

DIPLOMARBEIT AUS FINANZ- UND VERSICHERUNGSMATHEMATIK

# Über die Orderbuchmodellierung mit Markovschen Ketten in stetiger Zeit

Ausgeführt am Institut für  
Stochastik und Wirtschaftsmathematik  
der Technische Universität Wien

unter Anleitung von

**Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Friedrich Hubalek**

durch

Jakob Weissteiner B.Sc.

St.Georgen 46  
4710 Grieskirchen

Wien, November 2017

# Kurzfassung

Diese Diplomarbeit befasst sich mit der Modellierung des Limit Orderbuches durch einen zeitstetigen Markovprozess. Im ersten Kapitel erläutern wir zentrale Begriffe zu elektronischen Märkten und dem Limit Order Buch. Danach präsentieren wir eine allgemeine Modellbeschreibung des Orderbuchvolumens durch einen zeitstetigen  $2K$ -dimensionalen Markovprozess  $X$  unter Annahme eines konstanten Referenzpreises. Hierbei wird es uns gelingen die Existenz einer invarianten Verteilung von  $X$  unter gewissen Regularitätsvoraussetzungen zu beweisen. Des Weiteren werden grundlegende Eigenschaften wie Irreduzibilität, Explosivität und Rekurrenz des Markovprozesses  $X$  diskutiert. In Kapitel 3 präsentieren wir das erste Modell, welches das Orderbuch in Zeiten eines konstanten Referenzpreises beschreibt. Unter der Unabhängigkeitsannahme der einzelnen Komponenten, erhalten wir eine explizite Darstellung der invarianten Verteilung von  $X$  im Modell I. Hierbei diskutieren wir auch die Schätzung der Intensitäten des Markovprozesses und den Zusammenhang zu dessen assoziierten Sprungprozess. Eine einfache Schätzung des Referenzpreises aus gegebenen Daten wird in Kapitel 4 dargestellt. Kapitel 5 widmet sich der Beschreibung der LOBS-TER Datenbank, welche uns als Grundlage für die empirischen Studie dient. Dabei erklären wir auch die Methodik zur Schätzung der Intensitäten aus den zugrunde liegenden Daten. In Kapitel 6 führen wir eine empirische Studie zur Orderbuchmodellierung mittels Modell I für die beiden Aktien France Telekom und Microsoft Corporation durch. Wir präsentieren dazu die geschätzten Intensitäten und invarianten Verteilungen und vergleichen diese mit den empirisch beobachteten Werten aus den Daten. In Kapitel 7 erläutern wir einen Algorithmus zur Simulation der einzelnen Preislimits aus Modell I und besprechen als konkretes Anwendungsbeispiel die Berechnung der Wahrscheinlichkeit einer Exekution einer Limit Order. Im Kapitel 8 besprechen wir schlussendlich Modell III. Mittels diesem kann nun das Orderbuchvolumen für einen beliebigen Zeithorizont mit einem dynamischem Referenzpreis beschrieben werden. Somit stellt Modell III eine Erweiterung von Modell I dar. Zudem präsentieren wir einen Algorithmus zur Simulation des Orderbuchvolumens im Modell III und besprechen eine mögliche Variante zur Kalibrierung der Parameter. Im letzten Kapitel ist der verwendete Quellcode in **R** dargestellt.

# Abstract

This diploma thesis deals with the modeling of the Limit Order Book using a  $2K$ -dimensional markov process  $X$  in a continuous time setting. The first chapter serves as a general introduction to electronic markets and the limit order book. In chapter 2 we present a general framework of the model under the assumption of a fixed reference price. We will thereby prove, that under certain regularity conditions the markov process  $X$  admits an invariant distribution. Furthermore, properties such as irreducibility, explosivity and recurrence concerning the markov process  $X$  will be discussed. Chapter 3 presents the first specific model for the orderbook volume considering time periods with constant reference price under the additional assumption of componentwise independence. Within the framework of Model I it will be possible to obtain an explicit formula for the invariant distribution. Moreover we will discuss parameter estimation and the relation of  $X$  to the associated discrete time jump process  $Z$  in the setting of Model I. In chapter 4 a simple estimation for the reference price using empirical data is discussed. Chapter 5 is devoted to the presentation of the LOBSTER data base, that will be the foundation for all our empirical studies. As part of this chapter we will also explain the methodology of the parameter estimation from underlying data. In Chapter 6 we will conduct an empirical study for the two stocks France Telekom and Microsoft Corporation in the setting of Model I. Chapter 7 presents an algorithm for modelling an individual price limit within Model I. As an application we will discuss how to estimate the probability of an execution of a Limit Order. In Chapter 8 we finally present a model for the whole period of time that includes a dynamic reference price. In addition we present an algorithm for simulating the orderbook volume as specified in Model III and discuss how parameter calibration can be performed. In the last chapter we provide some source code in **R**.

# Inhaltsverzeichnis

<b>1</b>	<b>Elektronische Märkte und das Limit Order Buch</b>	<b>1</b>
1.1	Elektronische Märkte . . . . .	1
1.1.1	Arten von Finanzprodukten . . . . .	1
1.1.2	Teilnehmer an elektronischen Märkten . . . . .	2
1.2	Das Limit Order Buch . . . . .	2
1.2.1	Ordertypen . . . . .	3
1.2.2	Preis/Zeit Priorität . . . . .	4
1.2.3	Colocation (Serverhousing) . . . . .	5
1.2.4	Limit Orderbuch - Illustration . . . . .	6
<b>2</b>	<b>Allgemeine Modellbeschreibung des LOB</b>	<b>8</b>
2.1	LOB als zeitstetiger homogener Markovprozess . . . . .	9
2.2	Eigenschaften des Markovprozesses $X$ . . . . .	10
2.2.1	Irreduzibilität . . . . .	10
2.2.2	Explosivität . . . . .	12
2.2.3	Rekurrenz und Transienz . . . . .	14
2.3	Existenz einer invarianten Verteilung . . . . .	17
<b>3</b>	<b>Modell I - unabhängige Limits</b>	<b>19</b>
3.1	Modellbeschreibung - Modell I . . . . .	19
3.2	Schätzung der Intensitäten . . . . .	21
3.3	Konfidenzintervalle für die Intensitäten . . . . .	24
3.4	Berechnung der invarianten Verteilung . . . . .	26
3.4.1	Invariante Verteilung des Sprungprozesses . . . . .	28
<b>4</b>	<b>Schätzung des Referenzpreises</b>	<b>32</b>
<b>5</b>	<b>LOBSTER Datenbank</b>	<b>34</b>
5.1	Output Struktur . . . . .	34
5.1.1	Begriffsunterscheidung Trade - Exekution im Kontext von LOBSTER . . . . .	36
5.2	Methodik der Schätzung im Modell I . . . . .	37
<b>6</b>	<b>Empirische Studie Modell I - LOBSTER Daten</b>	<b>40</b>
6.1	France Telekom - FTE . . . . .	40
6.2	Microsoft Corporation - MSFT . . . . .	47
<b>7</b>	<b>Simulation des Orderbuches - Modell I</b>	<b>55</b>
7.1	Simulation eines einzelnen Preislimits . . . . .	55
7.1.1	Simulation Markovprozess Methode I . . . . .	56
7.2	Algorithmus zur Simulation eines Preislimits . . . . .	56
7.3	Empirische Studie zur Orderbuchmodellierung . . . . .	57
7.4	Anwendungsbeispiel - Wahrscheinlichkeit der Exekution einer Limit Order . . . . .	62
<b>8</b>	<b>Modell III - dynamischer Referenzpreis</b>	<b>65</b>
8.1	Modellbeschreibung - Modell III . . . . .	65
8.2	Simulation des Orderbuches - Modell III . . . . .	69

8.2.1	Simulation Markovprozess Methode II . . . . .	69
8.2.2	Kalibrierung der Parameter . . . . .	71
<b>9</b>	<b>Quellcode - R</b>	<b>74</b>
	<b>Anhang</b>	<b>88</b>
<b>A</b>	<b>Markovprozesse mit abzählbarem Zustandsraum</b>	<b>88</b>
A.1	Konvergenz gegen ein Äquilibrium . . . . .	89
A.2	Explosivitätskriterium . . . . .	89
A.3	Rekurrenz und Transienz . . . . .	89
A.4	Ergodensätze . . . . .	90
<b>B</b>	<b>Parameter Punktschätzung und Asymptotisches Verhalten</b>	<b>93</b>
<b>C</b>	<b>Wahrscheinlichkeitstheorie</b>	<b>94</b>
	<b>Literaturverzeichnis</b>	<b>95</b>
	<b>Abbildungsverzeichnis</b>	<b>96</b>
	<b>Tabellenverzeichnis</b>	<b>98</b>

# Kapitel 1

## Elektronische Märkte und das Limit Order Buch

Dieses Kapitel gibt einen kurzen Überblick über die wichtigsten Charakteristika von elektronischen Märkten und dem Limit Order Buch. Es folgt im Wesentlichen Ausschnitten von [2, Kapitel 1], [5, Abschnitt 2.1.1 und 2.1.2] und [9, Abschnitt 7.1, 7.2 und 7.3].

