Retailinvestoren in den USA auf ihre Abhängigkeiten von makroökonomischen Parametern untersucht. Einerseits werden die gefundenen Abhängigkeiten erfasst und mögliche Gründe dafür beschrieben, andererseits wird ein Modell erarbeitet und empfohlen, mit welchem die zukünftigen Veränderungen des Handelsvolumens von Retailinvestoren geschätzt werden können.

Die erhaltenen Resultate zeigen, dass die Volumina signifikant von einem Set von makroökonomischen Parametern abhängen, und dass vor allem die vergangenen Returns des Leitindex S&P 500 einen signifikanten Einfluss auf besagte Volumina ausüben.

Inhaltsverzeichnis

| | |
|---|-----------|
| 1. Einleitung | 1 |
| 1.1. Hintergrund und Relevanz | 1 |
| 1.2. Thema und Fragestellung..... | 2 |
| 1.3. Methodik und Analyse..... | 2 |
| 2. Literature review | 4 |
| 2.1. Individuelle Einflüsse auf Tradingaktivitäten von Privatpersonen..... | 4 |
| 2.2. Einflüsse der Coronapandemie auf das Verhalten von Retailinvestoren..... | 8 |
| 2.3. Einfluss von Retailinvestoren auf die Liquidität und Preissetzung | 9 |
| 2.4. Einflüsse auf das Tradingvolumen von Retailinvestoren | 11 |
| 2.5. Forschungslücke | 12 |
| 3. Definitionen | 13 |
| 3.1. Handelssystem zur Abwicklung von Retailaufträgen in den USA..... | 13 |
| 3.1.1. Marktstruktur für Tradeausführung in den USA | 13 |
| 3.1.2. Erträge der Online Broker | 14 |
| 3.2. Herleitung des aggregierten Marktvolumens von Retailinvestoren..... | 14 |
| 4. Empirischer Teil | 16 |
| 4.1. Hypothesen | 16 |
| 4.2. Datenanalyse..... | 17 |
| 4.2.1. Datenerhebung..... | 17 |
| 4.2.2. Analyse TRF Off Exchange Umsätze | 18 |
| 4.2.3. Analyse Routing Erträge Online Broker | 21 |
| 4.2.4. Deskriptive Statistik ausgewählter exogener Modellparameter | 22 |
| 4.3. Modelle | 24 |
| 4.3.1. OLS Regression..... | 24 |
| 4.3.2. SARIMAX-Modell..... | 25 |
| 4.3.3. Panel Regression mit fixed Effects..... | 26 |
| 4.4. Analyse des Handelsvolumens von Retailinvestoren | 26 |
| 4.4.1. Rollende Zeitfenster | 27 |
| 4.5. Analyse der Orderroutererträge | 28 |
| 4.6. Modellparameter | 29 |
| 4.7. Adfullertest auf Stationarität..... | 33 |
| 4.8. Überprüfung auf Vollständigkeit / Korrelation / Verteilung | 34 |
| 4.9. Lasso Regression zur optimalen Kombination von Modellparameter..... | 35 |
| 4.10. Hyperparametertuning für das SARIMAX-Modell..... | 36 |

| | |
|---|-----------|
| 5. Ergebnisse | 37 |
| 5.1. Analyse Ergebnisse OLS-Regression | 37 |
| 5.2. Analyse SARIMAX-Modell | 40 |
| 5.3. Rollende Zeitfenster für OLS-Regression und SARIMAX Modell | 44 |
| 5.4. Analyse der besten OLS und SARIMAX Regression | 46 |
| 5.5. Durchführung Panel-OLS-Regression | 47 |
| 6. Schlussteil | 50 |
| 6.1. Diskussion der Resultate..... | 50 |
| 6.2. Restriktion der Resultate..... | 52 |
| 6.3. Implikationen | 52 |
| 7. Quellenverzeichnis..... | 54 |
| 8. Anhang..... | 58 |
| 8.1. Korrelationsmatrix Modellparameter..... | 58 |
| 8.2. OLS-Regression alle Modellparameter (normalisierte Werte)..... | 59 |
| 8.3. OLS-Regression nach L1-Regularisierung (normalisierte Werte) | 59 |
| 8.4. SARIMAX-Modell vor Grid Search (normalisierte Werte)..... | 60 |
| 8.5. SARIMAX-Modell nach Grid Search (normalisierte Werte)..... | 60 |
| 8.6. Python Code..... | 61 |

Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1: Sechs Dimensionen der Haltung gegenüber Finanzfragen..... | 6 |
| Abbildung 2: Handelsaktivität Retailinvestoren in den USA | 19 |
| Abbildung 3: Dichte und Verteilung der Umsätze an Off-Exchange Börsen..... | 20 |
| Abbildung 4: Schätzungen vs. wahre Werte beste OLS Regression | 46 |
| Abbildung 5: Schätzungen vs. wahre Werte bestes SARIMAX-Modell..... | 47 |

Tabellenverzeichnis

| | |
|---|----|
| Tabelle 1: Aufteilung Investoren nach Charakteristiken)..... | 7 |
| Tabelle 2: Aufteilung Handelsvolumen TRF nach Börse und Tape..... | 18 |
| Tabelle 3: Erträge aus Orderrouting | 21 |
| Tabelle 4: Übersicht Modellparameter (Teil 1) | 22 |
| Tabelle 5: Übersicht Modellparameter (Teil 2) | 23 |
| Tabelle 6: Anzahl gemachte Predictions pro Zeitraum..... | 27 |
| Tabelle 7: Modellparameter und deren Relevanz | 32 |
| Tabelle 8: ADF Resultate und Klassifikation | 33 |
| Tabelle 9: Shapiro-Wilk-Test und Klassifikation..... | 34 |
| Tabelle 10: Resultate OLS Regression | 37 |
| Tabelle 11: Resultate SARIMAX-Modell | 42 |
| Tabelle 12: Durchschnittliche MSE der getesteten OLS-Regressionen | 44 |
| Tabelle 13: Durchschnittliche MSE der getesteten SARIMAX-Modelle | 45 |
| Tabelle 14: Panel-OLS-Regression der Orderroutingserträge von Online-Broker | 48 |

1. Einleitung

1.1. Hintergrund und Relevanz

Eine Hauptertragsquelle für Banken in den USA und weltweit ist neben dem klassischen Zinsdifferenzgeschäft, das bedeutet die Annahme von Kundeneinlagen und die Vergabe von Krediten, der Ertrag aus dem Kommissions- und Dienstleistungsgeschäft. Insgesamt erwirtschafteten alle US-Finanzunternehmen einen Kommissionsertrag aus «Securities Intermediation and Brokerage» von über 318.3 Milliarden USD im Jahr 2021 (US Census Bureau, 2023b). Diese Erträge erlebten in den letzten fünf Jahren ein konstantes Wachstum, so nahmen diese im Schnitt um jährlich 5.3% zu (US Census Bureau, 2023b). Diese Kommissionserträge umfassen Unternehmen, die hauptsächlich Kapital bei der Platzierung von Wertpapieren aufs Spiel setzen oder als Marketmaker für Wertpapiere und Waren fungieren. Ebenso umfassen sie Unternehmen, die als Agenten oder Makler zwischen Käufern und Verkäufern von Wertpapieren und Waren agieren, und in der Regel eine Provision erhalten (US Census Bureau, 2023a). Folglich sind diese Courtagenerträge sehr volatil und hängen mutmasslich mit der Stimmung der Privatanleger zusammen.

Ebenfalls unterscheiden sich Privatpersonen, welche an den Finanzmärkten aktiv sind, in verschiedenen Punkten von institutionellen Anlegern. So weisen sie kleinere Investitionsbeträge, Anlageressourcen, weniger Zugänge zu Analysten- und Researchberichten und professioneller Beratung auf (M. Talwar et al., 2021, S. 1). Nicht nur sind in den vergangenen Jahren die Erträge der Banken aus dem Geschäft mit dem Brokerage gestiegen, auch hat die Handelsaktivität von Privatanlegern in Europa und den USA generell zugenommen. Dementsprechend waren im Jahr 2021 zwischenzeitlich über 25% der Marktaktivität in den USA auf Retailinvestoren zurückzuführen (Roy, 2022, S. 4). Aufgrund der hohen Relevanz für die Erträge der Banken und den Einflüssen auf die globalen Finanzmärkte ist es deshalb wichtig, die Marktaktivität von Privatinvestoren schätzen zu können und die Einflussfaktoren auf diese zu verstehen.

1.2. Thema und Fragestellung

Aufgrund der substanziellen Zunahme von «Securities Intermediation and Brokerage» Erträgen von US-Banken über die letzten Jahre sowie die Relevanz von Handelsaktivitäten der Privatinvestoren auf die Kursentwicklung an den Finanzmärkten, ist es – wie im Abschnitt 1.1 aufgezeigt – von grosser Bedeutung, relevante Einflussfaktoren auf das Handelsverhalten von Retailinvestoren zu kennen und deren Auswirkungen zu verstehen. Diese Masterarbeit befasst sich deshalb mit der folgenden Fragestellung:

Wie können Schwankungen des Tradingvolumens von Privatpersonen in den USA erklärt werden?

Der Fokus liegt dabei auf dem US-amerikanischen Markt und den Implikationen, welche aus den Resultaten entstehen.

1.3. Methodik und Analyse

Die Fragestellung wird anhand einer quantitativen Analyse eines zu erhebenden Datensatzes untersucht. Die Schwierigkeit bei der Datenerhebung liegt darin, Transaktionen von privaten sowie institutionellen Finanzmarktteilnehmern zu trennen. Hierzu wird für die Marktaktivitäten in den USA die folgende Herangehensweise gewählt:

Um die Marktaktivitäten der Retailinvestoren in den USA zu eruieren, wird analog dem Paper von Boehmer et al. (2021) vorgegangen. Dieses Paper beruht auf der Untersuchung der Trade and Quote Datenbank der New York Stock Exchange (NYSE), in welcher alle Handelsabschlüsse für Finanzprodukte der NYSE, Nasdaq und regionaler US-Börsen erfasst werden. Als relevant wurden nur Handelsabschlüsse mit dem Code «D» für off-exchange qualifiziert. Die Beschränkung auf den Code «D» wird zur eigentlichen Qualifizierung als Retail Investor Trade genutzt, da in den USA Aufträge von Privatinvestoren durch Broker ausserbörslich abgewickelt werden (Boehmer et al., 2021, S. 2253). Um die aggregierten off-exchange Handelsvolumina zu errechnen, werden somit die täglichen Marktvolumina aller Börsen der USA aus der Datenbank der Chicago Board Options Exchange (CBOE) bezogen. Danach werden die Daten auf täglicher Basis vom 01.01.2009 – 31.12.2022 für Code «D»-Märkte aggregiert und somit die abgewickelten Turnover privater Investoren berechnet. Nachfolgend werden die Daten auf monatlicher Ebene zusammengefasst.

Der Datensatz (USA) wird mit länderspezifischen exogenen Daten ergänzt, welche in derselben Frequenz verfügbar sind. Dies sind die Renditen der zweijährigen

Staatsanleihen, die Renditen der zehnjährigen Staatsanleihen und die Kursentwicklung des wichtigsten Aktienindex (S&P500) sowie weitere makroökonomische Faktoren, welche einen Einfluss auf das Handelsvolumen von Retailinvestoren haben könnten. In der anschließenden Datenanalyse werden Datenlücken sowie weitere Herausforderungen hinsichtlich der Datengewinnung identifiziert. Die Daten werden auf statistische Interferenzen überprüft. Das daraus resultierende Datenset wird zudem mittels deskriptiver Statistik untersucht und plausibilisiert.

Die quantitative Modellanalyse, bei welcher der Zusammenhang zwischen der Marktaktivität von Privatinvestoren und Veränderungen der exogenen Einflüsse festgestellt werden soll, wird mittels einer OLS-Regression und einem Multivariate Seasonal Autoregressive Integrated Moving Average Model (SARIMAX) durchgeführt. Ein SARIMAX-Modell basiert auf der Annahme, dass Zeitreihendaten aus einer Kombination von Saisonalität, Autoregression sowie dem gleitenden Durchschnitt bestehen. Die Autoregression beschreibt eine Analyse, in welcher untersucht wird, wie die Werte einer Variablen von den vorhergegangenen Werten derselben Variablen abhängen. Der gleitende Durchschnitt bezeichnet eine Technik, mit welcher die Fluktuationen der Werte einer Variable geglättet, und somit der zugrundeliegende Trend erkannt wird. Zudem wird beim SARIMAX-Modell die Saisonalität der Werte berücksichtigt. Das Modell wird in der Untersuchung auf einen Cross Section Time Series Datensatz angewendet. Damit erhält es einen multivariaten Charakter. Dies bedeutet, dass mehrere exogene Faktoren herangezogen werden, um die abhängige Variable (Tradingvolumen von Privatinvestoren) zu erklären. Um das optimale Modell in Bezug auf Schätz- und Trainingszeitraum zu finden, werden im Anschluss an die Analyse der Abhängigkeiten der endogenen Variable Schätzungen mittels rollendem Zeitfenstern beim SARIMAX-Modell sowie der OLS-Regression durchgeführt.

Nach der Untersuchung der Volumina des Retailhandels und deren Abhängigkeiten zwischen den Variablen werden die Ergebnisse auf Orderrouteringträge von Online-Brokern in den USA angewendet. In den USA erwirtschaften Online-Broker ihre Erträge aus der Ausführung von Handelsaufträgen für Retailinvestoren nicht über Kommissionen und Gebühren, welche dem Investor belastet werden, sondern über Erträge aus dem gezielten Leiten von Handelsvolumen zu bestimmten Market-Makern. Diese entrichten dann eine Gebühr an das Online-Broker-Unternehmen. Um die Einflüsse der exogenen

makroökonomischen Variablen auf die Orderrouterträge zu untersuchen, wird eine Panel-OLS-Regression durchgeführt. Anschliessend können die gefundenen Abhängigkeiten und Ergebnisse mit den Erkenntnissen der zuvor gemachten Untersuchungen verglichen werden.

2. Literature review

2.1. Individuelle Einflüsse auf Tradingaktivitäten von Privatpersonen

Grundsätzlich geht die Forschung zur Mikro- sowie Verhaltensebene davon aus, dass Retailinvestoren sich von institutionellen Investoren unterscheiden. So haben sie eine viel kleinere Investitionsgrösse, weniger Ressourcen zur Verfügung und weniger professionelle Beratung (M. Talwar et al., 2021, S. 1). Jedoch kann das Verhalten von Retailinvestoren stark variieren: Während einzelne Kapitalpuffer aufbauen, um rationale Investmententscheidungen zu treffen, sind andere mehr vom Instinkt abhängig und treffen Entscheidungen kurzfristig (Fünfgeld & Wang, 2009, S. 123).

Ebenfalls zeigte sich, dass Investoren, welche sich selbst eine höhere Bildung in finanziellen Angelegenheiten zusprechen, höhere Renditen erzielen. Dementsprechend tendieren erfahrene Retailinvestoren dazu, bessere Anlageentscheidungen zu treffen. Sie investieren einen höheren Anteil ihres Vermögens in Aktien sowie komplexe Anlageprodukte. Zudem konzentrieren sie ihr Vermögen auf eine kleinere Anzahl von Wertpapieren und erreichen Diversifikation durch Investitionen in Fondsprodukte (Bellofatto et al., 2018, S. 178). Ebenfalls konnte Bellofatto et al. (2018) zeigen, dass die erfahrenen Anleger ihren höheren Return mittels einem besseren Sharpe Ratio erwirtschaften als unerfahrene Anleger. Zudem nahm das Risiko von Dispositionseffekten (Tendenz eines Investors, Vermögenswerte zu verkaufen, welche im Wert gestiegen sind) mit steigender Erfahrung ab (Bellofatto et al., 2018, S. 178). Ausserdem deuten verschiedene Studien darauf hin, dass sich das Verhalten von Retailinvestoren abhängig von Informationsvermittlung und Kommunikation von anderen Marktteilnehmern (zum Beispiel institutionelle Investoren) verändern kann (Seth et al., 2020, S. 12).

Auf der individuellen Ebene eines Privatinvestors ist nicht nur das Finanzwissen für das Volumen der Handelsaktivität ausschlaggebend, sondern auch die Fähigkeit, Entscheidungen im Zusammenhang mit Finanzfragen zu treffen (Shim et al., 2009, S. 711). Diese Haltung gegenüber Finanzfragen wird gemäss Talwar et al. (2021) durch sechs verschiedene Dimensionen (Abbildung 1) definiert (M. Talwar et al., 2021, S. 3). Auf die

verschiedene Ausprägung dieser sechs Dimensionen folgt die individuelle Aktivität der Retailinvestoren an den Finanzmärkten:

Die Dimension der finanziellen Furcht bezeichnet die Angst davor, Investitionsentscheidungen zu treffen. Diese kann dazu führen, dass keine Investitionen oder Investitionen in risikoärmere Vermögenswerte getätigt werden. In Konsequenz führt ein erhöhtes Mass an finanzieller Furcht zu einem tieferen Risikoprofil der Anlagen und einer niedrigeren Investitionstätigkeit (Kuhnen & Knutson, 2011, S. 607).

Die Dimension des Optimismus wirkt wiederum konträr zur finanziellen Furcht. So tendieren weniger optimistische Personen eher dazu, das Geld zu sparen, während Personen, welche der Zukunft optimistischer entgegen sehen, eher Investitionen tätigen (M. Talwar et al., 2021, S. 4).

Der Aspekt der finanziellen Sicherheit bezeichnet wiederum die finanziellen Umstände, in welcher sich eine Person befindet. Sind die Rahmenbedingungen gut, das heisst eine Person verfügt über ein hohes Vermögen, lässt dies darauf schliessen, dass die Aktivitäten an den Finanzmärkten höher ausfallen als bei Personen, welche finanziell weniger gut situiert sind (Keltner & Lerner, 2001, S. 148).

Das Mass des strukturierten Denkens bezeichnet die Fähigkeit eines strukturierten und geplanten Problemlösens. Investoren, welche über ein vergleichsweise höheres Mass an strukturierter Denkfähigkeit verfügen, tendieren eher dazu, rationale Entscheidungen zu treffen und weniger verzerrten Wahrnehmungen zu erliegen (M. Talwar et al., 2021, S. 4).

Gemäss den Erkenntnissen von Bellofatto et al (2018) führt das Interesse an Finanzthemen folglich zu einem höheren Wissenstand und im Umkehrschluss zu höheren Erfahrungswerten im Umgang mit Investitionen an den Finanzmärkten, was auf eine erhöhte Handelsaktivität schliessen lässt (M. Talwar et al., 2021, S. 5).

Die letzte Dimension bezeichnet das Bedürfnis nach Vorsorge. Personen, welche ein hohes Bedürfnis nach Vorsorge aufweisen, tendieren dazu, ihre Sparquote zu erhöhen. Nicht nur halten diese eine höhere Liquiditätsquote, sie investieren ebenfalls langfristig und verteilen ihre Handelsaktivitäten über einen längeren Zeitraum. Ein erhöhtes Bedürfnis nach privater Vorsorge führt also zu einer niedrigeren Handelsaktivität an den Finanzmärkten (M. Talwar et al., 2021, S. 5).

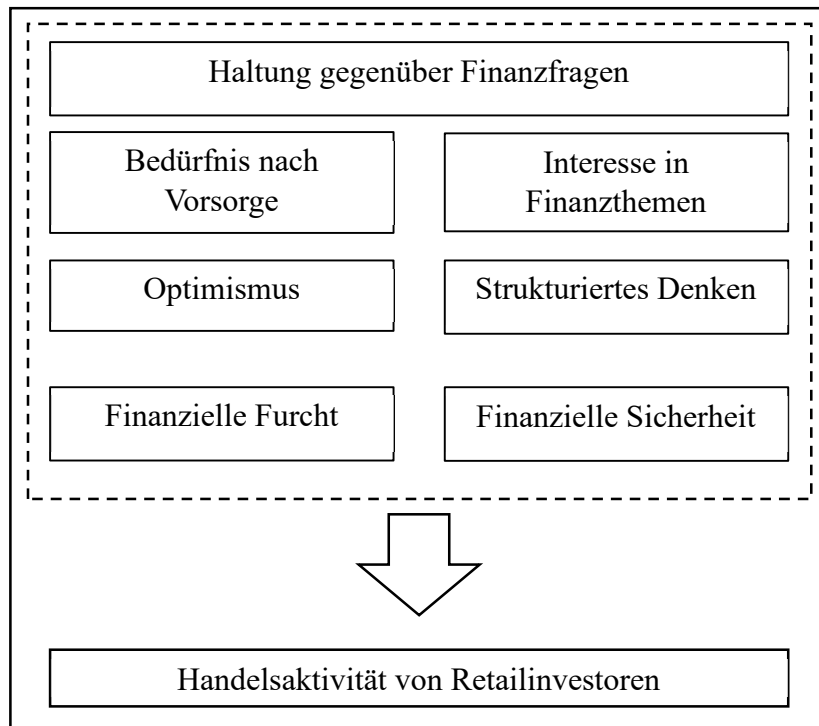


Abbildung 1: Sechs Dimensionen der Haltung gegenüber Finanzfragen (eigene Darstellung) (M. Talwar et al., 2021, S. 3).

Die vorhin beschriebenen Dimensionen lassen darauf schliessen, dass die Handelsaktivität von Retailinvestoren von einem individuellen Mix der verschiedenen Dimensionen abhängt. Je nach individueller Ausprägung der vorhin erläuterten Dimensionen fällt die Handelsaktivität eines Retailinvestors infolgedessen unterschiedlich aus.

Um die Aktivitäten der Privatanleger an den Finanzmärkten zu identifizieren ist es hilfreich, diese in relevante Gruppen einzuteilen und zu kategorisieren. Eine Studie der Bank of New York Mellon (2022) teilt die US-Privatinvestoren in vier verschiedene Gruppen (entstehende Investoren, traditionelle Investoren, etablierte Investoren, Investoren für die Altersvorsorge) ein (Tabelle 1). Die Gruppen werden anhand von Einkommen, Alter, Investitionserfahrung und Bildungsstand differenziert (The Bank of New York Mellon Corporation, 2022, S. 3).

Die Gruppe der Entstehenden Investoren (nascent investors) ist darauf fokussiert, über Investitionen zu lernen, zusätzliches Einkommen zu generieren und finanzielle Stabilität zu erreichen. Faktoren wie der verbesserte Zugang zu Informationen durch mobile Technologie, die einfache Eröffnung von Konten und die Möglichkeit, kleine Beträge zu investieren, haben ihre Marktaktivität beschleunigt (The Bank of New York Mellon Corporation, 2022, S. 7).

Die Gruppe der traditionellen Investoren betrachten Investitionen im Vergleich zu Nascent Investors weniger als zusätzliches Einkommen. Sie haben einen kurzen Anlagehorizont, wobei die meisten weniger als ein Jahr investieren. Sie besitzen mehr traditionelle Anlageklassen wie Einzelaktien, Investmentfonds und ETFs. Zudem legen traditionelle Investoren mehr Wert auf die Altersvorsorge und langfristige finanzielle Stabilität (The Bank of New York Mellon Corporation, 2022, S. 8).

Etablierte Investoren haben langjährige Erfahrung mit Investitionen und nehmen seit zehn oder mehr Jahren an den Finanzmärkten teil. Finanzexperten und Analystenberichte sind ihre Hauptinformationsquellen und sie überwachen die Bewertungen von Aktien und die Performance ihrer Portfolios regelmässig (The Bank of New York Mellon Corporation, 2022, S. 9).

