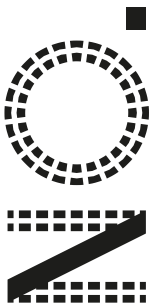




# TRADING INVESTING



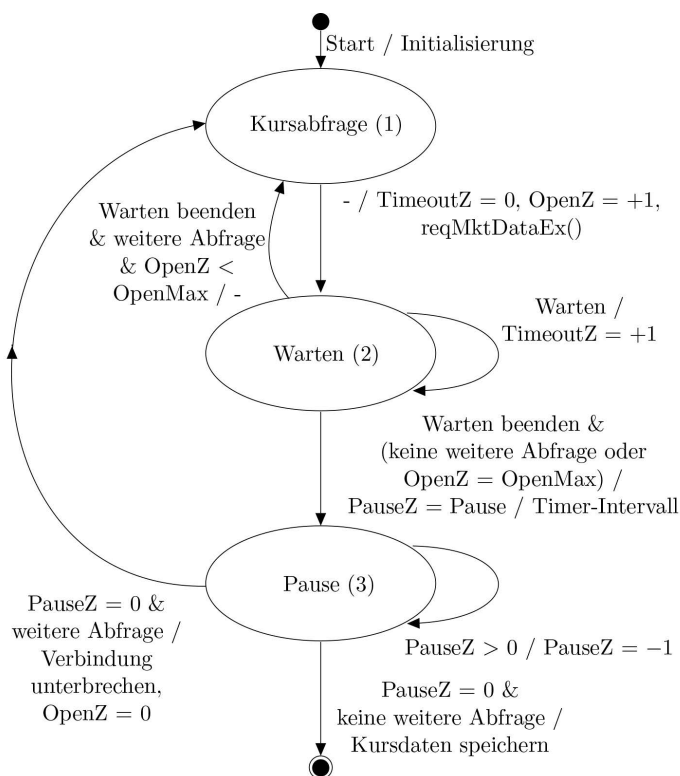
## Investieren unter Nebenbedingungen

### Teil 5: Auswahl von Aktien, Datenqualität und Statistik-Ergebnisse

Peter Vogel & Patrick Blättermann

Faculty of Media

Hochschule Duesseldorf, University of Applied Sciences



## Vorwort

In den ersten vier Teilen der Schriftenreihe „Trading“ liegt der Fokus auf der Entwicklung einer statistischen Beschreibung für das Investieren in Wertpapiere. Sie enthalten daher viele mathematische Formeln und Herleitungen. Um über mehrere Jahre hinweg experimentelle Ergebnisse zu erhalten, die gut miteinander verglichen werden können, wurden die Auswahl der Aktien und ihre Kurswerte in einer eigenen Kursdatenbank „eingefroren“. Diese Auswahl wird durch den Markt FEBRDUSA\_1 festgelegt, welcher Aktien enthält, die einem deutschen Aktienindex angehören. Einige dieser Aktien sind ebenfalls Bestandteil eines Aktienindex der USA. Die Kursdatenbank enthält zwangsläufig auch Aktien, die heute nicht mehr gehandelt werden. Sollen Aktien auch aktuell handelbar sein, stellt sich die Frage nach einer neuen Auswahl. Dies ist die Motivation des vorliegenden Teils 5 der Schriftenreihe. Die Orderausführung für diese Aktien wird in Teil 5 nicht behandelt. Vielmehr liegt der Fokus auf dem Auswahlprozess. Hierbei werden auch die Fragen aufgeworfen, welchen Einfluss die neu zu definierenden Märkte auf die Qualität der Kursdaten nimmt und welche Auswirkungen die Märkte auf Statistik-Ergebnisse für das Handeln mit Aktien haben.

Die Autoren danken der Bibliothek der Hochschule Düsseldorf für den freundlichen Support betreffend der Schriftenreihe und Frau Söhnitz für die Durchsicht des Manuskripts.

# Investieren unter Nebenbedingungen — Teil 5: Auswahl von Aktien, Datenqualität und Statistik-Ergebnisse

Peter Vogel \*      Patrick Blättermann †

15. März, 2021

## Kurzfassung

Im Teil 5 der Schriftenreihe werden dem bisherigen Markt FEBRDUSA\_1 Märkte gegenüber gestellt, die ausschließlich aktuell handelbare Aktien enthalten. Die Auswahl erfolgt nicht anhand von Aktien-Indizes, sondern mit Hilfe tagesaktueller Kursdaten wie Bid, Ask und Volume. Die Handelbarkeit soll durch eine Begrenzung des Spreads und einem Mindestumsatz sichergestellt werden.

Tagesaktuelle Kursdaten werden von Interactive Brokers bezogen. Die Deutsche Börse AG liefert hierbei das „Universum“ für die Aktien. Für diese erfolgt sequentiell der Kursdatenabruf. Dabei sollen Schwankungen des Serververhaltens nach Möglichkeit kompensiert werden. Der Auswahlprozess führt auf unterschiedliche Märkte aktuell handelbarer Aktien, abhängig von der Justierung des Auswahlprozesses. Die historischen Kursdaten dieser Märkte unterscheiden sich hinsichtlich der Datenqualität und der Statistik-Ergebnisse für das Handeln mit ihren Aktien.

### Schlagwörter:

Aktienauswahl, Datenqualität, Kursdaten

---

\*Hochschule Düsseldorf, Fachbereich Medien, University of Applied Sciences, e-mail: peter.vogel@hs-duesseldorf.de

†Hochschule Düsseldorf, Fachbereich Medien, University of Applied Sciences,  
e-mail: patrick.blaettermann@hs-duesseldorf.de

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>4</b>
<b>2</b>	<b>Mehrfach-Kursabfragen</b>	<b>5</b>
2.1	Vorbereitung der Mehrfach-Kursabfrage . . . . .	5
2.2	API von Interactive Brokers . . . . .	6
2.3	Implementierung einer Mehrfach-Kursabfrage . . . . .	7
2.4	Mehrfach-Kursabfrage ohne Unterbrechung . . . . .	9
2.5	Mehrfach-Kursabfrage mit Unterbrechung . . . . .	12
2.6	Abhängigkeit von Zeitpunkt und Zufall . . . . .	16
2.7	Abhängigkeit von Kursabfrage-Parametern . . . . .	19
<b>3</b>	<b>Marktdefinition</b>	<b>20</b>
3.1	Kursdatenquellen . . . . .	20
3.2	Mehrfach-Kurstabelle . . . . .	21
3.3	Definition zweier Märkte . . . . .	22
<b>4</b>	<b>Datenqualität</b>	<b>26</b>
4.1	Kursdatenbanken . . . . .	26
4.2	Datenbank-Diagnose: Definitionen . . . . .	29
4.2.1	Auffüllungen . . . . .	29
4.2.2	Repeats (Kurswiederholungen) . . . . .	29
4.2.3	Kurskorrekturen . . . . .	30
4.2.4	Fehler . . . . .	30
4.2.5	PennyTod und Tod . . . . .	30
4.2.6	Maximale Sprunghöhe . . . . .	30
4.3	Datenbank-Diagnose: Ergebnisse . . . . .	30
4.3.1	Anzahl Datensätze . . . . .	32
4.3.2	Anzahl Aktien . . . . .	32
<b>5</b>	<b>Abhängigkeit der Statistik vom Markt</b>	<b>33</b>
5.1	Zeitliche Verläufe . . . . .	34
5.1.1	Kursdatenbank_2015 und Markt FEBRDUSA_1 . . . . .	35
5.1.2	Kursdatenbank_2020 und Markt FEBRDUSA_1 . . . . .	36
5.1.3	Kursdatenbank_2020 und Markt FELynxMM_1_100 . . . . .	37
5.1.4	Kursdatenbank_2020 und Markt FELynxMM . . . . .	38
5.2	Gesamter Handelszeitraum . . . . .	39
<b>6</b>	<b>Zusammenfassung</b>	<b>41</b>
<b>A</b>	<b>Notation</b>	<b>43</b>
A.1	Tagesaktuelle Kursdaten . . . . .	43
A.2	Historische Kursdaten . . . . .	45
	<b>Literatur</b>	<b>47</b>

# 1 Einführung

Zur Erforschung des Handelns von Aktien wurde in den bisherigen Teilen 1-4 das Kursdaten-Abo des Börsenprogramms Market Maker (MM) von Lenz+Partner, ehemals Market Maker Software AG, benutzt. Mit Hilfe dieser Kursdaten wurde eine Kursdatenbank aufgebaut, welche unbereinigte Aktienkurse sowie Bereinigungsfaktoren enthält. Dies wurde durch den Export von Kursdaten aus MM in sowohl bereinigter als auch unbereinigter Form bewerkstelligt. Es handelt sich um historische Kursdaten, die bis zum vorherigen Handelstag reichen. Der Export tagesaktueller Kursdaten ist nicht mehr möglich, da dieser Dienst leider vor Jahren eingestellt wurde. Das Nachfolgeprodukt TAI-PAN von Lenz+Partner stellt zwar tagesaktuelle Kursdaten bereit, der Export von Aktienkursen in bereinigter und unbereinigter Form ist damit jedoch nicht möglich [1]. Die Bereitstellung kostenloser Kursdaten durch Yahoo Finance wurde in 2012 – zum Leidwesen vieler Kunden – ebenfalls eingestellt [2].

Alpha Vantage bietet wie Yahoo Finance den Kursdatenabruf simultan für viele Aktien an, jedoch verbunden mit Kosten für mehr als 500 Kursabfragen pro Tag [3]. Interactive Brokers (IB) bietet den Kursdatenabruf von Kursdaten mit 15 Minuten Verzögerung kostenlos an, wobei über deren API auch Orderaufgaben automatisiert werden können [4]. Daher wird der Bezug tagesaktueller Kursdaten durch IB bewerkstelligt. Der in Deutschland ansässige Online Broker Lynx stellt Kunden den Zugang zur Infrastruktur von IB zur Verfügung [5].

Der bisher verwendete Markt FEBRDUSA\_1 enthält viele Aktien, die aktuell nicht mehr gehandelt werden, sodass eine neue Marktdefinition erforderlich ist, bei der auch tagesaktuelle Kursdaten berücksichtigt werden. Die Selektion von ausschließlich Aktien erfolgt durch Integration der Kursdaten von IB und historischer Kursdaten von MM. Eine ausreichende Handelbarkeit der Aktien erfordert eine sorgfältige Definition des neuen Marktes. Außerdem stellt sich die Frage nach einer Bewertung der Datenqualität für den Markt. Darüber hinaus ist unklar, wie sich verschiedene Märkte auf das Handeln auswirken. Dies wird anhand der in den ersten drei Teilen der Schriftenreihe entwickelten statistischen Beschreibung untersucht.

## Aufbau der Arbeit:

- In Kap. 2 wird der Datenabruf tagesaktueller Kursdaten bei IB untersucht. Dabei stellt sich die Kursabfrage für eine Vielzahl von Wertpapieren als problematisch heraus und erfordert eine Justierung von Kursabfrage-Parametern.
- In Kap. 3 werden Märkte definiert, welche auf Kursabfragen bei IB an unterschiedlichen Handelstagen basieren.
- In Kap. 4 werden Kursdaten für die definierten Märkte analysiert, um den Einfluss eines Marktes auf die Kursqualität festzustellen.
- In Kap. 5 wird gezeigt, dass die Märkte einen signifikanten Einfluss auf statistische Größen besitzen. Die statistische Beschreibung erfolgt hierbei mit Hilfe des kumulativen Gewinnfaktors für einen mehrjährigen Handelszeitraum, sowie den Investitionsgrad und dem vom Investitionsgrad unabhängigen Moment erster Ordnung.

## 2 Mehrfach-Kursabfragen

Eine Mehrfach-Kursabfrage erfolgt für alle bei IB verfügbaren Wertpapiere für Aktien oder für eine Teilmenge davon, welche durch einen Markt definiert ist. Hierbei werden Kursdaten für mehrere Wertpapiere sequentiell abgefragt. Die Kursdaten umfassen die folgenden sieben *Kurstypen*: Die sechs Kurswerte Bid, Ask, Last, High, Low, Volume und einen Zeitstempel (Timestamp) für den Zeitpunkt des letzten Trades. Der Kurswert „Last“ ist der letzte gehandelte Preis bis zum aktuellen Zeitpunkt, die Kurswerte „High“ und „Low“ sind der höchste bzw. niedrigste Preis bis zum aktuellen Zeitpunkt und „Volume“ gibt das bislang gehandelte Volumen eines Tages an. Vom Server von IB erhaltene Werte für Kurstypen werden als Datensatz in einer Datenbank-Tabelle gespeichert. Werden Werte für alle 7 Kurstypen erhalten, wird der Datensatz als *vollständig* bezeichnet.

### 2.1 Vorbereitung der Mehrfach-Kursabfrage

Zur Vorbereitung einer Mehrfachkursabfrage werden die beiden folgenden Aufgaben durchgeführt, um das „Universum“ der an der Frankfurter Börse handelbaren Wertpapiere für Aktien zu definieren:

#### 1.) Deutsche Börse AG [6]

Es werden Metadaten bestehend aus

- Wertpapierkennnummer (WKN),
- Symbol des Wertpapiers bei der Deutsche Börse AG,
- ISIN zur eindeutigen Identifikation eines Wertpapiers

aller handelbaren Instrumente der Deutsche Börse AG für Aktien herunter geladen und gespeichert. Diese umfassen neben Aktien auch Aktienfonds. Die Abfrage der Daten der Deutsche Börse AG beinhaltet den Download einer CSV-Datei (Comma Separated Values), welche auf der Internetseite des Unternehmens heruntergeladen werden kann.

#### 2.) Symbole von IB

Die Wertpapiere in der CSV-Datei werden unter den bei IB verfügbaren Wertpapieren anhand ihrer ISIN gesucht. Für die gefundenen Wertpapiere werden die unter Punkt 1 bestimmte WKN, Symbol des Wertpapiers bei der Deutsche Börse AG und ISIN zusammen mit dem

- Symbol des Wertpapiers bei IB,
- Primärmarkt des Wertpapiers

in einer Symbol-Liste gespeichert.

## 2.2 API von Interactive Brokers

Die Einbettung der Kursabfragen in das eigene Börsenprogramm erfolgt über die ActiveX-Steuererelement Schnittstelle der TWS API (Version 9.71) von IB. Es werden die folgenden Methoden und Ereignisse für Mehrfach-Kursabfragen benutzt:

### Methoden

- *connect()* zum Verbindungsaufbau mit dem Server von IB
- *disconnect()* zur Beendigung der Verbindung
- *reqMktDataEx()* für die Kursabfrage eines einzelnen Wertpapiers
- *reqMktDataType()* zur Festlegung der Kursabfrage-Methode
- *reqContractDetailsEx()* für die Metadaten eines Wertpapiers

### Ereignisse (Nachrichten vom Server)

- *tickPrice()* für die Kurswerte Bid, Ask, Last, High und Low
- *tickSize()* für Volume
- *tickString()* für den Zeitstempel
- *tickSnapshotEnd()* für den Abschluss einer Abfrage
- *errMsg()* für eine Fehlermeldung
- *contractDetailsEx()* für die Metadaten eines Wertpapiers

### Randbedingungen für Kursabfragen

- Es können maximal 60 Abfragen pro Sekunde an den Server gestellt werden.
- Es werden maximal 100 offene Abfragen an den Server akzeptiert.

Hierbei wird eine Abfrage als „offen“ bezeichnet, wenn sie mit keiner der folgenden beiden Nachrichten des Servers,

- *tickSnapshotEnd()*,
- *errMsg()*

quittiert wurde.

### 2.3 Implementierung einer Mehrfach-Kursabfrage

In mehreren Entwicklungsstufen einer Mehrfach-Kursabfrage wurde festgestellt, dass eine batchweise Abfrage von Kursen serverseitig problematisch ist und zu nicht reproduzierbaren Ergebnissen führt. Außerdem wird wegen der Kursabfrage ohne kostenpflichtiges Abo für Aktienkurse auf das Prinzip von Ticker-Abos verzichtet. Stattdessen werden sogenannte *SnapShots* verwendet, welche anstelle von Datenstreams eine Momentaufnahme (ohne Abo 15 Minuten verzögert) der Kurse darstellen. Abb. 1/S. 7 zeigt die Steuerung einer Mehrfach-Kursabfrage mit Hilfe eines Mealy-Automaten in UML-Notation [7], [8]. Nach der ersten Kursabfrage legt der Automat den Zeitpunkt für alle weiteren Kursabfragen in Abhängigkeit von den Nachrichten des Servers fest. Während der Mehrfach-Kursabfrage werden Kursdaten mit Hilfe der Nachrichten des Servers laufend ergänzt bzw. vervollständigt. Kursdaten zu einem Wertpapier können eintreffen, wenn bereits viele Kursabfragen für weitere Wertpapiere erfolgt sind. Die Unterscheidung der Nachrichten nach Wertpapieren erfolgt hierbei durch eine Abfrage-ID. Sie wird durch die Reihenfolge der in der Symbol-Liste enthaltenen Wertpapiere mit vorhandenem IB-Symbol festgelegt.