### 1.1 Elektronische Märkte

Elektronische Märkte sind Märkte die durch moderne computerbasierte Kommunikationsnetzwerke miteinander verbunden sind. Teilnehmer an elektronischen Märkten müssen demnach nicht an demselben physischen Ort sein, um ein Geschäft abzuschließen. Ein prominentes Beispiel für einen elektronischen Markt ist die NASDAQ (National Association of Securities Dealers Automated Quotations). Sie stellt einen Spezialfall eines elektronischen Marktes, eine *elektronische Börse*, dar. Als solche ist die NASDAQ, gemessen an der Zahl der gelisteten Unternehmen, die größte ihrer Art. Eine weitere Art von elektronischen Märkten sind *Multilateral Trading Facilities* (MTF). Diese börsenähnlichen Handelssysteme werden vor allem von Finanzdienstleistern und Wertpapierfirmen betrieben und weisen weniger strenge Regeln als die elektronischen Börsen auf. Als letztes Beispiel elektronischer Märkte seien noch die sogenannten *Dark Pools* erwähnt. Als *Dark Pool* wird eine banken- oder börseninterne anonyme Handelsplattform bezeichnet. Das entscheidende Merkmal von *Dark Pools* ist die fehlende Transparenz beziehungsweise die Geheimhaltung der verfügbaren Liquidität. Es gibt eine Vielzahl von anderen elektronischen Märkten, wobei wir uns im Folgenden auf elektronische Börsen beschränken und den Begriff elektronischen Markt als Synonym dafür verwenden.

#### 1.1.1 Arten von Finanzprodukten

Heutzutage wird eine Vielzahl von verschiedenen Finanzprodukten an elektronischen Märkten gehandelt. In diesem Abschnitt werden wir in aller Kürze die wichtigsten davon auflisten.

Das wohl populärste Finanzprodukt sind *Aktien* (engl. share, stock). Eine Aktie ist ein Wertpapier, welches einen Anteil einer Aktiengesellschaft (AG) verbrieft. Aktien werden von der ausgebenden Gesellschaft genutzt, um Kapital zu generieren. In der einfachsten Form erlangt man durch einen Aktienkauf sowohl das Recht zu gleichen Teilen am Unternehmenserfolg zu profitieren (Dividenden), als auch das Recht auf Mitsprache im Unternehmen (Stimmrecht bei Hauptversammlung).

Es sei hier erwähnt, dass mit dem Erwerb einer Aktie neben diesen Rechten auch verschiedene Pflichten einhergehen.

Eine weitere Form der Unternehmensfinanzierung sind *Anleihen* (engl. bonds). Anleihen sind Finanzprodukte, bei denen sich das emittierende Unternehmen zu regelmäßigen Zinszahlungen (engl. interest) an zuvor vereinbarten Terminen (engl. coupon dates) und einer Rückzahlung am Ende der Laufzeit verpflichtet. Der entscheidende Unterschied im Vergleich zu Aktieninhabern ist: Aktieninhaber haben keine Garantie bezüglich der Höhe und Frequenz der Kompensationszahlungen (Dividenden) jedoch Stimmrechte, während Inhaber von Anleihen garantierte regelmäßige Kompensationszahlungen (Zinszahlungen) erhalten und kein Stimmrecht haben.

*Derivate* (engl. derivatives) sind Termingeschäfte (Vertragsabschluss und Vertragserfüllung liegen zeitlich getrennt), deren Wert sich aus dem Wert einer Referenzgröße ableitet. Die Referenzgröße wird Basiswert (engl. underlying) genannt. Basiswerte können beispielsweise sein: Aktien, Anleihen, Rohstoffe, Zinssätze, Indices. Beispiele von derivativen Finanzinstrumente sind: Futures und Forwards, Optionen, Swaps. Diese können beispielsweise zu Absicherungszwecken (engl. hedging) oder Spekulationsgeschäften eingesetzt werden.

Je nach Charakteristik des zugrunde liegenden Vermögenswertes (engl. underlying asset) werden Finanzprodukte in verschiedene Klassen eingeteilt. Hierbei seien einige weitere zu den oben genannten erwähnt: Fonds (engl. funds), Bargeld (engl. cash), Devisenmarkt (engl. foreign exchange FX), Rohstoffe (engl. commodities).

### 1.1.2 Teilnehmer an elektronischen Märkten

Teilnehmer an elektronischen Märkten (engl. traders) handeln entweder im eigenen Namen beziehungsweise auf eigene Rechnung: *Proprietary Traders* oder arrangieren Geschäfte für ihre Klienten: *Broker*, *Dealers* oder *Market Maker* sind einzelne Personen oder Firmen, die sich verpflichten verbindliche An- und Verkaufskurse für gewisse Finanzprodukte zu stellen und daher immerzu bereit sind mit anderen Teilnehmern zu handeln, beziehungsweise die jeweilige Gegenposition einzugehen. Börsen dienen als Plattform für den Austausch von Finanzprodukten. Man unterscheidet im groben zwischen *floor trading*, hierbei findet der Handel am Parkett der Börse statt, und *electronic trading*, wo Teilnehmer an elektronischen Märkten mittels elektronischen Systemen handeln.

## 1.2 Das Limit Order Buch

Es gibt verschiedene Arten, um einen elektronischen Markt zu realisieren. Der Grundgedanke, es Marktteilnehmern zu ermöglichen ihre Bereitschaft für eine Handelsgeschäft zu signalisieren und ein Matching-System bereitzustellen, bei dem Kauf- und Verkaufsangebote zusammengeführt werden, ist jedoch bei allen derselbe.

### 1.2.1 Ordertypen

Eine Order (Auftrag) repräsentiert eine Anweisung eines Marktteilnehmers. Dieser bestimmt im Zuge einer Order im Wesentlichen was, wieviel und wann etwas gehandelt werden soll. Mit einer Kauf Order (engl. bid or buy order) signalisiert ein Marktteilnehmer die Absicht ein bestimmtes Finanzgut zu erwerben. Eine Verkauf Order (engl. ask or sell order) widerspiegelt die Verkaufsabsicht eines bestimmten Finanzproduktes. Entsprechend sind der *Bid-* und der *Ask Preis*, zu dem ein Trader bereit ist zu handeln, zu verstehen. Der Preis, zu dem eine Order ausgeführt wird, wird *Transaktions- oder Exekutionspreis* genannt (engl. transaction or execution price). Der höchste (niedrigste) verfügbare Bid (Ask) Preis wird *best Bid Price (best Ask Price)* genannt. Die Differenz zwischen dem *best Bid* und dem *best Ask Price* wird *Bid Ask Spread* genannt. Im einfachsten Fall gibt es an elektronischen Märkten zwei Arten von Orders: *Market Orders* (MOs) und *Limit Orders* (LOs). MOs werden üblicherweise als aggressive Orders, die eine sofortige Ausübung (engl. execution) zum Ziel haben, beschrieben. Ein Marktteilnehmer, der eine MO abgibt, will unmittelbar eine bestimmte Menge eines Finanzgutes zum bestmöglichen Preis kaufen oder verkaufen. Hingegen werden LOs als passive Orders angesehen. Ein Marktteilnehmer, der eine LO abgibt, hat zum Ziel eine bestimmte Menge eines Finanzgutes zu einem fixen Preis zu kaufen oder zu verkaufen. Üblicherweise wird der festgelegte Preis einer Kauf LO niedriger sein als der beste Verkaufspreis von LOs (*best Ask Price*). Umgekehrt wird der Preis einer Verkauf LO höher sein als der beste Kaufpreis (*best Bid Price*) von LOs. Dies führt dazu, dass es zu keiner sofortigen Ausübung einer Limit Order kommen kann und diese solange verweilen muss, bis sie entweder gelöscht (engl. cancelled) oder gegen eine neue Order ausgeübt wird. Im seltenen Fall, dass eine Verkauf (Kauf) Limit Order zu einem niedrigeren (höheren) oder gleichen Preis als der *best Bid Price (best Ask Price)* abgegeben wird, wird diese wie eine Market Order sofort ausgeführt und befüllt, solange das vorgegebene Preislimit nicht leer ist. Dieser Spezialfall wird *marketable Order* genannt. Falls eine LO nicht sofort ausgeführt wird, kommt sie ins Limit Order Buch: LOB (engl. Limit Order Book). Dort werden alle nicht ausgeführten LOs nach einer Preis/Zeitpriorität angeführt. Je größer der Abstand des Preislimits einer Verkauf (Kauf) LO im LOB zum *best Ask Price (best Bid Price)* ist, desto unwahrscheinlicher ist eine zeitnahe Ausübung derer.

Es gibt eine Vielzahl an weiteren Ordertypen, die je nach Börse variieren können. Im Folgenden werden wir die Wichtigsten in aller Kürze auflisten:

- *Market to limit Order*: ist eine MO die lediglich zum *best Bid* oder *best Ask Price* ausgeführt wird. Das restliche Volumen wird in das LOB eingetragen.
- *Stop Order*: Mit dieser Order wird der Kauf/Verkauf des Wertpapiers ausgelöst falls ein bestimmtes Stop-Limit über- beziehungsweise unterschritten wird.
- *Day Order*: Gültigkeit nur für den aktuellen Handelstag.
- *Non Routable Orders*: Orders, die nur für die aktuelle Börse gelten und nicht automatisch an andere Börsen weitergeleitet werden.
- *Pegged Order*: dynamische Orders, deren Angebotspreis sich automatisch an gewissen Kennzahlen anpasst (z.B. bester Marktpreis).
- *Hidden Order*: Order, bei dem das Ordervolumen nicht veröffentlicht wird.
- *Iceberg Order*: Order, die nur teilweise ihr Ordervolumen veröffentlicht.

- *Immediate Or Cancel*: Order, die sofort und soweit wie möglich zum besten Preis ausgeführt und danach wieder gelöscht werden.
- *Fill or Kill Order*: Order, die entweder vollständig oder gar nicht ausgeführt wird.
- *Good-Till-Time*: Order, deren Gültigkeit bis zu einem bestimmten Zeitpunkt besteht. Diese wird danach gelöscht.
- *Discretionary*: Order, welche nur einen Preis veröffentlicht, jedoch möglicherweise zu anderen aggressiveren Preisen auch ausgeführt wird.