Investoren für die Altersvorsorge sind eine Untergruppe der etablierten Investoren. Sie legen Wert auf langfristige finanzielle Sicherheit und sind in ihrer Anlagestrategie relativ konservativ. Der Fokus von Investoren für die Altersvorsorge liegt in der Einkommengenerierung durch Anlageinstrumente wie Anleihen (The Bank of New York Mellon Corporation, 2022, S. 10).

| Kriterien | Entstehende Investoren (nascent Investors) | Traditionelle Investoren | Etablierte Investoren | Investoren für die Altersvorsorge |
|------------------------------|---|---------------------------------|------------------------------|--|
| Soziale Gruppe: | Generation Z, Millennials | Millenials, Generation X | Generation X, Baby Boomers | - |
| Alter: | 18 - 29 Jahre | 25 – 40 Jahre | 40 – 60 Jahre | 60+ Jahre |
| Einkommen | Weniger als 75tsd USD | 50 – 100tsd | 100 – 200tsd | - |
| Investierbare Vermögenswerte | 10 – 100tsd USD | 10 – 100tsd USD | 100tsd – 1mio | 100 – 500tsd |
| Investitionserfahrung | Weniger als 5 Jahre | Weniger als 5 Jahre | Mehr als 10 Jahre | Mehr als 10 Jahre |
| Bildungsstand | Universitätsabschluss/Studium | Universitätsabschluss | Höherer Abschluss | - |

Tabelle 1: Aufteilung Investoren nach Charakteristiken (eigene Darstellung) (The Bank of New York Mellon Corporation, 2022, S. 3)

2.2. Einflüsse der Coronapandemie auf das Verhalten von Retailinvestoren

Im Zusammenhang mit der Corona Pandemie und den darauffolgenden erhöhten Handelsaktivitäten von Retailinvestoren in den USA wirft Talwar et al. (2021) die Frage auf, ob sich die generelle Haltung von Privatpersonen gegenüber finanziellen Herangehensweisen ab dem Frühjahr 2020 geändert haben. Die Studie von Ortmann et al. (2020), welche das Verhalten der Investoren während der Coronapandemie beleuchtet, zeigt, dass, während sich das Coronavirus um die Welt verbreitete und diverse Länder sich in den Lockdown begaben, die Handelsaktivitäten von Retailinvestoren signifikant zunahmen. Einerseits eröffneten mehr Personen einen Investorenaccount bei einer Brokerfirma, andererseits nahm die Handelsaktivität von bestehenden Accounts ebenfalls zu (Ortmann et al., 2020, S. 9).

Die Zunahme der Handelsaktivitäten von Retailinvestoren während der Coronapandemie und den nachfolgenden zwei Jahren konnte S.Talwar et al. (2021) mit einer verstärkten Herdenmentalität begründen. So führt die verstärkte Herdenmentalität der Retailinvestoren zu einer höheren Handelsaktivität aufgrund des externen Events der Pandemie (S. Talwar et al., 2021, S. 2155). Das verstärkte Herding während der Pandemie ist einerseits darauf zurückzuführen, dass Millenials dazu tendieren, andere Personen zu konsultieren, wenn sie eine Investmententscheidung treffen. Folglich sind die Gruppen der traditionellen sowie der Nascent Investoren (s.Tabelle 1) eher dazu bereit, Aktien zu kaufen oder zu verkaufen, wenn andere dasselbe tun (S. Talwar et al., 2021, S. 2156). Yoon & Oh (2022) konnten aufzeigen, dass das erhöhte Herding während den Jahren 2020 und 2021 zum Teil auf abnormale Social Media-Aktivitäten zurückzuführen sind. Die erhöhte Abdeckung von Finanzthemen, einzelnen Aktien oder Indizes auf den sozialen Netzwerken kann so zu einem erhöhten Herding in diese Finanztitel geführt haben, was das Handelsvolumen von Retailinvestoren ansteigen liess. (Yoon & Oh, 2022, S. 15). Die Resultate der Studie implizieren, dass Social Media die Haupttreiber der verstärkten Herdenmentalität sind. Ein verstärktes Abdecken eines bestimmten Unternehmens in sozialen Netzwerken führte zu einer erhöhten Handelsaktivität in diesem Titel sowie zu steigenden Kursen (Yoon & Oh, 2022, S. 16). Die Studie von Rakowski et al. (2021), welche die Auswirkungen von Aufmerksamkeit, die durch die Social-Media-Plattform Twitter erzeugt wird, auf die Finanzmärkte untersuchte, konnte nachweisen, dass Twitter-Aktivitäten eine bedeutsame Quelle für Investorenaufmerksamkeit sind. Die Aktivitäten weisen eine statistisch und wirtschaftlich signifikante Verbindung mit dem Handelsvolumen und den Vermögenspreisen aus (Rakowski et al., 2021, S. 33). Diesen Einfluss erklärten

Rakowski et al (2021) damit, dass Twitter die Investorenaufmerksamkeit erhöht und die Verbreitung von Informationen zwischen den Investoren auf eine Weise erleichtert, die von traditionellen Messgrößen nicht erfasst wird. Da sich der Effekt für weniger sichtbare Aktien am grössten herausstellte, wird in der Studie davon ausgegangen, dass diese mit Twitter verbundenen Muster dazu beitragen, die in Aktien enthaltenen informationsbezogenen Asymmetrien zu beseitigen. Zudem stellten sie fest, dass Privatanleger den Twitter-Effekt antreiben, was konsistent mit der Forschung von Yoon & Oh (2022) ist (Rakowski et al., 2021, S. 33).

Ein weiterer Grund für die verstärkte Aktivität von Kleinanlegern an den Finanzmärkten kann das verstärkte Auftreten des Hindsight-Bias sein. Der Hindsight-Bias beschreibt die Tendenz von Personen, ein unerwartetes Ereignis im Rückblick als leicht vorhersehbar zu qualifizieren. Der verstärkte Rückschaufehler bei den Investoren korrelierte positiv mit der Tradingaktivität während der Coronapandemie. Dies impliziert, dass die Investoren annahmen, in der Lage zu sein, den Crash an den Finanzmärkten im Frühjahr 2020 vorzusehen. Diese Bestätigung führte im Nachgang zu einer erhöhten Handelsaktivität (S. Talwar et al., 2021, S. 2156).

Der dritte Bias, welcher zu einer Veränderung des Handelsmusters von Investoren führte, ist eine verzerrte Selbstzuschreibung und Selbstüberschätzung. Der starke Wertanstieg der Investments während der Coronapandemie führte dazu, dass die Investoren meinten, der Wertzuwachs sei hauptsächlich ihren Fähigkeiten zuzuschreiben und sich für fähig hielten, den Markt zu schlagen (S. Talwar et al., 2021, S. 2155).

2.3. Einfluss von Retailinvestoren auf die Liquidität und Preissetzung

Abudy (2020) konnte nachweisen, dass Retailinvestoren, welche einzeln ein niedriges Handelsvolumen aufweisen, in ihrer Gesamtheit zur Liquidität an den Finanzmärkten beitragen (Abudy, 2020, S. 10). Diese Resultate decken sich mit der Untersuchung von Barrot et al (2016). Gemäss dieser Untersuchung erhöht der Handel von Individualinvestoren die Liquidität in bestimmten Märkten vor allem dann, wenn die Unsicherheit sehr gross ist. So trugen Kleininvestoren während der Finanzkrise von 2008 – 2009 signifikant zur Liquidität an den Märkten bei. Die Retailinvestoren treten somit als Gruppe an den Finanzmärkten auf und haben dementsprechend einen Einfluss auf die Gesamtliquidität des Marktes, jedoch profitieren sie nicht monetär von diesem Service (Barrot et al., 2016, S. 165). Die Studie konnte zwei Ursachen identifizieren, weshalb Retailinvestoren keinen Gewinn aus der Liquiditätsversorgung ziehen können: Einerseits werden ihre Aufträge

durch erfahrene institutionelle Marktteilnehmer während der Handelsabwicklung ausgenutzt. Andererseits kehren die Kleininvestoren ihre Handelsaufträge nicht schnell genug wieder um. Damit werden die Erträge aus der Zurverfügungstellung der Liquidität nicht abgeholt (Barrot et al., 2016, S. 165).

Ebenfalls zeigte Abudy (2020), dass Privatinvestoren, welche über ein grösseres und besser diversifiziertes Portfolio verfügen, überwiegend mit Titel handeln, bei welchen die Liquidität bereits hoch ist. Nehmen jedoch Personen am Handel teil, welche besonders oft handeln, platzieren diese ihre Aufträge auch in Wertschriften, welche eine niedrigere Liquidität haben. Zudem zeigte sich, dass Retailinvestoren bei Kaufentscheidungen eine höhere Liquidität verlangen als bei Verkaufsentscheidungen (Abudy, 2020, S. 10).

Obwohl angenommen wird, dass die Aktivität von Retailinvestoren an den Finanzmärkten hauptsächlich von der vergangenen Performance der Märkte abhängt, kamen Barrot et al (2016) bei ihrer Untersuchung der Handelsaktivität von Retailinvestoren in Frankreich während der Finanzkrise zu einem differenzierteren Bild. So veräusserten zwar ein signifikanter Teil der Privatinvestoren ihre Anlagefonds, da das Vertrauen in die Finanzmärkte schwand. Jedoch nahm die Aktivität von aktiven Retailinvestoren im selben Zeitraum signifikant zu. Diese aktiven Investoren vergrösserten ihr Exposure gegenüber Aktien und gewährleisteten so eine erhöhte Liquidität an den Finanzmärkten. Aus diesen Erkenntnissen schliessen Barrot et al. (2016) dass, wenn die Marktliquidität aufgrund erhöhter Nachfrage von institutionellen Investoren oder Restriktionen von professionellen Market Makers abnimmt, diese von Retailinvestoren gewährleistet wird. In Konsequenz sind Retailinvestoren als ein integralen Bestandteil effizienter Finanzmärkte anzusehen (Barrot et al., 2016, S. 166).

Pagano et al. (2021) untersuchten ebenfalls den Einfluss der Retailinvestoren auf die Marktqualität. Ganzheitlich betrachtet zeigten ihre Resultate, dass sich die Marktqualität signifikant verbessert, wenn Retailinvestoren verstärkt am Markt teilnehmen. Dieser Effekt zeigt sich vor allem in einem stabilen Marktumfeld. In instabilen Zeiten wie z. B. während der Coronapandemie dagegen blieb der Effekt aus. Anders als Barrot et al. (2016) schloss Pagano et al (2021) , dass während einem Stressszenario der Einfluss auf die Marktqualität durch Retailinvestoren sogar negativ sein kann, und diese zu höheren Marktverwerfungen beitragen können (Pagano et al., 2021, S. 9). Ebenfalls zeigten die Resultate von Pagano et al (2021), dass Robinhood-Investoren schnell auf Neuigkeiten reagieren und sowohl Momentum- wie auch konträre Strategien anwenden.

Des Weiteren zeigte sich, dass Retailinvestoren die Finanzmärkte signifikant beeinflussen können, besonders während Marktkrisen (Pagano et al., 2021, S. 9). Betrachtet man die Preissetzung von Aktien und den Einfluss von Retailinvestoren auf die Preise, konnte die Studie von Berger et al. (2022) die Hypothese, dass Privatinvestoren den Preis von Vermögenswerten positiv beeinflussen, bestätigen. Es zeigte sich, dass der Einfluss von Retailinvestoren auf die Preisgestaltung von Vermögenswerten grundsätzlich assymetrisch ist. Da Privatinvestoren keine oder nur sehr beschränkte Möglichkeiten haben, Leerverkäufe zu tätigen, zeigte sich, dass vorherrschender Pessimismus der Privatinvestoren keinen grossen preiszersetzenden Einfluss auf den Preis einer Aktie oder Obligation ausüben kann. Gleichzeitig kann ein hoher Optimismus unter den Privatinvestoren zu einem verstärkten Preisschub bei Vermögenswerten führen, da es zu einer verstärkten Nachfrage nach den Titel kommen kann (Berger, 2022, S. 76).

2.4. Einflüsse auf das Tradingvolumen von Retailinvestoren

Eine Studie der Bank of New York Mellon (2022) führt den beobachteten Anstieg der Retailvolumen während und nach der Coronapandemie auf drei Hauptfaktoren zurück. Dies sind die Pandemie, technologische Entwicklungen im Handelsbereich und die Inflation (The Bank of New York Mellon Corporation, 2022, S. 4). So fand die Studie Hinweise, dass die durch die Pandemie ausgelöste Volatilität an den Finanzmärkten sich auf die Investorentätigkeit signifikant ausgewirkt hat. Die Unsicherheiten bezüglich der Wirtschaftsentwicklung, den Unternehmensergebnissen und den Supply-Chain-Unterbrüchen und den damit verbundenen erhöhten Preisniveaus führten zu einer starken Korrektur der Finanzmärkte. Diese Korrektur wurde gemäss der Studie von vielen Retailinvestoren als Kaufmöglichkeit wahrgenommen. Insbesondere neueintretende Investoren, welche noch wenig oder keine Erfahrung an den Finanzmärkten aufwiesen (vgl. Tabelle 1), wurden vermehrt aktiv.

Auch die Aktivität in den anderen Investorengruppen nahm während und nach der Coronapandemie signifikant zu (The Bank of New York Mellon Corporation, 2022, S. 5). Xu et al. (2022) konnte aufzeigen, dass individuelle Anleger während des Marktabschwungs als Nettokäufer agierten. Ausserdem kauften männliche und jüngere Anleger aggressiver als andere Anleger. Die höhere Kaufintensität dieser Anleger könnte gemäss Xu et al. (2022) darauf zurückzuführen sein, dass sie eine höhere Risikotoleranz haben und sich nicht bewusst sind, wenn sich Risiken stark verändern. Ausserdem nutzten die Anleger die vergangenen Renditen des Leitindex als Indikator für eine Trendumkehr

am Finanzmarkt und investierten im Folgemonat in Wertpapiere (Xu et al., 2022, S. 20). So stieg die durchschnittliche wöchentliche Handelsaktivität während der Coronapandemie um 13.9 Prozent (Ortmann et al., 2020, S. 9). Gemäss Ortmann et al (2020) investierten die Investoren ihr Vermögen vermehrt in Aktien sowie Exchange Traded Funds, nicht aber in sichere Häfen wie Gold. Oder aber sie investierten in sehr riskante Anlagen wie CFD-Optionen.

Das Wachstum an Investoren und Handelsaktivität führte Ortmann et al. (2020) im Gegensatz zu Xu et. al. (2022) hauptsächlich auf ältere männliche Investoren zurück (Ortmann et al., 2020, S. 9). Generell scheinen erstmalige Erfahrungen, welche Investoren an den Finanzmärkten machen, eine kausale und signifikante Auswirkung auf zukünftige Risikobereitschaft und Investitionen zu haben (Andersen et al., 2019, S. 116). Chen (2012) zeigte, wie die Volumina des Retailhandels mit den Renditen des Leitindex S&P 500 korrelierten. Ungeachtet dessen, ob ein Bärenmarkt oder ein Bullenmarkt vorliegt, sind die Renditen und die Handelsvolumen der Retailinvestoren positiv korreliert. Chen (2012) erklärt dies damit, dass in einem Bärenmarkt bereits beträchtliche Verluste bei den Investoren angefallen sind, und diese deshalb auch nicht am Aktienhandel teilnehmen. In einem Bullenmarkt hingegen kann das Selbstvertrauen der Retailinvestoren mit langanhaltendem Erfolg auf dem Markt wachsen, was zu einer starken positiven Rendite-Volumen-Korrelation führt (Chen, 2012, S. 1788).

2.5. Forschungslücke

Umfangreich analysiert sind bereits die individuellen Faktoren, welche sich auf die Handelsaktivität von Retailinvestoren auswirken respektive von welchen Kriterien die Entscheidungsfindung eines Investors abhängt, in welcher Weise er Kapital an den Finanzmärkten investiert. Zudem existiert bereits umfangreiche Forschungsliteratur zu den Auswirkungen auf die Finanzmärkte, welche durch den Handel der Privatinvestoren ausgelöst wird, und was sie dazu verleitet, ihre Handelsaktivitäten in bestimmte Titel zu verschieben. Die vorgenommenen Untersuchungen beschränken sich jedoch auf die Analyse der Verschiebung der Aufmerksamkeit und der Aktivitäten hin zu einzelnen Titeln am Finanzmarkt. Auch ein Zusammenhang zwischen der vergangenen Performance und der Aktivität von Retailinvestoren konnte bereits hergestellt werden, jedoch nicht, welche makroökonomischen, endogenen Einflüsse dazu führen könnten, dass vermehrt Retailinvestoren am Gesamtmarkt teilnehmen.

Im Rahmen dieser Arbeit wird die wissenschaftliche Forschung anhand einer Analyse der Einflüsse von makroökonomischen Faktoren auf das Handelsvolumen von Retailinvestoren ergänzt. Dieser Aspekt wurde von der Forschung bisher nicht detailliert untersucht. Zudem wurde bis jetzt noch keine Untersuchungen betreffend Orderroutingenerträgen von Online-Brokern und den damit verbundenen Einflussgrößen durchgeführt. Hier wird ergänzend überprüft, ob die makroökonomischen Einflussgrößen dieselben Zusammenhänge zu den Orderroutingenerträgen aufweisen wie zum Handelsvolumen der Retailinvestoren.

3. Definitionen

3.1. Handelssystem zur Abwicklung von Retailaufträgen in den USA

In den USA hat sich der Zugang zu Börsen für Retailinvestoren in den letzten Jahren stark ausgeweitet. So können Privatpersonen via Online-Broker wie Robinhood relativ einfach ein Depot eröffnen und über dieses kostenfrei handeln. Robinhood hat insgesamt über 23 Millionen Nutzer (Robinhood, 2023). Über 77 Prozent der Transaktionserträge von total 186 Millionen USD (Robinhood, 2023) wurden durch Orderrouting der Handelsaufträge von Retailkunden generiert. Bei diesen Erträgen handelt es sich um Zahlungen von Market Makern (zum Beispiel Citadel), welche unter der Bedingung bezahlt werden, dass der Onlinebroker die Handelsaufträge direkt an den Market Maker weiterleitet, anstatt diese an den öffentlichen Börsen zu platzieren (SEC, 2023, S. 3).

3.1.1. Marktstruktur für Tradeausführung in den USA

Generell haben in den USA die Online-Broker die Pflicht zur bestmöglichen Ausführung eines Auftrags ihrer Kunden. Die Finanzaufsichtsbehörde in den USA (FINRA) verlangt von den Online-Brokern, dass diese eine angemessene Sorgfalt bei der Ausführung der Kundenaufträge walten lassen und für ihre Kunden den besten Preis an den verschiedenen Finanzmärkten erzielen. Allerdings ist es in den USA sehr komplex, den besten Preis für ein Wertpapier zu ermitteln, denn der Handel findet auf verschiedenen Finanzplätzen und Plattformen statt. Ebenfalls erschwert die physische Distanz die optimale Marktsuche, da sich die Preise sehr schnell ändern können (Ernst & Spatt, 2022, S. 9). Um zu messen, ob ein Broker fähig ist, für die Kunden den besten Preis zu erzielen, werden die realisierten Per-Share-Preisverbesserungen berechnet. Diese Berechnungen der Verbesserung setzt sich aus dem Handelspreis, dem National Best Offer (NBO) für Kaufaufträge und dem National Best Bid (NBB) für Verkaufsaufträge zusammen. Die NBO und NBB repräsentieren die besten verfügbaren Kauf- und Verkaufspreise auf dem Gesamtmarkt. Erzielt

der Onlinebroker eine positive Preisverbesserung zeigt dies, dass der Kunde die Wertpapiere zu einem besseren Preis erhalten hat, als diese zum selben Zeitpunkt am Markt verfügbar waren. Dies kann für den Kunden zu Kosteneinsparungen führen. Indem sie eine maximale Preisverbesserung anstreben, versuchen Broker, die Kosten für ihre Kunden bei der Ausführung von Aufträgen zu minimieren (Wah et al., 2018, S. 93). Jedoch können sich in der modernen Finanzwelt die Kurse während der Übertragungszeit ändern. Daher muss die Routing-Entscheidung eines Maklers nicht nur den jetzigen Markt, sondern auch die Schwankungen im aktuellen Preis sowie den möglichen Markt in wenigen Millisekunden in der Zukunft berücksichtigen. Angesichts dieser Herausforderung, das beste Börsenangebot zu finden, können Makler zu einem Handelsplatz außerhalb der Börse routen (oder den Auftrag intern abwickeln). Hier bieten sich Market-Maker wie Citadel an, welcher den Kunden der Onlinebroker reduzierte Spreads anbietet (Ernst & Spatt, 2022, S. 9).

3.1.2. Erträge der Online Broker

Wie im Absatz 3.1.1 beschrieben, verfügen Online-Broker in den USA über die Flexibilität, die Handelsaufträge ihrer Kunden individuell an einen Marktteilnehmer zu routen. Sie haben dabei aber die Verpflichtung, den besten Preis für diesen zu erzielen. Grosse Market-Maker haben die Möglichkeit, Aufträge zu poolen und diese intern abzuwickeln. So können sie den Spread schliessen und müssen die Aufträge nicht an die Börsen weiterleiten. Um einen genügend grossen Orderflow zu haben, versuchen die Market-Maker, möglichst viele Online-Broker dazu zu bewegen, die Kundenaufträge zu ihnen zu routen. Hierfür bieten die Market-Maker den Online-Broker Zahlungen im Sub-Pennybereich an (Ernst & Spatt, 2022, S. 10). Diese Erträge aus dem Orderrouting stellen dann, wie oben beschrieben, die Hauptertragsquellen für Online-Broker wie Robinhood dar.

3.2. Herleitung des aggregierten Marktvolumens von Retailinvestoren

Das aggregierte Marktvolumen von Retailinvestoren wird determiniert durch die individuelle Partizipationsrate, Portfoliorebalancing sowie Verhaltensfaktoren (Alderighi, 2018, S. 152). Nimmt man einen komplett friktionslosen Handel an, müssten alle Privatpersonen sich an diesem beteiligen. Der Anteil des Vermögens, welcher in Aktien und andere Vermögenswerte investiert wird, wird in diesem Fall durch die individuelle Risikoaversion sowie die erwartete Rendite bestimmt. In der Praxis nehmen jedoch nur wenige Haushalte am Aktienhandel teil und investieren auch nur einen sehr kleinen Teil ihres Vermögens in gehandelte Wertpapiere. Aus dem Anteil der Haushalte, welche am

Aktienhandel teilnehmen, folgt die Partizipationsrate. Das Portfoliorebalancing wird vollzogen, wenn ein Retailinvestor bereits im Aktienhandeltätig ist. Sobald sich die Anteile im Portfolio durch passive Transaktionen verschieben, wird durch aktives Handeln der Ursprungszustand wieder hergestellt (Alderighi, 2018, S. 156). Folglich bewegt sich das aggregierte Marktvolumen von Privatinvestoren entweder durch Veränderung des Verhaltens, der Partizipationsrate oder der Rebalancingaktivitäten.

Die Schwierigkeit bei der Analyse des Verhaltens von Investoren liegt in der Abgrenzung der Handelsaktivitäten. Es ist nicht einfach, die getätigten Markthandlungen institutionellen oder privaten Akteuren zuzuordnen. Viele Forschende nutzten bisher nur relativ kleine Datenmengen von Brokerfirmen, welche nicht öffentlich zugänglich sind (Boehmer et al., 2021, S. 2249). So bezogen Barrot et al. (2016) die Daten für ihre Untersuchung nur von einem französischen Broker, welcher rund 2% des Volumens des französischen Marktes ausmachte (Barrot et al., 2016, S. 148). In anderer Forschungsliteratur wird die Abgrenzung zwischen retail- und institutionellen Investoren anhand der Grösse eines getätigten Trades festgelegt. So nutzten Campell et al. (2009) die Daten des Trade and Quote Systems der NYSE, um eine mittels Regression variierende Tradegröße zu definieren und so zu erkennen, ob ein Trade von einem institutionellen oder einem Retailinvestor getätigt wurde (Campbell et al., 2009, S. 70). Aufgrund des an den Finanzmärkten durch automatisierte Handelssoftware immer häufiger auftretenden Slicings, das heisst dem Aufteilen eines grossen Auftrags in viele kleine Aufträge, ist die Identifikation des Investors mittels der vorher beschriebenen Methode nicht mehr möglich (Boehmer et al., 2021, S. 2250).