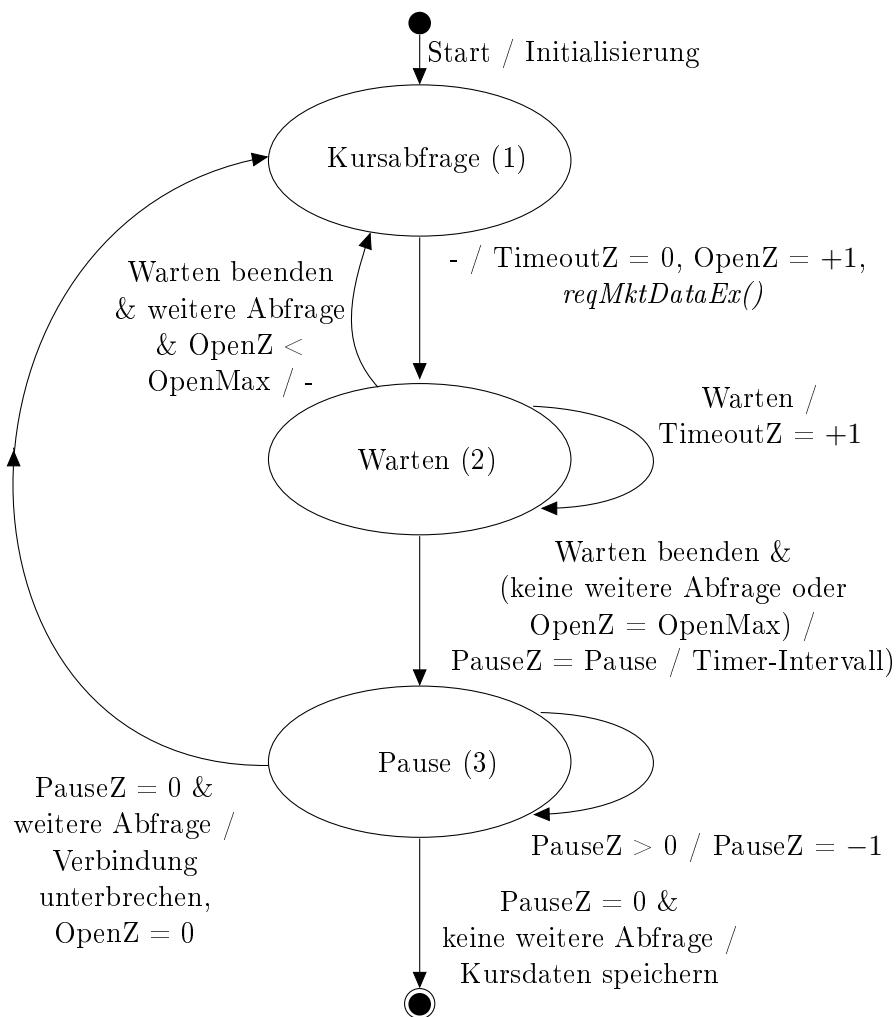


Abbildung 1: Steuerung einer Mehrfach-Kursabfrage mit einem Automaten. Die Notation „B / A“ bedeutet, dass unter der Bedingung B neben der Einstellung des Folgezustandes eine Aktion A durchgeführt wird. „+1“ bedeutet hoch zählen eines Zählers um 1 (OpenZ und TimeoutZ), „-1“ bedeutet herunter zählen um 1 (PauseZ).



## Erläuterungen

Der Automat beginnt im Zustand 1 mit der Kursabfrage für das erste Wertpapier, gefolgt von einem Wechsel in den Zustand 2. Spätestens nach Ablauf einer einstellbaren Zeitspanne erfolgt die Rückkehr in den Zustand 1 und danach die Kursabfrage für das nächste Wertpapier. Wann die nächste Kursabfrage startet, hängt von den vom Server empfangenen Nachrichten ab. Sie werden in den beiden komplementären Bedingungen „Warten“ und „Warten beenden“ berücksichtigt und werden periodisch ausgewertet, wobei die Periodendauer durch das Intervall eines Timers festgelegt ist. Dieses *Timer-Intervall* ist einer der Kursabfrage-Parameter. Zustandsänderungen werden nach Ablauf des Timer-Intervalls überprüft. Die Unterbrechung der Verbindung zum Server beim Verlassen des Zustands 3 stellt eine Besonderheit dar, auf die in Abschn. 2.5 weiter eingegangen wird.

**Zustand 1** Wertpapierkurse werden mit der Methode *reqMktDataEx()* abgefragt. Dies erfolgt im Zustand 1 (Kursabfrage) des Automaten. Im Zustand 1 erfolgt stets, d.h. ohne Bedingung, ein Wechsel in den Zustand 2 (warten auf Abfrageergebnis). Hierbei wird ein sog. *timeout-Zähler* auf 0 gesetzt (initialisiert).

**Zustand 2** Nach jedem Timer-Intervall wird der Timeout-Zähler hochgezählt. Wenn der Zähler eine einstellbare Grenze *Grenze Timeout* erreicht, wird das Warten beendet, um eine nächste Kursabfrage zu ermöglichen. In diesem Fall erfolgt der Zustandswechsel von 2 nach 1, sofern eine weitere Kursabfrage ansteht. Der Zustandswechsel von 2 nach 1 ist auch vorher möglich, wenn beispielsweise die Kursabfrage mit *tickSnapShotEnd()* vom Server beendet wird. Die vollständige Bedingung für die Beendigung des Wartens lautet:

### Bedingung „Warten beenden“

1. *tickSnapShotEnd()* erhalten oder
2. *errMsg()* erhalten oder
3. Datensatz ist vollständig (alle 7 Kurstypen erhalten) oder
4.  $\text{TimeoutZ} = \text{Grenze Timeout}$  (Kursabfrage-Parameter)

Ist die Grenze für Timeout gleich 0 gesetzt, ist mit Case 4 die Bedingung „Warten beenden“ erfüllt. In diesem Fall erfolgt nach dem Zustandswechsel von 1 nach 2 sofort nach Ablauf eines ersten Timer-Intervalls der Wechsel von Zustand 2 zurück in den Zustand 1. Im Zustand 1 erfolgt nach Ablauf eines zweiten Timer-Intervalls eine weitere Kursabfrage. Die Zeitspanne zwischen zwei Kursabfragen beträgt in diesem Fall das Zweifache des Timer-Intervalls. Dies ist die Mindestzeit zwischen zwei Kursabfragen.

**Zustand 3** Eine Ausnahmesituation stellt der Wechsel von Zustand 2 in den Zustand 3 dar, wenn die Anzahl offener Abfragen die Grenze *OpenMax* (Kursabfrage-Parameter) erreicht oder keine weiteren Kursabfragen anstehen. In diesem Fall wird die Mehrfach-Kursabfrage für eine bestimmte Zeit (Pause) unterbrochen. Während der Pause werden folglich nur Kursdaten auf bisherige Kursabfragen abgewartet. Die Pause wird mit dem Kursabfrage-Parameter *Pause/Intervalle* als Vielfaches des Timer-Intervalls angegeben. Stehen keine weiteren Kursabfragen an, wird die Mehrfach-Kursabfrage mit dem Speichern aller bis zu diesem Zeitpunkt eingesammelten Kursdaten in die Mehrfach-Kurstabelle beendet. Stehen weitere Kursabfragen an, wird die Verbindung zum Server kurzzeitig mit „*disconnect()*, *connect()*“ unterbrochen und dann in den Zustand 1 zurückgekehrt. Der Open-Zähler (OpenZ) für die Anzahl offener Abfragen wird hierbei auf 0 zurückgesetzt.

## 2.4 Mehrfach-Kursabfrage ohne Unterbrechung

Abb. 2/S. 9 zeigt den Screenshot einer Mehrfach-Kursabfrage, die am 21.10.2020 vorgenommen wurde. Es werden alle bei IB verfügbaren Wertpapiere für Aktien abgefragt. Links befinden sich Zählerstände und rechts oben befinden sich die Einstellungen für alle vier Kursabfrage-Parameter. Der Abfragezähler „aktuelle Abfrage“ beginnt beim Wert 0, sodass insgesamt 6185 Abfragen durchgeführt wurden. Die Einstellungen 20/50 für Timer-Intervall und Grenze Timeout sind Standardeinstellungen. Das Timer-Intervall garantiert hierbei eine Mindestzeit von  $2 \times 20 = 40$ ms zwischen zwei Kursabfragen. Damit erfolgen maximal 25 Kursabfragen pro Sekunde, womit die Grenze von maximal 60 Abfragen pro Sekunde sicher eingehalten wird (vgl. Abschn. 2.2). Der Grenzwert für offene Abfragen ist auf den Wert `openMax=1000` gesetzt, um eine Unterbrechung der Verbindung zum Server zu verhindern. Es erfolgt somit eine Mehrfach-Kursabfrage ohne Unterbrechung der Verbindung zum Server. Die Anzahl offener Abfragen überschreitet den Wert 100 bereits für die Abfrage-ID 2850. Insgesamt ergeben sich 705 offene Abfragen und es finden 0 (keine) Unterbrechungen der Verbindung zum Server statt.

Mehrfach-Kursabfrage		Einstellungen	
letzte Abfrage:	6184	Timer-Intervall:	20
aktuelle Abfrage:	6184	Grenze Timeout:	50
für ID:	22784	Pause/Intervalle:	500
offene Abfragen:	705 / 0	openMax:	1000
101 erreicht für ID	2850	Börsenplatz:	Smart
Pause Countdown	0	CancelMktData nach Timeout:	<input type="checkbox"/>
Full gesamt	573	Bid:	-1
FullEnd	515	Ask:	-1
Reject	1991	Last:	24,15
Timeout	3403	High:	24,315
ShortTimeout	276	Low:	24,13
ShortEnd	0	Volume:	260

Abbildung 2: Mehrfach-Kursabfrage ohne Unterbrechung am 21.10.2020

Unten links befinden sich Zähler für Statuswerte. Statuswerte sind aus den folgenden Bestandteilen zusammengesetzt:

- Full: Alle 7 Kurstypen erhalten
- End: `tickSnapShotEnd()` erhalten
- Reject: `errMsg()` erhalten
- Short: 1–6 Kurstypen erhalten
- Timeout (Tout): Eingestellte Zeitspanne für Timeout ist abgelaufen (Einstellung: Timeout nach  $20 * 50 \text{ ms} = 1\text{s}$ )

**Datenbank-Abfrage:**

Einen genaueren Einblick in aufgetretene Statuswerte zeigt Tab. 1/S. 10. Alle Statuswerte setzen sich aus den 5 Bestandteilen für Statuswerte zusammen. Die Tabelle wurde durch die folgende Datenbankabfrage für die Mehrfach-Kurstabelle gewonnen:

```
Select status, count(*) as Anz, min(Dauer) as min, max(Dauer) as max from Lynx_Kurse_Gesamt group by status
```

Als Ergebnis der Datenbankabfrage erhält man 13 verschiedene Statuswerte. Die Spalten „min“ und „max“ stellen den kleinsten und größten Wert der Dauer zur Bildung der Statuswerte in Sekunden dar. Tab. 1/S. 10 entnimmt man: Rejects besitzen die kleinste Dauer im Bereich 90ms–1.7s. Vollständige Datensätze dauern 400ms–9.2s.

Nr	Status	Anzahl	Min	Max
1	FullEnd	515	0.406	6.469
2	Reject	1991	0.093	1.75
3	ShortTout	78	1.562	1.672
4	ShortToutEnd	125	11.297	13.219
5	ShortToutFullEnd	31	4.891	2702
6	ShortToutShort	4	1.562	1.563
7	ShortToutShortEnd	38	4.953	5001
8	Tout	619	1.562	2.093
9	ToutEnd	2722	11.234	31.875
10	ToutFullEnd	27	1.578	9.25
11	ToutReject	9	1.562	1.984
12	ToutShort	4	1.562	1.563
13	ToutShortEnd	22	11.14	4418

Tabelle 1: Anzahl aller aufgetretenen Status-Werte für die Mehrfach-Kursabfrage ohne Unterbrechung am 21.10.2020

Mit den Anzahl-Werten in der Tabelle können die Zählerstände in Abb. 2/S. 9 überprüft werden:

- FullEnd = 515 (Statuswert 1),
- Full gesamt = 573 (Statuswerte 1, 5, 10),
- Reject = 1991 (Statuswert 2),
- Tout (Timeout) = 3403 (Statuswerte 8 – 13),
- ShortTout = 276 (Statuswerte 3 – 7),
- Offene Abfragen = 705 (Statuswerte 3, 6, 8, 12).

Es sind nur Full gesamt = 573 der insgesamt 6185 Datensätze vollständig. Der Grund dafür ist die frühzeitige Überschreitung der Grenze 100 für die Anzahl der offenen Abfragen, wie Abb. 3/S. 11 demonstriert. Sie stellt jeden der 13 auftretenden Statuswerte zusammen mit der Nummer des Datensatzes (1—6185) als Punkt dar. Hierbei geben die Datensatz-Nummern die Reihenfolge für die Kursabfragen wieder.

Die Abfrage-ID und die Datensatz-Nummer für den letzten erhaltenen Statuswert sind in Tab. 2/S. 11 dargestellt. Zur Ermittlung dieser Abfrage-ID wurde die Datenbankabfrage

```
Select status, Max(Id) as MaxId from Lynx_Kurse_Gesamt group by status
```

durchgeführt.

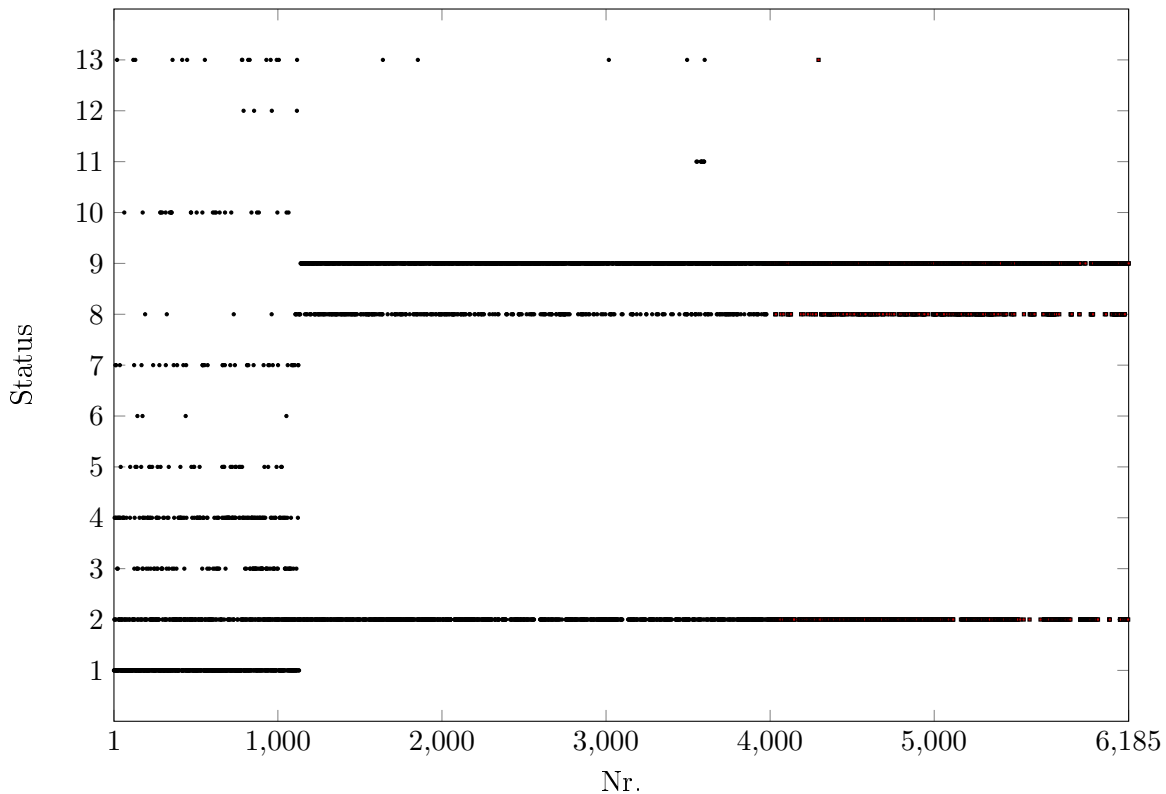


Abbildung 3: Zeitliche Verteilung der Statuswerte bei der Mehrfach-Kursabfrage ohne Unterbrechung

Nr.	Status	MaxId	DS-Nr.
1	FullEnd	2910	1129
2	Reject	22780	6182
3	ShortTout	2888	1113
4	ShortToutEnd	2901	1122
5	ShortToutFullEnd	2730	1024
6	ShortToutShort	2792	1052
7	ShortToutShortEnd	2904	1125
8	Tout	22722	6162
9	ToutEnd	22784	6185
10	ToutFullEnd	2814	1065
11	ToutReject	10587	3598
12	ToutShort	2894	1116
13	ToutShortEnd	15943	4295

Tabelle 2: MaxID ohne Unterbrechung

**Statuswerte 1, 3–7, 10 und 12:** Für diese Statuswerte liegt die Abfrage-ID für den letzten erhaltenen Statuswert in der Nähe der ID 2850, für die der open-Zähler den Wert 100 überschreitet. Die Statuswerte enthalten insbesondere die Werte 1, 5 und 10 für vollständige Datensätze. Dies bedeutet, dass vollständige Datensätze kurz vor oder nach Erreichen des Wertes 100 für die Anzahl offener Abfragen abbrechen.