### 1.2.2 Preis/Zeit Priorität

Die Dynamik der Orders wird durch ein Matching-System und das Limit Order Buch geregelt und dokumentiert. Das LOB führt Buch über ankommende und abgehende Orders. Das Matching-System verwendet einen bestimmten Algorithmus, welcher je nach Börse beziehungsweise elektronischem Markt variieren kann, um zu bestimmen, ob eine Order tatsächlich ausgeführt wird beziehungsweise zu welchem Preis.

Die meisten elektronischen Märkte weisen MOs eine höhere Priorität zu als LOs, wobei deren Ausführung nach der sogenannten *Preis/Zeit Priorität* erfolgt:

Im Falle einer neuen Kauf Market Order wird folgendermaßen vorgegangen: die MO wird gegen diejenigen Verkaufs LOs ausgeführt, welche den besten Preis (*best Ask Price*) anbieten. Ist das Volumen der MO kleiner als jenes der besten LOs werden nur die ältesten LOs gegen die MO ausgeführt. (engl. first in first out rule). Ist das Volumen der MO größer als jenes der besten LOs, wird, nachdem alle LOs am besten Preis gegen die MO ausgeführt worden sind, mit den Verkaufs LOs, welche den zweitbesten Preis anbieten, fortgefahren. Dieses Prozedere wird solange fortgesetzt, bis das Volumen der MO gänzlich ausgeführt wurde.

Im Falle einer neuen Verkauf Market Order wird analog vorgegangen. Man muss lediglich anstatt Verkauf LOs, Kauf LOs und anstatt dem *best Ask Price* den *best Bid Price* heranziehen.

Der Vorgang, bei dem eine MO gegen LOs, die tiefer (engl. deeper) im LOB stehen ausgeführt wird, wird *walking the book* genannt. Hierbei müssen die Auftraggeber der MO für Teile ihrer Order, da sie nicht vollständig gegen die LOs am besten Preis ausgeführt werden, einen Aufpreis zahlen (engl. *price concession*). Diese resultierenden Preisveränderungen einer Market Order werden *market impact* genannt. Diese steigen meist mit dem abgegeben Ordervolumen und bilden neben dem *Bid Ask Spread* einen wichtigen Teil der Transaktionskosten einer Order.

Der oben besprochene Preis-/Zeitprioritätsansatz ist nicht der einzige Weg, um einen elektronischen Markt zu organisieren. Als weiteren Ansatz sei hier die *Prorata* Regel genannt. Bei dieser werden MOs weiterhin gegen die verfügbaren LOs am besten Preis ausgeführt. Jedoch werden die LOs nicht anhand der Dauer, die sie im LOB stehen, bevorzugt, sondern diese werden proportional zu den abgegebenen LOs am selben Preislevel gegen die MO ausgeführt. Viele Börsen nutzen zusätzlichen zum Intraday-Handel Auktionen zu bestimmten Zeitpunkten z.B. Eröffnungsauktion, Schlussauktion. Bei einer Auktion werden von potenziellen Teilnehmern Kauf- und Verkaufsangebote abgegeben. Danach wird nach einem bestimmten Prozedere, welches von Börse zu Börse variieren kann, entschieden, welche Angebote einen Zuschlag erhalten haben. Unmittelbar nach Ende der Auktion werden die Teilnehmer über die abgewickelten Abschlüsse informiert. Während

einer Auktion ist das Orderbuch geschlossen.

Um sich gegen etwaige Kursmanipulation zu schützen, haben Börsen verschiedene Schutzmechanismen eingerichtet. Dazu sei exemplarisch die Volatilitätsunterbrechung erwähnt. Hierbei wird für jedes Wertpapier ein fixer Preiskorridor festgelegt. Eine Volatilitätsunterbrechung wird bei Verlassen dieses Korridors ausgelöst. Dadurch wird den Marktteilnehmer ermöglicht ihre Order zu modifizieren und es kommt zu einer Auktion.

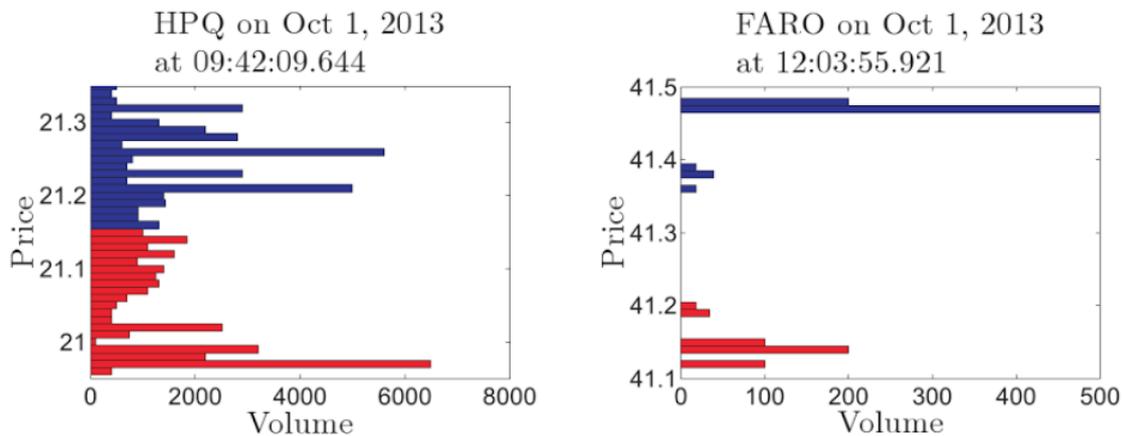


Abbildung 1.1: Schnappschuss des LOB an der NASDAQ nach dem 10000 Event des Tages.  
Quelle: [2]

Abbildung 1.1 zeigt eine Momentaufnahme der LOBs an der NASDAQ für die beiden Aktien HPQ (Hewlett Packard Inc.) und FARO (FARO Technologies Inc.). Blaue Balken stellen die Verkauf LOs zum jeweiligen Limit dar. Rote Balken symbolisieren die Kauf LOs. Anhand der Grafik lässt sich ein deutlicher Unterschied zwischen den beiden Aktien erkennen. Das LOB von HPQ weist an jedem dargestellten Limit LOs auf, während die LOs von FARO viele Lücken aufweisen. HPQ ist im Gegensatz zu FARO, welches ein selten gehandeltes und illiquides Asset ist, ein häufig gehandeltes und liquides Wertpapier. Dies erklärt den Unterschied in der Struktur der beiden LOBs.

### 1.2.3 Colocation (Serverhousing)

Ein weiterer wichtiger Punkt ist der Grad an Informationen und die Datenqualität, welche von den Börsen zur Verfügung gestellt wird. Der öffentliche verdichtete Datenfeed ist vergleichsweise günstig und enthält nur wenig Information. Will man hingegen einen speziellen (originalen) Datenfeed (engl. proprietary data feed), muss man dafür mehr bezahlen. In Zeiten, wo die durchschnittliche Halteperiode von Finanzprodukten immer kürzer wird, spielt nicht zuletzt die physische Nähe zur Börse beziehungsweise zum Matching-System eine wichtige Rolle. Börsen vermieten daher Computer- und Serverstellen, die in direkter Nähe zu ihrer Matching Maschine positioniert sind. Diesen Vorgang nennt man *Colocation*. Die Vereinheitlichung der Hardwareverbindungen innerhalb eines Colocationstandortes (z.B. Netzkabellänge) stellt einen gleich schnellen Zugang und somit keine Geschwindigkeitsnachteile für die Mieter solcher Serverstellen dar. Dieser Service seitens der Börse ist jedoch kostenpflichtig und es mehrt sich Kritik daran, da es nicht allen Marktteilnehmern möglich ist diesen Service zu nutzen. Trader, die keinen solchen direkten

Serverplatz besitzen, haben demnach einen Geschwindigkeitsnachteil. Hier ist es nun Aufgabe der Aufsichtsbehörden diesen Zugang zu überwachen und hinreichend wettbewerbsfähig zu gestalten.

### 1.2.4 Limit Orderbuch - Illustration

Wie bereits in den vorigen Abschnitten erwähnt sind die wesentlichen Bestandteile eines elektronischen Marktes das LOB und ein Matching-System/Algorithmus. Ein Limit Order Buch ist anhand eines fixierten diskreten Gitter von Preisen definiert (Preislimits oder Limits). Der Abstand zweier Gitterpunkte wird *ticksiz*e genannt. Diese hängt vom zugrunde liegenden Wertpapier, dessen Wert und der Börse, an dem es gehandelt wird, ab, z.B. in den Vereinigten Staaten von Amerika ist die minimale *ticksiz*e 1 Cent für alle Aktien, deren Preis über einem US Dollar liegt. Neben dem bereits erläuterten *Bid Ask Spread* (manchmal auch *quoted spread*) ist der *Mid Price* eine weitere wichtige Kennzahl. Dieser ist das arithmetische Mittel des *best Bid-* und *best Ask Price* und liefert eine einfache Approximation für den „wahren“ Preis des Wertpapiers. Im Folgenden werden wir die Funktionsweise des LOB anhand zweier konstruierter Beispiele erläutern.