Um trotzdem die Marktaktivität von Retailinvestoren bestimmen zu können, nutzten Boehmer et al. (2021) den Umstand aus, dass in den USA die meisten Handelsaufträge von Klein-, respektive Retailinvestoren nicht über die öffentlichen Handelsplätze getätigt werden, sondern ausserbörslich durch Broker ausgeführt werden (Boehmer et al., 2021, S. 2251). Diese Broker wickeln die Aufträge der Kunden mit Wertpapieren aus dem eigenen Bestand ab. Dies führt dazu, dass diese Transaktionen nicht in das Handelsvolumen der Börsen einfließen. Sie müssen aber nach der Abwicklung durch den Broker an die Financial Industry Regulatory Authority (FINRA) gemeldet werden. Diese Meldung wird dann in sogenannten Trade Reporting Facilitys (TRF) gespeichert. Alle aktiven sowie ehemaligen TRF weisen einen Exchange Code «D» aus. Anhand dieser Informationen können Retail-Aufträge identifiziert werden (Boehmer et al., 2021, S. 2251). Für den

Zeitraum der Untersuchung (2009 – 2022) betrifft dies die Alternative Display Facility (ADF), NYSE TRF und NASDAQ TRF.

4. Empirischer Teil

4.1. Hypothesen

Basierend auf der beschriebenen Forschungslücke und der in deren Zusammenhang entwickelten Forschungsfrage «Wie können Schwankungen des Tradingvolumens von Privatpersonen in den USA erklärt werden?» werden jeweils zwei Nullhypothesen (H0) sowie zwei Alternativhypothesen (H1) getestet. Die erste Hypothese befasst sich mit den Einflüssen von Makroparametern auf das Handelsvolumen von Retailinvestoren, während sich die zweite Hypothese damit befasst, ob die Makroparameter dieselben Zusammenhänge mit die Orderrouteringserträgen zeigen wie das Handelsvolumen der Retailinvestoren.

Hypothese 1:

H0: Das Tradingvolumen der Retailinvestoren hängt nicht signifikant von den gewählten Modellparametern ab.

H1: Mittels den gewählten Modellparametern lassen sich Schwankungen des Handelsvolumens von Retailinvestoren erklären.

Hypothese 2:

H0: Die Orderrouteringserträge der Online-Broker hängen nicht signifikant von den gewählten Modellparametern ab.

H1: Mittels den gewählten Modellparametern lassen sich Schwankungen der Orderrouteringserträge von Onlinebroker erklären und die festgestellten Einflüsse sind deckungsgleich mit den Einflüssen auf das Handelsvolumen der Retailinvestoren

Die Hypothese 1 wird mittels einer OLS-Regression sowie einem SARIMAX-Modell untersucht, während zur Überprüfung der Hypothese 2 eine Panel-OLS-Regression angewandt wird.

4.2. Datenanalyse

4.2.1. Datenerhebung

Die Grundlage zur Untersuchung der Fragestellung der Sensitivität der Handelsaktivität von Privatinvestoren auf exogene Ursachen bilden die täglichen Umsätze, welche in den Off-Exchange Börsen der USA zwischen den Jahren 2009 – 2022 stattgefunden haben. Hierfür wurden tägliche Daten der Chicago Board Options Exchange (CBOE) genutzt. Um den Datensatz zu erstellen, werden im relevanten Zeitraum alle Umsätze von Off-Exchange Börsen, welche einen Börsencode «D» besitzen, aggregiert. Über diese Off-Exchange Handelsplätze müssen alle Trades, die nicht über die üblichen Börsenplätze, d.h. welche nicht öffentlich abgewickelt werden, gemeldet werden (siehe Abschnitt 3.2). Die gehandelten Umsätze inkludieren alle im National Market System (NMS) enthaltenen Wertpapiere. Das NMS bestimmt, wie die US-amerikanischen Börsen für die zugrundeliegenden Wertpapiere Handelsaktivitäten reporten und ausführen müssen (*FINRA*, 2023).

Danach werden die täglichen Handelsumsätze auf Monatsebene aggregiert und mit makroökonomischen Zeitreihen ergänzt. Diese wurden mittels Refinitiv, dem Investor Sentiment Index der American Association of Individual Investors (AII) oder dem Datenportal des St Louis FED (FRED) erhoben.

Die bereinigten Gesamtdaten umfassen schliesslich die aggregierten Monatsumsätze der drei Off-Exchange Börsen, den Sentiment Index von Retailinvestoren, Monatsend-, Monatstiefst- und Monatshöchstdaten des Leitindex S&P500, Lagged Datenreihen für ein, zwei und drei Monate des S&P500, und diverse andere makroökonomische Zeitreihen, welche sich auf die Inflation, die Arbeitslosigkeit sowie die amerikanische Wirtschaftsleistung beziehen (s. Abschnitt 4.6 zu den Modellparametern). Des Weiteren wurden die Returns auf Auffälligkeiten und weitere Datenlücken geprüft und allfällig aus dem Datenset entfernt. Insgesamt stehen für 169 Monate vollständige Daten für die Handelsaktivität sowie die makroökonomischen Datenreihen zur Verfügung.

Für die Untersuchung der Erträge aus den Routinggeschäft der vier grössten US-Online Broker (TD-Ameritrade, Robinhood, E*Trade sowie Charles Schwab) wurden die quartalsweisen SEC 606a Reports (Held NMS Socks and Options Order Routing Public Report) der Unternehmen für den Zeitraum von 2020 – 2022 von den entsprechenden Firmenplattformen heruntergeladen. Danach wurden die Positionen der Zahlungen, welche die Unternehmen von den verschiedenen Brokerunternehmen für S&P500 Aktien,

anderen Wertpapieren sowie Optionen in den jeweiligen Monaten erhalten haben, summiert. Zusätzlich wurden die Erträge auf Auffälligkeiten und weitere Datenlücken geprüft und allfällig aus dem Datenset entfernt. Das Datenset der Erträge des Routinggeschäfts wurde anschliessend mit denselben makroökonomischen Datenreihen ergänzt.

4.2.2. Analyse TRF Off Exchange Umsätze

Wie die Tabelle 2 zeigt, umfasst das für die Untersuchung der Handelsaktivität von Retailinvestoren relevante Datenset Transaktionsdaten der drei Off Exchange Börsen Display Facility (ADF), NYSE TRF und NASDAQ TRF. Die im Datenset gemeldeten Transaktionen werden anhand drei verschiedener Tapes aufgeschlüsselt. Dabei betreffen Tape A Trades Titel, welche an der NYSE gelistet sind, Tape B Trades Titel, welche an der Amex oder regionalen Börsen gelistet sind und Tape C Titel, welche an der NASDAQ gelistet sind. Auffallend ist, dass ein Grossteil der Titel, welche ausserbörslich gehandelt werden, über die TRF der NASDAQ gemeldet werden. Diese Trade Report Facility macht circa 88 Prozent des Handelsvolumens aus, 11.79 Prozent werden über die NYSE gemeldet und nur 0.22 Prozent über die Alternative Display Facility. Die Art der Titel, die gehandelt werden, ist relativ ausgeglichen, so weisen Titel, welche an der NYSE kotiert sind, mit 40.68 Prozent die höchste Häufigkeit auf, während Titel, welche an der Amex kotiert sind, mit 23.31 Prozent am wenigsten gehandelt werden. Titel der NASDAQ werden mit 36.01 Prozent ebenfalls häufig gehandelt.

| Off-Exchange Börse | Tape A | Tape B | Tape C | Total | Durchschn. Handelsvolumen pro Trade (USD) |
|--------------------|--------|--------|--------|--------|---|
| ADF | 0.27% | 0.18% | 0.17% | 0.22% | 6948.84 |
| NYSE TRF | 10.50% | 8.51% | 15.38% | 11.79% | 9903.38 |
| NASDAQ TRF | 89.23% | 91.31% | 84.45% | 87.99% | 12015.47 |
| Total | 40.68% | 23.31% | 36.01% | 100 | 11702.70 |

Tabelle 2: Aufteilung Handelsvolumen TRF nach Börse und Tape (eigene Darstellung)

Auffallend ist, dass das durchschnittliche Handelsvolumen pro Trade bei der ADF am niedrigsten ausfällt, ein Trade an dieser Off-Exchange Börse hat ein durchschnittliches Volumen von 6948 USD über den gesamten Betrachtungszeitraum von 2009 bis 2022.

An der NYSE TRF beträgt das durchschnittliche Volumen pro Trade 9903 USD und am höchsten ist das Volumen bei der NASDAQ TRF mit 12015 USD. Im Schnitt über alle Börsen hinweg berechnet beträgt das durchschnittliche Volumen 11702 USD. Die verzeichneten durchschnittlichen Volumina an den drei Börsen lassen darauf schliessen, dass es sich bei diesen Aufträgen grossmehrheitlich um Aufträge von Retailinvestoren handelt. Betrachtet man die Entwicklung der aggregierten Handelsvolumina an den Off-Exchange Börsen über den betreffenden Zeitraum (Abbildung 2), kann man erkennen, dass das Handelsvolumen von Retailinvestoren sich über den gesamten Zeitraum in etwa vervierfacht hat.

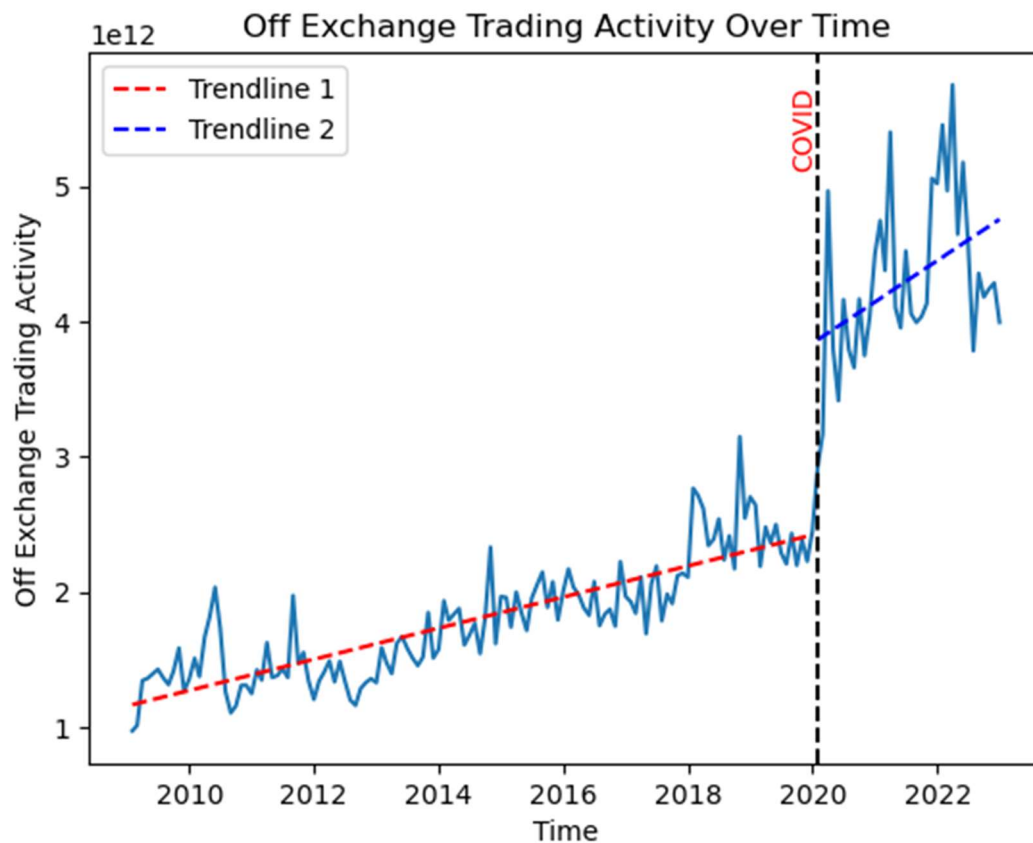


Abbildung 2: Handelsaktivität Retailinvestoren in den USA (eigene Darstellung)

Das Datenset lässt sich in zwei Perioden mit verschiedenen Trends einteilen. So nahm das Handelsvolumen zwischen 2009 und 2019 von circa einer Billion USD pro Monat auf über zwei Billionen USD pro Monat zu. Zu Beginn des Jahres 2020 lässt sich ein Bruch im Volumen der über die Off-Exchange Börsen gehandelten Volumen erkennen. Zum Zeitpunkt der Coronapandemie nahm dieses Sprunghaft zu und schwankte danach zwischen dreieinhalb und fünf Billionen USD pro Monat. Dieser Bruch lässt erkennen,

dass während und nach der Corona Pandemie mehr Retailinvestoren an den Finanzmärkten investierten oder die bereits bestehenden Investoren ihr Exposure signifikant erhöhten. Interessant ist ebenfalls, dass sich die Steigung des Trends nach dem Bruch der Coronapandemie veränderte. Sie zeigte sich im Zeitraum von 2020 bis 2022 steiler als in der Periode von 2010 bis 2020. Für die Untersuchung der Verteilung der monatlichen Veränderung der Off-Exchange Volumina wurde, wie Abbildung 3 zeigt, das Histogramm sowie die Dichtefunktion grafisch dargestellt. Der daraufhin durchgeführte Shapiro-Wilk-Test deutet darauf hin, dass die Returns normalverteilt sind. Während positive Schiefe von 0.1188 gegen eine Normalverteilung spricht, deutet die Kurtosis von 0.529 auf diese hin. Somit die Datenimplizieren eine leicht rechtsschiefe und platykurtische Verteilung.

Histogramm & density function Change of total notional traded by private individuals

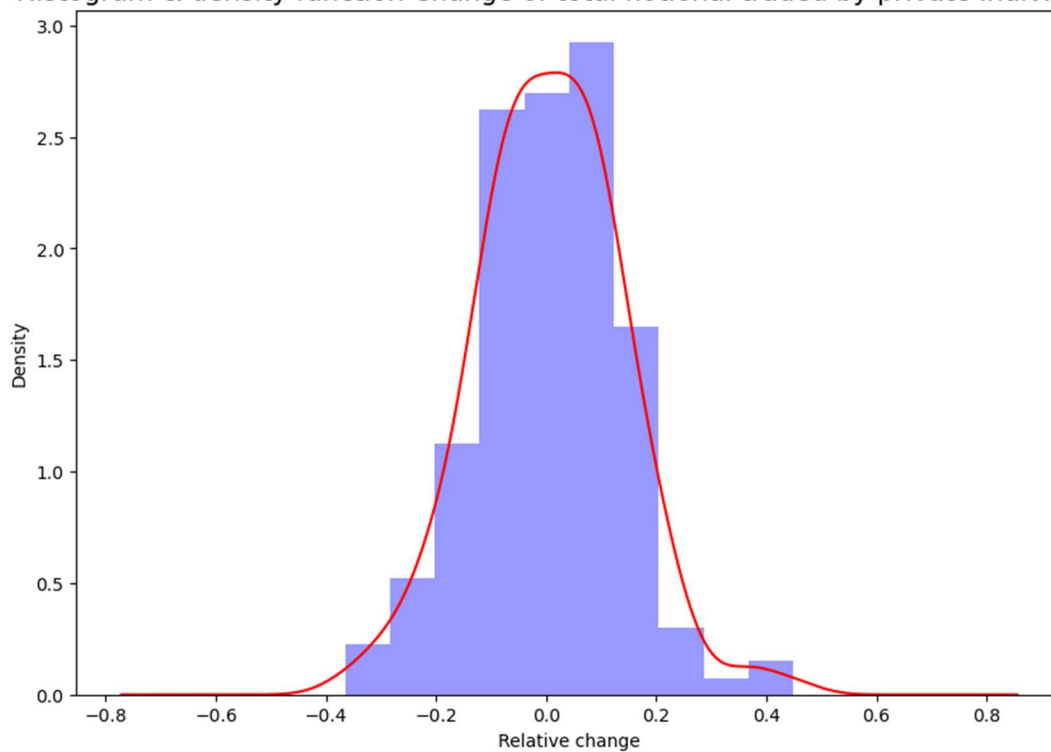


Abbildung 3: Dichte und Verteilung der Umsätze an Off-Exchange Börsen (eigene Darstellung)

4.2.3. Analyse Routing Erträge Online Broker

Das zu untersuchende Datenset enthält die Erträge, welche die Online Broker E*Trade, Charles Schwab, TD Ameritrade sowie Robinhood aus dem Orderrouting an die Market Maker im Zeitraum von 2020 bis 2022 erzielten.

| (in Mio. USD) | Aggregiert | TD Ameritrade | Robinhood | E*Trade | Charles Schwab |
|--------------------------------|------------|---------------|-----------|---------|----------------|
| <i>Durchschn. Ertrag/Monat</i> | 223.80 | 103.39 | 61.71 | 34.16 | 24.54 |
| <i>Maximaler Ertrag/Monat</i> | 342.70 | 148.50 | 120.8 | 47.74 | 32.27 |
| <i>Minimaler Ertrag/Monat</i> | 109.88 | 52.47 | 19.37 | 23.50 | 14.45 |
| <i>25%-Quantil des Ertrags</i> | 197.20 | 94.26 | 47.14 | 30.71 | 21.07 |
| <i>50%-Quantil des Ertrags</i> | 214.61 | 100.28 | 58.01 | 34.24 | 25.00 |
| <i>75%-Quantil des Ertrags</i> | 244.13 | 119.94 | 71.04 | 36.32 | 27.99 |
| <i>Standardabweichung</i> | 48.09 | 19.96 | 21.76 | 5.55 | 4.40 |

Tabelle 3: Erträge aus Orderrouting (eigene Darstellung)(The Bank of New York Mellon Corporation, 2022, S. 3)

Wie Tabelle 3 zeigt, weist TD Ameritrade das höchste Ertragsniveau mit einem Durchschnitt von 103.39 Mio. USD pro Monat aus dem Orderrouting aus. Die beiden Unternehmen, welche die grössten Erträge aus dem Orderrouting TD Ameritrade sowie Robinhood (mit einem Durchschnitt von 61.71 Mio. USD) erwirtschaften, weisen ebenfalls die inkonstantesten Erträge aus. So schwanken die Erträge bei TD Ameritrade zwischen 52.47 Mio. USD und 148.50 Mio. USD, also um circa den Faktor drei und bei Robinhood zwischen 19.37 Mio. USD und 120.80 Mio. USD, also um den Faktor 6.

Die beiden anderen Unternehmen, welche ein Orderrouteringsertrag auf einem vergleichsweise tieferen Niveau ausweisen, Charles Schwab mit einem Durchschnitt von 24.54 Mio. USD und E*Trade mit einem Durchschnitt von 34.16 Mio. USD, können diese Erträge viel stabiler halten als die anderen beiden Unternehmen. Daraus lässt sich schliessen, dass Robinhood sowie TD Ameritrade die Orderrouteringserträge zu optimieren versuchen, während Charles Schwab und E*Trade ihren Ertrag mittels anderer Einkommensquellen generieren möchten.

4.2.4. Deskriptive Statistik ausgewählter exogener Modellparameter

Um die Entwicklung der jeweiligen exogenen Modellparameter besser zu verstehen und nachvollziehen zu können, ist es sinnvoll, sich einen kurzen Überblick über die jeweiligen Entwicklungen im Untersuchungszeitraum von 2009 bis 2022 zu verschaffen.

| | RDI <i>(Milliarden)</i> | S&P500 | S&P500 Low | S&P500 High | Sentiment- BullBear | 2YR UST |
|-------------|-----------------------------------|-------------------|---------------------------|----------------------------|--------------------------------|--------------------|
| mean | 13633.68 | 2301.13 | 2222.03 | 2358.66 | -0.064% | 1.027% |
| min | 11538.10 | 735.09 | 676.53 | 832.86 | -32.64% | 0.11% |
| 25% | 12360.93 | 1404.41 | 1350.83 | 1417.75 | -9.15% | 0.325% |
| 50% | 13518.30 | 2082.96 | 2040.14 | 2116.44 | -1.45% | 0.67% |
| 75% | 14893.93 | 2917.10 | 2815.02 | 2947.76 | 9.60% | 1.332% |
| max | 19213.90 | 4766.18 | 4567.00 | 4796.56 | 41.6% | 4.51% |

Tabelle 4: Übersicht Modellparameter (Teil 1)(eigene Darstellung)

Erwähnenswert ist, dass das verfügbare Einkommen der US-Amerikaner im beobachteten Zeitraum generell gestiegen ist, so liegt die durchschnittliche monatliche Zunahme des verfügbaren Einkommens bei 21 Mrd USD. Der Höchstwert von über 19213 Mrd USD wurde während der Coronapandemie im März 2021 erreicht (FRED, 2023). Im Zeitraum von 2009-2022 hat sich der Leitindex ebenfalls positiv entwickelt, so nahm dieser ab 2009 stetig zu und erreichte seinen Höchststand von 4796.56 Punkten im Januar 2022.

Interessant ist, dass die Monatstiefstkurse stärker schwanken als die Monatshöchstkurse. Dementsprechend betragen die Werte der monatlichen prozentualen Veränderung des ersten und dritten Quartals bei den Monatstiefstwerten -1.08 Prozent und 3.37 Prozent, während diese bei den Monatshöchstwerten -0.22 Prozent und 2.63 Prozent betragen (Refinitiv, o. J.). Dies lässt darauf schliessen, dass neue Höchstwerte langsamer erreicht werden als durch mögliche Krisen ausgelöste Tiefstwerte.

Die Gefühlslage («sentiment index») der Investoren ist über den gesamten Zeitraum gesehen gesunken. Dies ist interessant, da die Börsenkurse in der gleichen Zeit stark angestiegen sind. Die durchschnittliche monatliche Veränderung sowie der Median sind negativ mit -0.064 Prozent und -1.45 Prozent (AAII, 2023). Die Renditen auf zweijährige US-Staatsanleihen haben im Schnitt 1.027% betragen, wobei das Minimum von 0.11 Prozent im Juli 2020 im Zuge der Unsicherheiten der Coronapandemie erreicht wurde. Das

Maximum von 4.51 Prozent wurde im Oktober 2022 erreicht und ist auf die inflationsbedingten Zinserhöhungen der US-Notenbank (FED) zurückzuführen (FRED, 2023).

| | Geldm. M3 <i>(Billionen)</i> | Delta GDP | Delta PersCons | Home Sales <i>(Tausend)</i> | Delta CPI YoY | Unemp. Rate |
|-------------|--|------------------|-----------------------|---------------------------------------|----------------------|--------------------|
| mean | 13.3366 | 2.00% | 0.37% | 532.57 | 0.04% | 6.33% |
| min | 8.3031 | -77.00% | -12.50% | 270 | -1.20% | 3.50% |
| 25% | 10.1039 | 0.65% | 0.20% | 385.75 | -0.20% | 4.20% |
| 50% | 12.5133 | 2.26% | 0.30% | 525 | 0.00% | 5.80% |
| 75% | 14.8591 | 3.78% | 0.60% | 636 | 0.25% | 8.23% |
| max | 21.7032 | 45.99% | 8.70% | 1029 | 2.00% | 14.70% |

Tabelle 5: Übersicht Modellparameter (Teil 2) (eigene Darstellung)

Die Geldmenge M3 ist im Zeitraum von 2009 – 2022 konstant angestiegen. Das Minimum von 8.3031 Billionen USD fiel zu Beginn der Datenreihe im Januar 2009 an. Das Maximum von 21.70 Billionen USD wurde im Juli 2022 erreicht. Spannend ist, dass die Geldmenge während der Coronakrise massiv ausgeweitet wurde. Die Geldmenge M3 belief sich im Februar 2020 auf 15.45 Billionen USD. Bereits im Juni 2020 betrug sie 18.1 Billionen USD, dies ist eine Zunahme von über 17 Prozent in einem Zeitfenster von nur vier Monaten.