**Statuswerte 2, 8 und 9:** Dies sind die einzigen Statuswerte, die auch am Ende der Mehrfach-Kursabfrage auftreten. Sie enthalten keine Kursinformation.

**Statuswerte 11 und 13:** Diese Statuswerte treten nur vereinzelt auf und ihre Kursinformation ist vernachlässigbar.

## 2.5 Mehrfach-Kursabfrage mit Unterbrechung

Die Abb. 4/S. 13 und Tab. 3/S. 13 zeigen eine Mehrfach-Kursabfrage am 4.5.2020 mit Unterbrechungen. Im Unterschied zu Abschn. 2.4 wird die Mehrfach-Kursabfrage jedes mal dann unterbrochen, wenn sich die Anzahl der offenen Abfragen der Grenze 100 nähert.

### Unterbrechung und offene Abfragen:

Aufgrund der Einstellungen erfolgt eine Unterbrechung bei einem Zählerstand von `openMax=95`. Die Pause beträgt  $500 \times 20 \text{ ms} = 10\text{s}$ . Gemäß der Anzeige für „offene Abfragen“ wird die Mehrfach-Kursabfrage insgesamt viermal unterbrochen und am Ende der Mehrfach-Kursabfrage bleiben 39 Kursabfragen offen. Bei jeder der vier Unterbrechungen erreicht der Zähler für offene Abfragen den Wert 95, woraus sich

$$4 \cdot 95 + 39 = 419$$

offene Abfragen ergeben.

Die Anzahl aller offenen Kursabfragen am Ende der Mehrfach-Kursabfrage ergibt sich aus der Anzahl der Statuswerte, die weder ein `Reject` noch ein `End` beinhalten, also für `ShortTout`, `ShortToutShort`, `Tout` und `ToutShort` zu

$$\text{Offene Abfragen} = 256 + 65 + 61 + 18 = 400 .$$

Die Diskrepanz zu 419 offenen Abfragen folgt daraus, dass in den Pausen ein Teil der offenen Abfragen abgeschlossen wird. Sie beträgt  $419 - 400 = 19$ .

### Vollständige Datensätze:

Im Unterschied zu Abschn. 2.2 sind viel mehr vollständige Datensätze erhalten worden. Der Zähler für „Full gesamt“ zeigt 3046 Datensätze, die sich wie folgt aufschlüsseln:

- FullEnd: 2874
- ShortToutFullEnd: 42
- ToutFullEnd: 130

Die Unterbrechungen der Verbindung zum Server sind somit erfolgreich.

Mehrfach-Kursabfrage für alle Titel

letzte Abfrage:	6184	Timer-Intervall:	20
aktuelle Abfrage:	6184	Grenze Timeout:	50
für ID:	20822	Pause/Intervalle:	500
offene Abfragen:	39 / 4	openMax:	95
101 erreicht für ID		Börsenplatz:	Smart
Pause Countdown	0	CancelMktData nach Timeout:	<input type="checkbox"/>
Full gesamt	3046	Bid:	12,36
FullEnd	2874	Ask:	12,658
Reject	1962	Last:	0
Timeout	257	High:	10,14
ShortTimeout	1092	Low:	10,14
ShortEnd	0	Volume:	40

Abbildung 4: Mehrfach-Kursabfrage mit Unterbrechung am 4.5.2020

Nr	Status	Anzahl	Min	Max
1	FullEnd	2874	0.75	7.109
2	Reject	1962	0.093	1.485
3	ShortTout	256	1.562	1.969
4	ShortToutEnd	512	6.031	13.031
5	ShortToutFullEnd	42	5.156	692
6	ShortToutShort	65	1.562	1.828
7	ShortToutShortEnd	217	11.266	831
8	Tout	61	1.562	1.969
9	ToutFullEnd	130	1.578	301
10	ToutShort	18	1.562	1.641
11	ToutShortEnd	48	11.11	14.219

Tabelle 3: Anzahl aller aufgetretenen Status-Werte für die Mehrfach-Kursabfrage mit Unterbrechung am 4.5.2020

Abb. 5/S. 14 zeigt die zeitlichen Verläufe der auftretenden Statuswerte. Im Unterschied zu Abb. 3/S. 11 brechen Statuswerte, insbesondere für vollständige Kursdaten (1, 5, 9), nicht ab.

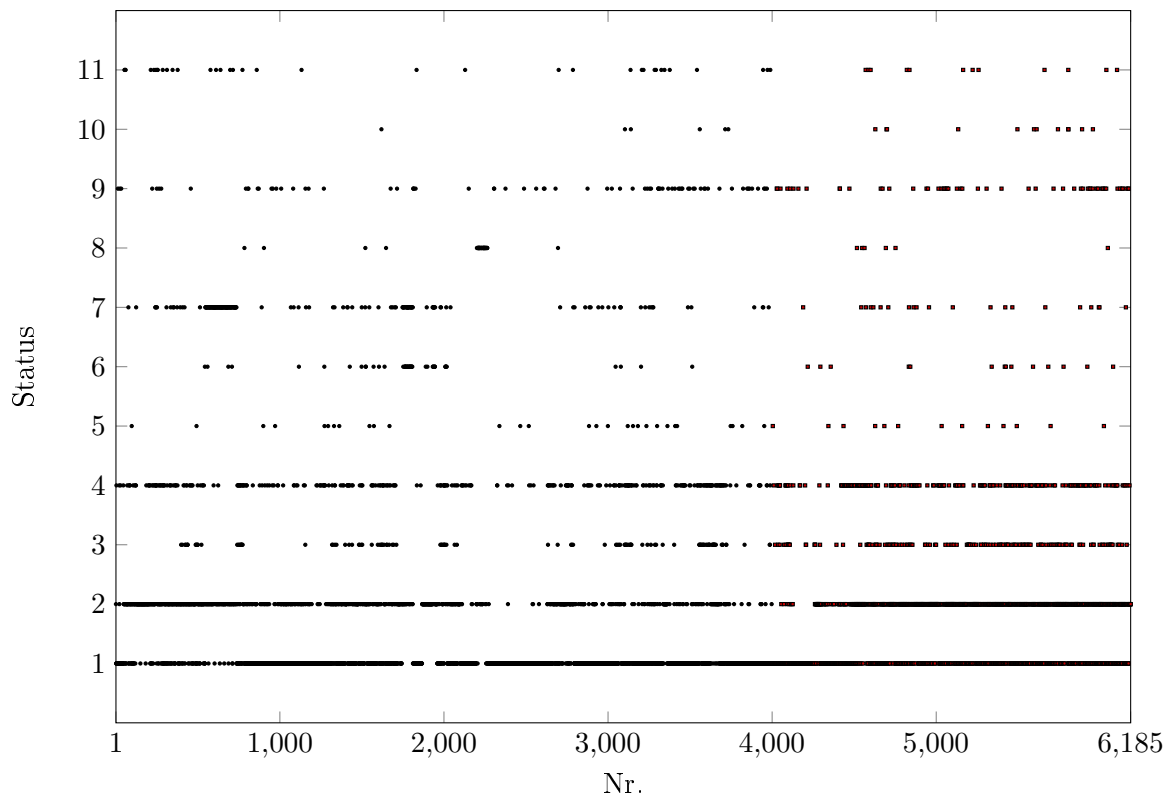


Abbildung 5: Zeitliche Verteilung der Status-Werte bei der Mehrfach-Kursabfrage mit Unterbrechung (4.5.2020)

Die Dauer für vollständige Datensätze kann für die Statuswerte 5 und 9 sehr große Werte annehmen. Abb. 6/S. 15 und Abb. 7/S. 15 zeigen, wie sich die Dauer für vollständige und zurückgewiesene Datensätze auf die Kursabfragen aufteilen. Auf der x-Achse ist die Dauer aufgetragen, auf der y-Achse die relative Häufigkeit für eine Dauer  $\leq x$ . Abb. 6/S. 15 zeigt, dass nach etwa 7s praktisch alle vollständige Datensätze erhalten wurden. Abb. 7/S. 15 zeigt, dass die Zurückweisung einer Kursabfrage in der Regel deutlich unter 1s erfolgt.

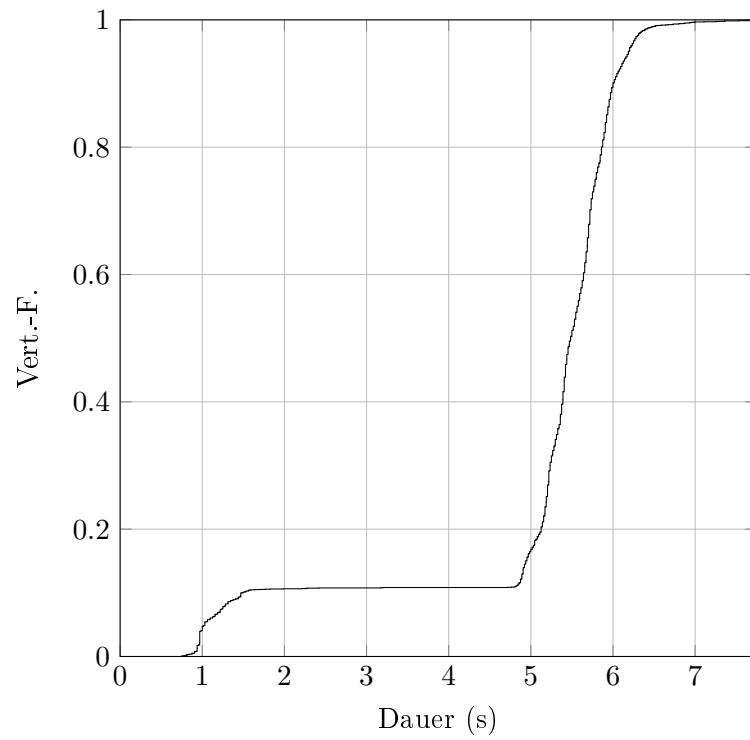


Abbildung 6: Verteilungsfunktion für vollständige Datensätze (Fulls)

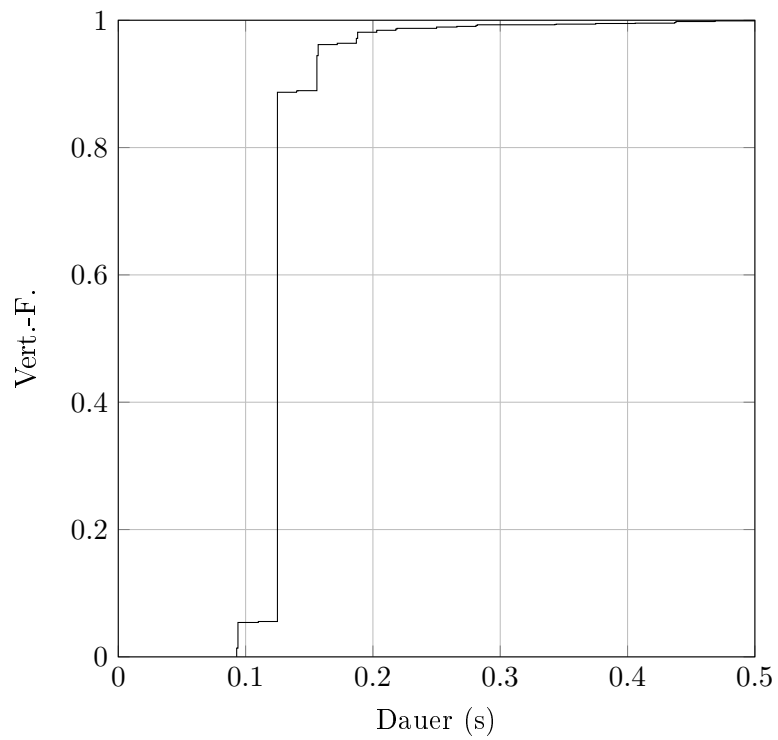


Abbildung 7: Verteilungsfunktion für zurückgewiesene Abfragen (Rejects)



## 2.6 Abhängigkeit von Zeitpunkt und Zufall

In Tab. 4/S. 17 – Tab. 6/S. 18 werden viele Mehrfach-Kursabfragen zu verschiedenen Uhrzeiten an mehreren Tagen vorgenommen, um sowohl den Einfluss des Zeitpunkts der Abfrage als auch den Einfluss des Zufalls auf das Ergebnis der Abfrage zu untersuchen. Im ersten Fall sollen Mehrfach-Kursabfragen in kurzen Zeitabständen gestartet werden und vor der nächsten Abfrage bereits beendet sein. Im zweiten Fall sind Abfragen in möglichst kurzen Zeitabständen vorzunehmen, um den Einfluss der Zeit zu reduzieren. Daher wurde erstens der Markt FEBRDUSA\_1 verwendet und zweitens die Grenze für den Timeout von 50 auf 10 gesenkt bei gleichem Timer-Intervall von 20ms. Die Abfrage-Parameter für die Mehrfach-Kursabfragen sind somit 20/10. Kursabfragen werden daher bereits nach wenigen Minuten beendet. Dargestellt werden

- die Anzahl der Wertpapiere mit vollständigen Datensätzen (Fulls),
- die Anzahl der Wertpapiere mit vollständigen und gültigen Datensätzen,
- die Anzahl der Wertpapiere mit  $\text{Bid} > 0$  und  $\text{Ask} > 0$ .

Hierbei bedeutet „vollständig und gültig“, dass alle sieben Kurstypen vom Server erhalten werden und darüber hinaus die sechs Kurswerte  $> 0$  sind und der Zeitstempel eine gültige Angabe von Datum und Uhrzeit enthält.

### Ergebnis für Tab. 4/S. 17 (23.6.—2.7.2020):

- Die meisten vollständigen Datensätze sind auch gültig.
- Alle Anzahl-Werte wachsen — abgesehen von Ausnahmen — mit fortschreitender Uhrzeit beständig an. Am Vormittag ist das Wachstum am größten. Kurz vor 20:00 geht die Anzahl brauchbarer Abfragen stark zurück.
- Auffallend ist der 2.7.2020. Bereits Mittags erreicht die Anzahl der Wertpapiere mit gültigem Bid und Ask den stationären Wert 567 bzw. 568. Am 1.7.2020 schwanken diese Anzahl-Werte auch Nachmittags.
- Die Anzahl-Werte lassen sich nicht reproduzieren, d.h. sie unterliegen Zufallsschwankungen: Zum Beispiel der Einbruch am 23.6.2020 um 9:45 oder die Unterschiede an verschiedenen Tagen zur gleichen Uhrzeit, z.B. am 23. und 24.6.2020 um 12:15.

### Ergebnis für Tab. 5/S. 18 (10.7.2020):

- Die Anzahl für  $\text{Bid} > 0$  und  $\text{Ask} > 0$  ist ab 10:34 nahezu konstant bei 544. Nach 20:00 erhält man keine gültigen Werte für Bid- und Ask (-1).
- Die Anzahl vollständiger Datensätze steigt bis auf den Wert 480 um 19:33, danach fällt sie ab.
- Vor 20:00 sind die meisten vollständigen Datensätze auch gültig.

### Ergebnis für Tab. 6/S. 18 (13.7.2020):

- Bis um 11:05 ist die Anzahl für  $\text{Bid} > 0$  und  $\text{Ask} > 0$  etwa konstant um 507, danach gibt es einen Sprung auf den Wert 565.
- Die Anzahl vollständiger Datensätze liegt über den entsprechenden Anzahl-Werten am 10.7.2020, während die Anzahl-Werte für  $\text{Bid} > 0$  und  $\text{Ask} > 0$  bis um 11:05 unter den Werten vom 10.7.2020 liegen.