#### 1. Neu hinzukommende Kauf Limit Order:

In diesem Kontext haben wir bereits die Funktionsweise der Preis/Zeitprioritätsregel ausführlich diskutiert. Diese wird nun anhand von Abbildung 1.2 illustriert. In Abbildung 1.2 werden LOs anhand von Blöcken, die ihrem Volumen entsprechen, dargestellt. Die neue Kauf LO (lila) hat ein Preislimit von \$23.09. Diese wird nun zu den bereits bestehenden Blöcken am unteren Ende hinzugefügt. LOs, die in Abbildung 1.2 den kleinsten Abstand zur y-Achse haben, sind als letztes zu diesem Limit hinzugefügt worden und werden daher entsprechend der Zeitprioritätsregel (engl. first in first out) als letztes ausgeführt.

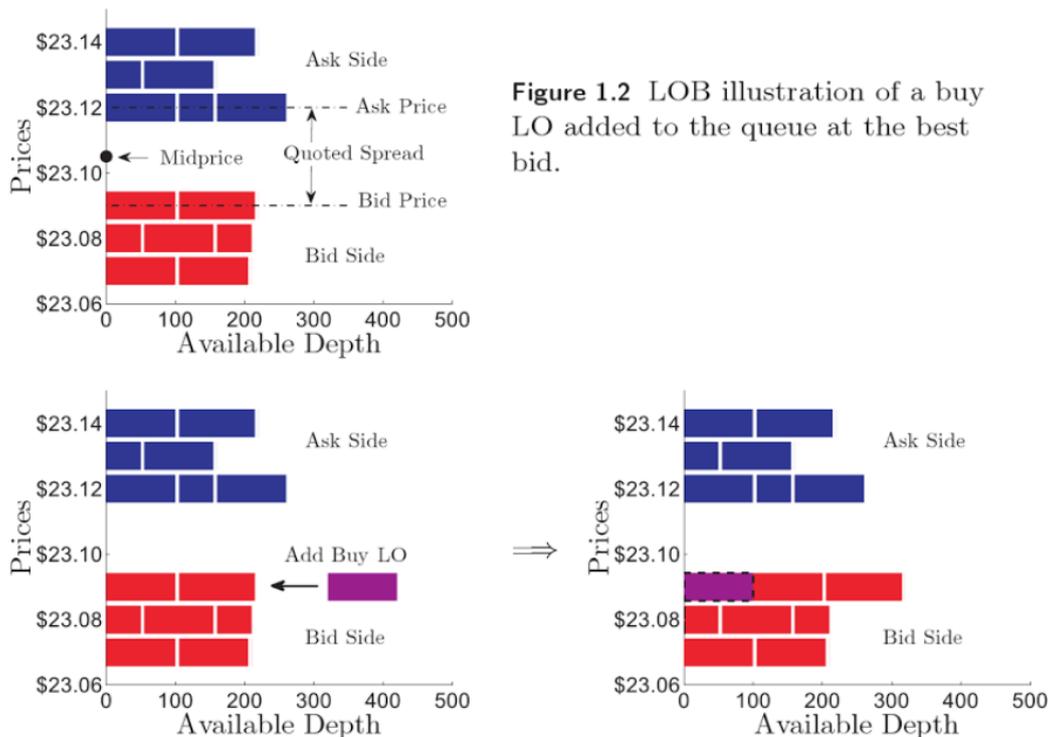


Abbildung 1.2: Quelle [2]

## 2. Market Order walking the Book:

Eine Verkauf Market Order mit Gesamtvolumen von 250 Aktien trifft im LOB ein (grün). Diese wird nun wie bereits besprochen gegen LOs beginnend mit den LOs am *best Bid Price* ausgeführt, solange deren Volumen erschöpft ist. Im LOB sind zwei LOs mit einem Volumen von 100 pro Order am *best Bid Price* von \$23.09 situiert. Daher werden die ersten 200 Aktien der MO zum *best Bid Price* ausgeführt. Die restlichen 50 Aktien werden zu einem Preis von \$23.08 gegen die älteste LO an diesem Preislimit ausgeführt. Die rechte Grafik zeigt schlussendlich das aktualisierte LOB nach Abschluss des Handels.

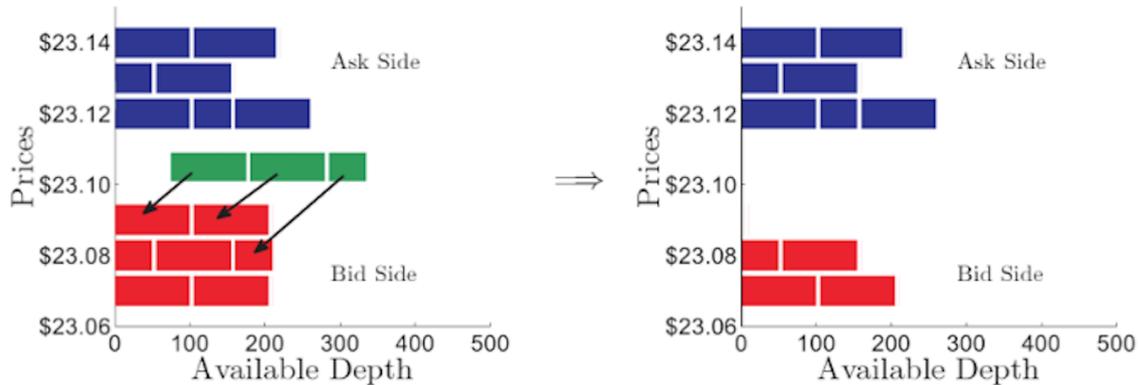


Abbildung 1.3: Illustration einer neu hinzukommenden Market Order, Quelle [2]

## Kapitel 2

# Allgemeine Modellbeschreibung des LOB

In diesem Kapitel widmen wir uns der Dynamik des Limit Order Book (LOB) in einer Periode eines konstanten Referenzpreises. Dieses Kapitel folgt im Wesentlichen dem Abschnitt 2.1 von [6]. Der aktuelle Zustand des LOB wird hierbei als  $2K$ -dimensionaler Vektor angesehen, wobei  $K$  die Anzahl der zur Verfügung stehenden Preislimits auf der *Bid*- beziehungsweise *Ask-Side* beschreibt. Hierbei wird nicht ausgeschlossen, dass das Volumen eines Limits leer ist. Wir betrachten für dieses Modell die folgenden drei Typen von Orders:

1. **Limit Orders:** Abgabe einer neuen *Sell*- oder *Buy Limit Order* an eines der zur Verfügung stehenden Limits auf der *Ask*- beziehungsweise *Bid-Side*. Eine *Sell Limit Order* (*Buy Limit Order*) wird in unserem Kontext ausschließlich zu einem höheren (geringeren) Preis als dem *best Bid Price* (*best Ask Price*) abgegeben.
2. **Cancellation Orders:** Stornierung einer bereits existenten Limit Order.
3. **Market Orders:** Abgabe einer unmittelbar auszuführenden *Buy*- oder *Sell Order* zum bestmöglichen Preis.

Als *ticksize*  $\Delta_{tick}$  (Notierungssprung) bezeichnet man die kleinstmögliche Preisänderung eines Wertpapierkurses. Diese kann je nach Art des Wertpapiers und der jeweiligen Börse, an der es gehandelt wird, variieren.

Im Folgenden sei nun  $P_i \in \Delta_{tick} \cdot \mathbb{N}$ ,  $i \in \{-K, \dots, -1, 1, \dots, K\}$  das  $i$ -te Preislimit. Die Position der Preislimits  $P_i$  wird durch einen konstanten Referenzpreis  $p_{ref}$  bestimmt. Dieser definiert das Zentrum des  $2K$ -dimensionalen Vektors  $(P_{-K}, \dots, P_{-1}, P_1, \dots, P_K)^T$ . Folgendermaßen beschreibt  $P_i$  für  $i > 0$  jenes Preislimit, das einen Abstand von  $(i - 0.5) \cdot \Delta_{tick}$  Ticks zu  $p_{ref}$  nach rechts aufweist und für  $i < 0$  beschreibt  $P_i$  das Limit, welches den Abstand von  $-(i + 0.5)$  zu  $p_{ref}$  nach links aufweist. Speziell gilt daher  $\frac{P_{-1} + P_1}{2} = p_{ref}$ . Der Referenzpreis teilt unser Orderbuch in zwei Bereiche: Die *Bid-Side* :  $(P_{-K}, P_{-K+1}, \dots, P_{-1})$  und die *Ask-Side* :  $(P_1, P_2, \dots, P_K)$ .

Die Anzahl der Orders an einem Limit  $P_i$  wir mit  $x_i \in \mathbb{N}_0$  bezeichnet.<sup>1</sup>

---

<sup>1</sup> $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$

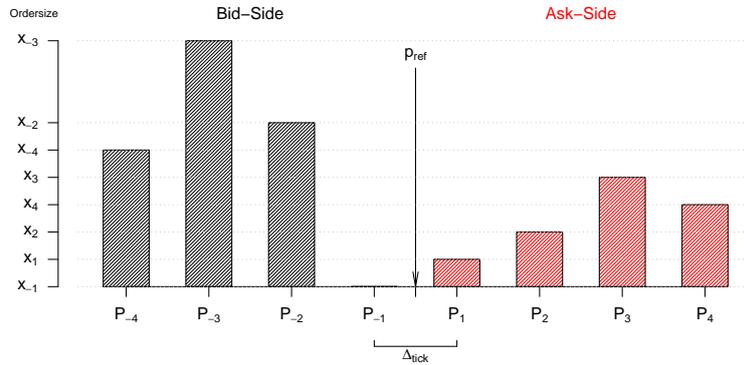


Abbildung 2.1: Limit Order Book: schematische Darstellung

## 2.1 LOB als zeitstetiger homogener Markovprozess

In diesem Abschnitt sei  $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  der zugrunde liegende filtrierte Wahrscheinlichkeitsraum. Das Ordervolumen des Limit Order Books (LOB) wird durch den stochastischen Prozess  $X$  gegeben durch:

$$X : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{N}_0^{2K}$$

$$(\omega, t) \mapsto X_t(\omega) = (X_t^{-K}(\omega), \dots, X_t^{-1}(\omega), X_t^1(\omega), \dots, X_t^K(\omega))$$

beschrieben.