Nachdem das Maximum der M3 Geldmenge im Juli 2022 erreicht wurde, begann diese leicht zu sinken auf ein Niveau von 21.35 Billionen USD im Dezember 2022. Dies ist auf die geldpolitische Straffung des FED zurückzuführen (FRED, 2023). Die monatliche Veränderung des Bruttoinlandproduktes (GDP) im Vergleich zum Vormonat war positiv mit einem Durchschnitt von 2 Prozent und einem Median von 2.26 Prozent. Der grösste Rückgang von -77 Prozent des Vorjahresmonates ist auf die verschiedenen Lockdowns im April 2020 zurückzuführen. Die Zunahme von +45.99 Prozent im Juli 2020 dagegen auf Lockerung ebendieser (Refinitiv, o. J).

Im Durchschnitt wurden in den USA 532.57 tausend neue Häuser pro Monat verkauft. Der Tiefstwert von 270 tausend Häuser wurde im Nachgang an die US-Finanzkrise im Februar 2011 erreicht, während der Höchststand von 1.029 Millionen während der Coronapandemie im August 2020 verkauft wurde. Dies ist möglicherweise auf die

Homeoffice-Beschränkungen und eine generell erhöhte Nachfrage nach Wohnraum in dieser Zeit zurückzuführen (FRED, 2023).

Die Veränderung des Konsumpreisindex (CPI) blieb während dem Beobachtungszeitraum weitgehend konstant. Der Median beträgt 0 Prozent und der Mittelwert 0.04 Prozent. Dies bedeutet, dass es der US-Notenbank während dem Beobachtungszeitraum meistens gelungen ist, Preisstabilität zu gewährleisten, obwohl der CPI ab dem März 2021 von 2.6 Prozent auf 9.1 Prozent im Juni 2022 angestiegen ist (FRED, 2023).

Die Arbeitslosenrate betrug im Mittel 6.3 Prozent und erreichte ihren Tiefstwert im September 2019 mit 3.5 Prozent, was nahezu einer Vollbeschäftigung entspricht. Der Höchstwert von 14.7 Prozent wurde während den Lockdowns im April 2020 verzeichnet. Die Arbeitslosenrate sank im Nachgang aber rasch wieder und erreichte einen Wert von 3.6 Prozent im April 2022 (FRED, 2023).

Zusammenfassend lässt sich festhalten, dass die Indikatoren während eines Grossteils des Betrachtungszeitraums stabil geblieben sind und sich die US-Wirtschaft zwischen 2009 und dem Frühjahr 2022 sehr positiv entwickelt hat. Die Extremwerte des Datensatzes sind hauptsächlich auf die Coronapandemie und die damit verbundenen ökonomischen Implikationen zurückzuführen.

4.3. Modelle

Für die Analyse der Einflüsse exogener Faktoren auf die Veränderung der Handelsaktivität von Retailinvestoren wurden eine OLS-Regression sowie ein SARIMAX-Modell angewendet. Für die Analyse der Orderrouteringträge von Onlinebrokern in den USA wurde eine Panel-OLS-Regression mit fixed Effects angewendet.

4.3.1. OLS Regression

Die Ordinary Least Squares (OLS)-Regression ist eine statistische Methode zur Modellierung der Beziehung zwischen einer abhängigen Variablen und einer oder mehreren unabhängigen Variablen. Bei der OLS-Regression wird angenommen, dass die Beziehung zwischen den Variablen linear ist, das heißt, eine lineare Funktion kann verwendet werden, um die abhängige Variable basierend auf den unabhängigen Variablen vorherzusagen. Das Ziel besteht darin, die Koeffizienten dieser linearen Funktion so anzupassen, dass die Summe der quadrierten Unterschiede zwischen den tatsächlichen beobachteten Werten und den vorhergesagten Werten minimiert wird. Dies wird als Methode der kleinsten Quadrate bezeichnet (Newbold et al., 2019, S. 431). Der OLS-

Regressionschätzer berechnet die besten Schätzwerte für die Koeffizienten der linearen Funktion, indem er die Residuen, das sind die Abweichungen zwischen den tatsächlichen und den vorhergesagten Werten, quadriert und minimiert. Die OLS-Regression ermöglicht es, verschiedene statistische Informationen über die Beziehung zwischen den Variablen zu erhalten. Dazu gehören die geschätzten Koeffizienten, die die Richtung und Stärke des Einflusses der unabhängigen Variablen auf die abhängige Variable angeben (Newbold et al., 2019, S. 436).

4.3.2. SARIMAX-Modell

Das SARIMAX-Modell (Seasonal Autoregressive Integrated Moving Average) ist ein statistisches Modell zur Analyse und Vorhersage von Zeitreihendaten. Es kombiniert drei Komponenten: Die autoregressive (AR)-Komponente, die integrierte (I)-Komponente und die gleitenden Durchschnitt (MA)-Komponente (Tiao, 2000, S. 54).

Die autoregressive (AR)-Komponente basiert auf der Annahme, dass zukünftige Werte einer Zeitreihe von vergangenen Werten abhängig sind. Sie modelliert die Abhängigkeit der aktuellen Beobachtung von den vorherigen Beobachtungen. Der Grad der Abhängigkeit wird durch den sogenannten AR-Parameter bestimmt. Ein $AR(p)$ -Modell verwendet die p vorherigen Werte, um den aktuellen Wert zu erklären (Tiao, 2000, S. 60).

Die integrierte (I)-Komponente befasst sich mit der Behandlung von Trends oder nicht-stationären Mustern in den Zeitreihendaten. Nicht-stationäre Daten weisen eine Veränderung im Mittelwert oder in der Varianz über die Zeit auf. Die I-Komponente verwendet Differenzierungstechniken, um die Daten zu stationär zu machen, das heißt den Trend zu entfernen und eine konstante Varianz zu erreichen. Die Anzahl der Differenzierungen wird durch den I-Parameter angegeben (Box et al., 2008a, S. 73).

Die gleitende Durchschnitt (MA)-Komponente modelliert den Einfluss früherer Fehler oder Residuen auf die aktuelle Beobachtung. Sie glättet die Daten, indem sie den Einfluss von zufälligen Variationen reduziert. Der Grad des gleitenden Durchschnitts wird durch den MA-Parameter angegeben. Ein $MA(q)$ -Modell verwendet die q vorherigen Residuen, um den aktuellen Wert zu erklären (Box et al., 2008a, S. 73).

Die Kombination dieser drei Komponenten ergibt das $ARIMA(p, d, q)$ -Modell. Zusätzlich wird das ARIMA-Modell noch durch eine saisonale Komponente in einer Zeitreihe ergänzt. Diese bezieht sich auf wiederkehrende Muster oder Schwankungen, die sich mit

einer bestimmten Periodizität wiederholen. Die saisonale Komponente kann durch saisonale Effekte verursacht werden (Box et al., 2008b, S. 375).

Die X Komponente bezeichnet die exogenen Modellparameter, welche noch hinzugefügt werden.

Das SARIMAX-Modell kann demzufolge verwendet werden, um vergangene Daten zu analysieren, Muster zu identifizieren und zukünftige Werte zu prognostizieren.

4.3.3. Panel Regression mit fixed Effects

Die Panelregression mit Fixed Effects ist eine statistische Methode zur Analyse von Paneldaten, die aus wiederholten Beobachtungen über dieselben Einheiten oder Individuen über den betreffenden Zeitraum besteht. Diese Methode zielt darauf ab, individuelle Effekte zu kontrollieren. In Paneldaten besteht einerseits die individuelle Dimension und andererseits die zeitliche Dimension. Die Fixed-Effects-Spezifikation zielt darauf ab, individuelle Unterschiede, die über die Zeit konstant bleiben, zu erfassen und zu kontrollieren. Diese individuellen Effekte werden als feste Effekte bezeichnet, da sie als konstant für jede einzelne Einheit betrachtet werden. Der grundlegende Ansatz bei der Verwendung von Fixed Effects besteht darin, die individuellen Effekte zu subtrahieren, um die individuellen Abweichungen von der Durchschnittsbeziehung zu eliminieren. Durch diese Subtraktion werden die individuellen Effekte zu Konstanten und verschwinden aus der Regression. Der verbleibende Teil der Regression beschreibt im Anschluss die dynamische Beziehung zwischen den unabhängigen Variablen und der abhängigen Variablen, wobei individuelle Unterschiede kontrolliert werden. (Best & Wolf, 2015).

4.4. Analyse des Handelsvolumens von Retailinvestoren

Für die Anwendung der zwei Schätzungsmethoden (OLS-Regression und SARIMAX-Modell) wird das Datenset in einen Trainings- und Testsatz unterteilt. Um Overfitting zu vermeiden und die optimale Anzahl und Selektion von Modellparametern für die Prognose und Analyse des Handelsvolumens von Retailinvestoren zu finden, wird zuerst eine Lasso-Regression durchgeführt. Die Lasso-Regression reduziert die Anzahl Parameter, bis der Modeloutput ein Optimum erreicht. Darauffolgend kann die OLS-Regression auf dem Trainingssatz trainiert und dann auf dem Testsatz getestet werden. Anschliessend können die in der OLS-Regression gemessenen Einflüsse auf das Handelsvolumen von Privatpersonen analysiert und interpretiert werden. Das SARIMAX Modell wird mit denselben exogenen Variablen trainiert und getestet welche sich mittels Lasso Regression als

am effektivsten erwiesen haben. Um die optimalen SARIMAX Modellparameter zu eruieren, wird zudem ein Grid Search durchgeführt. Mittels diesem Grid-Search wird innerhalb eines vordefinierten Wertebereichs aller SARIMAX-Parameter die optimale Kombination dieser SARIMAX-Parameter definiert. Anschliessend werden die Einflüsse der Parameter analysiert und mit den in der OLS-Regression erhaltenen Resultaten verglichen.

4.4.1. Rollende Zeitfenster

Der zweite Schritt der Analyse liegt in der Anwendung von rollenden Zeitfenstern. Der gesamte Datensatz wird auf fünf verschiedene Zeitfenster getestet (2 Monate, 3 Monate, 6 Monate, 9 Monate und 12 Monate). Der Testdatensatz wird rollend für unterschiedliche Zeitfenster im Datensatz von 2009 bis 2022 eingesetzt und die Trainingsdaten entsprechend neu gewählt. Das Fenster der Trainingsdaten wird von 12 Monate bis 120 Monate festgelegt (jeweils in 12 Monatsschritte). Die beiden Schätzungsmethoden der OLS-Regression sowie des SARIMAX-Modells werden anhand von diesen rollenden Zeitfenster (jeweils eine Verschiebung des Trainingsdatensatzes um einen Monat) getestet.

| | 2 Monate | 3 Monate | 6 Monate | 9 Monate | 12 Monate |
|------------|----------|----------|----------|----------|-----------|
| 12 Monate | 150 | 149 | 146 | 143 | 140 |
| 24 Monate | 138 | 137 | 134 | 131 | 128 |
| 36 Monate | 126 | 125 | 122 | 119 | 116 |
| 48 Monate | 114 | 113 | 110 | 107 | 104 |
| 60 Monate | 102 | 101 | 98 | 95 | 92 |
| 72 Monate | 90 | 89 | 86 | 83 | 80 |
| 84 Monate | 78 | 77 | 74 | 71 | 68 |
| 96 Monate | 66 | 65 | 62 | 59 | 56 |
| 108 Monate | 54 | 53 | 50 | 47 | 44 |
| 120 Monate | 42 | 41 | 38 | 35 | 32 |
| 132 Monate | 30 | 29 | 26 | 23 | 20 |

Tabelle 6: Anzahl gemachte Predictions pro Zeitraum (eigene Darstellung)

Wie in Tabelle 6 zu sehen, werden für die OLS-Regression insgesamt 4708 Schätzungen durchgeführt, um das optimale Modell zu eruieren. Beim SARIMAX-Modell werden aufgrund einer höheren benötigten Datenmenge nur die rollende Zeitfensteranalyse für ein Trainingsdatensatz von 120 und 132 Monaten durchgeführt. Folglich wird das SARIMAX-Modell auf insgesamt 316 Zeiträume getestet.

4.5. Analyse der Orderrouterträge

Um im dritten Schritt zu untersuchen, ob die in der Analyse des Handelsvolumens der Retailinvestoren gemachten Beobachtungen auch für die Orderrouterträge der Online-Broker anwendbar sind, wird eine Panel-Regression durchgeführt. Die Panel-Regression wird angewendet, da erwartet wird, dass die im Modell verwendeten erklärenden Variablen mit den unterschiedlichen Eigenschaften einzelner Banken in Verbindung stehen. Bei der Analyse der Orderrouterträge wird der Datensatz nicht in ein Training und Testdatensatz unterteilt, da es bei der Untersuchung hauptsächlich um das Erkennen der Zusammenhänge zwischen dem bereits durchgeführten SARIMAX-Modell und OLS-Regression geht.

4.6. Modellparameter

Zur Erarbeitung des Modells wurden folgende Parameter herangezogen:

| Variable | Definition | Relevanz für das Modell |
|--|---|--|
| <i>Zu erklärende Variablen: Untersuchung Handelsvolumen Retailinvestoren</i> | | |
| Umsatz Off-Exchange Börsen | Monatlich erzielte Umsätze der drei TRF Off-Exchange Börsen der USA (ADF, NYSE TRF, NASDAQ TRF) | Dienen als Proxy für die erzielten Umsätze von Retailinvestoren in den USA. |
| <i>Zu erklärende Variablen: Untersuchung Umsätze aus Orderrouting</i> | | |
| Orderrouting Erträge | Monatlich erzielte Erträge aus Orderrouting für SP 500 Aktien, Non-S&P 500 Aktien und Optionen. | Zeigt die Erträge, welche die Onlinebroker Unternehmen mittels Orderrouting erzielt haben. |
| <i>Unabhängige Variablen</i> | | |
| Real Disposable Income | Das inflationsbereinigte, verfügbare Einkommen in den USA pro Haushalt. | Das Einkommen, welche US-Haushalte monatlich zur Verfügung haben. |
| S&P500 Months End | Schlusskurs des S&P 500 per Monatsende | Der Monatschlusskurs des S&P500, dem wichtigsten Aktienindex in den USA |

| | | |
|--|---|---|
| S&P 500 Month Low | Monatstiefstwert des S&P 500 | Der Monatstiefstwert des S&P500, dem wichtigsten Aktienindex in den USA |
| S&P 500 Month High | Monatshöchstwert des S&P 500 | Der Monatshöchstwert des S&P500, dem wichtigsten Aktienindex in den USA |
| DeltaSentimentIndexBullBear-Spread | Änderung des Bull-Bear-Spread im AII Sentiment Index | Die Veränderung des Bull-Bear Spreads, der durch die American Association of Individual Investors erhobene Sentiment Index. |
| UST 2Y Rate | Zinssatz der zweijährigen US-Staatsanleihe | Mittelfristzinsen in den USA |
| UST 10Y Rate | Zinssatz der zehnjährigen US-Staatsanleihe | Langfristzinsen in den USA |
| Geldmenge M3 US | Gesamtmenge der verfügbaren Währung in den USA (M3) | Geldmenge, welche in den USA im Umlauf ist. |
| Delta GDP USA | Monatliche Veränderung des Bruttoinlandsprodukts, der USA | Veränderung der gesamtwirtschaftlichen Leistung der USA. |
| US Personal Consumption Expenditure Change | Änderung der Konsumausgaben in den USA | Veränderung der in den USA erzielten Konsumausgaben. |

| | | | |
|--|--------|---|---|
| New Home Sales-Units | | Anzahl der verkauften Neubauten in den USA | Verkaufte Neubauten in den USA, ist ein Indikator für den Zustand der US-amerikanischen Wirtschaft. |
| Delta CPI, % Year on year | | Montaliche Veränderung der jährlichen Änderungsrate des Verbraucherpreisindex (CPI) | Veränderung der Inflationsrate |
| Core CPI Index | | Kern-Verbraucherpreisindex ohne Nahrung und Energie | Kerninflationsrate, ohne stark schwankende Güter wie Energie und Nahrung. |
| 5 Year Inflation Expectations (median), in % | | Median der 5-Jahres-Inflationserwartungen | Die durch die US-Bevölkerung erwartete Inflation, hat Einfluss auf Konsumausgaben und tatsächliche Inflation. |
| Retail Sales YoY | | Jährliche Änderungsrate des Einzelhandelsumsatzes | Veränderung der Einzelhandelsumsätze in den USA |
| Industrial Production Index, % YoY | | Jährliche Änderungsrate des Industrieproduktionsindex | Veränderung des Indexes für Industrieproduktion, Indikator für die Wirtschaftsleistung in den USA. |
| UnemploymentRate | | Arbeitslosenquote. | Anzahl der Arbeitslosen in Prozent der Bevölkerung. |
| 1MLagged_S&P500 End | Months | S&P500 Monatsendwert, um einen Monat verzögert | Der Stand des S&P500 zum Monatsendwertes vor einem Monat. |
| 2MLagged_S&P500 End | Months | S&P500 Monatsendwert, um zwei Monate verzögert | Der Stand des S&P500 zum Monatsendwertes vor zwei Monaten. |

| | | | |
|------------------------|--------|--|---|
| 3MLagged_S&P500 End | Months | S&P500 Monatsendwert, um drei Monate verzögert | Der Stand des S&P500 zum Monatsendwertes vor drei Monaten |
| COVID19 | | Dummyvariable welche ab dem Januar 2020 den Wert 1 annimmt | Gibt die Präsenz von COVID19 an, welche in der deskriptiven Analyse der Off-Exchange Umsätze als struktureller Bruch identifiziert wurde. |

Tabelle 7: Modellparameter und deren Relevanz (eigene Darstellung)

4.7. Adfullertest auf Stationarität

Um die Modellparameter auf Stationarität zu überprüfen, wird für jede Variable ein Adfullertest (ADF) durchgeführt. Wenn der P-Wert des ersten Lags (Verzögerung) einen Wert kleiner als 0.05 annimmt, wird die Variable als stationär betrachtet. Für die Modellparameter ergeben sich folgende P-Werte:

| Modellparameter | P-Wert | Klassifikation |
|--|--------|-----------------|
| Umsatz Off-Exchange Börsen | 0.79 | Nicht stationär |
| Real Disposable Income | 0.93 | Nicht stationär |
| S&P500 Months End | 0.99 | Nicht stationär |
| S&P 500 Month Low | 0.95 | Nicht stationär |
| S&P 500 Month High | 1.00 | Nicht stationär |
| DeltaSentimentIndexBullBearSpread | 0.00 | Stationär |
| UST 2Y Rate | 0.89 | Nicht stationär |
| UST 10Y Rate | 0.41 | Nicht stationär |
| Geldmenge M3 US | 1.00 | Nicht stationär |
| Delta GDP USA | 0.00 | Stationär |
| US Personal Consumption Expenditure Change | 0.00 | Stationär |
| New Home Sales-Units | 0.66 | Nicht stationär |
| Delta CPI, % Year on year | 0.00 | Stationär |
| Core CPI Index | 1.00 | Nicht stationär |
| 5 Year Inflation Expectations (median), in % | 0.65 | Nicht stationär |
| Retail Sales YoY | 0.00 | Stationär |
| Industrial Production Index, % YoY | 0.00 | Stationär |
| UnemploymentRate | 0.17 | Nicht stationär |
| 1MLagged_ S&P500 Months End | 0.99 | Nicht stationär |
| 2MLagged_ S&P500 Months End | 0.99 | Nicht stationär |
| 3MLagged_ S&P500 Months End | 0.98 | Nicht stationär |

Tabelle 8: ADF Resultate und Klassifikation (eigene Darstellung)

Aufgrund der mittels ADF-Tests erzielten Resultate werden alle nicht stationären Modellparameter mittels Log-Differenz in stationäre Variablen umgewandelt. Somit werden anstatt der absoluten Werte der einzelnen Datenpunkten die relativen Differenzen zwischen den ursprünglichen Datenpunkten als neuer Parameterwert für die Modelle herangezogen.

4.8. Überprüfung auf Vollständigkeit / Korrelation / Verteilung

Um die Untersuchungen durchführen zu können, müssen die Modellparameter noch auf ihre Vollständigkeit und ihre Verteilung überprüft werden. Nachdem alle unvollständigen Datenpunkte ausgeschlossen wurden, bleiben total 164 Beobachtungen übrig.

Mittels Korrelationsmatrix (Anhang 1) werden die Modellparameter auf Korrelation beziehungsweise negative Korrelation überprüft. Die durchgeführte Korrelationsanalyse zeigt, dass folgende Parameter eine hohe (negative) Korrelation zueinander aufweisen:

- Delta UST 2Y Rate zu Delta UST 10Y Rate → eine Korrelation von 0.7014
- Delta GDP zu Delta Personal Consumption → eine Korrelation von 0.7232
- Delta GDP zu Delta Unemployment → eine Korrelation von -0.7770
- Delta Retailsales zu Delta Industrial Production → eine Korrelation von 0.6905

Aufgrund der beobachteten Korrelationen werden die Modellparameter, Delta UST 10Y, Delta Personal Consumption, Delta Unemployment, Delta Industrial Production aus der Untersuchung ausgeschlossen. Die übriggebliebenen Modellparameter werden anschließend mittels Shapiro-Wilk-Test auf ihre Normalverteilung überprüft. Falls der aus dem Shapiro-Wilk-Test ermittelte P-Wert über dem Alpha von 0.05 zu liegen kommt, werden die Modellparameter als normalverteilt klassifiziert:

| Modellparameter | P-Wert | Klassifikation |
|--|--------|----------------------|
| Delta Umsatz Off-Exchange Börsen | 0.5520 | Normalverteilt |
| Delta Real Disposable Income | 0.0000 | Nicht normalverteilt |
| Delta S&P500 Months End | 0.0035 | Nicht normalverteilt |
| Delta S&P 500 Month Low | 0.0000 | Nicht normalverteilt |
| Delta S&P 500 Month High | 0.0000 | Nicht normalverteilt |
| DeltaSentimentIndexBullBearSpread | 0.3183 | Normalverteilt |
| Delta UST 2Y Rate | 0.0000 | Nicht normalverteilt |
| Delta Geldmenge M3 US | 0.0000 | Nicht normalverteilt |
| Delta GDP USA | 0.0000 | Nicht normalverteilt |
| Delta New Home Sales-Units | 0.0004 | Nicht normalverteilt |
| Delta CPI, % Year on year | 0.0000 | Nicht normalverteilt |
| Delta Core CPI Index | 0.0000 | Nicht normalverteilt |
| Delta 5 Year Inflation Expectations (median), in % | 0.0016 | Nicht normalverteilt |
| Delta Retail Sales YoY | 0.0000 | Nicht normalverteilt |
| Delta 1MLagged_S&P500 Months End | 0.0041 | Nicht normalverteilt |
| Delta 2MLagged_S&P500 Months End | 0.0054 | Nicht normalverteilt |
| Delta 3MLagged_S&P500 Months End | 0.0016 | Nicht normalverteilt |

Tabelle 9: Shapiro-Wilk-Test und Klassifikation (eigene Darstellung)

Aufgrund der in der Untersuchung festgestellten Nichtnormalverteilung werden die Train- und Testsets für die Ermittlung der optimalen Modellparameter standardisiert, um die numerische Stabilität und die «Equal Feature Importance» zu gewährleisten.