Zeitpunkt (1-29)	vollständig	vollständig und gültig	Bid > 0 und Ask > 0
23.6.2020 9:30	316	312	423
23.6.2020 9:30	316	312	423
23.6.2020 9:45	254	254	326
23.6.2020 10:15	330	330	408
23.6.2020 11:15	351	349	410
23.6.2020 12:15	365	361	416
23.6.2020 13:55	439	435	564
23.6.2020 14:20	436	431	519
24.6.2020 12:15	433	427	564
24.6.2020 13:00	432	424	564
30.6.2020 13:57	441	436	519
30.6.2020 14:14	442	439	567
1.7.2020 8:22	217	209	529
1.7.2020 8:35	234	231	475
1.7.2020 9:07	274	268	540
1.7.2020 9:32	298	257	508
1.7.2020 10:05	355	310	487
1.7.2020 10:23	359	315	445
1.7.2020 11:00	375	350	515
1.7.2020 12:00	395	361	500
1.7.2020 13:00	425	417	516
1.7.2020 14:00	442	438	567
1.7.2020 15:45	456	449	566
1.7.2020 18:00	463	445	518
1.7.2020 19:40	476	469	562
2.7.2020 11:54	421	416	567
2.7.2020 12:28	426	420	567
2.7.2020 12:54	432	425	567
2.7.2020 13:37	441	432	567
2.7.2020 14:09	444	439	567
2.7.2020 15:04	447	440	567
2.7.2020 15:54	468	463	568
2.7.2020 17:00	483	478	568
2.7.2020 18:03	492	478	560
2.7.2020 19:03	497	486	562
2.7.2020 19:43	446	435	511
2.7.2020 20:17	391	137	147

Tabelle 4: Mehrfach-Kursabfrage für FEBRDUSA\_1 von 23.6.2020-2.7.2020

Zeitpunkt (1-29)	vollständig	vollständig und gültig	Bid > 0 und Ask > 0
8:04 (1)	0	0	29
8:33	258	251	464
9:03	299	290	501
9:33	364	357	534
10:03	383	378	537
10:34	402	394	546
11:03	412	405	550
11:33	426	419	552
12:03	427	417	536
12:33	439	432	538
13:03	443	436	539
13:33	448	439	538
14:03	452	446	541
14:33	454	447	543
15:03	461	452	543
15:33	459	451	543
16:03	460	454	545
16:33	468	462	545
17:03	475	470	548
17:33	479	473	549
18:03	479	470	545
18:33	479	469	545
19:03	478	469	546
19:33	480	470	545
20:03	475	465	546
20:33	465	0	0
21:03	452	0	0
21:33	444	0	0
22:03 (29)	437	0	0

Tabelle 5: Mehrfach-Kursabfrage für FEBRDUSA\_1 am 10.7.2020

Zeitpunkt	vollständig	vollständig und gültig	Bid > 0 und Ask > 0
10:39	445	443	507
10:43	445	443	507
10:48	446	442	506
10:52	449	446	507
10:56	448	446	507
11:01	446	444	508
11:05	450	448	508
11:26	452	449	565
11:34 (life-Trading)	456	453	564

Tabelle 6: Mehrfach-Kursabfrage für FEBRDUSA\_1 am 13.7.2020: 5-Minuten Abstände

## 2.7 Abhängigkeit von Kursabfrage-Parametern

In Tab. 7/S. 19 sind Mehrfach-Kursabfragen vorgenommen worden, wobei die Kursabfrage-Parameter variieren. Dabei wird der Timeout gegenüber der Standardeinstellung

$$\text{Timeout} = \text{Timer-Intervall} \cdot \text{Grenze} \quad \text{Timeout} = 20 \cdot 50\text{ms} = 1000\text{ms} = 1\text{s} \quad (1)$$

verringert.

Die ersten vier Kursabfragen sind für den Markt FEBRDUSA\_1 vorgenommen, die Kursabfrage um 15:59 betrifft alle 6185 bei IB verfügbaren Wertpapiere für Aktien.

Uhrzeit	Timer-Intervall (ms)	Grenze Timeout (ms)	vollständig	vollständig und gültig	Bid > 0 und Ask > 0	Offene Abfragen
Markt FEBRDUSA_1						
16:28	20	50	479	473	562	3/0
16:33	20	10	481	475	565	3/0
16:36	20	1	480	474	566	0/2
16:50	10	1	483	477	567	7/5
Alle Wertpapiere						
15:59	20	1	2918	2882	3825	1/17

Tabelle 7: Mehrfach-Kursabfrage für FEBRDUSA\_1 mit verschiedenen Parametern am 13.7.2020

Für die Einstellung 20/10 treten, entsprechend der Standardeinstellung 20/50, für den Markt FEBRDUSA\_1 keine Pausen auf. Bei der Einstellung 20/1 dagegen treten zwei Pausen auf, bei der Einstellung 10/1 sogar 5 Pausen. Bei der letzten Mehrfach-Kursabfrage um 15:59 treten 17 Pausen auf, da alle Wertpapiere abgefragt werden. Eine zu kleine Grenze für den Timeout führt demnach zu vielen Pausen, was durch längeres Warten im Zustand 2 des Automaten in Abb. 1/S. 7 verhindert werden kann.

Details für die letzte Mehrfach-Kursabfrage sind in Tab. 8/S. 19 dargestellt. Wegen des kurzen Timeouts nach  $20 \cdot 1\text{ms} = 20\text{ms}$  werden alle Nachrichten des Servers erst nach Ablauf eines Timeouts erhalten. Den 2918 vollständigen Datensätzen für die Einstellung 20/1 stehen gemäß Abb. 4/S. 13 Full gesamt=3046 vollständige Datensätze für die Einstellung 20/50 gegenüber. Dafür ist die Mehrfach-Kursabfrage für 20/1 wesentlich schneller beendet.

Status	Anzahl
Tout	58
ToutEnd	11
ToutFullEnd	2320
ToutFullEndEnd	584
ToutFullEndEndEnd	14
ToutReject	1975
ToutShort	95
ToutShortEnd	1048
ToutShortEndEnd	80

Tabelle 8: Mehrfach-Kursabfrage am 13.7.2020 um 15:59 (alle Wertpapiere)

### 3 Marktdefinition

Die Auswahl von Aktien und die damit verbundene Reduktion des Universums aller verfügbarer Wertpapiere zur Konstruktion von Portfolios ist ein wesentlicher und herausfordernder Prozess des quantitativen Investment Managements. Die Börsenplätze (Xetra/Frankfurt) der Deutsche Börse AG beinhalten derzeit beispielsweise über 15000 verfügbare Aktien [6]. Die quantitativen Methoden zur Findung von Investitionsentscheidungen beruhen im Allgemeinen auf Daten- und Modellbasierten Ansätzen [9]. In Verbindung mit der computergestützten Analyse von (historischen) Daten wurde bereits eine Vielzahl unterschiedlicher quantitativer Methoden für die Auswahl von profitablen Aktien entwickelt und untersucht. Ökonometrische Techniken, wie z.B. die lineare Regression, gehören zu den klassischen Ansätzen, welche insbesondere im Umfeld dynamischer Modelle und bedingt durch verrauschte Daten, Multikollinearität der Einflussfaktoren und nichtlineare Beziehungen zwischen Aktienrenditen und Faktoren eher selten erfolgreich sind [10]. Der Einsatz von Algorithmen und Optimierungsverfahren aus anderen Disziplinen kann ebenfalls zur Selektion von Aktien bzw. Prädiktion von Aktienrenditen genutzt werden und liefert teilweise bessere Erfolge im Vergleich zu den klassischen Modellen. Die Dateneinhüllanalyse, welche hauptsächlich zur Effizienzanalyse von Organisationseinheiten innerhalb des Operations Research verwendet wird, stellt einen solchen Ansatz dar [11]. Durch die seit ca. 1970 stark zunehmende Rechenleistung von Computern und die exponentiell steigende Verfügbarkeit von Daten sind in diesem Zusammenhang insbesondere auch Algorithmen aus dem Bereich des maschinellen Lernens, wie die Verwendung von Entscheidungsbäumen, Gegenstand wissenschaftlicher Untersuchungen [12]. Es hat sich hierbei gezeigt, dass beispielsweise Support Vector Machines und genetische Algorithmen genutzt werden können, um die Ergebnisse der Prädiktion und der Optimierung von Portfolios zu verbessern [13]. Mithilfe neuronaler Netze kann, unter anderem auch bedingt durch die Berücksichtigung nichtlinearer Zusammenhänge, der Erfolg bei der Auswahl von Aktien im Sinne eines profitablen Investment Managements ebenfalls gesteigert werden [10]. Die Datengrundlage der verwendeten Verfahren besteht im Wesentlichen aus Kurs- bzw. Finanzdaten und Fundamentaldaten und kann zusätzlich die Erwartung von Experten (Analysten) beinhalten. Untersuchungen unterstützen die Kombination der verschiedenen Datentypen als Basis der Optimierungsverfahren [14].

Im Folgenden wird das verwendete Verfahren zur Definition von Märkten im Sinne einer Vorauswahl von Aktien dargestellt. Es beinhaltet eine vorgeschaltete Selektion von Aktien, um die Ausführbarkeit einer Order in der Praxis sicherzustellen.

#### 3.1 Kursdatenquellen

Die drei Datenquellen zur Definition der in der Arbeit verwendeten Märkte sind:

1. Deutsche Börse AG: WKN, Symbol und ISIN für an der Frankfurter Börse gehandelte Wertpapiere für Aktien, gespeichert in einer Symbol-Liste (s. Abschn. 2.1).
2. Interactive Brokers (IB): Tagesaktuelle Kursdaten von Wertpapieren für Aktien, gespeichert in einer Mehrfach-Kurstabelle (s. Kap. 2).
3. Market Maker (MM): Die Wertpapierkennnummern der an der Frankfurter Börse gehandelte Aktien in der Währung EUR.

### 3.2 Mehrfach-Kurstabelle

Die Mehrfach-Kurstabelle wurde durch die folgenden drei Mehrfach-Kursabfragen erzeugt:

1. Erste Kursabfrage am 4.5.2020, 12:45 (s. Abschn. 2.5).
2. Zweite Kursabfrage am 23.6.2020, 18:15 (erste Ergänzung).
3. Dritte Kursabfrage am 24.6.2020, 17:10 (zweite Ergänzung).

Die erste Mehrfach-Kursabfrage ist in Abschn. 2.5 beschrieben. Bei der zweiten und dritten Kursabfrage wurde die Mehrfach-Kurstabelle schrittweise ergänzt, indem ein neuer Datensatz übernommen wird, wenn er mehr gültige Kurstypen enthält als der in der Mehrfach-Kurstabelle bereits gespeicherte Datensatz. Gültige Kurstypen sind positive Kurswerte für Bid, Ask, Last, High, Low, Volume und eine gültige Angabe für den Zeitstempel. Die Kursabfrage-Parameter sind für alle drei Abfragen gleich. Insbesondere ist daher das Timer-Intervall gleich 20ms und die Grenze für den Timeout beträgt 50.

Tab. 9/S. 21 zeigt die Entwicklung der Statuswerte nach der ersten, zweiten und dritten Mehrfach-Kursabfrage. Für  $n=1$  sind die Anzahl-Werte der Tab. 3/S. 13 entnommen. Die Anzahl-Werte für  $n=2$  und  $n=3$  ergeben sich aus einer Datenbankabfrage der Mehrfach-Kurstabelle nach der ersten und zweiten Ergänzung.

Nr	Status	Anzahl nach n-ter Abfrage		
		$n = 1$	$n = 2$	$n = 3$
1	FullEnd	2874	3106	3215
2	Reject	1962	1936	1936
3	ShortTout	256	202	169
4	ShortToutEnd	512	373	298
5	ShortToutFullEnd	42	44	47
6	ShortToutShort	65	56	56
7	ShortToutShortEnd	217	203	197
8	Tout	61	61	61
9	ToutFullEnd	130	152	157
10	ToutShort	18	13	13
11	ToutShortEnd	48	39	36
vollständig (1+5+9)		3046	3302	3419
vollständig und gültig		2985	3225	3333

Tabelle 9: Entwicklung der Statuswerte nach der  $n$ -ten Mehrfach-Kursabfrage

### 3.3 Definition zweier Märkte

#### Der alte Markt FEBRDUSA\_1:

Der Markt FEBRDUSA\_1 besitzt 867 Aktien. Es handelt sich um Aktien, die mindestens bis Ende 2013 Kurswerte besitzen mit einem Eröffnungskurs von mindestens 1 EUR. Um eine ausreichende Handelbarkeit sicher zu stellen, wurde die Zugehörigkeit zu einem der deutschen Aktienindizes DAX, MDAX, TECDAX, SDAX, HDAX, CDAX, Technology All Share, Prime All Share, GEX oder einem der beiden amerikanischen Aktienindizes S&P500 und Nasdaq gefordert.

#### Ausgangspunkt für die neuen Märkte:

Ausgangspunkt für neue Markt-Definitionen ist die Mehrfach-Kurstabelle nach Abschn. 3.2. Sie enthält nach Tab. 9/S. 21 3333 vollständige und gültige Datensätze. Die Wertpapiere sind Aktien oder Wertpapiere für Aktienfonds, gespeichert im Filterergebnis FELynxAlle. Wie beim Markt FEBRDUSA\_1 bezeichnet „FE“ ein Filterergebnis, d.h. eine Menge von Wertpapieren.

#### Der Markt FELynx\_1\_100:

Zur Definition neuer Märkte wird ebenfalls ausreichende Handelbarkeit gefordert, wobei aber jetzt ein anderer Ansatz verfolgt wird. Anstelle der Zugehörigkeit zu einem Aktienindex werden ein maximaler Spread sowie ein Mindestumsatz gefordert. Spread und Umsatz werden aus den in der Mehrfach-Kurstabelle gespeicherten Kursdaten Bid, Ask, Volume und Low gemäß

$$Spread = 100 \cdot (Ask - Bid) / Bid, \quad Umsatz = Volume \cdot Low$$

berechnet. Die folgende Tabelle zeigt den Einfluss der Begrenzung von Spread und Umsatz auf die Anzahl der Wertpapiere in der Mehrfach-Kurstabelle.

Zeile	Spread < (in %)	Umsatz > (in Tausend EUR)	Last > (in EUR)	Anzahl in FELynx
Bedingung unwirksam				
1	1000	0	0	3333
Last > 1 EUR, keine Bedingung für Umsatz				
2	100	0	1	3070
3	5	0	1	3004
4	2	0	1	2659
5	1	0	1	2255
6	0.5	0	1	1739
7	0.2	0	1	1023
Spread < 1%, Last > 1 EUR				
8	1	1	1	2133
9	1	5	1	1963
10	1	10	1	1861
11	1	50	1	1535
12	1	100	1	1339
13	1	200	1	1109
14	1	500	1	815
15	1	1000	1	614
16	1	2000	1	461
17	1	5000	1	282

Tabelle 10: Einfluss einer oberen Grenze für den Spread und eines Mindestumsatzes auf die Anzahl der Wertpapiere mit vollständigen und gültigen Datensätzen in der Mehrfach-Kurstabelle

In der ersten Zeile sind keine Begrenzungen wirksam. Die Anzahl der Wertpapiere entspricht daher FELynxAlle.

Ab Zeile 2 wird für Last eine untere Grenze von 1 EUR gefordert. Dies führt bereits zu einer deutlichen Verringerung der Anzahl. Zeile 2 zeigt, dass FELynxAlle  $3333 - 3070 = 263$  Pennystocks enthält.

In den Zeilen 3 bis 7 wird die Grenze für den Spread schrittweise verringert. Dies führt zu einer weiteren deutlichen Abnahme der Anzahl.

In Zeile 8 wird ein Mindest-Umsatz von 1000 EUR gefordert bei einem Spread unterhalb 1%. Mindestumsätze über 1 Mio. EUR für Nr. 15-17 führen auf kleine Märkte FELynx deutlich unter 1000 Wertpapiere.

Um einen möglichst großen Markt mit ausreichender Handelbarkeit zu definieren, wird die Parameterwahl in Zeile 12 gewählt. Der Spread liegt daher unter 1% und der Umsatz liegt über 100000 EUR. Die zugehörigen 1339 Wertpapiere sind im Filterergebnis FELynx\_1\_100 gespeichert:

$$\text{FELynx\_1\_100} := \{ \text{FELynxAlle} \mid \text{Spread} < 1\%, \text{Umsatz} > 100\text{T EUR}, \text{Last} > 1 \text{ EUR} \}. \quad (2)$$



Einen Überblick über die bisherigen und weiteren Definitionen enthält Tab. 11/S. 24.

Nr.	Filterergebnis	Anzahl Wertpapiere	Definition	Bemerkung
1	FELynxAlle	3333 (Stand 24.6.2020)	Alle Wertpapiere in der Mehrfach-Kurstabelle Lynx_Kurse_Gesamt_2020_6_24 mit vollständigen und gültigen Datensätzen	Enthält auch Aktienfonds. Alle 7 Kurstypen sind vorhanden und gültig.
2	FELynx_1_100	1339	Alle Wertpapiere von FELynxAlle mit Spread < 1%, Umsatz > 100T EUR, Last > 1 EUR	
3	FEMM_FfM_2020	7386	Alle Aktien in MM, gehandelt an der Frankfurter Börse in der Währung EUR mit Kurs am 2.1.2020 $\geq$ 1 EUR	
4	FELynxMM	1868	Durchschnitt von FELynxAlle und FEMM_FfM_2020	Das FE enthält alle bei IB handelbaren Aktien, die in MM verfügbar sind mit Börsenplatz FfM und Kurs $\geq$ 1 EUR am 2.1.2020.
5	FELynxMM_1_100	779	Durchschnitt von FELynx_1_100 und FEMM_FfM_2020	„Gut“ in Deutschland handelbare Aktien bei IB. Kursdaten in MM sind verfügbar.