Als Filtrierung wird die natürliche Filtration des Prozesses  $X$  gewählt d.h.  $\mathcal{F}_t = \sigma(X_s, s \leq t)$ . Im Folgenden sei  $(X_t)_{t \geq 0}$  ein zeitstetiger homogener rechtsstetiger Markovprozess mit Sprunghöhe  $\pm 1$ , abzählbarem Zustandsraum  $\mathbb{N}_0^{2K}$ , Anfangsverteilung  $\lambda$  und Intensitätsmatrix  $Q = (Q_{x,y})_{(x,y) \in \mathbb{N}_0^{2K} \times \mathbb{N}_0^{2K}}$ :  $X \sim \text{Markov}(\lambda, Q)$ .

Da der Zustandsraum des Prozesses diskret ist, können wir diesen in natürlicher Weise mit der Potenzmenge ausstatten und erhalten den messbaren Raum:  $(\mathbb{N}_0^{2K}, \mathcal{P}(\mathbb{N}_0^{2K}))$ , welchen wir in Zukunft immer als Bildbereich des Prozesses betrachten werden.

**Annahme 1** Die Anzahl der Orders in Aktien ist in unserem Modell für  $t \geq 0$  gegeben als  $A : \Omega \times \mathbb{R}_+ \rightarrow \mathbb{N}_0^{2K}$ , wobei  $A_t^i := AES_i \cdot X_t^i$ . Ist das Volumen<sup>2</sup>  $X_t^i$  zum Zeitpunkt  $t \geq 0$  am Preislimit  $P_i$  gleich  $x_i \in \mathbb{N}_0$  so entspricht das einem Volumen in Aktien von  $a_i$  mit

$$a_i = AES_i \cdot x_i \in AES_i \cdot \mathbb{N}_0$$

wobei  $AES_i$  die durchschnittliche Ordergröße für am Preislimit  $P_i$  abgegebene Order, bei einem der oben genannten Typen von Ordnern beschreibt (engl. Average Event Size). Das bedeutet, es wird eine konstante, vom jeweiligen Limit abhängige Ordergröße angenommen. Ein Sprung des Markovprozesses um  $\pm 1$  am Limit  $i$  bedeutet eine Veränderung des Orderbuchvolumens von  $\pm 1 \cdot AES_i$  in Aktien. Wir werden uns im weiteren Verlauf der Arbeit meist auf die Diskussion des

<sup>2</sup>Im Folgenden werden wir die Begriffe *Volumen* und *Anzahl der Orders* an einem Preislimit  $P_i$  als Synonym gebrauchen.

Prozesses  $X$  beschränken.

**Definition 2.1 (Intensitätsmatrix)** Sei  $x, y \in \mathbb{N}_0^{2K}$  und <sup>3</sup>  $e_i := (\delta_{i,-K}, \dots, \delta_{i,-1}, \delta_{i,1}, \dots, \delta_{i,K})$ , dann ist die Intensitätsmatrix  $Q$  des Markovprozesses  $X$  folgendermaßen definiert:

$$Q_{x, x+e_i} = f_i(x) \quad (2.1)$$

$$Q_{x, x-e_i} = g_i(x) \quad (2.2)$$

$$Q_{x,x} = - \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} < \infty \quad (2.3)$$

$$Q_{x,y} = 0 \quad , y \notin \{x, x + e_i, x - e_i\} \quad (2.4)$$

wobei  $f_i, g_i : \mathbb{N}_0^{2K} \rightarrow (0, \infty), \forall i \in \{-K, \dots, -1, 1, \dots, K\}$ .

**Bemerkung 2.2** Die Bedingung (2.3) in der obigen Definition besagt, dass die Zeilensumme der unendlich dimensionalen Matrix  $Q$  gleich null ist. Markovprozesse mit dieser Eigenschaft bezeichnet man als konservativ.

Ferner sei  $\mathcal{M}$  der Raum der beschränkten und messbaren Funktionen  $V : \mathbb{N}_0^{2K} \rightarrow \mathbb{R}$ . Die Intensitätsmatrix kann als Abbildung  $Q : \mathcal{M} \rightarrow \mathcal{M}$  aufgefasst werden. Sie operiert auf Funktionen aus  $\mathcal{M}$  in folgender Weise:

$$V \in \mathcal{M} : (QV)(\cdot) := \sum_{x \in \mathbb{N}_0^{2K}} Q_{\cdot, x} V(x)$$

## 2.2 Eigenschaften des Markovprozesses $X$

### 2.2.1 Irreduzibilität

Bezeichne  $X$  weiterhin den Markovprozess mit der Intensitätsmatrix  $Q$  aus Definition 2.1.

Im weiteren Verlauf sei  $(P(t))_{t \geq 0}$  die Übergangsmatrix des zeitstetigen homogenen Markovprozesses  $X$ . Diese ist die (komponentenweise) minimale nicht negative Lösung folgendes unendlich dimensional Systems von Differentialgleichungen (Rückwärtsgleichung) :

$$P'(t) = QP(t), \quad P(0) = I \quad (2.5)$$

und bildet eine Halbgruppe (*Chapman-Kolmogorow-Gleichung*):

$$P(s+t) = P(t)P(s) \quad \forall s, t \geq 0 \quad (2.6)$$

(Beweis: Siehe [8, Theorem 2.8.3])

Für  $x, y \in \mathbb{N}_0^{2K}$  sei somit  $P_{x,y}(t) := \mathbb{P}[X_t = x | X_0 = y]$ .

---

<sup>3</sup>Hier bezeichnet  $\delta_{i,j}$  das Kronecker Delta und ist definiert als  $\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

Unter obigen Voraussetzungen können wir zeigen, dass der Markovprozess  $X$  irreduzibel ist. Davor wiederholen wir noch einige Begriffe.

Im Folgenden sei  $x, y \in \mathbb{N}_0^{2K}$ . Wir sagen der Zustand  $x$  führt zum Zustand  $y$  und schreiben  $x \rightarrow y$  falls Folgendes gilt:

$$\mathbb{P}[X_t = y \text{ für ein } t > 0 \mid X_0 = x] > 0.$$

Wir sagen die Zustände  $x$  und  $y$  *kommunizieren* und schreiben  $x \longleftrightarrow y$ , falls  $x \rightarrow y$  und  $y \rightarrow x$ .

**Theorem 2.3** (siehe [8], Theorem 3.2.1) *Sei  $x \neq y \in \mathbb{N}_0^{2K}$ . Dann sind folgende Aussagen äquivalent*

1.  $x \rightarrow y$
2.  $x \rightarrow y$  für den assoziierten Sprungprozess (siehe 2.8)
3.  $Q_{q_1, q_2} \cdot \dots \cdot Q_{q_{n-1}, q_n} > 0$  für  $q_i \in \mathbb{N}_0^{2K}$  und  $q_1 = x$ ,  $q_n = y$
4.  $P_{x,y}(t) > 0$  für alle  $t > 0$
5.  $P_{x,y}(t) > 0$  für ein  $t > 0$

**Beweis.** Beweis siehe [8] □

**Korollar 2.4** *Die Relation  $\longleftrightarrow$  ist eine Äquivalenzrelation.*

**Beweis.**

- Reflexivität:  $x \longleftrightarrow x$  klar.
- Symmetrie:  $x \longleftrightarrow y \iff y \longleftrightarrow x$  klar.
- Transitivität: Gelte  $x \rightarrow y$  und  $y \rightarrow z$ . Mittels Theorem 2.3 (3) folgt weiters für alle  $t > 0$ :

$$P_{x,y}(t) > 0 \quad \text{und} \quad P_{y,z}(t) > 0$$

komponentenweise Anwendung der Chapman-Kolmogorow-Gleichung (2.6) liefert nun für  $t > s > 0$ :

$$P_{x,z}(t) = \sum_{p \in \mathbb{N}_0^{2K}} P_{x,p}(s) P_{p,z}(t-s) > P_{x,y}(s) P_{y,z}(t-s) > 0$$

Daraus folgt wiederum mittels Theorem 2.3, dass  $x \rightarrow z$ . Analog folgt auch  $z \rightarrow x$  und damit die gewünschte Transitivitätseigenschaft. □

Da  $\longleftrightarrow$  eine Äquivalenzrelation auf  $\mathbb{N}_0^{2K}$  festlegt, teilt sie den Zustandsraum des Markovprozesses  $X$  in Äquivalenzklassen ein.

**Definition 2.5 (Irreduzibilität)** *Sei  $Y$  ein zeitstetiger homogener Markovprozess mit Intensitätsmatrix  $R$ , Anfangsverteilung  $\lambda: Y \sim \text{Markov}(\lambda, R)$  und abzählbarem Zustandsraum  $E$ . Dann heißt der Markovprozess  $Y$  irreduzibel, falls  $E$  bezüglich der Äquivalenzrelation  $\longleftrightarrow$  eine einzelne Klasse ist.*

**Bemerkung 2.6** Bei einem irreduziblen Markovprozess ist also jeder Zustand von jedem anderen Zustand aus mit positiver Wahrscheinlichkeit erreichbar. In unserem Fall muss demnach für den Markovprozess  $X$  gelten: für alle  $x, y \in \mathbb{N}_0^{2K}$  gilt  $x \rightarrow y$ .