4.9. Lasso Regression zur optimalen Kombination von Modellparameter

Um ein Overfitting in der OLS Regression zu verhindern und die optimale Auswahl der Modellparameter zu wählen, wird eine Lasso Regression auf den normalisierten Datensatz durchgeführt. Um die Verbesserung des Modells zu messen, wird das Modell mit allen Parametern sowie das mittels Lasso Regression eingeschränkte Modell auf einem Trainset von 2009 – 2021 trainiert und auf ein Testset mit dem Datensatz von 2022 appliziert. Danach wird der MSE dieser beiden Schätzungen berechnet und verglichen.

Die Lasso Regression, welche auf dem normalisierten Trainset durchgeführt wurde, reduzierte die Anzahl Modellparameter von 16 auf neun. Folgende Modellparameter wurden als optimal ausgeschieden:

- Delta S&P500Low
- Delta S&P500High
- Delta UST 2Y Rate
- Delta GDP
- Delta New Home Sales-Units
- Delta CPI Core
- Delta 5 Year Inflation Expectations (median), in %
- Delta Retail Sales YoY
- Delta 1MLagged_S&P500 Months End

Die erste OLS Regression mit allen Modellparametern erreicht auf dem normalisierten Trainset ein R-Quadrat von 0.394 und auf dem normalisierten Testset einen MSE von 1.01593. Das mittels Lasso-Regression eingeschränkte Modell erreichte ein R-Quadrat von 0.389 und einen MSE von 1.00956. Trotz der Reduzierung der Modellparameter um fast die Hälfte sank das R-Quadrat nur um fünf Prozent, während sich die Prognoseleistung signifikant verbesserte. Daraus kann geschlossen werden, dass die ausgeschlossenen Modellparameter nur einen kleinen respektive nicht signifikanten Einfluss auf die Leistung des Modells haben.

4.10. Hyperparametertuning für das SARIMAX-Modell

Das SARIMAX-Modell wird mit denselben Modellparametern durchgeführt, welche sich in der OLS-Regression am effektivsten erwiesen haben. Um die optimalen SARIMAX-Modellparameter zu eruieren, wird zuerst der Datensatz (2009 – 2022) in ein Trainingsdatensatz (2009 – 2021) und einen Testdatensatz (2022) aufgeteilt, analog dem OLS-Hyperparametertuning. Anschliessend wird ein SARIMAX-Modell mit Basisparametern (Ordnung (1,0,0), Saisonal (1,0,0,12)) auf den normalisierten Trainingsdatensatz (2009 – 2020) ausgeführt und dann auf den Testdatensatz (2022) getestet. Anschliessend wird ein Grid-Search in nachfolgendem Raster durchgeführt:

$$\begin{array}{ll} p = (0, 1, 2, 3) & P = (0, 1, 2) \\ d = (0) & D = (0,1) \\ q = (0, 1, 2, 3) & Q = (0, 1, 2) \\ & S = (12) \end{array}$$

Der Grid-Search ist so konzipiert, dass alle möglichen Varianten durchgerechnet werden und schliesslich die Variante mit dem tiefsten MSE als optimale Parameterkonfigurationsvariante gewählt wird. Der Grid-Search erzielte mit dem SARIMAX-Modell mit der Konfiguration: (3,0,3) (1,0,2,12) den besten MSE. Das Modell mit diesem Parameter erzielte einen MSE von 0.60748 im Vergleich zum MSE von 1.09080 des Basismodells. Somit werden diese Parameter als optimal angenommen und für die weiteren Untersuchungen verwendet.

5. Ergebnisse

5.1. Analyse Ergebnisse OLS-Regression

Um die Einflüsse der verschiedenen exogenen Variablen auf das Handelsvolumen der Retailinvestoren in den USA zu ermitteln, wurde eine OLS-Regression durchgeführt. Als exogene Faktoren wurden die mittels Lasso-Regularisierung eingeschränkten Modellparameter gewählt. Die mittels der OLS-Regression auf dem Traindatensatz von 2009 bis 2021 errechneten Koeffizienten zeigen die Einflüsse, welche auf das Handelsvolumen der Retailinvestoren wirken. Die OLS-Regression (Tabelle 10) erzielt einen P-Wert der F-Statistik von kleiner als einem Prozent und indiziert somit, dass das gesamte Regressionsmodell als statistisch signifikant klassifiziert werden kann. Das Modell zeigt zudem, dass ein relativ grosser Anteil der Änderungen des Handelsvolumens von Retailinvestoren auf die Verschiebungen in den Modellparametern zurückzuführen sind. So errechnete die durchgeführte OLS-Regression ein R-Quadrat von 38.9 Prozent. Dies ist so zu interpretieren, dass die Schwankungen des Retailhandelsvolumens in den USA zu knapp 40 Prozent auf die ausgewählten Modellparameter erklärt werden können. Das adjustierte R-Quadrat (R-Quadrat bereinigt um Modellkomplexität), welches einen Wert von 35 Prozent aufweist, bestätigt die Relevanz der gewählten Parameter.

| | | | |
|--------------------------|---------------|----------------------------|----------|
| Dep. Variable: | ChngOffExchng | R-squared: | 0.389 |
| Model: | OLS | Adj. R-squared: | 0.350 |
| Method: | Least Squares | F-statistic: | 10.04 |
| No. Observations: | 152 | Prob (F-statistic): | 7.56e-12 |
| Df Residuals: | 142 | Log-Likelihood: | 127.74 |
| Df Model: | 9 | AIC: | -235.5 |
| Covariance Type: | nonrobust | BIC: | -205.2 |

| | coef | std err | t | P> t | [0.025 | 0.975] |
|---------------------|---------|---------|--------|-------|--------|--------|
| const | -0.0046 | 0.013 | -0.360 | 0.719 | -0.030 | 0.021 |
| ChngS&P500Low | -2.0858 | 0.241 | -8.647 | 0.000 | -2.563 | -1.609 |
| ChngS&P500High | 1.9775 | 0.470 | 4.207 | 0.000 | 1.048 | 2.907 |
| Chng2YR | -0.0239 | 0.041 | -0.584 | 0.560 | -0.105 | 0.057 |
| ChngGDP | -0.0017 | 0.001 | -1.508 | 0.134 | -0.004 | 0.001 |
| ChngNewHomes | 0.2092 | 0.108 | 1.934 | 0.055 | -0.005 | 0.423 |
| ChngCPICore | 9.2304 | 5.759 | 1.603 | 0.111 | -2.154 | 20.615 |
| ChngInflExp | 0.3080 | 0.158 | 1.953 | 0.053 | -0.004 | 0.620 |
| ChngRetailSales | -0.0020 | 0.002 | -1.237 | 0.218 | -0.005 | 0.001 |
| 1MLagged_ChngS&P500 | 0.9411 | 0.271 | 3.468 | 0.001 | 0.405 | 1.477 |

| | | | |
|-----------------------|--------|--------------------------|----------|
| Omnibus: | 3.998 | Durbin-Watson: | 2.693 |
| Prob(Omnibus): | 0.135 | Jarque-Bera (JB): | 2.565 |
| Skew: | -0.112 | Prob(JB): | 0.277 |
| Kurtosis: | 2.405 | Cond. No. | 7.04e+03 |

Tabelle 10: Resultate OLS Regression (eigene Darstellung)

Aufgrund der beobachteten Signifikanz des Modells kann die H0 Hypothese der Hypothese 1 verworfen werden. Demzufolge lässt sich festhalten, dass die gewählten makroökonomischen Parameter einen signifikanten Einfluss auf das Handelsvolumen der Retailinvestoren haben.

Mit Blick auf die einzelnen Modellparameter lässt sich schliessen, dass vor allem die Parameter, welche im Zusammenhang mit dem Leitindex S&P 500 stehen, signifikant sind und einen starken Einfluss auf das Handelsvolumen der Retailinvestoren ausüben. So erzielten die Parameter «Delta S&P 500 Low», «Delta S&P 500 High» und «Delta 1MLagged_S&P500 Months End» Signifikanzniveaus von null beziehungsweise nahe null. Der Parameter und «Delta 1MLagged_S&P500 Months End» weist einen Koeffizienten von 0.9411 auf. Dies bedeutet, dass, falls der S&P500 im vergangenen Monat einen Anstieg um ein Prozent erlebte, ein Anstieg des Handelsvolumens von Retailinvestoren um 0.94 Prozent folgte. Dieser Umstand ist plausibel, da davon ausgegangen werden kann, dass vergangene positive Renditen des Leitindex dazu führen, dass Retailinvestoren Finanzanlagen vermehrt als attraktiv betrachten und zunehmend ihre Ersparnisse an den Finanzmärkten investieren.

Ebenfalls einen stark positiven Effekt auf das Handelsvolumen der Retailinvestoren, mit einem Koeffizienten von 1.9775 hat der Parameter «Delta S&P 500 High». Hier lässt sich schliessen, dass sich Kleininvestoren ähnlich dem Parameter «Delta 1MLagged_S&P500 Months End» durch die erfahrene Performance in ihren Handelsaktivitäten leiten lassen. So können starke Anstiege von Monatshöchstwerten des S&P 500 durch die Retailinvestoren als Signal verstanden werden, dass es an den Börsen ein Momentum gibt. Umgekehrt führen sinkende Höchstwerte zu einer Abnahme der Handelsaktivität von privaten Anlegern. In diesem Zusammenhang ist interessant, wie sich der Parameter «Delta S&P 500 Low» verhält. Dieser weist einen Koeffizienten von -2.0858 auf. Durch das negative Vorzeichen führt ein Anstieg des Monatstiefstwertes zu einer Abnahme des Handelsvolumens. Eine Erklärung für diesen Einfluss kann sein, dass ein ständiger Aufwärtstrend an den Börsen Aktien und weitere Finanzanlagen als teuer erscheinen lassen. Wiederum kann ein Absinken der Kurse bei den Retailinvestoren dazu führen, die Marktlage als Einstiegsmöglichkeit zu sehen. Somit wirkt der Parameter «Delta S&P 500 Low» im Zusammenspiel mit dem Parameter «Delta S&P 500 High» dämpfend. Ein paralleler Anstieg der Monatstiefstkurse und Monatshöchstkurse führt dazu, dass sich die beiden Effekte fast komplett aufheben. Derselbe Effekt findet bei einem parallelen Kursrückgang

statt. Variieren die Kurse in einem Monat jedoch sehr stark, das heisst der Monatstiefstwert sinkt und der Monatshöchstwert steigt, führt dies zu einem starken Anstieg des Handelsvolumens. Somit können die beiden Parameter als Proxy für die Marktvolatilität verstanden werden. Ist die Volatilität an den Märkten hoch, nimmt auch das Handelsvolumen der Retailinvestoren zu. Eine mögliche Erklärung dieser Beobachtung ist, dass Retailinvestoren durch die Volatilität die Möglichkeit für gegeben erachten, attraktive Renditen zu erwirtschaften und deshalb bereit sind höhere Risiken einzugehen.

Als ebenfalls relativ signifikant erwiesen sich mit circa fünf Prozent die Parameter «Delta 5 Year Inflation Expectations (median), in %» und «Delta New Home Sales-Units». Der Koeffizient für die neu in den USA verkauften Häuser beträgt 0.2092. Der Makroindikator «New Home Sales Units» in den USA gibt Auskunft über den Verkauf von neu gebauten Wohnimmobilien. Er misst die Anzahl der verkauften Einheiten von neu errichteten Häusern, die an Käufer veräussert wurden. Dieser Indikator wird oft als eine wichtige Kennzahl betrachtet, um die Stärke und Gesundheit des US-Immobilienmarktes zu bewerten. Privatinvestoren verstärken hiernach ihre Aktivität an den Finanzmärkten leicht, wenn sie davon ausgehen, dass die US-Wirtschaft auf einem soliden Fundament beruht.

Der Koeffizient für die Inflationserwartung über fünf Jahre ist positiv mit einem Wert von 0.3080. Dabei ist bemerkenswert, dass ein Anstieg der Inflationserwartung um ein Prozent nur zu einem Anstieg des Handelsvolumens von 0.3 Prozent führt. Tendenziell sollten Finanzanlagen bei einer steigenden Inflationserwartung attraktiver werden, da aufgrund der Entwertung der Bankeinlagen der Anreiz, das Vermögen zu investieren, steigen sollte. Der Parameter «Delta 5 Year Inflation Expectations (median), in %» muss jedoch im Kontext mit dem Parameter «Delta CPI Core» betrachtet werden. Dieser verfügt über eine tiefere Signifikanz (0.111), einen im Vergleich zu den anderen Parametern hingegen einen sehr hohen Koeffizienten von 9.23. Folglich ist die Inflation relevant für die Anlageentscheidungen der Retailinvestoren und führt dazu, dass Investoren Wege suchen, um eine Entwertung ihres Vermögens zu verhindern. Aus dem Vergleich der beiden Parameter lässt sich schliessen, dass sich Privatinvestoren bei ihren Anlageentscheidungen stark auf die realen Inflationserfahrungen stützen und sich nicht im gleichen Ausmass mit den Inflationserwartungen befassen. Aufgrund ihrer Erfahrungen mit der Inflation und dem Erlebnis stark steigender Preise ist könnte daher ein Antrieb sein, ihre Finanzanlagen entsprechend umzuschichten, um einen inflationsbedingten Vermögensverzehr entgegenzuwirken.

Der Parameter des «Delta GDP» mit einer Signifikanz von 0.134 und einem Koeffizienten von -0.0017 besitzt im Vergleich mit den anderen makroökonomischen Variablen eine tiefere Signifikanz. Auch der Koeffizient ist mit einem Wert von -0.0017 ist nahe Null. Demgemäss wirken sich Veränderungen des Bruttoinlandprodukts nur gering auf das Handelsvolumen von Privatinvestoren aus. Eine mögliche Erklärung hierfür ist, dass der Parameter aufgrund des Wirtschaftseinbruchs während der Coronapandemie und der gleichzeitigen massiven Hilfsleistungen der US-Regierung an die Privathaushalte an Gewicht verloren hat. So hatten die Privathaushalte trotz schlechter Wirtschaftslage und hoher Arbeitslosigkeit keine Angst, in Armut abzurutschen, und blieben an den Finanzmärkten weiter aktiv.

Der gleiche Effekt ist beim Parameter «Delta Retail Sales YoY» mit einem Koeffizienten von -0.0020 und einer Signifikanz von 0.218 anzunehmen. Der starke Einbruch der Retail Sales führte während der Coronapandemie nicht zu einem Einbruch der Handelsaktivitäten, anders als dies bei «gewöhnlichen» Krisen zu beobachten wäre.

Der Einflussfaktor «Delta UST 2Y Rate», welcher für die Mittelfristzinsen in den USA steht, zeigt sich mit einem Wert von 0.560 als nicht signifikant und auch der Koeffizient von -0.0239 indiziert keinen starken Einfluss auf die Handelsaktivität von Retailinvestoren. Möglicherweise sind entsprechende Zusammenhänge bereits mit den Parametern «Delta CPI Core» und «Delta 5 Year Inflation Expectations (median), in %» abgedeckt. Die Parameter weisen jedoch untereinander nur geringe Korrelationen auf.

5.2. Analyse SARIMAX-Modell

Um die in den OLS-Regressionen erzielten Ergebnisse zu verifizieren, wurde zusätzlich ein SARIMAX-Modell (Tabelle 11) für Time-Series Daten auf dem Trainingsset von 2009 bis 2021 angewandt. Das SARIMAX-Modell bezieht auch die vergangenen Daten der endogenen Variable (Delta Umsatz Off-Exchange Börsen) selbst in die Kalkulation mit ein. Dies ist vor allem für die Vorhersage von zukünftigen Änderungen des Handelsvolumens von Retailinvestoren hilfreich.

Die Koeffizienten und Signifikanzen im errechneten SARIMAX zeigen ähnliche Werte wie die zuvor durchgeführte OLS-Regression. So sind die die stärksten Signifikanzen ebenfalls bei den drei Variablen, welche sich auf den S&P 500 beziehen, zu beobachten. Diese betragen für den Parameter «Delta S&P500Low» und «Delta S&P500High» 0.000 und für den Parameter «Delta 1MLagged_S&P500 Months End» 0.001. Diese Signifikanzen sind übereinstimmend mit der OLS-Regression.

Mit Blick auf die Koeffizienten ist festzustellen, dass diese für die drei Modellparameter ebenfalls ähnlich ausfallen und die gleichen Vorzeichen wie bei der OLS-Regression besitzen. Die Werte für die beiden Parameter «Delta S&P500Low» und «Delta S&P500High» sind mit -1.8930 (OLS -2.0858) und 1.8646 (OLS 1.9775) leicht tiefer, während der Koeffizient für den Parameter «Delta 1MLagged_S&P500 Months End» mit 1.1055 (OLS 0.9411) leicht höher ausfällt. Folglich wird im SARIMAX Modell der vergangenen Performance des S&P 500 ein höheres Gewicht beigemessen.

Der Parameter «Delta New Home Sales-Units» weist im SARIMAX-Modell eine tiefere Signifikanz von 0.301 (OLS 0.055), aber einen vergleichbaren Koeffizienten von 0.1079 (OLS 0.2092) aus. Dies bestätigt, die Annahme, dass die Retailinvestoren ihr Handelsvolumen steigern, wenn sie davon ausgehen können, dass die Wirtschaft solide ist.

Die drei Einflussgrößen, welche einen Zins- oder Inflationsbezug haben («Delta 5 Year Inflation Expectations (median), in %», «Delta CPI Core» und «Delta UST 2Y Rate») unterscheiden sich im SARIMAX-Modell in Bezug auf die Signifikanzen sowie die Koeffizienten zum Teil stark von den Werten, welche die OLS-Regression berechnet hat.

Die Signifikanzen sind bei den Parametern «Delta 5 Year Inflation Expectations (median), in %» (SARIMAX 0.036, OLS 0.053) und «Delta UST 2Y Rate» (SARIMAX 0.089, OLS 0.560) höher ausgefallen, während sich die Signifikanz bei «Delta CPI Core» (SARIMAX 0.971, OLS 0.111) stark verschlechtert hat.

Während die Koeffizienten bei «Delta 5 Year Inflation Expectations (median), in %» (SARIMAX 0.3995, OLS 0.3080) und «Delta UST 2Y Rate» %» (SARIMAX -0.0385, OLS -0.0239) in etwa stabil geblieben sind, hat sich der Koeffizient von «Delta CPI Core» (SARIMAX -0.2095, OLS 9.2304) stark abgesenkt.

Vor allem die starke Veränderung der Signifikanz und des Koeffizienten beim Parameter «Delta CPI Core» sind bemerkenswert. Einerseits kann dies darauf zurückzuführen sein, dass der CPI Core sich im Trainingszeitraum von 2009 – 2020 stabil gehalten hat und dann im Jahr 2021 stark angestiegen ist. Dies ist möglicherweise ein Grund, wieso der Parameter sich zwischen den beiden Modellen stark verändert hat. Ein positiver Einfluss der Inflation auf die Handelsaktivitäten von Privatinvestoren ist jedoch aus bereits genannten Gründen plausibler als der dämpfende Effekt, welcher im SARIMAX-Modell erzielt wurde.

Die Werte der Koeffizienten für die letzten beiden Parameter «Delta GDP» (SARIMAX -0.0013, OLS -0.0017) und «Delta Retail Sales YoY» (SARIMAX -0.0007, OLS -0.0020) blieben annähernd konstant.

Die Werte der Signifikanzen verbesserten sich bei «Delta GDP» von 0.134 (OLS) auf 0.071 (SARIMAX) während sie sich diese bei «Delta Retail Sales YoY» von 0.218 (OLS) auf 0.471 (SARIMAX) verschlechterten.

| | | | |
|-------------------------|-------------------------------------|--------------------------|----------|
| Dep. Variable: | ChngOffExchng | No. Observations: | 152 |
| Model: | SARIMAX(3, 0, 3)x(0, 1, [1, 2], 12) | Log Likelihood | 130.554 |
| Covariance Type: | opg | AIC | -223.108 |
| Sample: | 05-31-2009 | BIC | -171.457 |
| Sample: | - 12-31-2021 | HQIC | -202.151 |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|--------------------------------|---------|--------------------------|--------|-------|---------|--------|
| intercept | 0.0010 | 0.001 | 1.256 | 0.209 | -0.001 | 0.003 |
| ChngS&P500Low | -1.8930 | 0.226 | -8.390 | 0.000 | -2.335 | -1.451 |
| ChngS&P500High | 1.8646 | 0.372 | 5.015 | 0.000 | 1.136 | 2.593 |
| Chng2YR | -0.0385 | 0.023 | -1.703 | 0.089 | -0.083 | 0.006 |
| ChngGDP | -0.0013 | 0.001 | -1.807 | 0.071 | -0.003 | 0.000 |
| ChngNewHomes | 0.1079 | 0.104 | 1.035 | 0.301 | -0.096 | 0.312 |
| ChngCPICore | -0.2095 | 5.671 | -0.037 | 0.971 | -11.324 | 10.905 |
| ChngInflExp | 0.3995 | 0.191 | 2.092 | 0.036 | 0.025 | 0.774 |
| ChngRetailSales | -0.0007 | 0.001 | -0.721 | 0.471 | -0.003 | 0.001 |
| 1MLagged_ChngS&P500 | 1.1055 | 0.325 | 3.404 | 0.001 | 0.469 | 1.742 |
| ar.L1 | -0.3315 | 0.248 | -1.336 | 0.181 | -0.818 | 0.155 |
| ar.L2 | -0.4185 | 0.231 | -1.808 | 0.071 | -0.872 | 0.035 |
| ar.L3 | 0.3896 | 0.146 | 2.667 | 0.008 | 0.103 | 0.676 |
| ma.L1 | -0.3269 | 0.263 | -1.245 | 0.213 | -0.842 | 0.188 |
| ma.L2 | 0.0301 | 0.228 | 0.132 | 0.895 | -0.417 | 0.478 |
| ma.L3 | -0.5976 | 0.233 | -2.562 | 0.010 | -1.055 | -0.140 |
| ma.S.L12 | -0.5719 | 0.218 | -2.626 | 0.009 | -0.999 | -0.145 |
| ma.S.L24 | -0.2657 | 0.198 | -1.342 | 0.180 | -0.654 | 0.122 |
| sigma2 | 0.0052 | 0.001 | 4.125 | 0.000 | 0.003 | 0.008 |
| Ljung-Box (L1) (Q): | 0.10 | Jarque-Bera (JB): | 0.02 | | | |
| Prob(Q): | 0.75 | Prob(JB): | 0.99 | | | |
| Heteroskedasticity (H): | 1.97 | Skew: | -0.01 | | | |
| Prob(H) (two-sided): | 0.04 | Kurtosis: | 2.94 | | | |

Tabelle 11: Resultate SARIMAX-Modell (eigene Darstellung)

Betrachtet man die SARIMAX spezifischen Parameter, lassen sich zudem weitere Rückschlüsse ziehen. Die Parameter «ar.L1» bis «ar.L3» beziehen sich auf die Parameter der autoregressiven Komponenten des Modells, jeweils mit Verzögerung (Lag) eins bis drei. Der Koeffizient von -0.3315 für «ar.L1» besagt, dass eine Erhöhung um eine Einheit in der Zeitreihe zu einer Abnahme von 0.3315 Prozent im aktuellen Zeitpunkt führt, vorausgesetzt, alle anderen Faktoren bleiben gleich. Im Modell sind die Koeffizienten «ar.L1» und «ar.L2» nicht statistisch signifikant auf dem 5-Prozent-Niveau, während «ar.L3»

statistisch signifikant ist. Dies deutet darauf hin, dass die Werte von vor drei Zeiteinheiten einen signifikanten Einfluss auf den aktuellen Wert, mit einem Koeffizienten von 0.3896, haben, während die Werte von vor ein oder zwei Zeiteinheiten das nicht tun, zumindest nicht auf dem 5-Prozent-Niveau der Signifikanz.