Tabelle 11: Entwicklung der Märkte FELynxMM\_1\_100 und FELynxMM

**Die Märkte FELynxMM und FELynxMM\_1\_100:**

Wegen des Bezugs historischer Kursdaten mit dem Börsenprogramm Market Maker (MM) wird eine Einschränkung auf Aktien vorgenommen, die in Market Maker verfügbar sind, gespeichert in FEMM\_FfM\_2020. Für die Wertpapiere dieses Marktes wird außerdem ein Kurs von mindestens 1 EUR am 2.1.2020 gefordert. Erstens wird damit die Handelbarkeit der Wertpapiere bis mindestens Anfang 2020 gewährleistet. Zweitens sollen durch diese Forderung Pennystocks vom Handel weitgehend ausgeschlossen werden.

Als Ergebnis erhält man das Filterergebnis

$$\text{FELynxMM} = \text{FELynxAlle} \cap \text{FEMM\_FfM\_2020} \quad (3)$$

mit 1868 Aktien. Mit der Bezeichnung „LynxMM“ wird deutlich gemacht, dass dieser Markt durch eine Kombination der beiden Datenquellen IB bzw. Lynx und MM gebildet wird.

Analog zum Markt FELynMM erhält man den Markt

$$\text{FELynxMM\_1\_100} = \text{FELynx\_1\_100} \cap \text{FEMM\_FfM\_2020} . \quad (4)$$

**Beziehungen zu FEBRDUSA\_1:**

Die folgende Tabelle zeigt Durchschnitte des Marktes FEBRDUSA\_1 mit verschiedenen Filterergebnissen.

Zeile	Durchschnitt mit	Anzahl Wertpapiere	Bemerkung
1	FELynxMM_1_100	251	Die meisten Aktien von FEBRDUSA_1 ( $616 = 867 - 251$ ) sind nicht in FELynxMM_1_100 enthalten
2	FELynxMM	514	$353 = 867 - 514$ Aktien von FEBRDUSA_1 sind nicht in FELynx_MM enthalten
3	FEMM_FfM_2020	657	$210 = 867 - 657$ Aktien in FEBRDUSA_1 sind nicht in FEMM_FfM_2020 enthalten

Tabelle 12: Beziehung der Lynx-Märkte zum Markt FEBRDUSA\_1

Der Durchschnitt des Marktes FEBRDUSA\_1 mit den zwei Lynx-Märkten (Zeilen 1 und 2) enthält weniger Aktien als vielleicht erwartet. Selbst der Markt FELynxMM mit 1868 Aktien umfasst nicht FEBRDUSA\_1, denn 353 Wertpapiere von FEBRDUSA\_1 sind nicht in FELynxMM enthalten.

**Erklärung:**

Ein Grund dafür wird in Zeile 3 der Tabelle aufgedeckt. Demnach sind 210 Aktien von FEBRDUSA\_1 nicht im großen Markt FEMM\_FfM\_2020 bestehend aus 7386 Aktien enthalten, obwohl die Märkte FEBRDUSA\_1 und FEMM\_FfM\_2020 beide aus MM-Daten gewonnen wurden. Ein Unterschied zwischen diesen Märkten besteht darin, dass FEMM\_FfM\_2020 ausschließlich Aktien enthält, die am 2.1.2020 gehandelt wurden, während der Markt FEBRDUSA\_1 viele Aktien enthält, die zu diesem Zeitpunkt nicht mehr gehandelt werden konnten. Folglich fehlen diese Wertpapiere auch in den Märkten der Zeilen 1 und 2.

## 4 Datenqualität

Eine Datenreinigung als ein Vorverarbeitungsschritt ist beispielsweise für Data-Mining-Methoden essenziell, z.B. [15]. Hierbei werden in den Datensätzen befindliche Ausreißer (outliers) korrigiert, fehlende Werte geschätzt, inkonsistente Werte beseitigt und es werden ganze Datensätze verworfen. Fehlende oder fehlerhafte Daten sind in verschiedenen Forschungsgebieten wie z.B. der Medizin und den Sozialwissenschaften allgegenwärtig [16]. Neben diverser ad hoc Methoden haben sich im Falle fehlender Daten die Maximum-Likelihood-Methode und die Multiple Imputation als geeignete statistische Verfahren zur Vervollständigung der Daten erwiesen [17]. Für den Erfolg eines Handelssystems ist eine sorgfältige Datenreinigung ebenfalls von großer Bedeutung, z.B. [18]. Das Ziel der Vorverarbeitung ist dabei eine Verbesserung der Datenqualität.

### 4.1 Kursdatenbanken

Es wird im Folgenden die Datenqualität für die Lynx-Märkte und den Markt FEBRDUSA\_1 anhand der historischen Kurswerte

Eröffnung, Kassa, Schluss, Hoch und Tief

untersucht. Da die zu untersuchenden Märkte per Definition in MM verfügbar sind, werden die Kurswerte aus MM verwendet. Dabei werden Kurswerte sowohl in unbereinigter als auch in bereinigter Form exportiert, um Bereinigungsfaktoren zu bilden. Die unbereinigten Kurswerte werden zusammen mit den Bereinigungsfaktoren in eine (eigene) Kursdatenbank importiert.

#### A) Datenbankimport der Kurswerte:

Negative Kurswerte werden gelöscht, indem sie gleich 0 gesetzt werden. Damit Kurswerte eines Handelstages für ein bestimmtes Wertpapier als Datensatz in die Kursdatenbank übernommen werden, muss wenigstens einer der fünf Kurswerte  $> 0$  sein. Außerdem wird überprüft, ob Kurse unter dem Tiefkurs oder über dem Hochkurs liegen (Tiefunterschreitungen und Hochüberschreitungen). Wenn dies der Fall ist, wird durch das Löschen eines einzelnen Kurswertes versucht, Fehlerfreiheit herbeizuführen. Falls dies nicht gelingt, wird der Datensatz nicht in die Kursdatenbank übernommen.

#### B) Auslesen der Kurswerte aus der Kursdatenbank:

Beim Auslesen von Kurswerten aus der Kursdatenbank findet zunächst eine Bereinigung der unbereinigten Kurswerte mit Hilfe der Bereinigungsfaktoren statt. Dann erfolgt eine Analyse der bereinigten Kurswerte. Darauf basierend werden zwei unterschiedliche Arten der Kursänderung vorgenommen:

- Auffüllungen: Fehlender Kurswerte werden durch Kurswerte des gleichen Handelstages oder vorheriger Tage ersetzt,
- Kurskorrekturen: Kursausreißer werden korrigiert.

Als Ergebnis erhält man für jedes Wertpapier bereinigte, aufgefüllte und korrigierte Kurswerte. Vorgenommene Auffüllungen und Kurskorrekturen und weitere Auffälligkeiten in den Kurswerten eines Wertpapiers werden bei einer *Datenbank-Diagnose* genau aufgeschlüsselt. Sie werden als wertvolle Information über die Qualität der Kursdaten interpretiert, worauf im nächsten Abschnitt 4.2 näher eingegangen wird.

#### Zwei Kursdatenbanken:

Bezüglich der Kursdaten werden zwei verschiedene Kursdatenbanken verwendet. Die erste Kursdatenbank enthält Aktien des bisher ausschließlich verwendeten Marktes FEBRDUSA\_1. Sie wurde im Jahr 2015 neu aufgebaut und wird daher als *Kursdatenbank\_2015* bezeichnet. Die Lynx-Märkte wurden erst in 2020 definiert, danach erfolgte der Neuaufbau einer Kursdatenbank mit den Kursdaten für diese Märkte. Diese Kursdatenbank wird daher als *Kursdatenbank\_2020* bezeichnet.

### A) Kursdatenbank\_2020:

Diese Kursdatenbank wurde mit den MM-Kursdaten vom 21.7.2020 für die Lynx-Märkte neu aufgebaut. Zu Vergleichszwecken wurden auch die Kursdaten für den Markt FEBRDUSA\_1 neu aufgebaut und in der Kursdatenbank\_2020 gespeichert. Beim Neuaufbau ist zu berücksichtigen, dass für die Aktualisierung der Kursdaten innerhalb von MM die empfohlene Einstellung

„Wertpapier-Löschungen durchsetzen“

für den Bezug von Kursdaten stets benutzt wurde. Diese Einstellung beinhaltet die Löschung von Wertpapieren in MM, die aktuell nicht mehr gehandelt werden. Daher enthält die Kursdatenbank\_2020 nur Aktien, die am 21.7.2020 gehandelt werden konnten. Da viele Aktien von FEBRDUSA\_1 vor dem 21.7.2020 enden, enthält die Kursdatenbank nicht alle 867 Aktien dieses Marktes.

### B) Kursdatenbank\_2015:

Im Gegensatz zur Kursdatenbank\_2020 enthält die Kursdatenbank\_2015 die Kursdaten aller 867 Aktien des Marktes FEBRDUSA\_1, da diese Aktien alle in 2015 gehandelt wurden. Die Kursdatenbank wurde nach 2015 für die Aktien von FEBRDUSA\_1 periodisch mit Kursdaten von MM bis zum 21.7.2020 ergänzt. Die Aktien von FEBRDUSA\_1 enden daher alle frühestens in 2015 und spätestens am 21.7.2020.

Die folgende Abbildung gibt die Verteilungsfunktion für das Enddatum der Aktien des Marktes FEBRDUSA\_1 an (Ende). Hierbei wird die Anzahl der Aktien, die bis zu einem Handelstag enden, bezogen auf die Gesamtzahl aller 867 Aktien von FEBRDUSA\_1 dargestellt. Der Vollständigkeit halber ist die Verteilungsfunktion des ersten Handelstages ebenfalls angegeben (Beginn). Einige Werte dieser Verteilungsfunktionen sind in Tab. 13/S. 28 dargestellt.

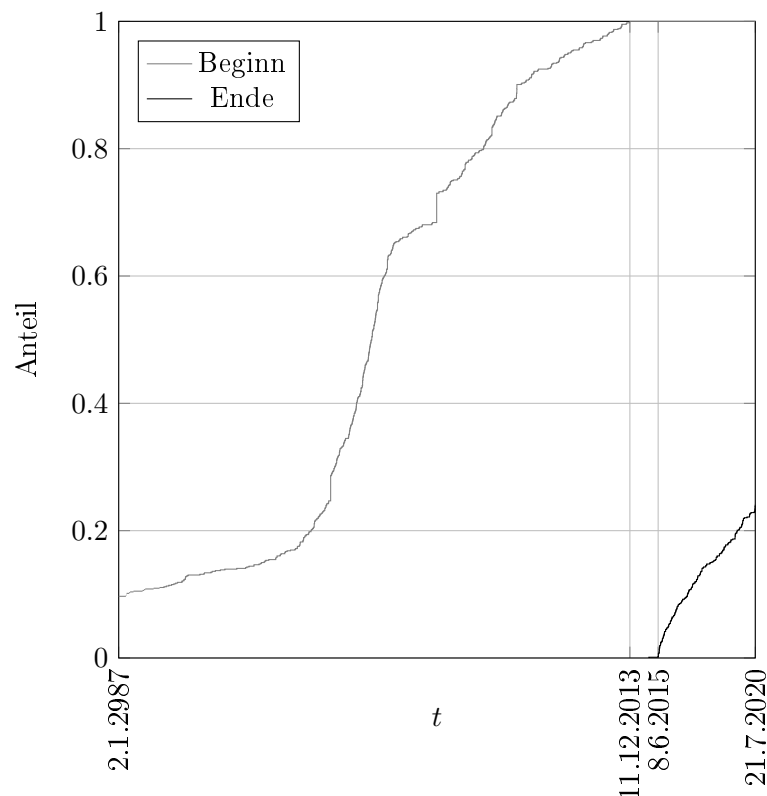


Abbildung 8: Verteilungsfunktion von Beginn und Ende für den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015

Datum	Beginn	Ende	Anzahl Aktien
2.1.1987	0.09689		84
11.12.2013	1		867
11.12.2014		0.001153	1
8.6.2015		0.002307	2
15.7.2020		0.233	202
16.7.2020		0.23414	203
17.7.2020		0.2376	206
20.7.2020		0.23991	208
21.7.2020		1	867

Tabelle 13: Einige Werte der Verteilungsfunktionen von Beginn und Ende für den Markt FEBRD-USA\_1 in der Kursdatenbank\_2015

#### Erläuterung:

- Am 2.1.1987 beginnen 84 Aktien (Anteil = 0.09689),
- der späteste Beginn ist der 11.12.2013,
- alle Aktien enden spätestens am 21.7.2020, da nur bis zu diesem Zeitpunkt Kursdaten in der Kursdatenbank\_2015 vorliegen,
- am 11.12.2014 endet ein erstes Wertpapier,
- am 8.6.2015 endet ein zweites Wertpapier,
- 4 Handelstage vor dem 21.7.2020 enden 202 Wertpapiere usw.

## 4.2 Datenbank-Diagnose: Definitionen

Bei der Datenbank-Diagnose erfolgt für jedes Wertpapier des untersuchten Marktes eine Analyse der historischen Kurswerte

Eröffnung, Kassa, Schluss, Hoch und Tief.

Analysiert werden

1. fehlende Kurswerte,
2. Kurswiederholungen, kurz *Repeats* genannt, die dadurch gekennzeichnet sind, dass alle Kurswerte eines Tages mit denen des Vortages identisch sind,
3. Kursausreißer in Form fehlerhafter Kurswerte, die beispielsweise durch einen fehlerhaften Dezimalpunkt zustande kommen [19],
4. große Abweichungen zwischen zwei Kurswerten an zwei aufeinanderfolgenden Handelstagen, sog. *Kurssprünge*,
5. Wertpapiere, deren Kurswerte vor dem Ende des Analysezeitraums enden.

Überarbeitungen der Kurswerte umfassen

- Auffüllungen unter Punkt 1,
- Kurskorrekturen von Kursausreißern unter Punkt 3.

### 4.2.1 Auffüllungen

Die im Zusammenhang mit Auffüllungen verwendeten Begriffe „vertikal“ und „horizontal“ ergeben sich aus der Vorstellung, dass die fünf historischen Kurswerte vertikal übereinander über der horizontal verlaufenden Zeitachse aufgetragen sind.

#### a) Vertikale Auffüllungen

Bei einer vertikalen Auffüllung werden fehlende Kurswerte eines Tages mit Hilfe von vorhandenen Kurswerten des gleichen Tages wie folgt berechnet:

- Ein fehlendes Hoch wird durch das Maximum der vorhandenen Kurswerte ersetzt,
- ein fehlendes Tief wird durch das Minimum der vorhandenen Kurswerte ersetzt,
- ein fehlender Kurswert für Eröffnung, Kassa, Schluss wird durch den Mittelwert vorhandener Kurswerte gebildet.

#### b) Horizontale Auffüllungen

Bei einer horizontalen Auffüllung werden für Tage ohne Kurswerte letzte vorherige Kurswerte verwendet. Auffüllungen für Kassa werden in der Datenbank-Diagnose nicht gezählt.

### 4.2.2 Repeats (Kurswiederholungen)

Gezählt wird die Anzahl der Tage, an denen alle Kurswerte mit denen des Vortages identisch sind. Solche Tage werden durch eine horizontale Auffüllung verursacht oder durch importierte Kurswerte, welche an zwei aufeinander folgenden Handelstagen identisch sind. Folglich setzt sich die Anzahl der Repeats aus zwei Anteilen zusammen:

- Horizontale Auffüllungen,
- importierte, sich wiederholende Kurswerte.

### 4.2.3 Kurskorrekturen

Die Kurskorrektur dient der Entfernung von Kursausreißern. Kursausreißer sind fehlerhafte Kurswerte, die beispielsweise durch einen fehlerhaften Dezimalpunkt zustande kommen. Es werden alle Kurswerte symmetrisch median gefiltert, indem für die fünf Kurswerte

$$s(t-2), s(t-1), s(t), s(t+1), s(t+2)$$

der Medianwert  $\text{MEDIAN}(t)$  gebildet wird. Eine größere Abweichung des Kurswertes  $s(t)$  vom Medianwert gemäß

$$s(t) \geq 2 \cdot \text{MEDIAN}(t) \text{ oder } s(t) \leq 0.5 \cdot \text{MEDIAN}(t)$$

wird als Kursausreißer definiert. Die Korrektur des Kursausreißers besteht darin, dass der Kursausreißer  $s(t)$  durch den Medianwert  $\text{MEDIAN}(t)$  ersetzt wird.

### 4.2.4 Fehler

Treten trotz Kurskorrektur Kursverhältnisse

$$c := s(t)/s(t-1)$$

mit einer Sprunghöhe

$$\text{Sprunghöhe} := \max\{c, 1/c\}$$

von mindestens 5 auf, wird der Kurswert  $s(t)$  als fehlerhaft interpretiert. Der Kurswert  $s(t)$  wird jedoch übernommen.