Mittels Theorem 2.3, 3. lässt sich nun zeigen, dass der Markovprozess irreduzibel ist.

**Theorem 2.7** Der Markovprozess  $(X_t)_{t \geq 0}$  ist irreduzibel.

**Beweis.** Sei  $x, y \in \mathbb{N}_0^{2K}$  beliebig. Weiters definiere den Differenzvektor  $\Delta := x - y \in \mathbb{Z}^{2K}$ . Wir unterscheiden folgende zwei Fälle:

1.  $\Delta_i < 0$ : Dann beschreiben folgende Einträge der Intensitätsmatrix  $Q$  einen Weg von der  $i$ -ten Komponente von  $x$  zu der  $i$ -ten Komponente von  $y$ :

$$Q_{x, x+e_i} \cdot Q_{x+e_i, x+2e_i} \cdot \dots \cdot Q_{x+(|\Delta_i|-1)e_i, x+|\Delta_i|e_i} = \prod_{j=0}^{|\Delta_i|-1} f_i(x + je_i) > 0$$

2.  $\Delta_i > 0$ : Analog zu oben beschreiben folgende Einträge der Intensitätsmatrix  $Q$  einen Weg von der  $i$ -ten Komponente von  $x$  zu der  $i$ -ten Komponente von  $y$ :

$$Q_{x, x-e_i} \cdot Q_{x-e_i, x-2e_i} \cdot \dots \cdot Q_{x-(\Delta_i-1)e_i, x-\Delta_i e_i} = \prod_{j=0}^{\Delta_i-1} g_i(x - je_i) > 0$$

Schlussendlich liefert Multiplikation über die Komponenten  $i \in I := \{-K, \dots, -1, 1, \dots, K\}$  des Differenzvektors:

$$\prod_{i \in I} \prod_{j=0}^{|\Delta_i|-1} (\mathbb{1}_{\{\Delta_i=0\}} + \mathbb{1}_{\{\Delta_i < 0\}} f_i(x + je_i) + \mathbb{1}_{\{\Delta_i > 0\}} g_i(x - je_i)) > 0$$

Nach Theorem 2.3 3. ist dies äquivalent zu  $x \rightarrow y$ . Da  $x, y \in \mathbb{N}_0^{2K}$  beliebig gewählt war folgt die Aussage.  $\square$

## 2.2.2 Explosivität

In diesem Abschnitt werden wir zeigen, dass der Markovprozess  $X$  nicht explosiv ist.

Ein (rechtsstetiger) stochastischer Prozess  $(Y_t)_{t \geq 0}$  auf einem abzählbarem Zustandsraum wird *explosiv* genannt, falls dieser mit positiver Wahrscheinlichkeit in endlicher Zeit unendlich viele Sprünge erlaubt. Wir werden diesen Sachverhalt nun mathematisch formalisieren.

Bevor wir zeigen, dass der Prozess  $X$  nicht explosiv ist, erläutern wir noch einige zentrale Begriffe eines zeitstetigen stochastischen Prozesses mit abzählbarem Zustandsraum:

Sei  $(Y_t)_{t \geq 0}$  ein rechtsstetiger stochastischer Prozess auf einem abzählbarem Zustandsraum  $E$ .

**Definition 2.8 (Sprungzeiten)** Wir nennen eine Folge von Zeitpunkten  $(J_n)_{n \in \mathbb{N}}$  mit  $J_n \in \mathbb{R}_+ \cup \{\infty\}$  und

$$J_0 := 0 \quad J_{k+1} := \inf\{t \geq J_k : Y_t \neq Y_{J_k}\}$$

wobei  $\inf \emptyset := \infty$ , die Sprungzeiten (engl. jump times) des Prozesses  $Y$ .

Der zeitdiskrete Prozess  $Z_n := Y_{J_n}$ ,  $n \in \mathbb{N}_0$  ist der mit  $Y$  assoziierte **Sprungprozess** (engl. jump process). Weiters definieren wir mittels der Sprungzeiten:  $(J_n)_{n \in \mathbb{N}_0}$ , die Halteperioden des Prozesses  $Y$  folgendermaßen:

**Definition 2.9 (Halteperioden)** Bezeichne mit  $(J_n)_{n \in \mathbb{N}}$  die Sprungzeitpunkte des Prozesses  $Y$ . Wir bezeichnen die Folge  $(S_n)_{n \in \mathbb{N}}$  definiert durch:

$$S_n := \begin{cases} J_n - J_{n-1} & J_n < \infty \\ \infty & \text{sonst} \end{cases}$$

als die Halteperioden (engl. sojourn or holding times) des zugrunde liegenden Prozesses  $Y$ .

Wir definieren den Zeitpunkt der Explosion  $\zeta$  als:

$$\zeta := \sup_{n \in \mathbb{N}} J_n = \sum_{i=1}^{\infty} S_i$$

und nennen den Prozess  $Y$  **explosiv** falls:  $\mathbb{P}[\zeta < \infty] > 0$ . Andernfalls wird dieser **nicht explosiv** genannt.<sup>4</sup>

Im Kontext von Markovprozessen sagen wir, dass eine Intensitätsmatrix  $R$  explosiv ist, falls ein  $e \in E$  existiert, sodass für den zugehörigen Markovprozess  $Y$  gilt:

$$\mathbb{P}[\zeta < \infty | Y_0 = e] > 0$$

Um zu zeigen, dass der Markovprozess  $X$  beziehungsweise dessen Intensitätsmatrix  $Q$  aus Definition 2.1 nicht explosiv ist, verwenden wir Theorem A.5.

**Theorem 2.10** Die Intensitätsmatrix  $Q$ , definiert in (2.1), ist nicht explosiv.

**Beweis.** Sei  $\theta > 0$  beliebig,  $I := \{-K, \dots, -1, 1, \dots, K\}$  und wähle ein  $V : \mathbb{N}_0^{2K} \rightarrow [-1, 1]$  d.h.  $|V| \leq 1$ . Unter Verwendung der Definition 2.1 folgt dann weiters für  $x \in \mathbb{N}_0^{2K}$ :

$$\begin{aligned} (QV)(x) &= \sum_{y \in \mathbb{N}_0^{2K}} Q_{x,y} V(y) = \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} V(y) + Q_{x,x} V(x) \stackrel{(2.3)}{=} \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} (V(y) - V(x)) \stackrel{(2.1)}{=} \\ &= \sum_{i \in I} [f_i(x)(V(x + e_i) - V(x)) + g_i(x)(V(x - e_i) - V(x))] \end{aligned}$$

Gelte nun für alle  $x \in \mathbb{N}_0^{2K}$ :  $\theta V(x) = (QV)(x)$ . Wir müssen zeigen, dass  $V \equiv 0$ . Angenommen  $\exists x^* \in \mathbb{N}_0^{2K}$  mit  $V(x^*) \neq 0$ . Dann würde folgen:

$$\begin{aligned} \theta |V(x^*)| &= |(QV)(x^*)| = \left| \sum_{i \in I} [f_i(x^*)(V(x^* + e_i) - V(x^*)) + g_i(x^*)(V(x^* - e_i) - V(x^*))] \right| \leq \\ &\leq 2 \sum_{i \in I} [f_i(x^*) + g_i(x^*)] =: C < \infty \end{aligned}$$

<sup>4</sup>Wir werden uns im weiteren Verlauf der Arbeit nicht damit auseinandersetzen was mit dem Prozess  $Y$  nach dem Explosionszeitpunkt  $\zeta$  passiert. Deshalb fügen wir einen neuen Zustand:  $\infty$  zu  $E$  hinzu und definieren  $\forall t \geq \zeta$   $Y_t := \infty$ . Prozesse mit dieser Eigenschaft werden *minimal* genannt.

Damit gäbe es ein  $\tilde{C}$  unabhängig von  $\theta$  so dass  $\theta < \tilde{C} < \infty$ . Daraus ergibt sich ein Widerspruch, da  $\theta > 0$  beliebig.  $\square$

Damit haben wir gezeigt, dass in unserem Modell der Markovprozess  $X$  beziehungsweise das Ordervolumen  $X_t^i$  an einem beliebigen Limit  $i$  in endlicher Zeit nicht explodiert.

### 2.2.3 Rekurrenz und Transienz

Wir kommen nun in diesem Abschnitt zu einer weiteren wichtigen Eigenschaft bezüglich Markovprozessen. Die Begriffe Rekurrenz und Transienz geben uns in gewisser Weise Aufschluss über das Langzeitverhalten eines Markovprozesses. Wie wir weiter sehen werden, ist die positive Rekurrenz eine hinreichende und notwendige Bedingung für die Existenz einer eindeutigen invarianten Verteilung (siehe Abschnitt 2.3). Zuerst wiederholen wir einige Begriffe für einen allgemeinen zeitstetigen homogenen Markovprozess  $Y$  mit abzählbarem Zustandsraum.

Im Folgenden sei  $(Y_t)_{t \geq 0} \sim \text{Markov}(\lambda, R)$  auf einem abzählbarem Zustandsraum  $E$ .

Wir sagen ein Zustand  $e \in E$  ist *rekurrent* falls:

$$\mathbb{P}_e[\{t \geq 0 : Y_t = e\} \text{ ist unbeschränkt}] := \mathbb{P}[\{t \geq 0 : Y_t = e\} \text{ ist unbeschränkt} | Y_0 = e] = 1$$

Weiters ist ein Zustand  $e \in E$  ist *transient* falls gilt:

$$\mathbb{P}_e[\{t \geq 0 : Y_t = e\} \text{ ist unbeschränkt}] = 0$$

Ein rekurrenter Zustand ist demnach dadurch charakterisiert, dass der Markovprozess zu ihm immer wieder zurückkehrt, während ein transienter Zustand nach einer gewissen Zeit nie wieder angenommen wird. Wir sagen der Markovprozess  $Y$  ist rekurrent oder transient, falls dies auf jeden Zustand  $e \in E$  zutrifft.