Die Parameter «ma.L1», «ma.L2» und «ma.L3» beziehen sich auf die Moving Average (MA) Komponenten des SARIMAX-Modells mit Verzögerungen (Lags) von eins bis drei Zeiteinheiten. Im Kontext des Modells repräsentieren diese Koeffizienten den Einfluss, den der Vorhersagefehler (die Differenz zwischen der beobachteten und der vorhergesagten Größe) vor einem, zwei, beziehungsweise drei Zeitschritten auf den aktuellen Vorhersagefehler hat. Beispielsweise bedeutet der Koeffizient von -0.3269 für «ma.L1», dass ein Anstieg des Fehlers um eine Einheit in der Zeitreihe zu einer Abnahme des Fehlers um 0.3269 Prozent im aktuellen Zeitpunkt führt, vorausgesetzt, alle anderen Faktoren bleiben gleich. Der Koeffizient «ma.L3» ist signifikant mit einem P-Wert von 0.010, während «ma.L1» (0.213) und «ma.L2» (0.895) eine niedrigere Signifikanz ausweisen. Dies zeigt, dass der Fehler von vor drei Zeiteinheiten einen signifikanten Einfluss (-0.5976) auf den aktuellen Vorhersagefehler hat, während dies bei den Parametern «ma.L1» und «ma.L2» nicht der Fall ist. Die beiden Einflussgrößen «ma.S.L12» und «ma.S.L24» beziehen sich auf die saisonalen Moving Average (MA) Komponenten des SARIMAX-Modells. Die Interpretation dieser beiden Parameter ist ähnlich zu den Parametern anderen MA-Parametern, mit dem Unterschied, dass diese Komponenten saisonal sind und periodische Muster in den Daten darstellen. Der Parameter «ma.S.L12» (0.009) ist signifikant auf einem Niveau von fünf Prozent. Demnach haben Vorhersagefehler von vor 12 Zeiteinheiten einen signifikanten Einfluss (mit einem Wert von -0.5719) auf den aktuellen Fehler. Beim Parameter «ma.S.L12» ist dies nicht der Fall. Der letzte ausgewiesene Parameter «sigma2» repräsentiert die geschätzte Varianz der Residuen im SARIMAX-Modell. Somit ist «sigma2» das geschätzte Mass der Streuung der Daten um die vom Modell vorhergesagten Werte. Aufgrund der Signifikanz ist dieser Parameter ein Indiz, dass das Modell einige Aspekte der Daten nicht vollständig erklärt, was dazu führt, dass die Fehler mehr streuen, als sie es tun würden, wenn das Modell perfekt wäre.

Wendet man die OLS-Regression sowie das SARIMAX-Modell auf den Testdatensatz von 2022 an, erhält man einen Mean Squared Error (MSE) Wert für die OLS-Regression von 0.01801 und beim SARIMAX-Modell einen Wert von 0.01783. Folglich schätzt das SARIMAX-Modell für den vorliegenden Zeitraum die Handelsumsätze der

Retailinvestoren besser als die OLS-Regression. Für eine bessere Interpretierbarkeit kann auch noch der dazugehörige Root Mean Squared Error (RMSE) herangezogen werden. Dieser liegt bei der OLS-Regression bei 0.1342 und bei dem SARIMAX Modell bei 0.1335. Dies bedeutet, dass das die beiden Schätzverfahren im Durchschnitt etwa 13.42 Prozent beziehungsweise 13.35 Prozent von den tatsächlichen Werten abweichen.

5.3. Rollende Zeitfenster für OLS-Regression und SARIMAX Modell

Um die Robustheit der im vorherigen Abschnitt errechneten OLS-Regression sowie des SARIMAX-Modells zu überprüfen, werden diese auf unterschiedliche Zeiträume trainiert und getestet (siehe Abschnitt 4.4.1). Die Zeiträume für die Testung betragen zwei bis zwölf Monate, während das Trainingsset 12 bis 132 Monate beträgt. Die Modelle werden jeweils rollend, das heisst sie verschieben sich jeweils um einen Monat, durchgeführt. Die rollende Anwendung der OLS-Regression zeigt folgendes Bild:

| | 2 Monate | 3 Monate | 6 Monate | 9 Monate | 12 Monate |
|------------|----------|----------|----------|----------|-----------|
| 12 Monate | 0.81171 | 2.00953 | 1.48766 | 1.17426 | 0.99031 |
| 24 Monate | 0.04405 | 0.04066 | 0.03598 | 0.03572 | 0.04562 |
| 36 Monate | 0.02306 | 0.02282 | 0.02694 | 0.03219 | 0.03579 |
| 48 Monate | 0.02257 | 0.02165 | 0.01940 | 0.01834 | 0.01812 |
| 60 Monate | 0.02298 | 0.02133 | 0.01951 | 0.01832 | 0.01815 |
| 72 Monate | 0.02365 | 0.02155 | 0.02076 | 0.02082 | 0.02089 |
| 84 Monate | 0.01963 | 0.01887 | 0.01948 | 0.02018 | 0.02030 |
| 96 Monate | 0.01932 | 0.01793 | 0.01735 | 0.01806 | 0.01838 |
| 108 Monate | 0.01849 | 0.01750 | 0.01809 | 0.01838 | 0.01855 |
| 120 Monate | 0.01936 | 0.01792 | 0.01877 | 0.01993 | 0.02045 |
| 132 Monate | 0.02169 | 0.02032 | 0.02081 | 0.02076 | 0.02249 |

Tabelle 12: Durchschnittliche MSE der getesteten OLS-Regressionen (eigene Darstellung)

Aus den erhaltenen durchschnittlichen MSE für die OLS-Regression lassen sich die Einflüsse der Konditionalität und einer genügend grossen Datenmenge auf das Model erkennen. Die Konditionalität bezeichnet den Zusammenhang zwischen möglichst kurzfristigen aktuellen Daten, um eine möglichst präzise Schätzung durchzuführen. Ein möglichst kurzer Trainingszeitraum führt dazu, dass zeitabhängige Veränderungen in den Beziehungen zwischen den unabhängigen Variablen und der abhängigen Variable möglichst ausgeschlossen werden. Jedoch muss die Datenmenge im Gegenzug genügend gross sein. Eine zu geringe Datenmenge im Verhältnis zur Anzahl der unabhängigen Variablen kann zu Problemen wie geringer Präzision, überangepassten Modellen oder statistischer Unzuverlässigkeit führen. Diese Effekte lassen sich in den Ergebnissen der durchgeführten rollenden OLS-Regressionen (Tabelle 12) erkennen. So sind die OLS-Schätzungen mit

einem kleinen Testdatensatz von 12 Monaten sehr unpräzise. Diese erzielen eine MSE-Bandbreite von 0.81171 (Testzeitraum zwei Monate) bis 2.00953 (Testzeitraum drei Monate). Rechnet man diese in einen RMSE um, so lagen die Werte im Schnitt mindestens um über 90 Prozent neben den wirklich erzielten Werten. Wird die Datenmenge erhöht, das heisst führt man die Regressionen mit einem längeren Trainingszeitraum durch, verbessern sich die Schätzungen der Regression. So nehmen die MSE bis zu einem Trainingszeitraum von 96 Monaten laufend ab. Führt man die OLS-Regression jedoch auf einen zu grossen Trainingsdatensatz durch, ist die Konditionalität der Daten nicht mehr gegeben, was im OLS-Modell ab einem Trainingszeitraum von 120 Monaten wieder zu einem generell erhöhten MSE führt.

Wie die durgeführten rollenden Regressionen zeigen, liegt der optimale Trainingszeitraum für das im Abschnitt 5.1 errechnete OLS-Regressionsmodell zwischen 96 und 108 Monaten. So finden sich die fünf tiefsten durchschnittlichen MSE in diesem Zeitraum. Als optimal erwies sich ein Testzeitraum zwischen drei und neun Monaten, wobei ein Testzeitraum von drei und sechs Monaten vier der besten fünf Schätzungen lieferte. Der tiefste durchschnittliche MSE verzeichnete die OLS-Regression mit einem Trainingszeitraum von 96 Monaten und einem Testzeitraum von 6 Monaten. Der zweittiefste und dritt-tiefste durchschnittliche MSE erreichten die OLS-Regressionen mit einem Testzeitraum von drei Monaten und einem Trainingszeitraum von 96 beziehungsweise 108 Monaten. Die beste Schätzung (Train 96 Monate/Test 6 Monate) erzielte einen durchschnittlichen MSE von 0.1735 oder einen RMSE von 0.1317. Damit weichen die Schätzungen durch das Modell um durchschnittlich 13.17 Prozent von den wahren Werten ab. Das Modell, ist bei 62 durchgeführten Regressionen und 372 geschätzten Werten, als relativ robust zu betrachten.

Zeitreihenmodelle wie SARIMAX benötigen mehr Datenpunkte als lineare Regressionen, da diese auf vergangenen Beobachtungen basieren und Muster und Zusammenhänge in den Daten erfassen. Die Anwendung der rollenden Zeitfenster auf das im Abschnitt 5.2 errechnete SARIMAX-Modell zeigte nur ab einem Trainingszeitfenster von 120 Monaten akzeptable Ergebnisse (durchschnittlicher MSE > 10). Aus diesem Grund wurden die rollenden Zeitfenster erst ab einem Trainingsdatensatz von 120 Monaten durchgeführt.

| | 2 Monate | 3 Monate | 6 Monate | 9 Monate | 12 Monate |
|------------|----------|----------|----------|----------|-----------|
| 120 Monate | 0.02528 | 0.02350 | 0.02134 | 0.02011 | 0.02103 |
| 132 Monate | 0.01898 | 0.01796 | 0.01742 | 0.01675 | 0.01733 |

Tabelle 13: Durchschnittliche MSE der getesteten SARIMAX-Modelle

Wie in Tabelle 13 ersichtlich, verbesserte sich die Leistung des SARIMAX-Modells mit einem zunehmenden Trainingsdatensatz. Lagen alle durchschnittlichen MSE der verschiedenen Schätzungen bei einem Trainingsdatensatz von 120 Monaten über 0.02011, sanken diese bei einem Trainingsdatensatz von 132 Monaten merklich. Das beste Resultat der rollenden Zeitfenster erzielte ein Testdatensatz von 9 Monaten mit einem durchschnittlichen MSE von 0.01675 und einem RMSE von 0.1284. Folglich lagen im Schnitt über alle 23 Regressionen und 207 geschätzten Werten um 12.84 Prozent daneben.

5.4. Analyse der besten OLS und SARIMAX Regression

Zieht man die errechneten Schätzungen der rollenden OLS-Regression mit einem Trainingsdatensatz von 96 Monaten und einem Testzeitraum von 6 Monaten heran und stellt diese den wahren beobachteten Änderungsraten des Handelsvolumens von Retailinvestoren gegenüber, erhält man folgenden Graphen (Abbildung 4):

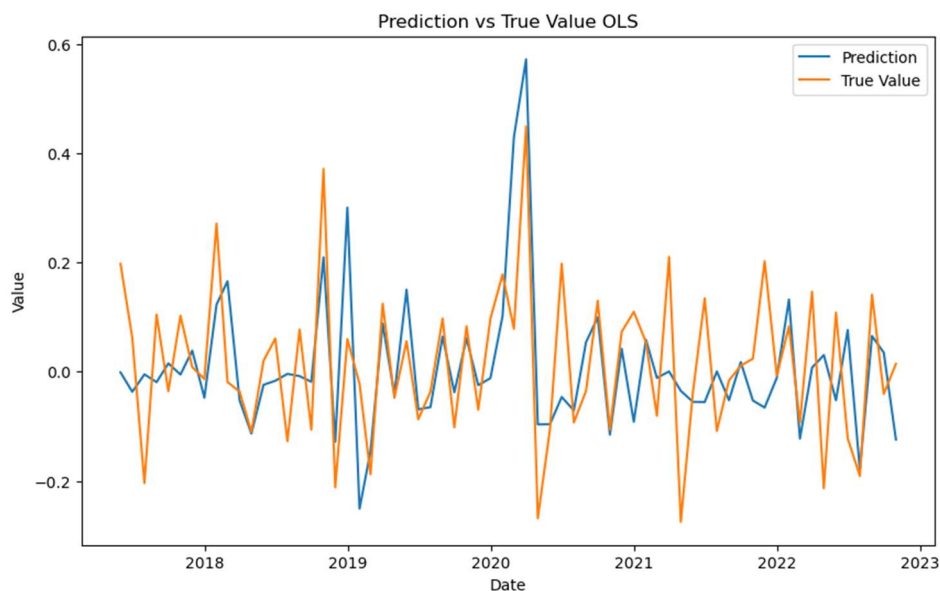


Abbildung 4: Schätzungen vs. wahre Werte beste OLS Regression (eigene Darstellung)

Wie in Abbildung 4 zu sehen ist, konnte das Modell die Wendepunkte erkennen. Es ist also möglich, mit den angewandten Parametern zu ermitteln, ob das Handelsvolumen der Retailinvestoren in den USA steigen oder sinken wird. Im Zeitraum von Ende 2018 bis kurz nach dem Beginn der Coronapandemie (Frühjahr) 2020 erzielte das Modell die besten Werte. Hier gelang es nicht nur die Wendepunkte zu ermitteln, sondern auch die jeweilige Veränderung des Handelsvolumens der Retailinvestoren.

In den Perioden von 2017 bis dem Frühjahr 2018 und zwischen dem Herbst 2021 und Frühjahr 2022 erzielte das Modell eine schlechte Performance. Vor allem das schlecht performende Zeitfenster im Jahr 2021 bis zum Frühjahr 2022 könnte auf den Meme-Stock

Hype im Jahr 2021 zurückzuführen sein (*finanzen.net*, o. J.). Die hohe Medienaufmerksamkeit und die damit verbundene nicht durch das Modell erklärbare Volatilität im Tradingvolumen von Retailinvestoren könnte die Ursache dafür sein.

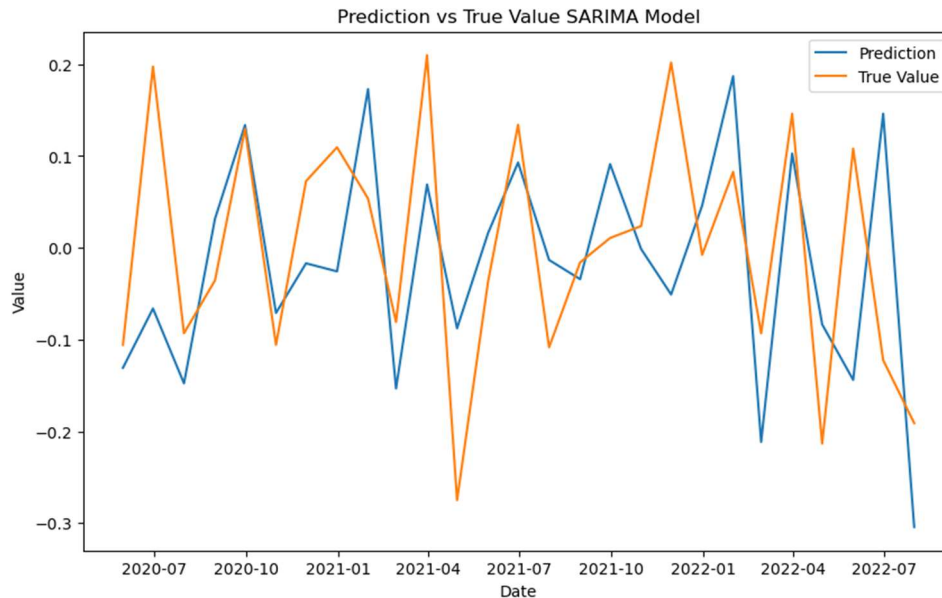


Abbildung 5: Schätzungen vs. wahre Werte bestes SARIMAX-Modell (eigene Darstellung)

In der Abbildung 5 sind die ermittelten Werte des SARIMAX-Modells den wahren Werten der Veränderungen der Handelsaktivität von Privatinvestoren gegenübergestellt. Aufgrund des höheren Trainingszeitraums von 132 Monaten konnten nur Werte ab Frühjahr 2020 geschätzt werden. Analog den Schätzungen der OLS-Regression lassen sich mit dem Modell die Wendepunkte des Tradingvolumens ermitteln. Entsprechend kann geschätzt werden, ob das Handelsvolumen der Privatinvestoren eher Steigen oder sinken wird.

5.5. Durchführung Panel-OLS-Regression

Um nachvollziehen zu können, ob die Orderrouting Erträge der Online-Broker-Unternehmen ähnlichen Einflüssen unterliegen wie das Handelsvolumen der Retailinvestoren wurde eine Panel-OLS-Regression durchgeführt. Dafür wurde die endogene Variable des Handelsvolumens von Retailinvestoren durch die Orderrouting erträge der Onlinebroker ersetzt. Die exogenen Faktoren wurden übernommen.

Wie die Tabelle 14 zeigt, erzielte das Panel-OLS-Regression mit den Daten aus den SEC 606a Reports für den Zeitraum von 2020 bis 2022 ein R-Quadrat von 30.78 Prozent. Somit erklären dieselben Variablen, welche die Handelsumsätze der Retailinvestoren zu 38.9 Prozent erklärten, die Veränderungen der Orderrouting Erträge zu knapp 8 Prozent weniger. Das Modell weist jedoch einen F-Statistikwert von 6.2740 und einen damit

verbundenen P-Wert von 0.000 aus. Demnach ist das Modell signifikant und die gewählten Variablen können einen Teil der Schwankungen der Orderroutingerträge der Onlinebroker erklären. Die Differenz zwischen den P-Werten könnte darauf zurückzuführen sein, dass die Orderrouting-Erträge noch zusätzliche Einflüsse beinhalten, welche nicht in den exogenen Modellparametern enthalten sind.

| | | | |
|--------------------------|----------------------|------------------------------|----------|
| Dep. Variable: | Chng Routing Revenue | R-squared: | 0.3078 |
| Estimator: | PanelOLS | R-squared (Between): | 0.0000 |
| No. Observations: | 140 | R-squared (Within): | 0.3078 |
| Cov.Estimator: | Unadjusted | R-squared (Overall): | 0.3062 |
| | | Log-likelihood | 68.878 |
| Entities: | 4 | F-statistic: | 6.2740 |
| Avg Obs: | 35.000 | P-value | 0.0000 |
| Min Obs: | 35.000 | Distribution: | F(9,127) |
| Max Obs: | 35.000 | | |
| Time periods: | 36 | F-statistic (robust): | 6.2740 |
| Avg Obs: | 3.8889 | P-value | 0.0000 |
| Min Obs: | 0.0000 | Distribution: | F(9,127) |
| Max Obs: | 4.0000 | | |

| | Parameter | Std. Err. | T-stat | P-value | Lower CI | Upper CI |
|---------------------|-----------|-----------|---------|---------|----------|----------|
| const | 0.0644 | 0.0235 | 2.7441 | 0.0069 | 0.0180 | 0.1109 |
| ChngS&P500Low | -1.2240 | 0.2902 | -4.2176 | 0.0000 | -1.7983 | -0.6497 |
| ChngS&P500High | 2.8624 | 0.8031 | 3.5640 | 0.0005 | 1.2731 | 4.4517 |
| Chng2YR | 0.0244 | 0.0504 | 0.4840 | 0.6292 | -0.0753 | 0.1241 |
| ChngGDP | -0.0030 | 0.0011 | -2.7494 | 0.0068 | -0.0051 | -0.0008 |
| ChngNewHomes | 0.4512 | 0.1589 | 2.8392 | 0.0053 | 0.1367 | 0.7657 |
| ChngCPICore | 8.3124 | 7.5334 | 1.1034 | 0.2719 | -6.5949 | 23.220 |
| ChngInflExp | -0.0601 | 0.3048 | -0.1970 | 0.8441 | -0.6632 | 0.5431 |
| ChngRetailSales | -0.0068 | 0.0018 | -3.6752 | 0.0003 | -0.0105 | -0.0031 |
| 1MLagged_ChngS&P500 | -0.4756 | 0.3747 | -1.2692 | 0.2067 | -1.2172 | 0.2659 |

Tabelle 14: Panel-OLS-Regression der Orderroutingerträge von Online-Broker

Betrachtet man die die Signifikanzen sowie die Koeffizienten der verschiedenen Modellparameter und vergleicht diese mit der OLS-Regression, welche im Abschnitt 5.1 durchgeführt wurden, lässt sich feststellen, dass diese ähnlichen Einflüsse ausweisen. Analog zur OLS-Regression weisen die beiden Parameter «Delta S&P500Low» und «Delta S&P500High» die höchsten Signifikanzen aus. Die Koeffizienten dieser beiden Parameter verhalten sich leicht anders als die Koeffizienten in der OLS-Regression, die Vorzeichen bleiben jedoch gleich. Dies zeigt, dass die Volatilität im Leitindex S&P 500 einen signifikanten Einfluss auf die Orderroutingerträge ausweisen.

Der Modellparameter «Delta 1MLagged_S&P500 Months End» zeigt im Vergleich der beiden Regressionsmodelle jedoch diametral verschiedene Werte der Signifikanzen (Panel-OLS 0.2067, OLS 0.001) und der Koeffizienten (Panel-OLS -0.4756, OLS 0.9411). Folglich hat eine vergangene Steigerung des Leitindex auf die

Orderrouteringträge einen dämpfenden Einfluss, während dieser bei den Handelsvolumen der Retailinvestoren einen unterstützenden Einfluss aufweist. Dies könnte darauf zurückzuführen sein, dass die Volatilität der Märkte hier einen grösseren Einfluss auf die Erträge ausweist als bei den Handelsvolumina der Retailinvestoren.

Vergleichbare Koeffizienten im Vergleich der beiden Modelle weisen auch die Parameter «Delta GDP» (Panel-OLS -0.0030, OLS -0.0017), «Delta New Home Sales-Units» (Panel-OLS 0.4512, OLS 0.2092), «Delta CPI Core» (Panel-OLS 8.3124, OLS 9.2304) und «Delta Retail Sales YoY» (Panel-OLS -0.0068, OLS -0.0020) auf. Hier sind generell dieselben Einflüsse und Ursachen auf die Orderrouteringträge zu beobachten wie auf die Handelsvolumina der Retailinvestoren.

Unterschiede sind bei den Koeffizienten der Modellparameter «Delta UST 2Y Rate» (Panel-OLS 0.0244, OLS -0.0239) und «Delta 5 Year Inflation Expectations (median), in %» (Panel-OLS -0.0601, OLS 0.3080) zu beobachten. Bei diesen beiden Variablen änderten sich unter anderem die Vorzeichen, so wechselten sie im Vergleich von der OLS-Regression zur Panel-OLS-Regression die Vorzeichen. Beim Parameter «Delta UST 2Y Rate» ist ein möglicher Grund für diesen Wechsel die tiefe Signifikanz des Parameters (Panel-OLS 0.6292, OLS 0.560) in den beiden Regressionsmodellen. Zudem ist die absolute Abweichung des Koeffizienten von «Delta UST 2Y Rate» zwischen den beiden Modellen mit 0.0483 relativ tief. Für den Parameter und «Delta 5 Year Inflation Expectations (median), in %» kann der Grund für die Abweichung der Koeffizienten ebenfalls in der tiefen Signifikanz (0.8441) im Panel-OLS Modell liegen. Ausserdem ist das Konfidenzintervall (KI) für diesen Parameter mit -0.6632 und 0.5431 sehr gross und schliesst ebenfalls den Wert des Koeffizienten der OLS-Regression mit ein.

Zusammenfassend lässt sich feststellen, dass die Orderrouteringträge zum Teil auf die gleichen Einflüsse zurückzuführen sind wie das Retailhandelsvolumen an den Finanzmärkten in den USA. Der tiefere P-Wert könnte auf Einflussgrössen zurückzuführen sein, welche handelsvolumenunabhängig sind. Dies könnte zum Beispiel eine sich verändernde Vergütung pro gerouteten Umsatz der Market Maker sein. Aufgrund der erzielten Befunde lässt sich bei der Hypothese 2 die H0 Hypothese verwerfen. Somit haben die gewählten Parameter einen signifikanten Einfluss auf die Orderrouteringträge der Retailinvestoren, denn sie zeigen ähnliche Muster und bestätigen somit die gemachte H1 Hypothese.