### 4.2.5 PennyTod und Tod

Ein *Tod* liegt vor, wenn die Kurse spätestens vier Tage vor dem Ende des Analysezeitraums enden. Beträgt der letzte vorhandene Eröffnungskurs außerdem weniger als 1 EUR, liegt definitionsgemäß ein *Penny-Tod* vor. Ein Tod entsteht beispielsweise bei einer Änderung der WKN. Die unter der alten WKN gespeicherten Kursdaten enden dann vorzeitig. Ein anderes Beispiel ist die Beendigung des Handels für ein Wertpapier. Ein kleiner Kurs am Ende des Handelns, d.h. ein Penny-Tod spricht in diesem Fall für einen Totalverlust.

### 4.2.6 Maximale Sprunghöhe

Es wird die Anzahl der Wertpapiere gezählt, für die trotz Korrektur eine maximale Sprunghöhe von mindestens 2, 5 oder 10 auftritt. Maximiert wird über alle Tage und die fünf historischen Kurswerte.

## 4.3 Datenbank-Diagnose: Ergebnisse

Die Auswertung erfolgt für die Aktien-Märkte

1. FEBRDUSA\_1,
2. FeLynxMM\_1\_100,
3. FELynxMM.

Der bisherige Markt FEBRDUSA\_1 wird für die Kursdatenbank\_2015 und Kursdatenbank\_2020 ausgewertet (siehe Tab. 14/S. 31). Für die beiden Lynx-Märkte ist nur eine Auswertung in der Kursdatenbank\_2020 möglich, da in der Kursdatenbank\_2015 nur Aktien des Marktes FEBRDUSA\_1 aufgenommen sind. Der Auswertungszeitraum umfasst die Handelstage

2.1.2002 bis zum 21.7.2020 (Auswertungszeitraum).

Markt	FEBRDUSA_1	FEBRDUSA_1	FELynxMM_1_100	FELynxMM
Anzahl Aktien	867	867	779	1868
Kursdatenbank	2015	2020	2020	2020
Anzahl Datensätze				
Kurswerte vorhanden (100 %)	3,546,972	2,885,204	2,731,243	6,549,665
Auffüllungen vertikal	16,972	8182	2880	8575
Auffüllungen horizontal	16,960	16,859	24,159	59,409
Repeats in Prozent	107,842 3%	75,998 2.6%	56,315 2.1%	204,610 3.1%
Korrekturen	242	116	68	172
Fehler	25	9	15	42
Anzahl Aktien				
Kurswerte vorhanden (100 %)	867	670	776	1857
Penny-Tod in Prozent	26 3%	0	1	3
Tod in Prozent	202 23%	9 1.3%	5	15
Max. Sprunghöhe ≥ 2 in Prozent	106 12.2%	64 9.6%	44 5.7%	174 9.4%
≥ 5 in Prozent	15 1.7%	6	6	25 1.3%
≥ 10	5	2	4	12

Tabelle 14: Datenbank-Diagnose für die Märkte FEBRDUSA\_1, FELynxMM\_1\_100 und FELynxMM. Auswertungszeitraum ist der 2.1.2002 bis 21.7.2020. Nicht angegebene Prozentwerte liegen unter 1%.

Der obere Teil der Tabelle gibt die Anzahl der Datensätze an. Dabei liefert jede Aktie des Marktes und jeder Handelstag einen Datensatz. Hierbei werden nur solche Handelstage berücksichtigt, die sowohl im Handelszeitraum der Aktie liegen als auch in den Analysezeitraum fallen. Der untere Teil gibt die Anzahl der Aktien an, welche Kurswerte besitzen, die in den Auswertungszeitraum fallen. Die Anzahl aller Datensätze bzw. Aktien ist unter „Kurswerte vorhanden“ angegeben. Die Angabe von 100% bedeutet, dass sich Prozentangaben auf diese Anzahl als Grundgröße beziehen. Grundsätzlich sind nur Prozentwerte angegeben, wenn sie mindestens 1% betragen.



### 4.3.1 Anzahl Datensätze

Dem oberen Teil der Tabelle entnimmt man, dass nur Repeats einen nennenswerten Anteil an Datensätzen besitzen.

#### Auffüllungen und Repeats:

Die Anzahl horizontaler Auffüllungen ist deutlich kleiner als die Anzahl der Repeats. Daher werden Repeats hauptsächlich durch importierte Kurswerte verursacht, die an aufeinander folgenden Handelstagen identisch sind, und nicht durch horizontale Auffüllungen. Die Rate für Repeats liegt bei etwa 3% oder etwas darunter. Die kleinste Rate besteht für den Markt FELynxMM\_1\_100. Dieser Markt besitzt auch die kleinste Anzahl vertikaler Auffüllungen.

#### Korrekturen und Fehler:

Die Korrektur von Kursausreißern erfolgt selten. Noch seltener verbleiben nach der Korrektur Fehler, d.h. Kurssprünge mit einer Sprunghöhe  $\geq 5$ . Für die Kursdatenbank\_2015 beispielsweise erfolgen 242 Korrekturen und danach verbleiben 25 Datensätze mit einer Sprunghöhe  $\geq 5$ . Sie werden von 15 Aktien verursacht, welche eine maximale Sprunghöhe  $\geq 5$  besitzen, wie der untere Teil der Tabelle zeigt.

### 4.3.2 Anzahl Aktien

Dem unteren Teil der Tabelle entnimmt man, dass nur Sprunghöhen  $\geq 2$  und teilweise Sprunghöhen  $\geq 5$  sowie Tode in der Kursdatenbank\_2015 einen nennenswerten Anteil an Aktien besitzen.

#### Kurswerte vorhanden:

Die Anzahlwerte für „Kurswerte vorhanden“ wird durch die Anzahl aller Aktien des Marktes begrenzt, welche in der Tabelle oben angegeben sind. Ein gewisser Schwund ist darauf zurückzuführen, dass die Marktdefinitionen in 2020 vor dem 21.7.2020 vorgenommen wurden. Für den Markt FEBRDUSA\_1 in der Kursdatenbank\_2020 liegt die Anzahl mit 670 deutlich unter der Anzahl von 867 Aktien des Marktes. Dies liegt daran, dass  $867 - 670 = 197$  Aktien des Marktes FEBRDUSA\_1 am 21.7.2020 nicht mehr bei MM vertreten waren.

#### Penny-Tod und Tod:

In der Kursdatenbank\_2015 erleiden 202 Aktien einen Tod. Laut Definition enden diese Aktien spätestens vier Tage vor dem Ende des Analysezeitraums (21.7.2020), laut Tab. 13/S. 28 also bis zum 15.7.2020. Für dieses Datum findet man in Tab. 13/S. 28 als Bestätigung 202 Aktien, die bis zu diesem Zeitpunkt enden.

Für die Kursdatenbank\_2020 ist die Anzahl der Aktien, die einen Tod erleiden, gering und Aktien enden als Pennystocks noch seltener. Der geringe Schwund an Aktien ist der Tatsache geschuldet, dass Marktdefinitionen mit Hilfe der zu Beginn 2020 in MM verfügbaren Aktien vorgenommen wurde, von denen nur ein geringer Teil am 21.7.2020 nicht mehr in MM verfügbar waren.

Die Kursdatenbank\_2020 enthält für den Markt FEBRDUSA\_1 keine Aktien, die vorzeitig als Pennystocks enden, denn die Anzahl der Penny-Tode ist 0. Auch für die beiden Lynx-Märkte ist die Anzahl der Penny-Tode sehr gering. Andererseits sind sie für eine Analyse von Kursdaten wichtig, da sich unter ihnen Aktien befinden, die einen Totalverlust erlitten haben. Sie fehlen in der Kursdatenbank\_2020, da diese mit den Kursdaten von MM neu aufgebaut wurde, welche diese Aktien zum Zeitpunkt des Neuaufbaus nicht mehr enthielt. Die Kursdatenbank\_2015 dagegen enthält solche Aktien. In dieser Hinsicht bildet die Kursdatenbank\_2015 Kursdaten realistischer ab als die Kursdatenbank\_2020.

#### Maximale Sprunghöhe:

Die häufigsten Sprunghöhen  $\geq 2$  weist der Markt FEBRDUSA\_1 in der Kursdatenbank\_2015 auf (12.2%). Den kleinsten Anteil besitzt der Markt FELynxMM\_1\_100 (5.7%). Sprunghöhen  $\geq 10$  treten sehr selten auf.

## 5 Abhängigkeit der Statistik vom Markt

Die statistische Beschreibung beruht auf einer Segmentierung des Handelszeitraums. Ein Segment stellt hierbei  $\tau_h$  aufeinander folgende Handelstage dar, wobei  $\tau_h$  die Haltefrist bezeichnet, welche im Folgenden auf den Wert

$$\tau_h = 30 \quad (5)$$

gesetzt ist. Die statistische Auswertung erfolgt wie in Tab. 14/S. 31 für den Markt FEBRDUSA\_1 und die Kursdaten in der Kursdatenbank\_2015 sowie für drei Märkte und die Kursdaten in der Kursdatenbank\_2020. Hierbei werden die zeitliche Gewinnentwicklung sowie der Gewinn und andere statistische Größen für den gesamten Handelszeitraum dargestellt. Alle statistischen Größen sind in Teil 3 der Schriftenreihe beschrieben. Eine statistische Auswertung erfolgte bereits für den Markt FEBRDUSA\_1 und die Kursdatenbank\_2015 für den Handelszeitraum 23.4.2002 – 2.6.2017 [20, Tab. 1 und Abb. 9].

### Kumulativer Gewinnfaktor:

Der kumulative Gewinnfaktor beschreibt die zeitliche Gewinnentwicklung durch die Entwicklung des Gesamtkapitals  $C_{tot}(q \cdot \tau_h)$  am Ende des  $q$ -ten Segments gemäß [20, Gl. (18)],

$$C_{tot}((q+1) \cdot \tau_h) = G_{cum}(q) \cdot C_{tot}(1 \cdot \tau_h), \quad 1 \leq q \leq Q. \quad (6)$$

Hierbei ist  $Q$  die Anzahl der Segmente des Handelszeitraums bis auf das letzte Segment, dass keinen Erlös im Handelszeitraum bewirkt.  $C_{tot}(1 \cdot \tau_h)$  ist das Gesamtkapital am Ende des ersten Segments und stellt das Barkapital zu Beginn des Handelns dar.

### Handelszeitraum:

Der jetzt zugrunde gelegte Handelszeitraum umfasst  $Q+1 = 155$  Segmente mit  $\tau_h = 30$  Handelstagen und somit  $155 \times 30 = 4650$  Handelstagen insgesamt. Er reicht vom 23.4.2002 bis zum August 2020 und beinhaltet daher den „Corona-Crash“ und eine erste Phase der darauf folgenden Kurserholung,

Handelszeitraum Kursdatenbank\_2015: 23.4.2002 – 21.8.2020 .

Auswertungen für die Kursdatenbank\_2020 werden für einen davon geringfügig abweichenden Handelszeitraum vorgenommen:

Handelszeitraum Kursdatenbank\_2020: 23.4.2002 – 20.8.2020 .

Der Unterschied ergibt sich aus geringfügig abweichenden Zeitbasen der beiden Kursdatenbanken. Nähere Einzelheiten sind den Rohdaten (Kap. 3) zu entnehmen.

### Kauf- und Verkaufsbedingung:

Die Kaufbedingung ist die *Momentum-Strategie* aus [20, Gl. (92)] und lautet

$$1.5 \leq c_1(t) < 1.625, \quad s^*(t) \geq 2. \quad (7)$$

Das Merkmal  $c_1(t)$  setzt den Kurs einer Aktie am Handelstag  $t$  mit dem gleitenden Durchschnitt der Kurse der letzten 200 Handelstage ins Verhältnis.  $s^*(t)$  stellt den unbereinigten Kurswert der Aktie am Handelstag  $t$  dar.

## 5.1 Zeitliche Verläufe

Dargestellt werden die Verläufe des kumulativen Gewinnfaktors für die ersten  $Q = 154$  Segmente. Neben den Plots werden Rückgänge des kumulativen Gewinnfaktors  $\geq 20\%$  in der letzten Spalte von Tab. 15/S. 35 – Tab. 17/S. 38 angegeben. Ein Rückgang von einem hohen Kurswert zu einem tiefen Kurswert ergibt sich aus

$$\text{Rückgang} := [\text{Hoch} - \text{Tief}] / \text{Hoch} .$$

Die dritte Tabellenspalte gibt den (von der Zeitbasis der Kursdatenbank abhängigen) letzten Handelstag des  $q$ -ten Segments an.

### Diskussion der Ergebnisse:

**Gewinnentwicklung:** Der Verlauf des kumulativen Gewinnfaktors wird deutlich vom Markt beeinflusst. Erstens unterscheiden sich die Gewinnfaktoren am Handelsende beträchtlich, worauf gesondert im nächsten Abschnitt 5.2 eingegangen wird. Zweitens fallen Rückgänge nach einem Hoch unterschiedlich aus.

#### FEBRDUSA\_1 (Abschn. 5.1.1 und 5.1.2):

Rückgänge sind für den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015 mit 37.3% am größten. Der Kursrückgang nach dem 20.1.2020 ist auf den „Corona-Crash“ zurückzuführen, fällt aber mit 20.1% vergleichsweise gering aus. Andererseits findet keine Erholung bis zum Handelsende am 21.8.2020 statt. Die Kursdatenbanken haben einen großen Einfluss auf die Gewinnentwicklung. So liegen Rückgänge in der Kursdatenbank\_2020 unter 20%. Diese Kursdatenbank vermittelt somit ein zu optimistisches Bild über die Gewinnentwicklung für den Markt FEBRDUSA\_1.

**Lynx-Märkte (Abschn. 5.1.3 und 5.1.4):** Die beste Gewinnentwicklung besteht für den Markt FELynxMM\_1\_100. Die Rückgänge fallen vergleichsweise moderat aus und der kumulative Gewinnfaktor steigt rasch und beständig bis zum Handelsende an. Dagegen weist der größere Markt FELynxMM mehrere Rückgänge von über 20% auf, welche den Anstieg des Gewinnfaktors „bremsen“. Die Forderung für den Markt FELynxMM\_1\_100 nach Aktien mit noch vertretbarem Umsatz und Spread wirkt sich somit positiv auf die Gewinnentwicklung aus.

**Corona-Crash:** Der Corona-Crash ist nicht in allen Märkten deutlich spürbar. Für den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015 äußert er sich in einem Rückgang von 20.1% ohne Erholung zu Handelsende. Für den Markt FELynxMM\_1\_100 beträgt der Rückgang 20.6% und die Erholung bis Handelsende ist ausgeprägt. Bei dem Markt FELynxMM liegt der Rückgang etwas unter 20%, weswegen er in Tab. 17/S. 38 nicht dargestellt ist. Die Erholung ist etwas schwächer als für den Markt FELynxMM\_1\_100.