Theorem A.6 liefert uns, dass Rekurrenz und Transienz Klasseigenschaften bezüglich der Äquivalenzrelation  $\longleftrightarrow$  sind und dass jeder Zustand eines Markovprozesses entweder rekurrent oder transient ist.

Im Folgenden sei die erste Ankunftszeit eines Zustandes  $e$  definiert als:  $T_e := \inf\{t \geq J_1 : Y_t = e\}$ . Anhand von Theorem A.7 können wir erkennen, dass aus  $\mathbb{P}_e[T_e < \infty] = 1$  folgt, dass der Zustand  $e$  rekurrent ist.

Wir nennen einen Zustand *positiv rekurrent* falls gilt:  $\mathbb{E}_e[T_e] := \mathbb{E}[T_e | Y_0 = e] < \infty$ . Anhand der Definitionen ist klar ersichtlich, dass aus positiver Rekurrenz immer die Rekurrenz eines Zustandes folgt. Die Umkehrung gilt im Allgemeinen jedoch nicht.

Im Folgenden bezeichne  $X \sim \text{Markov}(\lambda, Q)$  weiterhin den Markovprozess mit der Intensitätsmatrix  $Q$  aus Definition 2.1 und einer beliebigen Anfangsverteilung  $\lambda$ . Wir werden nun zeigen, dass unter bestimmten Regularitätsvoraussetzungen an die Intensitätsmatrix  $Q$  jeder Zustand  $x \in \mathbb{N}_0^{2K}$  des Markovprozesses  $X$  positiv rekurrent ist.

Zuerst formulieren wir die beiden Regularitätsvoraussetzungen:

**Annahme 2 (negativer individual Drift)** Es existiert ein  $C \in \mathbb{N}$  und ein  $\delta > 0$  sodass für alle  $i \in \{-K, \dots, -1, 1, \dots, K\}$  und für alle  $x \in \mathbb{N}_0^{2K}$  folgendes gilt:

Falls  $x_i > C$  dann gilt:

$$f_i(x) - g_i(x) < -\delta \quad (2.7)$$

**Annahme 3 (beschränkte Ankunftsintensität)** Es existiert ein  $H > 0$  sodass:

$$\sum_{i \in \{-K, \dots, -1, 1, \dots, K\}} f_i(x) \leq H \quad (2.8)$$

für alle  $x \in \mathbb{N}_0^{2K}$ .

**Bemerkung 2.11** Annahme 2 kann folgendermaßen interpretiert werden: Falls das Ordervolumen an einem Limit eine gewissen Schranke überschreitet, tendiert es wieder dazu zu sinken (mean reverting effect). Annahme 3 garantiert uns, dass die Ankunftsintensität von Orders für jeden beliebigen Zustand  $x \in \mathbb{N}_0^{2K}$  beschränkt bleibt.

**Theorem 2.12** Der Markovprozess  $(X_t)_{t \geq 0}$  ist positiv rekurrent.

**Beweis.** Sei  $z > 1$ ,  $C \in \mathbb{N}$ ,  $\delta$  laut Annahme 2 gewählt und  $H > 0$  laut Annahme 3 gewählt. Weiters definieren wir  $I := \{-K, \dots, -1, 1, \dots, K\}$  und folgende Funktion  $V$ :

$$V : \mathbb{N}_0^{2K} \rightarrow [2K, \infty) \\ V(x) := \sum_{i \in I} z^{(x_i - C)^+}$$

Als erstes zeigen wir, dass die oben definierte Funktion  $V$  kompakt ist (siehe Definition A.8).

Sei  $d > 0$  beliebig, und definiere  $A := \{x \in \mathbb{N}_0^{2K} : V(x) < d\}$ .

• **1.Fall:** Sei  $d \leq 2K$ , dann folgt dass  $|A| = |\emptyset| = 0 < \infty$ .

• **2.Fall:** Sei  $d > 2K$ . Wir unterscheiden wiederum zwei Fälle:

- 2.1: Betrachte  $x \in A \cap B$  mit  $B := \{x \in \mathbb{N}_0^{2K} : x_i \leq C, \forall i \in \{-K, \dots, -1, 1, \dots, K\}\}$   
 $\Rightarrow |A \cap B| = (C + 1)^{2K} < \infty$ .
- 2.2: Betrachte  $x \in A \cap B^c$  mit  $B^c := \{x \in \mathbb{N}_0^{2K} : \exists i \in \{-K, \dots, -1, 1, \dots, K\} : x_i > C\}$

Sei zuallererst  $i^* \in \{-K, \dots, -1, 1, \dots, K\} : x_{i^*} > C$  und  $x_i = 0$  für  $i \neq i^*$ . Setze  $\tilde{B} := \{x \in \mathbb{N}_0^{2K} : \exists i^* \in \{-K, \dots, -1, 1, \dots, K\} : x_{i^*} > C, x_i = 0 \forall i \neq i^*\}$ . Damit  $V(x) < d$  muss gelten  $x_{i^*} \in \left[ C + 1, \left\lfloor \frac{\ln(d - 2K + 1)}{\ln(z)} + C \right\rfloor \right]^5$ .

Daher ist die Anzahl der Möglichkeiten für dieses  $x_{i^*}$  gegeben durch:  $\left\lfloor \frac{\ln(d - 2K + 1)}{\ln(z)} \right\rfloor$ .

Dementsprechend bekommen wir für  $|\tilde{B}| = 2K \cdot \left\lfloor \frac{\ln(d - 2K + 1)}{\ln(z)} \right\rfloor$ . Damit wurde jedoch lediglich der Fall behandelt, bei dem ein  $x_i > C$  ist und alle anderen gleich null

<sup>5</sup> $\lfloor \cdot \rfloor$  bezeichnet hierbei die Gaußklammer und ist definiert für  $x \in \mathbb{R}$  als  $\lfloor x \rfloor := \max\{n \in \mathbb{N} : n \leq x\}$

sind. Doch wir können die Anzahl der Möglichkeiten im allgemeinen Fall 2.2 mit diesen Überlegungen nach oben abschätzen indem wir für jedes  $x_i > C$  die Möglichkeiten mit  $\left\lfloor \frac{\ln(d-2K+1)}{\ln(z)} \right\rfloor$  nach oben hin abschätzen und erhalten dass in 2.2 gilt:

$$|A \cap B^c| \leq \sum_{i=1}^{2K} \binom{2K}{i} \cdot \left\lfloor \frac{\ln(d-2K+1)}{\ln(z)} \right\rfloor^i \cdot (C+1)^{2K-i} < \infty.$$

Da  $d > 0$  beliebig war ist A in jedem Fall endlich und V kompakt.

Sei nun  $x \in \mathbb{N}_0^{2K}$  beliebig, dann folgt

$$\begin{aligned} (QV)(x) &= \sum_{y \in \mathbb{N}_0^{2K}} Q_{x,y} V(y) = \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} V(y) + Q_{x,x} V(x) \stackrel{(2.3)}{=} \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} (V(y) - V(x)) = \\ &= \sum_{i \in I} [Q_{x, x+e_i} (V(x+e_i) - V(x)) + Q_{x, x-e_i} (V(x-e_i) - V(x))] \stackrel{*}{=} \end{aligned}$$

unter Verwendung von :

$$V(x \pm e_i) - V(x) = \sum_{j \in I} z^{(x_j \pm (e_i)_j - C)^+} - \sum_{j \in I} z^{(x_j - C)^+} = z^{(x_i \pm 1 - C)^+} - z^{(x_i - C)^+}$$

folgt weiters:

$$\stackrel{*}{=} \sum_{i \in I} \left[ f_i(x) (z^{(x_i+1-C)^+} - z^{(x_i-C)^+}) + g_i(x) (z^{(x_i-1-C)^+} - z^{(x_i-C)^+}) \right] \stackrel{**}{=}$$

$$\text{falls } x_i < C \Rightarrow x_i + 1 \leq C \Rightarrow (x_i + 1 - C)^+ = (x_i - C)^+ = 0 \Rightarrow z^{(x_i+1-C)^+} - z^{(x_i-C)^+} = 0$$

$$\text{falls } x_i < C + 1 \Rightarrow x_i \leq C \Rightarrow (x_i - 1 - C)^+ = (x_i - C)^+ = 0 \Rightarrow z^{(x_i-1-C)^+} - z^{(x_i-C)^+} = 0$$

Daher können wir uns auf die Mengen  $\{x_i \geq C\}$  beziehungsweise  $\{x_i \geq C + 1\}$  einschränken und es folgt weiters:

$$\begin{aligned} &\stackrel{**}{=} \sum_{i \in I} f_i(x) \mathbb{1}_{\{x_i \geq C\}} \left( z^{(x_i-C)^+} \cdot (z-1) \right) + g_i(x) \mathbb{1}_{\{x_i \geq C+1\}} \left( z^{(x_i-C)^+} \cdot \underbrace{\left( \frac{1}{z} - 1 \right)}_{= -\frac{z-1}{z}} \right) = \\ &= (z-1) \cdot \sum_{i \in I} z^{(x_i-C)^+} \cdot \left( f_i(x) \mathbb{1}_{\{x_i \geq C\}} - \frac{\mathbb{1}_{\{x_i \geq C+1\}} g_i(x)}{z} \right) = \\ &= (z-1) \left[ \underbrace{\sum_{i \in I: x_i = C} f_i(x)}_{\leq H} \right] + (z-1) \left[ \sum_{i \in I: x_i > C} \left( f_i(x) - \frac{g_i(x)}{z} \right) z^{(x_i-C)^+} \right] \stackrel{***}{\leq} \end{aligned}$$