6. Schlussteil

6.1. Diskussion der Resultate

Das Ziel dieser Masterarbeit war es, die folgende Forschungsfrage zu beantworten:

Wie können Schwankungen des Tradingvolumens von Privatpersonen in den USA erklärt werden?

Um die Fragestellung zu beantworten, wurde ein Datensatz auf monatlicher Basis mittels Volumina von Off-Exchange Börsen erstellt und mit endogenen makroökonomischen Variablen ergänzt. Im Nachgang wurde das Datenset mit einer OLS-Regression sowie einem SARIMAX-Modell analysiert und das beste Modell zur Vorhersage der zukünftigen Veränderung evaluiert. Als zusätzliche Untersuchung wurden Orderrouterträge von Online-Broker ebenfalls untersucht, um nachvollziehen zu können, ob diese dieselben Einflüsse aufweisen wie das Tradingvolumen von Privatpersonen.

Die gemachte Untersuchung zeigte, dass Veränderungen des Tradingvolumens von Privatpersonen signifikant von den gewählten exogenen makroökonomischen Modellparametern abhängt. Mit Blick auf die einzelnen Parameter lässt sich feststellen, dass das Handelsvolumen von Retailinvestoren signifikant durch die vergangene Performance des Leitindex S&P 500 beeinflusst wird. Dieser Fakt ist übereinstimmend mit den Forschungen von Chen (2012) und Talwar et al. (2021). Talwar et al. (2021) begründete dies mit einer verstärkten Herdenmentalität, welche die Volumina zu immer neuen Höchstwerten trieb. Zudem erwies sich auch die Volatilität im S&P 500 als signifikant. Desto grösser die Veränderungen der Kurse im Leitindex, desto höher das Volumen des Handels von Retailinvestoren. Weitere starke Einflüsse zeigte die Kerninflationsrate. Deren starke Einfluss konnte jedoch nur bei der OLS-Regression und nicht bei dem SARIMAX-Modell beobachtet werden. Jedoch zeigte sich der Effekt auch bei der Untersuchung der Orderrouterträge der Online-Brokerunternehmen. Dies deckt sich mit der Studie der Bank of New York Mellon (2022), welche Hinweise dafür fand, dass die durch die Pandemie ausgelöste Volatilität an den Finanzmärkten sowie die steigenden Inflationsraten sich auf die Investorentätigkeit signifikant ausgewirkten. Die weiteren Parameter zeigten zum Teil verschiedene Signifikanzen und Koeffizienten, bei denen sich die Einflüsse nicht zweifelsfrei feststellen lassen.

Um eine Schätzung des Handelsvolumens vorzunehmen, wird aufgrund der gemachten Untersuchung folgendes OLS-Regressions Modell mit einem Trainingszeitraum von 96 Monaten und einem Schätzzeitraum von sechs Monaten vorgeschlagen:

$$\gamma = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \varepsilon$$

γ = *Delta Retailvolumen US*

x_1 = *Delta S&P500Low*

x_2 = *Delta S&P500High*

x_3 = *Delta UST 2Y Rate*

x_4 = *Delta GDP*

x_5 = *Delta New Home Sales – Units*

x_6 = *Delta CPI Core*

x_7 = *Delta 5 Year Inflation Expectations (median)*

x_8 = *Delta Retail Sales YoY*

x_9 = *Delta 1MLagged_S&P500 Months End*

Obwohl das SARIMAX-Modell in der Untersuchung einen tieferen durchschnittlichen MSE erreichte, ist eine Wahl des OLS-Regressions Modell unter Anbetracht der tieferen Modellkomplexität, dem einfacheren Verständnis der Einflüsse der Modellparameter und dem kleineren Datenbedarf sinnvoll.

Die gewählten Parameter können auch zur Schätzung und Erklärung des Handelsvolumens von Retailinvestoren herangezogen werden. Sie erklären die Veränderungen des Orderroutingertrags nur zu 30 Prozent im Vergleich zu 38.9 Prozent zum Retailhandelsvolumen. Die Regression ist aber in ihrer Gesamtheit signifikant. Somit können Rückschlüsse auf die Erträge von Online-Brokern gezogen werden, was eine gewisse praktische Relevanz besitzt.

Zusammenfassend lässt sich sagen, dass die vorliegende Arbeit aufzuzeigen vermochte, von welchen exogenen Faktoren die Handelsumsätze von Retailinvestoren abhängen. Somit konnte die Forschungsfrage beantwortet werden. Zudem liefert die Arbeit ein einfach anzuwendendes Modell, mit welchem einerseits die zukünftige Veränderung des Volumens von Privatinvestoren oder andererseits die Veränderung von Orderroutingerträge für die nächsten sechs Monate geschätzt werden können.

6.2. Restriktion der Resultate

Die empirische Untersuchung trägt dazu bei, ein Verständnis dafür zu entwickeln, welche makroökonomische Parameter Zusammenhänge mit dem Handelsvolumen von Retailinvestoren aufweisen und diese erklären können. Zudem konnte aufgezeigt werden, dass Orderrouterträge von Online-Broker ähnliche Zusammenhänge mit denselben Parametern aufweisen. Allerdings müssen die erhaltenen Ergebnisse aufgrund folgender Punkte limitiert werden:

Die gemachten Analysen beruhen auf den Off-Exchange Handelsdaten der drei Off-Exchange Börsen in den USA. Obwohl ein Grossteil des Handels von Retailinvestoren in den USA über dieses System läuft, sind bestimmte Handelsklassen (Krypto) davon ausgeschlossen. Infolgedessen kann eine Verzerrung der Resultate im Hinblick auf den gesamten Retailhandel mit (digitalen) Vermögenswerten auftreten.

Die Orderrouterträge wurden nur für vier Online-Brokerunternehmen in den USA erhoben und dies aufgrund von fehlenden Daten auch nur für einen Zeitraum von 2020 – 2022. Somit sind die gefunden Zusammenhänge und Resultate nur beschränkt repräsentativ für die gesamte Online-Brokerage Industrie in den USA. Zudem scheinen sich die Geschäftsmodelle der jeweiligen Unternehmen stark zu unterscheiden.

6.3. Implikationen

Die gemachte Untersuchung beschränkt sich nur auf US-amerikanische Retailinvestoren und die Beobachtungen können nur auf diese Gruppe von Individuen angewandt werden. Aufgrund dieses Umstandes könnte eine fortführende Arbeit den geographischen Fokus auf andere Weltgegenden (Europa, Asien) legen. Ausserdem erklärt das erarbeitete Modell nur beschränkt die Schwankungen des Handelsvolumens respektive der Orderrouterträge. Hier könnten zusätzliche Untersuchungen durchgeführt werden, um das Modell zu verbessern. Zudem könnte die Untersuchung der Orderrouterträge auf die gesamte Industrie ausgeweitet werden, nicht nur auf einen Bruchteil der im Markt aktiven Unternehmen.

Die Implikationen für die Praxis beziehen sich vor allem auf die Schätzung von zukünftigen Erträgen von börsenkotierten Online-Broker-Unternehmen. Die Erträge aus dem Orderrouting sind für einzelne Online-Broker sehr relevant und machen abhängig vom Geschäftsmodell einen grossen Anteil des Umsatzes aus. In Anbetracht dessen, dass ein Teil dieser Unternehmen börsenkotiert ist, können mittels des erarbeiteten Modells

zukünftige Erträge abgeschätzt werden. Dies kann dabei helfen bessere Investitionsentscheidungen bei diesen Unternehmen zu treffen.

7. Quellenverzeichnis

- AAII. (2023). *AAII Investor Sentiment Survey*. <https://www.aaii.com/sentimentsurvey>
- Abudy, M. M. (2020). Retail investors' trading and stock market liquidity. *The North American Journal of Economics and Finance*, 54, 1-12-
- Alderighi, S. (2018). The determinants of retail trading activity in emerging markets: A cross-market analysis. *Global Finance Journal*, 37, 152–167.
- Andersen, S., Hanspal, T., & Nielsen, K. M. (2019). Once bitten, twice shy: The power of personal experiences in risk taking. *Journal of Financial Economics*, 132(3), 97-117.
- Barrot, J.-N., Kaniel, R., & Sraer, D. (2016). Are retail traders compensated for providing liquidity?. *Journal of Financial Economics*, 120, 146–168.
- Bellofatto, A., D'Hondt, C., & De Winne, R. (2018). Subjective financial literacy and retail investors' behavior. *Journal of Banking & Finance*, 92, 168–181.
- Berger, D. (2022). Investor sentiment: A retail trader activity approach. *Review of Accounting and Finance*, 21(2), 61–82.
- Best, H., & Wolf, C. (Hrsg.). (2015). *The handbook of regression analysis and causal inference*. SAGE Reference.
- Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2021). Tracking Retail Investor Activity. *The Journal of Finance*, 76(5), 2249–2305.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008a). Linear Stationary Models. In *Time Series Analysis* (S. 47–91). John Wiley & Sons, Ltd.
- Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (2008b). Seasonal Models. In *Time Series Analysis* (S. 353–411). John Wiley & Sons, Ltd.
- Campbell, J. Y., Ramadorai, T., & Schwartz, A. (2009). Caught on tape: Institutional trading, stock returns, and earnings announcements. *Journal of Financial Economics*, 92(1), 66–91.

- Chen, S.-S. (2012). Revisiting the empirical linkages between stock returns and trading volume. *Journal of Banking & Finance*, 36(6), 1781–1788.
- Ernst, T., & Spatt, C. S. (2022). Payment for Order Flow and Asset Choice. *National Bureau of Economic Research*, 29883, 1–57.
- FINRA. (2023). *National Market System Plans*. von <https://www.finra.org/rules-guidance/guidance/national-market-system-plans>.
- FRED. (2023). *FRED Economic Data, St Louis FED*. <https://fred.stlouisfed.org/>
- Fünfgeld, B., & Wang, M. (2009). Attitudes and behaviour in everyday finance: Evidence from Switzerland. *International Journal of Bank Marketing*, 27(2), 108–128.
- Keltner, D., & Lerner, J. (2001). Fear, Anger, and Risk. *Journal of Personality and Social Psychology*, 81(1), 146–159.
- Kuhnen, C. M., & Knutson, B. (2011). The Influence of Affect on Beliefs, Preferences, and Financial Decisions. *Journal of Financial and Quantitative Analysis*, 46(3), 605–626.
- Newbold, P., Carlson, W., & Thorne, B. (2019). *Statistics for Business and Economics, Global Edition*. Pearson Education, Limited.
- Ortmann, R., Pelster, M., & Wengerek, S. T. (2020). COVID-19 and investor behavior. *Finance Research Letters*, 37(101717), 1-10.
- Pagano, M. S., Sedunov, J., & Velthuis, R. (2021). How did retail investors respond to the COVID-19 pandemic? The effect of Robinhood brokerage customers on market quality. *Finance Research Letters*, 43(101946), 1-11.
- Rakowski, D., Shirley, S. E., & Stark, J. R. (2021). Twitter activity, investor attention, and the diffusion of information. *Financial Management (Wiley-Blackwell)*, 50(1), 3–46.
- Robinhood. (2023). *Investor Relations*. <https://investors.robinhood.com/overview/default.aspx>.

- Roy, K. (2022). *Rise of the retail investor*. S&P Global Market Intelligence.
- SEC. (2023). *SEC Rule 606 and 607 Disclosure Report RHF Q4 2022*. Securities and Exchange Commission
- Seth, H., Talwar, S., Bhatia, A., Saxena, A., & Dhir, A. (2020). Consumer resistance and inertia of retail investors: Development of the resistance adoption inertia continuance (RAIC) framework. *Journal of Retailing and Consumer Services*, 55(102071), 1-14.
- Shim, S., Xiao, J. J., Barber, B. L., & Lyons, A. C. (2009). Pathways to life success: A conceptual model of financial well-being for young adults. *Journal of Applied Developmental Psychology*, 30(6), 708–723.
- Talwar, M., Talwar, S., Kaur, P., Tripathy, N., & Dhir, A. (2021). Has financial attitude impacted the trading activity of retail investors during the COVID-19 pandemic? *Journal of Retailing and Consumer Services*, 58, 1–11.
- Talwar, S., Talwar, M., Tarjanne, V., & Dhir, A. (2021). Why retail investors traded equity during the pandemic? An application of artificial neural networks to examine behavioral biases. *Psychology & Marketing*, 38(11), 2142–2163.
- The Bank of New York Mellon Corporation. (2022). *The State of the U.S. Retail Investor: Insights & Implications*.
- Tiao, G. C. (2000). Univariate Autoregressive Moving-Average Models. In *A Course in Time Series Analysis* (S. 53–85). John Wiley & Sons, Ltd.
- US Census Bureau. (2023a). *North American Industry Classification System (NAICS)*. <https://www.census.gov/naics/?input=5231&year=2017&details=5231>
- US Census Bureau, U. C. (2023b). *Service Annual Survey Latest Data (NAICS-basis): 2021*. <https://www.census.gov/data/tables/2021/econ/services/sas-naics.html>
- Wah, E., Feldmann, S., Chung, F., Bishop, A., & Aisen, D. (2018). A Comparison of Execution Quality across US Stock Exchanges. In *Global Algorithmic Capital*

Markets: High Frequency Trading, Dark Pools, and Regulatory Challenges (1., S. 91–146). Oxford University Press.

Xu, R., Liu, Y., Hu, N., & Guo, J. (Michael). (2022). What drives individual investors in the bear market? *The British Accounting Review*, 54(6), 1-22.

Yoon, J., & Oh, G. (2022). Investor herding behavior in social media sentiment. *Frontiers in Physics*, 10, 1-16.

8. Anhang

8.1. Korrelationsmatrix Modellparameter

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 1.000 | 0.070 | -0.195 | -0.355 | 0.110 | -0.161 | -0.099 | -0.085 | 0.081 | 0.012 | 0.063 | 0.036 | 0.042 | 0.073 | 0.101 | -0.024 | -0.038 | -0.097 | 0.153 | 0.105 | 0.065 |
| 2 | 0.070 | 1.000 | -0.147 | -0.179 | -0.083 | -0.047 | -0.040 | 0.005 | 0.157 | -0.148 | -0.249 | -0.190 | -0.104 | -0.064 | -0.092 | -0.211 | -0.044 | -0.043 | -0.041 | -0.007 | 0.067 |
| 3 | -0.195 | -0.147 | 1.000 | 0.602 | 0.484 | 0.494 | 0.107 | 0.144 | -0.085 | 0.065 | 0.003 | 0.213 | 0.069 | -0.118 | 0.063 | -0.043 | -0.145 | 0.107 | -0.127 | -0.092 | 0.036 |
| 4 | -0.355 | -0.179 | 0.602 | 1.000 | 0.587 | 0.340 | 0.088 | 0.153 | -0.211 | 0.176 | 0.195 | 0.256 | 0.120 | -0.131 | 0.053 | -0.028 | -0.166 | 0.024 | 0.411 | -0.159 | -0.129 |
| 5 | 0.110 | -0.083 | 0.484 | 0.587 | 1.000 | 0.208 | 0.001 | 0.032 | -0.181 | 0.348 | 0.316 | 0.163 | 0.233 | -0.033 | -0.015 | 0.058 | -0.110 | -0.285 | 0.583 | 0.017 | -0.025 |
| 6 | -0.161 | -0.047 | 0.494 | 0.340 | 0.208 | 1.000 | -0.022 | 0.112 | -0.043 | 0.062 | 0.000 | 0.035 | 0.038 | -0.086 | -0.073 | -0.005 | -0.038 | 0.022 | -0.262 | -0.102 | 0.125 |
| 7 | -0.099 | -0.040 | 0.107 | 0.088 | 0.001 | -0.022 | 1.000 | 0.701 | -0.306 | 0.119 | 0.060 | -0.039 | 0.082 | 0.094 | 0.089 | 0.166 | 0.125 | -0.102 | 0.019 | 0.008 | 0.015 |
| 8 | -0.085 | 0.005 | 0.144 | 0.153 | 0.032 | 0.112 | 0.701 | 1.000 | -0.193 | 0.083 | 0.008 | 0.011 | 0.024 | -0.004 | 0.094 | 0.025 | -0.042 | 0.004 | -0.016 | 0.018 | -0.029 |
| 9 | 0.081 | 0.157 | -0.085 | -0.211 | -0.181 | -0.043 | -0.306 | -0.193 | 1.000 | -0.578 | -0.463 | -0.061 | -0.150 | -0.152 | 0.060 | -0.109 | -0.238 | 0.502 | -0.215 | -0.069 | 0.058 |
| 10 | 0.012 | -0.148 | 0.065 | 0.176 | 0.348 | 0.062 | 0.119 | 0.083 | -0.578 | 1.000 | 0.723 | 0.202 | 0.339 | 0.273 | -0.071 | 0.318 | 0.160 | -0.777 | 0.302 | 0.233 | 0.012 |
| 11 | 0.063 | -0.249 | 0.003 | 0.195 | 0.316 | 0.000 | 0.060 | 0.008 | -0.463 | 0.723 | 1.000 | 0.297 | 0.336 | 0.235 | -0.059 | 0.313 | 0.055 | -0.771 | 0.378 | 0.088 | -0.106 |
| 12 | 0.036 | -0.190 | 0.213 | 0.256 | 0.163 | 0.035 | -0.039 | 0.011 | -0.061 | 0.202 | 0.297 | 1.000 | -0.066 | -0.051 | -0.102 | -0.049 | -0.149 | -0.129 | 0.079 | 0.075 | -0.220 |
| 13 | 0.042 | -0.104 | 0.069 | 0.120 | 0.233 | 0.038 | 0.082 | 0.024 | -0.150 | 0.339 | 0.336 | -0.066 | 1.000 | 0.313 | 0.029 | 0.425 | 0.086 | -0.286 | 0.222 | 0.078 | 0.125 |
| 14 | 0.073 | -0.064 | -0.118 | -0.131 | -0.033 | -0.086 | 0.094 | -0.004 | -0.152 | 0.273 | 0.235 | -0.051 | 0.313 | 1.000 | -0.045 | 0.546 | 0.326 | -0.266 | -0.041 | 0.055 | 0.069 |
| 15 | 0.101 | -0.092 | 0.063 | 0.053 | -0.015 | -0.073 | 0.089 | 0.094 | 0.060 | -0.071 | -0.059 | -0.102 | 0.029 | -0.045 | 1.000 | -0.030 | -0.040 | 0.187 | 0.087 | -0.045 | -0.015 |
| 16 | -0.024 | -0.211 | -0.043 | -0.028 | 0.058 | -0.005 | 0.166 | 0.025 | -0.109 | 0.318 | 0.313 | -0.049 | 0.425 | 0.546 | -0.030 | 1.000 | 0.691 | -0.348 | 0.054 | 0.047 | 0.000 |
| 17 | -0.038 | -0.044 | -0.145 | -0.166 | -0.110 | -0.038 | 0.125 | -0.042 | -0.238 | 0.160 | 0.055 | -0.149 | 0.086 | 0.326 | -0.040 | 0.691 | 1.000 | -0.248 | -0.081 | -0.033 | 0.008 |
| 18 | -0.097 | -0.043 | 0.107 | 0.024 | -0.285 | 0.022 | -0.102 | 0.004 | 0.502 | -0.777 | -0.771 | -0.129 | -0.286 | -0.266 | 0.187 | -0.348 | -0.248 | 1.000 | -0.285 | -0.230 | 0.007 |
| 19 | 0.153 | -0.041 | -0.127 | 0.411 | 0.583 | -0.262 | 0.019 | -0.016 | -0.215 | 0.302 | 0.378 | 0.079 | 0.222 | -0.041 | 0.087 | 0.054 | -0.081 | -0.285 | 1.000 | -0.097 | -0.107 |
| 20 | 0.105 | -0.007 | -0.092 | -0.159 | 0.017 | -0.102 | 0.008 | 0.018 | -0.069 | 0.233 | 0.088 | 0.075 | 0.078 | 0.055 | -0.045 | 0.047 | -0.033 | -0.230 | -0.097 | 1.000 | -0.133 |
| 21 | 0.065 | 0.067 | 0.036 | -0.129 | -0.025 | 0.125 | 0.015 | -0.029 | 0.058 | 0.012 | -0.106 | -0.220 | 0.125 | 0.069 | -0.015 | 0.000 | 0.008 | 0.007 | -0.107 | -0.133 | 1.000 |

Legende: (1) ChngTNV, 2) ChngRDI, (3) ChngS&P500, (4) ChngS&P500Low, (5) ChngS&P500High, (6) ChngSentimentBullBear, (7) Chng2YR UST, (8) Chng10YR UST, (9) ChngM3, (10) ChngGDP, (11) ChngPersConsump, (12) ChngNewHomes, (13) ChngCPI, (14) ChngCPICore, (15) ChngInflExp, (16) ChngRetailSales, (17) ChngIndProd, (18) ChngUnemp, (19) 1MLagged_ChngS&P500, (20) 2MLagged_ChngS&P500, (21) 3MLagged_ChngS&P500

8.2. OLS-Regression alle Modellparameter (normalisierte Werte)

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|----------|-------|--------|--------|
| Dep. Variable: | ChngTNV | R-squared: | 0.400 | | | |
| Model: | OLS | Adj. R-squared: | 0.324 | | | |
| Method: | Least Squares | F-statistic: | 5.253 | | | |
| Date: | Mon, 12 Jun 2023 | Prob (F-statistic): | 9.50e-09 | | | |
| Time: | 11:30:10 | Log-Likelihood: | -176.86 | | | |
| No. Observations: | 152 | AIC: | 389.7 | | | |
| Df Residuals: | 134 | BIC: | 444.2 | | | |
| Df Model: | 17 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| const | 0 | 0.067 | 0 | 1.000 | -0.132 | 0.132 |
| ChngRDI | 0.0005 | 0.073 | 0.007 | 0.995 | -0.143 | 0.144 |
| ChngS&P500 | -0.0495 | 0.127 | -0.391 | 0.697 | -0.300 | 0.201 |
| ChngS&P500Low | -0.7572 | 0.115 | -6.598 | 0.000 | -0.984 | -0.530 |
| ChngS&P500High | 0.3840 | 0.123 | 3.134 | 0.002 | 0.142 | 0.626 |
| ChngSentimentBullBear | 0.0824 | 0.084 | 0.978 | 0.330 | -0.084 | 0.249 |
| Chng2YR | -0.0503 | 0.074 | -0.679 | 0.498 | -0.197 | 0.096 |
| ChngM3 | -0.1213 | 0.121 | -1.001 | 0.319 | -0.361 | 0.118 |
| ChngGDP | -0.1938 | 0.109 | -1.773 | 0.078 | -0.410 | 0.022 |
| ChngNewHomes | 0.1510 | 0.077 | 1.956 | 0.052 | -0.002 | 0.304 |
| ChngCPI | 0.0148 | 0.081 | 0.182 | 0.856 | -0.146 | 0.175 |
| ChngCPICore | 0.1223 | 0.083 | 1.473 | 0.143 | -0.042 | 0.287 |
| ChngInflExp | 0.1356 | 0.071 | 1.905 | 0.059 | -0.005 | 0.276 |
| ChngRetailSales | -0.1242 | 0.092 | -1.357 | 0.177 | -0.305 | 0.057 |
| 1MLagged_ChngS&P500 | 0.2935 | 0.128 | 2.290 | 0.024 | 0.040 | 0.547 |
| 2MLagged_ChngS&P500 | 0.0103 | 0.075 | 0.137 | 0.891 | -0.139 | 0.159 |
| 3MLagged_ChngS&P500 | -0.0092 | 0.073 | -0.126 | 0.900 | -0.154 | 0.136 |
| COVID19 | 0.1087 | 0.095 | 1.138 | 0.257 | -0.080 | 0.298 |
| Omnibus: | 6.983 | Durbin-Watson: | 2.723 | | | |
| Prob(Omnibus): | 0.030 | Jarque-Bera (JB): | 3.535 | | | |
| Skew: | -0.109 | Prob(JB): | 0.171 | | | |
| Kurtosis: | 2.286 | Cond. No. | 5.51 | | | |