## 5.1.1 Kursdatenbank\_2015 und Markt FEBRDUSA\_1

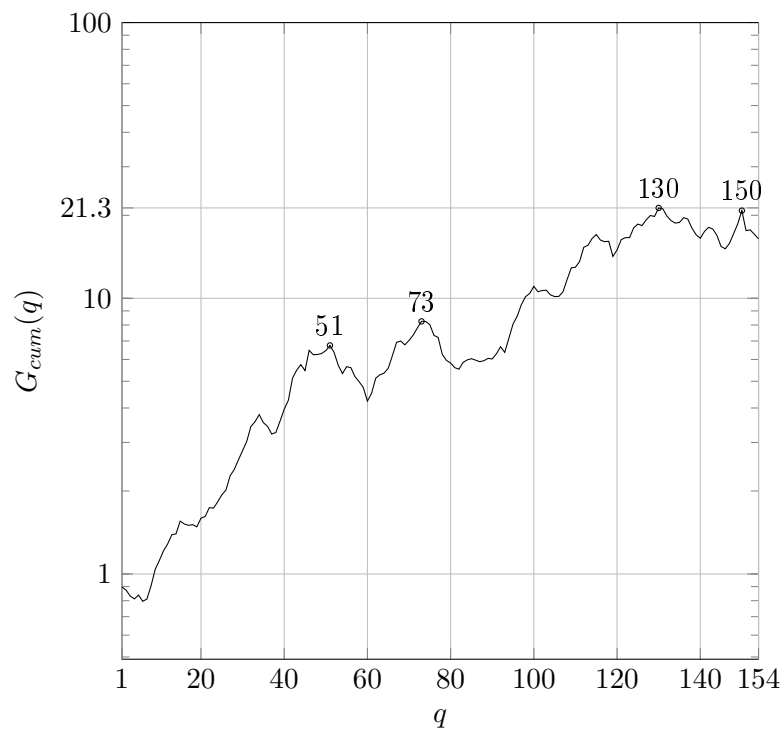


Abbildung 9: Segmentverlauf des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015

Hoch/Tief	Segment-Nr. $q$	Handelstag	$G_{cum}(q)$	Rückgang (%)
Hoch	51	30.4.2008	6.75	
Tief	60	26.5.2009	4.23	37.3%
Hoch	73	1.12.2010	8.25	
Tief	82	20.12.2011	5.53	33%
Hoch	130	28.8.2017	21.3	
Tief	146	26.7.2019	15.1	29.1%
Hoch	150	20.1.2020	20.8	
Tief	154	10.7.2020	16.4	20.1%

Tabelle 15: Hochs und Tiefs des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015

## 5.1.2 Kursdatenbank\_2020 und Markt FEBRDUSA\_1

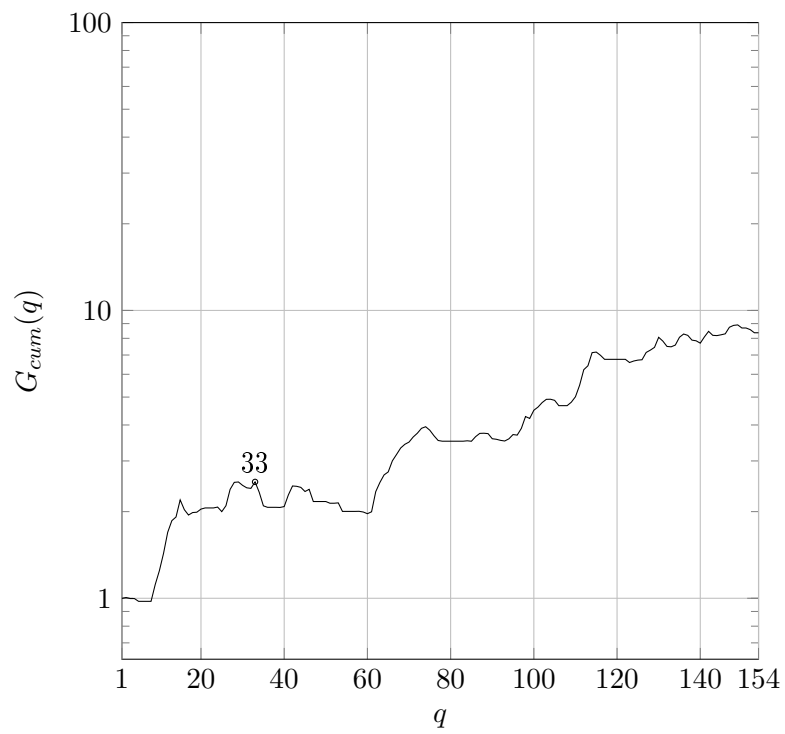


Abbildung 10: Segmentverlauf des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FEBRDUSA\_1 in der Kursdatenbank\_2020

Die Rückgänge liegen alle unter 20%.

## 5.1.3 Kursdatenbank\_2020 und Markt FELynxMM\_1\_100

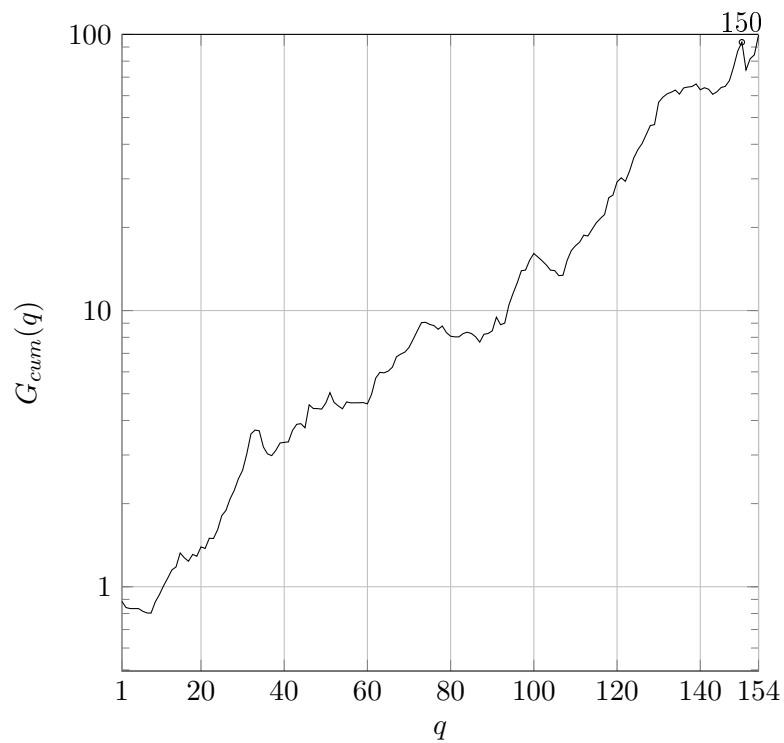


Abbildung 11: Segmentverlauf des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FELynMM\_1\_100 in der Kursdatenbank\_2020

Hoch/Tief	Segment-Nr. $q$	Handelstag	$G_{cum}(q)$	Rückgang (%)
Hoch	150	17.1.2020	93.7	
Tief	151	28.2.2020	74.4	20.6%

Tabelle 16: Hochs und Tiefs des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FELynxMM\_1\_100 in der Kursdatenbank\_2020

## 5.1.4 Kursdatenbank\_2020 und Markt FELynxMM

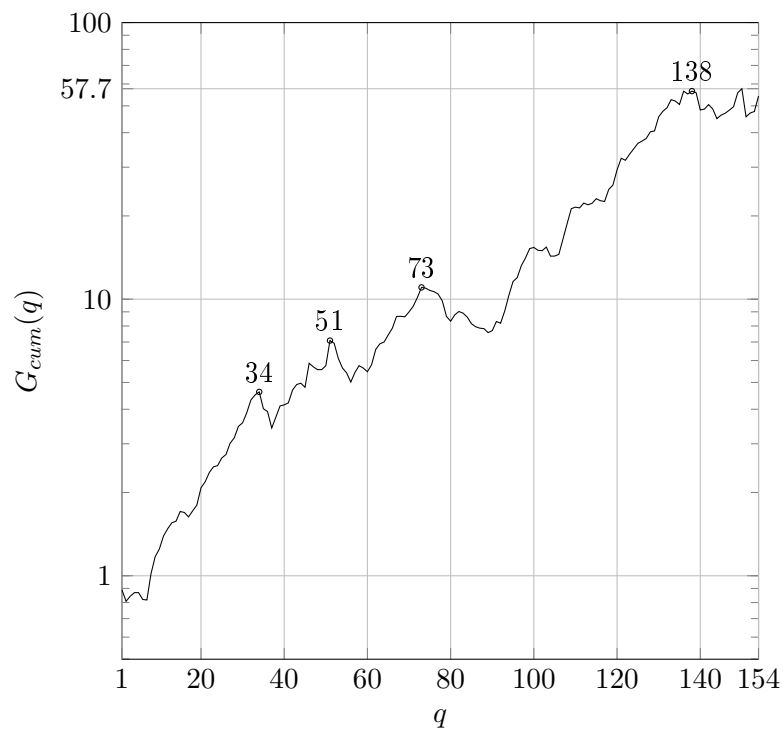


Abbildung 12: Segmentverlauf des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FELynxMM in der Kursdatenbank\_2020

Hoch/Tief	Segment-Nr. $q$	Handelstag	$G_{cum}(q)$	Rückgang (%)
Hoch	34	21.4.2006	4.6	
Tief	37	29.8.2006	3.4	26%
Hoch	51	30.4.2008	7.1	
Tief	56	27.11.2008	5.0	30%
Hoch	73	1.12.2010	11.0	
Tief	89	15.10.2012	7.6	20.1%
Hoch	138	9.8.2018	56.5	
Tief	144	30.4.2019	45	20%

Tabelle 17: Hochs und Tiefs des kumulativen Gewinnfaktors für die Momentum-Strategie und den Markt FELynxMM in der Kursdatenbank\_2020

## 5.2 Gesamter Handelszeitraum

Tab. 18/S. 40 zeigt den Gewinnfaktor, den Investitionsgrad und das Moment  $M_1$  erster Ordnung für den gesamten Handelszeitraum. Alle drei Statistikgrößen findet man in Teil III der Schriftenreihe.

### Gewinnfaktor:

Der Gewinnfaktor für den gesamten Handelszeitraum setzt das Gesamtkapital am Ende des Handels mit dem zu Handelsbeginn gemäß [20, Gl. (19)],

$$G = C_{tot}((Q + 1) \cdot \tau_h) / C_{tot}(1 \cdot \tau_h) \quad (8)$$

in Zusammenhang. Der Vergleich mit dem kumulativen Gewinnfaktor in (6) liefert den Zusammenhang

$$G = G_{cum}(Q), \quad (9)$$

d.h. der Gewinnfaktor für den gesamten Handelszeitraum entspricht dem kumulativen Gewinnfaktor für das letzte Segment  $Q$  in Handelszeitraum, das einen Erlös bewirkt. Der Zusammenhang mit den Gewinnfaktoren für Segmente in [20, Gl. (14c)],

$$g(q) = C_{tot}((q + 1) \cdot \tau_h) / C_{tot}(q \cdot \tau_h), \quad 1 \leq q \leq Q \quad (10)$$

lautet [20, Gl. (12)]

$$G = G_{cum}(Q) = \prod_{q=1}^Q g(q). \quad (11)$$

### Investitionsgrad:

Der Investitionsgrad für den gesamten Handelszeitraum ist der mittlere Investitionsgrad am Ende der ersten  $Q$  Segmente gemäß [20, Gl. (6)],

$$IG = \frac{1}{Q} \sum_{q=1}^Q IG(q \cdot \tau_h). \quad (12)$$

### Moment erster Ordnung:

Das Moment  $M_1$  erster Ordnung stellt wie  $G$  einen Gewinnfaktor für den gesamten Handelszeitraum dar. Im Gegensatz zu  $G$  wird keine Reinvestition von Erlösen aus Verkäufen vorgenommen. Das Moment erster Ordnung kann auf den mittleren Investitionsgrad  $IG$  und den leichter zugänglichen arithmetischen Mittelwert

$$AMG = \frac{1}{Q} \sum_{q=1}^Q g(q) \quad (13)$$

zurückgeführt werden, denn es besteht der Zusammenhang [20, Gl. (24)]

$$AMG = 1 + IG \cdot [M_1 - 1]. \quad (14)$$

Bei Vollinvestition ( $IG = 1$ ) stimmen demnach  $M_1$  und  $AMG$  überein. Bei einem Investitionsgrad zwischen 0 und 1 dagegen liegt der  $AMG$  näher bei 1 als das Moment erster Ordnung, da der  $AMG$  den Investitionsgrad ( $IG$ ) berücksichtigt,  $M_1$  dagegen nicht.



Markt	FEBRDUSA_1	FEBRDUSA_1	FELynxMM_1_100	FELynxMM
Anzahl Aktien Kurswerte vorhanden	867	670	776	1857
Kursdatenbank	2015	2020	2020	2020
Gewinnfaktor $G$	16.4	8.36	99.4	54.3
Inv.-grad $IG$	79.8%	36.8%	76.1%	93.3%
Moment $M1$	1.025	1.041	1.042	1.031

Tabelle 18: Statistik-Ergebnisse für die Märkte FEBRDUSA\_1, FELynxMM\_1\_100 und FELynxMM. Auswertungszeitraum ist der 2.1.2002 bis 21.7.2020.

### Diskussion der Ergebnisse:

In der Tabelle sind oben als ein Maß für die Marktgröße die Anzahl der Aktien angegeben, für die Kurswerte in der Kursdatenbank vorhanden sind (s. Tab. 14/S. 31). Die Statistik-Ergebnisse unterscheiden sich gravierend:

#### Markt FEBRDUSA\_1:

Der Gewinnfaktor ist für die Kursdatenbank\_2015 etwa doppelt so groß wie für die Kursdatenbank\_2020. Der Grund ist der höhere Investitionsgrad in der Kursdatenbank\_2015. Das Moment  $M1$  dagegen wird nicht vom Investitionsgrad beeinflusst und zeigt einen deutlich größeren Wert für die Kursdatenbank\_2020. Die Kursdatenbank\_2020 führt folglich zu einem zu optimistischen Wert für  $M1$ .

#### Markt FELynxMM\_1\_100:

Den größten Gewinnfaktor unter allen Märkten weist der Markt FELynxMM\_1\_100 auf. Dies trifft auch für das Moment  $M1$  zu.

### Investitionsgrade:

Die Investitionsgrade folgen der Marktgröße, wenn die in der Tabelle angegebene Anzahl der Aktien mit vorhandenen Kurswerten zugrunde gelegt wird:

$$\begin{aligned}
 \text{FEBRDUSA}_1/2020 (670) &< \text{FELynxMM}_1_100 (776) \\
 &< \text{FEBRDUSA}_1/2015 (867) \\
 &< \text{FELynxMM} (1857) .
 \end{aligned}$$

## 6 Zusammenfassung

**Mehrfach-Kursabfragen (Kap. 2):** Der Abruf tagesaktueller Kursdaten vom Server von Interactive Brokers (IB) stößt schnell an Grenzen, wenn die Kurse vieler Wertpapiere sequentiell abgefragt werden. Eine Unterbrechung der Mehrfach-Kursabfrage, welche eine Unterbrechung der Verbindung zum Server beinhaltet, hilft weiter. Das Ergebnis einer solchen Mehrfach-Kursabfrage bemisst sich an der Anzahl erhaltener vollständiger Datensätze, welche alle sechs Kurstypen Bid, Ask Last, High, Low und Volume sowie einen Zeitstempel enthalten. Das Ergebnis variiert an unterschiedlichen Handelstagen und zu unterschiedlichen Uhrzeiten. Aus diesem Grund werden Mehrfach-Kursabfragen an unterschiedlichen Handelstagen vorgenommen, um Kursdaten schrittweise zu ergänzen.

**Marktdefinition (Kap. 3):** Durch Kursabfrage-Parameter werden unterschiedliche Anforderungen an den Server von IB gestellt. Es werden geeignete Einstellungen angegeben, die einen erfolgreichen Datenabruf auch für alle 6185 bei IB verfügbaren Wertpapiere für Aktien ermöglichen. Die Beschränkung auf vollständige Datensätze mit gültigen Kursdaten führt auf den Markt FELynxAlle mit 3333 Wertpapieren. Dieser Markt enthält neben Aktien auch andere Wertpapiere wie beispielsweise Aktienfonds. Die weitere Beschränkung auf die im Börsenprogramm Market Maker erhältlichen Aktien liefert 1868 Aktien. Die Einschränkung auf Aktien mit einem Spread unter 1%, Umsatz  $> 100T$  EUR und Last  $> 1$  EUR liefert den kleineren Markt FELynxMM\_1\_100 mit 779 Aktien. Die Lynx-Märkte enthalten ausschließlich Wertpapiere, die in 2020 gehandelt werden können. Darin unterscheiden sie sich grundsätzlich vom bisher verwendeten Markt FEBRDUSA\_1, der viele Aktien enthält, die nicht mehr in 2020 gehandelt werden.

**Datenqualität (Kap. 4):** Die Analyse der Qualität der Kursdaten wird in zwei Kursdatenbanken durchgeführt. Kursdaten des Marktes FEBRDUSA\_1 liegen seit 2015 vollständig in der Kursdatenbank\_2015 vor. Für die Lynx-Märkte musste die Kursdatenbank\_2020 neu aufgebaut werden. Neben den Lynx-Märkten enthält sie außerdem die Aktien des Marktes FEBRDUSA\_1, sofern diese in 2020 noch gehandelt wurden. In der Kursdatenbank\_2020 fehlen daher insbesondere Aktien, die vor 2020 einen Totalverlust erlitten.

Einen Hinweis auf solche Aktien enthält die signifikante Anzahl der Aktien in der Kursdatenbank\_2015, die einen *Penny-Tod* erleiden. Solche Aktien fehlen in der Kursdatenbank\_2020. Insofern bildet die Kursdatenbank\_2015 Kursdaten des Marktes FEBRDUSA\_1 realistischer ab als die Kursdatenbank\_2020.

*Repeats* (Kurswiederholungen) werden hauptsächlich durch importierte Kurswerte verursacht, die an aufeinander folgenden Handelstagen identisch sind, und nicht durch horizontale Auffüllungen. Die Rate für Repeats liegt bei etwa 3% oder etwas darunter. Die kleinste Rate besteht für den Markt FELynxMM\_1\_100. Dieser Markt besitzt auch die kleinste Anzahl vertikaler Auffüllungen.

Die Korrektur von *Kursausreißern* ist selten erforderlich. Noch seltener verbleiben nach der Korrektur *Kurssprünge* mit einer Sprunghöhe  $\geq 5$ .

Die häufigsten *Sprunghöhen*  $\geq 2$  weist der Markt FEBRDUSA\_1 in der Kursdatenbank\_2015 auf (12.2%). Den kleinsten Anteil besitzt der Markt FELynxMM\_1\_100 (5.7%).

**Statistik-Auswertung (Kap. 5):** Der kumulativen Gewinnfaktor wird deutlich vom Markt beeinflusst. Insbesondere unterscheiden sich die Gewinnfaktoren am Handelende beträchtlich und Rückgänge nach einem Hoch fallen unterschiedlich stark aus.

### A) Zeitliche Verläufe:

**FEBRDUSA\_1:** Zeitliche Verläufe des kumulativen Gewinnfaktors zeigen unterschiedliche Kursrückgänge. Rückgänge sind für den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015 mit 37.3% am größten. Rückgänge in der Kursdatenbank\_2020 liegen dagegen unter 20%. Die Kursdatenbank\_2020 vermittelt somit ein zu optimistisches Bild über die Gewinnentwicklung.

**Lynx-Märkte:** Die beste Gewinnentwicklung besteht für den Markt FELynxMM\_1\_100. Die Rückgänge fallen vergleichsweise moderat aus und der kumulative Gewinnfaktor steigt rasch und beständig bis zum Handelsende an. Dagegen weist der größere Markt FELynxMM mehrere Rückgänge von über 20% auf, welche den Anstieg des Gewinnfaktors „bremsen“. Die Forderung für den Markt FELynxMM\_1\_100 nach Aktien mit noch vertretbarem Umsatz und Spread wirkt sich somit positiv auf die Gewinnentwicklung aus.

**Corona-Crash:** Für den Markt FEBRDUSA\_1 in der Kursdatenbank\_2015 äußert sich der Corona-Crash in einem Rückgang von etwa 20% ohne vollständige Erholung bis zum 10.7.2020. Für den Markt FELynxMM\_1\_100 liegt der Rückgang etwas über 20% und die Erholung bis Handelsende ist ausgeprägt. Bei dem Markt FELynxMM liegt der Rückgang etwas unter 20% und die Erholung ist etwas schwächer als für den Markt FELynxMM\_1\_100.

### **B) Gesamter Handelszeitraum:**

Die Statistik-Ergebnisse unterscheiden sich gravierend: Für den Markt FEBRDUSA\_1 ist der Gewinnfaktor in der Kursdatenbank\_2015 etwa doppelt so groß wie in der Kursdatenbank\_2020. Der Grund ist der höhere Investitionsgrad in der Kursdatenbank\_2015. Das Moment  $M1$  erster Ordnung dagegen wird nicht vom Investitionsgrad beeinflusst und zeigt einen deutlich größeren Wert in der Kursdatenbank\_2020. Die Kursdatenbank\_2020 führt folglich zu einem zu optimistischen Wert für  $M1$ . Den größten Gewinnfaktor unter allen Märkten weist der Markt FELynxMM\_1\_100 auf. Dies trifft auch für das Moment  $M1$  zu. Für die Investitionsgrade wurde festgestellt, dass sie der Marktgröße folgen, wenn man als Marktgröße die Anzahl der Aktien mit vorhandenen Kurswerten im Handelszeitraum zugrunde legt.

## A Notation

Aktienmarkt	(Markt) Auswahl von Aktien, die bis zu einem bestimmten Tag handelbar waren
FE	Filterergebnis
IB	Interactive Brokers
MM	Börsenprogramm Market Maker

### A.1 Tagesaktuelle Kursdaten

Bid	bid price
Ask	ask price
Last	last traded price
High	highest price
Low	lowest price
Volume	traded volume (Stückzahl in Tausend)
Zeitstempel	(time stamp) time of the last trade (in Unix time)

Davon abgeleitet:

Spread	$Spread = 100 \cdot (Ask - Bid) / Bid$
Umsatz	$Umsatz = Volume \cdot Low$

Kursabfrage eines Wertpapiers

abgeschlossen	Kursabfrage wird vom Server mit <i>errMsg()</i> oder mit <i>tickSnapShotEnd()</i> beantwortet
offen	Kursabfrage ist nicht abgeschlossen, d.h. für die Kursabfrage wird weder <i>errMsg()</i> noch <i>tickSnapShotEnd()</i> vom Server erhalten
vollständig	Alle sieben Kurstypen werden vom Server erhalten. Kurstypen können auch ungültig sein.
vollständig und gültig	Alle sieben Kurstypen werden vom Server erhalten. Alle sechs Kurswerte sind gültig ( $> 0$ ) und der Zeitstempel enthält eine gültige Angabe von Datum und Uhrzeit.

Mehrfach-Kursabfrage Abfrage der Kurstypen für die Wertpapiere eines Marktes oder aller bei IB verfügbaren Wertpapiere für Aktien. Es sind im Allgemeinen Kursabfragen für mehrere Wertpapiere gleichzeitig offen.

Symbol-Liste	Enthält für jedes Wertpapier die WKN (Wertpapierkennnummer) und Symbol (Kürzel) bei der Deutsche Börse AG, ISIN, ferner Symbol (IB-Symbol) und Börsenplatz bei IB
Unterbrechung	Die Mehrfach-Kursabfrage wird für eine eingestellte Zeit (Pause) unterbrochen und danach die Verbindung zum Server kurzzeitig unterbrochen
Mehrfach-Kurstab.	Eine Mehrfach-Kurstabelle enthält das Ergebnis einer Mehrfach-Kursabfrage in Form von IB-Symbol, WKN, ISIN, Status, Dauer, alle 7 Kurstypen und Abfrage-ID

## Bestandteile für Statuswerte bei Mehrfach-Kursabfragen

Full	Alle 7 Kurstypen erhalten
End	<i>tickSnapShotEnd()</i> erhalten
Reject	<i>errMsg()</i> erhalten
Short	1–6 Kurstypen erhalten (unvollständig)
Tout	Eingestellte Zeitspanne für Timeout ist abgelaufen

## Zähler für Statuswerte bei Mehrfach-Kursabfragen

Full gesamt	(vollständig) Alle 7 Kurstypen erhalten
FullEnd	Alle 7 Kurstypen, gefolgt von <i>tickSnapShotEnd()</i> (kein Timeout)
Reject	Kursabfrage wird vom Server mit <i>errMsg()</i> abgelehnt
Timeout	Server antwortet nicht bis zum Timeout
ShortTimeout	Server antwortet bis zum Timeout mit 1–6 Kurstypen
ShortEnd	1–6 Kurstypen erhalten, gefolgt von <i>tickSnapShotEnd()</i> (kein Timeout)

## Parameter für Mehrfach-Kursabfragen

Timer-Intervall	Zustandsänderungen werden nach Ablauf einer Zeitspanne überprüft, gegeben durch das Timer-Intervall
Grenze Timeout	Erreicht im Wartezustand die Anzahl abgelaufener Timer-Intervalle diesen Schwellwert, wird das Warten beendet
Pause/Intervalle	Legt eine Pause als Vielfaches des Timer-Intervalls fest
OpenMax	Schwellwert für die Anzahl offener Abfragen. Beim Erreichen dieses Schwellwertes wird eine Pause eingelegt und danach die Verbindung zum Server kurzzeitig unterbrochen

## Markdefinition

FEBRDUSA_1	Bisher verwendeter Markt, bestehend aus 867 Aktien
FELynxAlle	Alle bei IB bzw. Lynx handelbaren Wertpapiere für Aktien (Aktien und Aktienfonds) mit vollständigen und gültigen Kurstypen
FELynx_1_100	Wertpapiere von FELynxAlle mit einem Spread < 1%, Umsatz > 100T EUR, Last $\geq 1$ EUR
FELynx_FfM_2020	Alle bei Market Maker verfügbaren Aktien mit einem Kurswert $\geq 1$ EUR am 2.1.2020
FELynxMM	Durchschnitt von FELynxAlle und FEMM_FfM_2020
FELynxMM_1_100	Durchschnitt von FELynx_1_100 und FEMM_FfM_2020

## A.2 Historische Kursdaten

Kurswerte	Eröffnung, Kassa, Schluss, Hoch und Tief eines Handelstages
Datenbanken	
_2015	Die Kursdatenbank_2015 enthält die Kursdaten für alle Aktien des Marktes FEBRDUSA_1. Die Kursdatenbank wurde mit Hilfe der Kursdaten von MM in 2015 aufgebaut. Nach 2015 wurden die Kursdaten für die Aktien von FEBRDUSA_1 bis zum 21.7.2020 laufend ergänzt.
_2020	Die Kursdatenbank wurde mit Hilfe der Kursdaten von MM am 21.7.2020 neu aufgebaut. Sie enthält die Kursdaten für die Aktien der Lynx-Märkte und des Marktes FEBRDUSA_1, sofern diese Aktien am 21.7.2020 gehandelt werden konnten.
Datenqualität	
Auffüllungen:	
a) vertikal	Fehlende Kurswerte eines Handelstages werden mit Hilfe von Kurswerten des gleichen Tages berechnet
b) horizontal	Für einen Handelstag ohne Kurswerte werden letzte vorherige Kurswerte verwendet
Repeats	(Kurswiederholungen) Anzahl der Handelstage, an denen alle Kurswerte mit denen der Vortages identisch sind
Kursausreißer	Ein Kurswert $s(t)$ wird als Kursausreißer bezeichnet, wenn er über dem doppelten bzw. unter dem halben Medianwert $Median(t)$ liegt, gebildet aus den Kurswerten $s(t+i)$ , $ i  \leq 2$
Kurskorrekturen	Dienen der Entfernung von Kursausreißern. Der Kursausreißer wird durch den Medianwert $Median(t)$ ersetzt (s. Kursausreißer)
Sprunghöhe	Die Sprunghöhe ist das Maximum von $c$ und $1/c$ , wobei $c$ das Kursverhältnis für die Kurse an zwei aufeinanderfolgenden Handelstagen ist
Fehler	Es verbleiben trotz einer Kurskorrektur Sprunghöhen $\geq 5$
Tod	Kurse eines Wertpapiers enden mindestens 4 Handelstage vor dem Ende des Analysezeitraums
PennyTod	Es liegt ein Tod vor und der letzte Kurs liegt unterhalb 1 EUR
Max. Sprunghöhe	Die maximale Sprunghöhe ist die größte Sprunghöhe eines Wertpapiers im Analysezeitraum
Statistik	
Segment	Besteht aus $\tau_h$ aufeinanderfolgenden Handelstagen, wobei $\tau_h$ die Haltefrist ist
$q$	Segmentnummer
$Q$	Anzahl der Segmente im Handelszeitraum bis auf das letzte Segment
$g(q)$	Gewinnfaktor für das $q$ -te Segment,

$$g(q) = C_{tot}((q+1) \cdot \tau_h) / C_{tot}(q \cdot \tau_h)$$

$G_{cum}(q)$  Der kumulative Gewinnfaktor beschreibt die Entwicklung des Gesamtkapitals am Segmentende gemäß

$$C_{tot}((q+1) \cdot \tau_h) = G_{cum}(q) \cdot C_{tot}(1 \cdot \tau_h) .$$

Hierbei entspricht das Gesamtkapital am Ende des ersten Segments  $C_{tot}(1 \cdot \tau_h)$  dem Anfangs-Barkapital. Es gilt die Darstellung

$$G_{cum}(q) = \prod_{q'=1}^q g(q') , \quad G_{cum}(0) := 1 .$$

$c_1(t)$  Setzt den Eröffnungskurs eines Wertpapiers am Handelstag  $t$  mit dem gleitenden Durchschnitt der Eröffnungskurse der letzten 200 Handelstage ins Verhältnis

$s^*(t)$  Unbereinigter Eröffnungskurs am Handelstag  $t$

Kaufbedingung  $1.5 \leq c_1(t) < 1.625 , \quad s^*(t) \geq 2$

$G$  Gewinnfaktor für den gesamten Handelszeitraum,

$$G = G_{cum}(Q) = \prod_{q=1}^Q g(q)$$

$IG$  Mittlerer Investitionsgrad für den gesamten Handelszeitraum,

$$IG = \frac{1}{Q} \sum_{q=1}^Q IG(q \cdot \tau_h)$$

$M_1$  Moment erster Ordnung für den gesamten Handelszeitraum.  $M_1$  kann auf den mittleren Investitionsgrad  $IG$  und den leichter zugänglichen arithmetischen Mittelwert

$$AMG = \frac{1}{Q} \sum_{q=1}^Q g(q)$$

gemäß des Zusammenhangs

$$AMG = 1 + IG \cdot [M_1 - 1]$$

zurückgeführt werden. Insbesondere stimmen bei Vollinvestition ( $IG = 1$ )  $M_1$  und  $AMG$  überein.

## Literatur

- [1] Tai-Pan, Lenz+Partner (2021), <https://www.lp-software.de/>
- [2] Google Finance API and Its Alternatives (Updated for 2021), <https://rapidapi.com/blog/google-finance-api-alternatives/>
- [3] Alpha Vantage (2021), <https://rapidapi.com/alphavantage/api/alpha-vantage/pricing>
- [4] Interactive Brokers Trader Workstation API (2021), <https://interactivebrokers.github.io/tws-api/>
- [5] Interactive Brokers und Lynx (2021), <https://www.lynxbroker.de/lp/p/interactive-brokers/>
- [6] Deutsche Börse Group (2021), <https://www.xetra.com/xetra-de/instrumente/alle-handelbaren-instrumente/boersefrankfurt>
- [7] H. Balzert (2009): *Lehrbuch der Softwaretechnik — Basiskonzepte und Requirements Engineering*. Springer Verlag.
- [8] D. Pilone (2006): *UML 2.0*. O'Reilly Verlag.
- [9] J. A. Cerniglia, F. J. Fabozzi, P. N. Kolm (2016): Best Practices in Research for Quantitative Equity Strategies. *The Journal of Portfolio Management*, Vol. 42, No. 5, 135–143.
- [10] K. C. Rasekhschaffe, R. C. Jones (2019): Machine Learning for Stock Selection. *Financial Analysts Journal*, Vol. 75, No. 3, 70–88.
- [11] H.-H. Chen (2008): Stock selection using data envelopment analysis. *Industrial Management & Data Systems*, Vol. 108, No. 9, 1255–1268.
- [12] E. H. Sorensen, K. L. Miller, C. K. Ooi (2000): The Decision Tree Approach to Stock Selection. *The Journal of Portfolio Management*, Vol. 27, No. 1, 42–52.
- [13] C.-F. Huang (2012): A hybrid stock selection model using genetic algorithms and support vector regression. *Applied Soft Computing*, Vol. 12, No. 2, 807–818.
- [14] J. B. Guerard, H. Markowitz, G. Xu (2015): Earnings forecasting in a global stock selection model and efficient portfolio construction and management. *International Journal of Forecasting*, Vol. 31, No. 2, 550–560.
- [15] D. T. Larose, C. D. Larose (2014): *Discovering Knowledge in Data — an Introduction to Data Mining*. John Wiley & Sons Inc.
- [16] C. K. Enders (2020) *Applied Missing Data Analysis*. The Guilford Press.
- [17] J. L. Schafer, J. W. Graham (2002): Missing Data: Our View of the State of the Art. *Psychological Methods*, Vol. 7, No. 2, 147–177.
- [18] P. Treleaven, M. Galas, V. Lalchand (2013): Algorithmic trading review. *Communications of the ACM*, Vol. 56, No. 11, 76–85.
- [19] W. Stier (2001): *Methoden der Zeitreihenanalyse*. Springer Verlag.
- [20] P. Vogel (2019): Investing under constraints — part 3: Extension of the statistics. Düsseldorf: University of Applied Sciences, *Trading*, No. 3, 1–37.



# IMPRINT

**Hochschule Düsseldorf  
University of Applied Sciences**

**Fachbereich Medien  
Münsterstraße 156  
40476 Düsseldorf**

**Mehr Informationen zur Reihe „Trading“ unter:  
[https://opus4.kobv.de/opus4-hs-duesseldorf/solrsearch/index/  
search/searchtype/series/id/10](https://opus4.kobv.de/opus4-hs-duesseldorf/solrsearch/index/search/searchtype/series/id/10)  
editor-in-chief Prof. Dr. Peter Vogel  
B.Eng. Patrick Blättermann  
ISSN 2567-2347**

**Trading ; 5 : Investing  
URN urn:nbn:de:hbz:due62-opus-28848**

**Die Deutsche Nationalbibliothek listet diese Publikation in der  
Deutschen Nationalbibliografie auf. Detaillierte bibliografische Daten  
finden Sie im Internet unter <http://dnb.dnb.de>.**

P. Vogel, P. Blättermann (2021): Investieren unter Nebenbedingungen -Teil  
5: Auswahl von Aktien, Datenqualität und Statistik-Ergebnisse. Düsseldorf:  
Hochschule Düsseldorf, Series Trading, No. 5 (Investing), online unter:  
<https://nbn-resolving.org/urn:nbn:de:hbz:due62-opus-28848>



Der Forschungsbericht erscheint unter der  
Creative-Commons-Lizenz CC-BY-NC-SA 4.0.