8.3. OLS-Regression nach L1-Regularisierung (normalisierte Werte)

| OLS Regression Results | | | | | | |
|------------------------|------------------|---------------------|----------|-------|--------|--------|
| Dep. Variable: | ChngTNV | R-squared: | 0.389 | | | |
| Model: | OLS | Adj. R-squared: | 0.350 | | | |
| Method: | Least Squares | F-statistic: | 10.04 | | | |
| Date: | Thu, 08 Jun 2023 | Prob (F-statistic): | 7.56e-12 | | | |
| Time: | 15:05:51 | Log-Likelihood: | -178.26 | | | |
| No. Observations: | 152 | AIC: | 376.5 | | | |
| Df Residuals: | 142 | BIC: | 406.8 | | | |
| Df Model: | 9 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| const | 0 | 0.066 | 0 | 1.000 | -0.130 | 0.130 |
| ChngS&P500Low | -0.7261 | 0.084 | -8.647 | 0.000 | -0.892 | -0.560 |
| ChngS&P500High | 0.3815 | 0.091 | 4.207 | 0.000 | 0.202 | 0.561 |
| Chng2YR | -0.0395 | 0.068 | -0.584 | 0.560 | -0.173 | 0.094 |
| ChngGDP | -0.1207 | 0.080 | -1.508 | 0.134 | -0.279 | 0.037 |
| ChngNewHomes | 0.1349 | 0.070 | 1.934 | 0.055 | -0.003 | 0.273 |
| ChngCPICore | 0.1267 | 0.079 | 1.603 | 0.111 | -0.030 | 0.283 |
| ChngInflExp | 0.1311 | 0.067 | 1.953 | 0.053 | -0.002 | 0.264 |
| ChngRetailSales | -0.0978 | 0.079 | -1.237 | 0.218 | -0.254 | 0.058 |
| 1MLagged_ChngS&P500 | 0.2814 | 0.081 | 3.468 | 0.001 | 0.121 | 0.442 |
| Omnibus: | 3.998 | Durbin-Watson: | 2.693 | | | |
| Prob(Omnibus): | 0.135 | Jarque-Bera (JB): | 2.565 | | | |
| Skew: | -0.112 | Prob(JB): | 0.277 | | | |
| Kurtosis: | 2.405 | Cond. No. | 2.63 | | | |

8.4. SARIMAX-Modell vor Grid Search (normalisierte Werte)

```

SARIMAX Results
=====
Dep. Variable:          ChngTNV      No. Observations:      152
Model:                 SARIMAX(0, 0, 1)  Log Likelihood         -153.498
Date:                  Mon, 12 Jun 2023  AIC                    330.996
Time:                  11:11:54        BIC                    367.124
Sample:                05-31-2009      HQIC                   345.674
                    - 12-31-2021
Covariance Type:      opg
=====
              coef  std err      z  P>|z|    [0.025    0.975]
-----
intercept      0.0117    0.017    0.706    0.480    -0.021    0.044
ChngS&P500Low -0.7987    0.087   -9.136    0.000    -0.970   -0.627
ChngS&P500High  0.5042    0.090    5.581    0.000    0.327    0.681
Chng2YR        -0.0020    0.046   -0.043    0.966    -0.092    0.088
ChngGDP        -0.1418    0.058   -2.450    0.014    -0.255   -0.028
ChngNewHomes   0.1600    0.059    2.694    0.007    0.044    0.276
ChngCPICore    0.0391    0.057    0.686    0.493    -0.072    0.151
ChngInflExp    0.1523    0.090    1.693    0.090    -0.024    0.329
ChngRetailSales -0.0641    0.045   -1.416    0.157    -0.153    0.025
1MLagged_ChngS&P500  0.2819    0.099    2.843    0.004    0.088    0.476
ma.L1          -0.7401    0.063  -11.796    0.000   -0.863   -0.617
sigma2         0.4509    0.059    7.629    0.000    0.335    0.567
=====
Ljung-Box (L1) (Q):          1.74  Jarque-Bera (JB):          0.17
Prob(Q):                     0.19  Prob(JB):                  0.92
Heteroskedasticity (H):      0.96  Skew:                      0.07
Prob(H) (two-sided):         0.89  Kurtosis:                   3.07
=====

```

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

8.5. SARIMAX-Modell nach Grid Search (normalisierte Werte)

```

SARIMAX Results
=====
Dep. Variable:          ChngTNV      No. Observations:      152
Model:                 SARIMAX(3, 0, 3)x(0, 1, [1, 2], 12)  Log Likelihood         -87.598
Date:                  Mon, 12 Jun 2023  AIC                    213.196
Time:                  11:14:20        BIC                    264.847
Sample:                05-31-2009      HQIC                   234.152
                    - 12-31-2021
Covariance Type:      opg
=====
              coef  std err      z  P>|z|    [0.025    0.975]
-----
intercept      0.0056    0.004    1.271    0.204    -0.003    0.014
ChngS&P500Low -0.6461    0.078   -8.261    0.000   -0.799   -0.493
ChngS&P500High  0.4519    0.083    5.423    0.000    0.289    0.615
Chng2YR        -0.0340    0.041   -0.837    0.402    -0.114    0.046
ChngGDP        -0.1201    0.054   -2.244    0.025   -0.225   -0.015
ChngNewHomes   0.1023    0.062    1.648    0.099   -0.019    0.224
ChngCPICore    0.0216    0.074    0.291    0.771   -0.124    0.167
ChngInflExp    0.2194    0.070    3.148    0.002    0.083    0.356
ChngRetailSales -0.0504    0.043   -1.186    0.236   -0.134    0.033
1MLagged_ChngS&P500  0.2096    0.092    2.274    0.023    0.029    0.390
ar.L1          -0.4849    0.191   -2.541    0.011   -0.859   -0.111
ar.L2          -0.2659    0.197   -1.348    0.178   -0.652    0.121
ar.L3          0.4843    0.130    3.738    0.000    0.230    0.738
ma.L1          -0.1704    0.361   -0.473    0.636   -0.877    0.536
ma.L2          -0.3168    0.313   -1.011    0.312   -0.931    0.297
ma.L3          -0.5318    0.295   -1.800    0.072   -1.111    0.047
ma.S.L12       -0.6940    0.157   -4.412    0.000   -1.002   -0.386
ma.S.L24       -0.1238    0.199   -0.623    0.534   -0.513    0.266
sigma2         0.2503    0.112    2.231    0.026    0.030    0.470
=====
Ljung-Box (L1) (Q):          0.26  Jarque-Bera (JB):          0.83
Prob(Q):                     0.61  Prob(JB):                  0.66
Heteroskedasticity (H):      1.90  Skew:                      -0.09
Prob(H) (two-sided):         0.05  Kurtosis:                   3.38
=====

```

Warnings:

[1] Covariance matrix calculated using the outer product of gradients (complex-step).

8.6. Python Code

1. Laden des Datensatzes ¶

```
df = pd.read_excel('MonthlyData.xlsx')  
df = df.set_index(["Date"]) # Setzen des Indexes
```

Einfügen von Lagged Variablen für S&P 500

```
df['1MLagged_S&P500'] = df['S&P500 Months End'].shift(1)  
df['2MLagged_S&P500'] = df['S&P500 Months End'].shift(2)  
df['3MLagged_S&P500'] = df['S&P500 Months End'].shift(3)
```

2. Datenanalyse

```
df.dropna(axis='rows',inplace=True, how='any') #Zeilen, welche mind. 1 x NaN enthalten werden gelö
```

2.1. Auf Stationarität prüfen

```
variablestocheck=['Total Notional', 'Real Disposable Income', 'S&P500 Months End',  
                  'S&P 500 Low', 'S&P 500 High', 'DeltaSentimentIndexBullBearSpread',  
                  'UST 2Y', 'UST 10Y', 'Geldmenge M3 US', 'GDP',  
                  'US Personal Consumption Expenditure Change', 'New Home Sales-Units',  
                  'CPI, % Year on year', 'Core CPI Index',  
                  '5 Year Inflation Expectations (median), in %', 'Retail Sales YoY',  
                  'Industrial Production Index, % YoY', 'UnemploymentRate',  
                  '1MLagged_S&P500', '2MLagged_S&P500', '3MLagged_S&P500']  
  
#Loop  
for i in variablestocheck:  
    for lags in range(1,3):  
        print('Number of lags used:',lags)  
  
        adf_library = adfuller(df[i], maxlag=lags, regression='n',autolag=None)  
        adf_library_d = adfuller(np.diff(df[i]), maxlag=lags, regression='n',autolag=None)  
  
        print(i,'in Level','\n',  
              'ADF Stat (Library) for',i,':',"% .2f" % adf_library[0],'\n',  
              'ADF P-val (Library)',i,':',"% .2f" % adf_library[1],'\n')  
  
        print(i,'in 1st Difference','\n',  
              'ADF Stat (Library) for',i,':',"% .2f" % adf_library_d[0],'\n',  
              'ADF P-val (Library)',i,':',"% .2f" % adf_library_d[1],'\n')  
    print('_____','\n')
```

2.2 Stationär machen

```
df1 = df

# Spalten welche Stationär gemacht werden sollen
columns_to_transform = ['Total Notional', 'Real Disposable Income', 'S&P500 Months End',
                        'S&P 500 Low', 'S&P 500 High',
                        'UST 2Y', 'UST 10Y', 'Geldmenge M3 US',
                        'New Home Sales-Units',
                        'Core CPI Index',
                        '5 Year Inflation Expectations (median), in %',
                        'unemploymentRate',
                        '1MLagged_S&P500', '2MLagged_S&P500', '3MLagged_S&P500']

df1[columns_to_transform] = df1[columns_to_transform].apply(lambda x: np.log(x + 1e-8) - np.log(x.s

df1.dropna(axis='rows',inplace=True, how='any') #Zeilen, welche mind. 1 x NaN enthalten werden gelö

df1.columns = ['ChngTNV', "ChngRDI", "ChngS&P500", "ChngS&P500Low", "ChngS&P500High", "ChngSentimer
```

2.3 Einfügen der Dummyvariablen COVID 19

```
df1['COVID19'] = 0
df1.loc[df1.index >= '2020-01-31', 'COVID19'] = 1
```

2.4 Variablen auf Korrelation überprüfen

```
corr = df1.corr()
corr.style.background_gradient(cmap='coolwarm')
```

Überprüfung der Verteilung

```
Variablen = ['ChngRDI', 'ChngS&P500', 'ChngS&P500Low', 'ChngS&P500High',
            'ChngSentimentBullBear', 'Chng2YR', 'ChngM3', 'ChngGDP', 'ChngNewHomes',
            'ChngCPI', 'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
            '1MLagged_ChngS&P500', '2MLagged_ChngS&P500', '3MLagged_ChngS&P500',
            'COVID19']

for i in Variablen:
    print('Variable:',i)
    stat,p=sp.stats.shapiro(descriptive1[i])

    print("p-value for Shapiro-Wilk: ")
    print(p)

    alpha = 0.05
    if p > alpha:
        print('Normalverteilt')
    else:
        print('Nicht normalverteilt')
        if (stats.skew(descriptive1[i]))>0:
            print('rechtsschief')
        else:
            print('linksschief')

    print('_____','\n')
```

4. OLS Regression

4.1 OLS Regression mittels Traindaten von 2009 - 2021 und Testset von 2022 (inkl Lasso Regression)

```
train_data1 = df1[:, '2021']
test_data1 = df1[:, '2022']

from sklearn.preprocessing import StandardScaler #normalisierung der Daten aufgrund nicht normalver
scaler = StandardScaler()
df_train_scaled = pd.DataFrame(scaler.fit_transform(train_data1), columns=train_data1.columns, index=
df_test_scaled = pd.DataFrame(scaler.transform(test_data1), columns=test_data1.columns, index=test_

# Definieren der abhängigen (endog) und unabhängigen (exog) Variablen
exog_vars = ['ChngRDI', 'ChngS&P500', 'ChngS&P500Low', 'ChngS&P500High',
            'ChngSentimentBullBear', 'Chng2YR', 'ChngM3', 'ChngGDP', 'ChngNewHomes',
            'ChngCPI', 'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
            '1MLagged_ChngS&P500', '2MLagged_ChngS&P500', '3MLagged_ChngS&P500']
endog_var = ['ChngTNV']

X = df_train_scaled[exog_vars]
y = df_train_scaled[endog_var]

# Einfügen der Konstante zu den unabhängigen Variablen
X = sm.add_constant(X)

# Durchführen der OLS-Regression
model = sm.OLS(y, X)
results = model.fit()

# Ergebnisse
print(results.summary())
```

L1 regularisierung zur Einschränkung der relevanten Parameter

```
from sklearn.linear_model import LassoCV

# Definieren der unabhängigen (X) und abhängigen Variable (y)
X_train = df_train_scaled[exog_vars]
y_train = df_train_scaled[endog_var].values.ravel() # ravel wird verwendet, um die Dimension von y

# Erstellen und fitten des LassoCV-Modell
lasso = LassoCV(cv=5, random_state=0)
lasso.fit(X_train, y_train)

# Ausgabe der Koeffizienten
coefficients = lasso.coef_

# Verbinden der Namen der Merkmale mit ihren Koeffizienten
coef_with_names = list(zip(exog_vars, coefficients))

# Drucken der relevanten Merkmale (mit Koeffizienten ungleich Null)
relevant_features = [pair for pair in coef_with_names if pair[1] != 0]

print("Relevante Modellparameter (Feature, Koeffizient):")
for feature in relevant_features:
    print(feature)
```

4.2 Rollende Anwendung verschiedener Zeiträume des Test sowie Trainsets

Rollende OLS auf verschiedene Windows

```
M train_window_size = 132 #anpassung der Zeiträume
prediction_window_size = 12

# Rollendes Fenster
start_date = pd.to_datetime('2009-05-31')
end_date = pd.to_datetime('2022-12-31')
current_date = start_date

mse_list = [] # Liste für MSE

while current_date + pd.DateOffset(months=train_window_size + prediction_window_size) <= end_date:
    # Definition der Start und Enddaten
    train_start = current_date
    train_end = current_date + pd.DateOffset(months=train_window_size) - pd.DateOffset(days=1)
    test_start = current_date + pd.DateOffset(months=train_window_size)
    test_end = current_date + pd.DateOffset(months=train_window_size + prediction_window_size) - pd

    # Extrahieren des Zeitfensters
    train_data = df2[train_start:train_end]
    test_data = df2[test_start:test_end]

    # vorbereiten der Trainingsdaten
    X_train = train_data[['ChngS&P500Low', 'ChngS&P500High',
                          'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                          'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                          '1MLagged_ChngS&P500',]]
    y_train = train_data["ChngTNV"]
    X_train = sm.add_constant(X_train)

    # Modell trainieren
    model = sm.OLS(y_train, X_train).fit()

    # Vorbereiten der Testdaten
    X_test = test_data[['ChngS&P500Low', 'ChngS&P500High',
                        'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                        'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                        '1MLagged_ChngS&P500',]]
    y_test = test_data['ChngTNV']
    X_test = sm.add_constant(X_test)

    # Vorhersagen machen
    y_pred = model.predict(X_test)

    # MSE für jetzige Vorhersage
    mse = mean_squared_error(y_test, y_pred)
    mse_list.append(mse)

    # Nächstes Fenster
    current_date += pd.DateOffset(months=1)

# Berechnen des durchschnittlichen MSE
average_mse = sum(mse_list) / len(mse_list)
print(f"Average MSE: {average_mse}")
```

5. ARIMA Modell

5.1 SARIMAX mittels Traindaten von 2009 - 2021 und Testset von 2022

```
train_data2 = df2[:'2021']
test_data2 = df2['2022':]

from sklearn.preprocessing import StandardScaler #normalisieren

scaler = StandardScaler()
df_train_scaled = pd.DataFrame(scaler.fit_transform(train_data2), columns=train_data2.columns, index=train_data2.index)
df_test_scaled = pd.DataFrame(scaler.transform(test_data2), columns=test_data2.columns, index=test_data2.index)
```

Durchführen Basis SARIMAX-Modell

```
# Verwenden des Trainingsset im SARIMAX-Modell
model = sm.tsa.statespace.SARIMAX(df_train_scaled['ChngTNV'],
                                  exog=df_train_scaled[['ChngS&P500Low', 'ChngS&P500High',
                                                         'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                         'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                                                         '1MLagged_ChngS&P500']],
                                  order=(0, 0, 1),
                                  seasonal_order=(0,0,0, 12),
                                  trend='c',
                                  enforce_stationarity=False,
                                  enforce_invertibility=False)

# Trainieren des Modells mit den aktualisierten Daten
results = model.fit()

# Verwenden des Modells, um Vorhersagen für den Testzeitraum zu erstellen
predictions = results.predict(start='2022-01-31', end='2022-12-31',
                              exog=df_test_scaled[['ChngS&P500Low', 'ChngS&P500High',
                                                    'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                    'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                                                    '1MLagged_ChngS&P500']])

# Berechnen des mittleren quadratischen Fehler (Mean Squared Error, MSE)
mse_Sarima = ((predictions - df_test_scaled['ChngTNV']) ** 2).mean()

# Ausgabe des MSE
mse_Sarima
```

5.2 Durchführen Grid-Search für optimales SARIMAX-Modell

```
import itertools

# Definieren des Rasters für die Hyperparameter
p = [0, 1, 2, 3]
d = [0, 1]
q = [0, 1, 2, 3]
P = [0, 1, 2]
D = [0, 1]
Q = [0, 1, 2]
S = [12]

# Erzeugen aller möglichen Kombinationen der Hyperparameter
hyperparameters = list(itertools.product(p, d, q, P, D, Q, S))

best_mse = float('inf')
best_params = None

# Grid Search über alle Hyperparameterkombinationen
for params in hyperparameters:
    try:
        model = sm.tsa.statespace.SARIMAX(df_train_scaled['ChngTNV'],
                                          exog=df_train_scaled[['ChngS&P500Low', 'ChngS&P500High',
                                                                'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                                'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                                                                '1MLagged_ChngS&P500']],
                                          order=(params[0], params[1], params[2]),
                                          seasonal_order=(params[3], params[4], params[5], params[6]),
                                          trend='c',
                                          enforce_stationarity=False,
                                          enforce_invertibility=False)

        results = model.fit()
        predictions = results.predict(start='2021-01-31', end='2021-12-31',
                                     exog=df_test_scaled[['ChngS&P500Low', 'ChngS&P500High',
                                                         'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                         'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                                                         '1MLagged_ChngS&P500']],)

        mse = ((predictions - df_test_scaled['ChngTNV']) ** 2).mean()

        if mse < best_mse:
            best_mse = mse
            best_params = params

    except:
        continue

print("Beste Hyperparameter:", best_params)
print("Bestes MSE:", best_mse)
```

SARIMAX mit optimalen Parametern auf nicht normalisiertes Testset**

```
# Verwenden des aktualisierten Trainingsset im SARIMAX-Modell
model = sm.tsa.statespace.SARIMAX(train_data1['ChngTNV'],
                                  exog=train_data1[['ChngS&P500Low', 'ChngS&P500High',
                                                    'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                    'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                                                    '1MLagged_ChngS&P500']],
                                  order=(3, 0, 3),
                                  seasonal_order=(0,1, 2, 12),
                                  trend='c',
                                  enforce_stationarity=False,
                                  enforce_invertibility=False)

# Trainieren des Modells mit den aktualisierten Daten
results = model.fit()

# Verwenden des Modells, um Vorhersagen für den Testzeitraum zu erstellen
predictions = results.predict(start='2022-01-31', end='2022-12-31',
                              exog=test_data1[['ChngS&P500Low', 'ChngS&P500High',
                                                'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
                                                '1MLagged_ChngS&P500']])

# Berechnen des mittleren quadratischen Fehler (Mean Squared Error, MSE)
mse_Sarima = ((predictions - test_data1['ChngTNV']) ** 2).mean()

# Ausgabe des MSE
mse_Sarima
```


5.2 Rollende Anwendung verschiedener Zeiträume des Test sowie Trainsets

```
M df_train_2=df2
df_test_2=df2

M # Anzahl der Datenpunkte im Testzeitraum
test_length = 9

# Fenstergröße für den rollenden Trainingszeitraum
train_window = 132

df_train_2 = df_train_2.asfreq('M')
df_test_2 = df_test_2.asfreq('M')

# Liste zur Speicherung der MSE-Werte
mse_list = []

# Rollendes Modell
for i in range(len(df_test_2) - train_window - test_length + 1):
    # Start- und Endpunkte des aktuellen Trainingszeitraums
    train_start = df_train_2.index[i]
    train_end = df_train_2.index[i + train_window - 1]

    # Daten für das aktuelle Fenster
    train_data = df_train_2[train_start:train_end]
    test_start = df_train_2.index[i + train_window]
    test_end = df_train_2.index[i + train_window + test_length - 1]
    test_data = df_test_2[test_start:test_end]

    # SARIMAX-Modell für das aktuelle Fenster
    model = sm.tsa.statespace.SARIMAX(train_data['ChngTNV'],
                                      exog=train_data[['ChngS&P500Low', 'ChngS&P500High',
                                                         'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                         'ChngCPICore', 'ChngInfExp', 'ChngRetailSales',
                                                         '1MLagged_ChngS&P500']],
                                      order=(3, 0, 3),
                                      seasonal_order=(0, 1, 2, 12),
                                      trend='c',
                                      enforce_stationarity=False,
                                      enforce_invertibility=False)

    # Modelltraining
    with warnings.catch_warnings():
        warnings.filterwarnings("ignore", category=ConvergenceWarning)
        results = model.fit()

    # Vorhersagen für den aktuellen Testzeitraum
    predictions = results.predict(start=test_data.index[0], end=test_data.index[-1],
                                  exog=test_data[['ChngS&P500Low', 'ChngS&P500High',
                                                    'Chng2YR', 'ChngGDP', 'ChngNewHomes',
                                                    'ChngCPICore', 'ChngInfExp', 'ChngRetailSales',
                                                    '1MLagged_ChngS&P500']])

    # Berechnung des MSE
    mse = ((predictions - test_data['ChngTNV']) ** 2).mean()
    mse_list.append(mse)

    # Ausgabe der MSE für die aktuelle Vorhersage
    print(f"MSE for window {i+1}: {mse}")

# Durchschnittlicher MSE über alle Fenster
avg_mse = sum(mse_list) / len(mse_list)
print(f"Avg_MSE: {avg_mse}")
```

7. Durchführen der Panel OLS

```

M exog_vars = ['ChngS&P500Low', 'ChngS&P500High',
              'Chng2YR', 'ChngGDP', 'ChngNewHomes',
              'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
              '1MLagged_ChngS&P500']
exog = sm.add_constant(df2[exog_vars])
endogen = df2['Chng Routing Revenue']
model = PooledOLS(endogen, exog, check_rank=False)
pooled_res = model.fit()
print(pooled_res)

```

...

8. Adjustment zur Panel OLS mit Fixed Effects

```

M exog_vars = ['ChngS&P500Low', 'ChngS&P500High',
              'Chng2YR', 'ChngGDP', 'ChngNewHomes',
              'ChngCPICore', 'ChngInflExp', 'ChngRetailSales',
              '1MLagged_ChngS&P500']
exog = sm.add_constant(df2[exog_vars])
mod = PanelOLS(df2['Chng Routing Revenue'], exog, entity_effects=True)
fe_te_res = mod.fit()
print(fe_te_res)

```

...