Im Folgenden wollen wir nun die zweite Summe in obiger Gleichung abschätzen. Dazu sei  $\delta > 0$  so dass:  $f_i(x) - g_i(x) < -\delta$  falls  $x_i > C$ .

definiere  $r := H(z-1) + \delta \Rightarrow f_i(x) - g_i(x) < -r + (z-1)H$

daher folgt für  $z$  hinreichend nahe bei 1 :  $f_i(x) - \frac{g_i(x)}{z} < \frac{-r+(z-1)H}{z} =: r'$

Schließlich folgt dann:

$$\begin{aligned}
& \stackrel{***}{\leq} (z-1)H - (z-1)r' \sum_{i \in I: x_i > C} z^{(x_i - C)^+} \leq \\
& = (z-1)H - (z-1)r' \sum_{i \in I} z^{(x_i - C)^+} + (z-1)r' \sum_{i \in I: x_i \leq C} z^{(x_i - C)^+} \leq \\
& \leq (z-1)H - (z-1)r' \underbrace{\sum_{i \in I} z^{(x_i - C)^+}}_{=V(x)} + 2K(z-1)r' = \\
& = - \underbrace{(z-1)r'}_{=:b} V(x) + \underbrace{(z-1)[H + 2rK']}_{=:c}
\end{aligned}$$

Zusammenfassend haben wir nun für  $b, c > 0$  und  $x \in \mathbb{N}_0^{2K}$  und kompaktes  $V : \mathbb{N}_0^{2K} \rightarrow [2K, \infty)$  gezeigt:

$$(QV)(x) = \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} (V(y) - V(x)) \leq -bV(x) + c$$

Damit folgt die Aussage mittels Theorem A.9. □

## 2.3 Existenz einer invarianten Verteilung

Wir zeigen in diesem Kapitel, dass der Markovprozess  $X \sim \text{Markov}(\lambda, Q)$  mit der Intensitätsmatrix aus Definition 2.1 eine invariante Verteilung besitzt. Zuvor erläutern wir noch den Begriff der Ergodizität im Bezug auf einen allgemeinen zeitstetigen homogenen Markovprozess mit abzählbarem Zustandsraum.

**Definition 2.13 (ergodisch)** Sei  $Y$  ein zeitstetiger homogener Markovprozess mit Intensitätsmatrix  $R$ , Anfangsverteilung  $\lambda: Y \sim \text{Markov}(\lambda, R)$ , abzählbarem Zustandsraum  $E$  und Übergangsmatrix  $(P(t))_{t \geq 0}$ . Dann heißt  $Y$  ergodisch, falls ein Wahrscheinlichkeitsmaß  $\pi$  auf dem Messraum  $(E, \mathcal{E})$  existiert sodass:

$$\pi P(t) = \pi \quad \forall t \geq 0 \tag{2.9}$$

$$\lim_{t \rightarrow \infty} P_{e,h}(t) = \pi(h) \quad \forall e, h \in E \tag{2.10}$$

Die linke Seite von (2.9) ist ein Wahrscheinlichkeitsmaß auf  $(E, \mathcal{E})$  und ist definiert als:

$$(\pi P(t))(\cdot) := \sum_{e \in E} \pi(e) P_{e,\cdot}(t)$$

**Bemerkung 2.14** Ein Wahrscheinlichkeitsmaß  $\pi$ , welches Bedingung (2.9) in Definition 2.13 erfüllt wird invariantes Wahrscheinlichkeitsmaß, beziehungsweise im Falle, dass  $\pi$  die Anfangsverteilung des Markovprozesses  $X$  beschreibt, stationäre oder invariante Verteilung genannt (siehe A.1). In unserem Fall gilt  $E = \mathbb{N}_0^{2K}$  und  $\mathcal{E} = \mathcal{P}(\mathbb{N}_0^{2K})$ . Daher bedeutet Gleichung (2.9) für

$X \sim \text{Markov}(\pi, Q)$  *komponentenweise*:

$$\forall x \in \mathbb{N}_0^{2K}, \forall t \geq 0: \quad \mathbb{P}[X_t = x] = (\pi P(t))(x) = \pi(x) = \mathbb{P}[X_0 = x]$$

Wir kommen nun zur zentralen Aussage in diesem Kapitel. Unter den Voraussetzungen von Annahme 3 und Annahme 2 ist der Markovprozess  $X$ , welcher das Orderbuchvolumen modelliert, ergodisch und besitzt daher insbesondere eine invariante Verteilung.

**Korollar 2.15** *Sei  $X \sim \text{Markov}(\lambda, Q)$ , wobei  $Q$  die Intensitätsmatrix aus Definition 2.1 beschreibt. Es gelten die Annahmen 2 und 3. Dann ist der Markovprozess  $(X_t)_{t \geq 0}$  ergodisch. Insbesondere existiert ein eindeutiges invariantes Wahrscheinlichkeitsmaß  $\pi$ .*

**Beweis.** Aus Theorem 2.12 gemeinsam mit Theorem A.10 folgt die Existenz eines invarianten Wahrscheinlichkeitsmaßes  $\pi$ . Die Eindeutigkeit eines invarianten Wahrscheinlichkeitsmaßes liefert uns [8, Theorem 3.5.2].

Die Konvergenz von  $\lim_{t \rightarrow \infty} P_{x,y}(t)$  gegen  $\pi(y)$  folgt schließlich aus Theorem A.4. □

Die Aussage, dass  $Q$  nicht explosiv ist würde demnach auch aus Theorem 2.12 (positiven Rekurrenz des Markovprozesses  $X$ ) folgen (vergleiche Kapitel 2.2.2).

Wir werden im restlichen Teil der Arbeit die invariante Verteilung unseres Markovprozesses  $X$  mit  $\pi$  bezeichnen.

# Kapitel 3

## Modell I - unabhängige Limits

In diesem Kapitel widmen wir uns dem ersten Modell zur Orderbuchbeschreibung. Als Grundlage dient hierbei die allgemeine Modellbeschreibung in Kapitel 2. Wie in Kapitel 2 beschreiben wir in diesem Modell die Dynamik des LOB in Zeiten eines konstanten Referenzpreises  $p_{ref}$ . Dieses Kapitel basiert im Wesentlichen auf Abschnitt 2.3 von [6].

### 3.1 Modellbeschreibung - Modell I

Im Modell I nehmen wir an, dass die einzelnen Komponenten des Markovprozesses  $X_t(\omega) = (X_t^{-K}(\omega), \dots, X_t^{-1}(\omega), X_t^1(\omega), \dots, X_t^K(\omega))$  unabhängig sind. Des Weiteren fordern wir, dass die Intensitäten für beliebiges Limit  $P_i$  lediglich von der Ordergröße  $x_i \in \mathbb{N}_0$  an diesem Limit abhängen d.h.  $f_i(x) = f_i(x_i)$  beziehungsweise  $g_i(x) = g_i(x_i)$ .

Wir betrachten wie bisher die folgenden drei Ordertypen: *Market Order*, *Cancellation Order* und *Limit Order*. Zur Vereinfachung nehmen wir an, dass *Market Orders* direkt an jedes Limit  $P_i$  gesendet werden können. Das heißt sie konsumieren nicht automatisch zum *best Bid-* oder *best Ask Price*. Es ist somit beispielsweise möglich eine *Market Order* am zweiten Limit  $P_{\pm 2}$  abzugeben im Falle eines nichtleeren ersten Limits  $P_{\pm 1}$ .

Im Folgenden spalten wir die Intensitäten additiv in die verschiedenen Typen von Ordnern auf: Wir bezeichnen mit  $\lambda_i^L(n)$  die Intensität für eine *Limit Order* am Preislimit  $P_i$ , falls für die dazugehörige Ordergröße  $x_i$  gilt  $x_i = n \in \mathbb{N}_0$ . Analog bezeichnen  $\lambda_i^C(n)$  und  $\lambda_i^M(n)$  die Intensitäten für eine *Cancellation* respektive eine *Market Order*. Zusammenfassend haben wir<sup>1</sup>:

$$\lambda_i^k : \mathbb{N}_0 \rightarrow \mathbb{R}_+ \quad k \in \{L, C, M\}$$

**Annahme 4** Als eine weitere Vereinfachung in diesem Modell setzen wir eine gewisse *Symmetrieeigenschaft* des LOB voraus:

$$\lambda_i^k(\cdot) = \lambda_{-i}^k(\cdot) \quad (k, i) \in \{L, C, M\} \times \{-K, \dots, -1, 1, \dots, K\}$$

und verlangen, dass die Ankunftsprozesse für diese drei Typen von Ordnern an einem Limit  $i$ , be-

---

<sup>1</sup> $\lambda_i^C(0) = \lambda_i^M(0) := 0$

dingt auf einen Zustand des Orderbuches an diesem Limit, d.h. bedingt auf  $\{X_t^i = n\}$ , unabhängig voneinander sind.

Damit ist die Intensitätsmatrix  $Q$  (siehe Definition 2.1) des Markovprozesses  $X$  für  $x, y \in \mathbb{N}_0^{2K}$  gegeben durch:

$$\begin{aligned} Q_{x, x+e_i} &= f_i(x_i) := \lambda_i^L(x_i) \\ Q_{x, x-e_i} &= f_i(x_i) := \lambda_i^C(x_i) + \lambda_i^M(x_i) \end{aligned} \tag{3.1}$$

$$\begin{aligned} Q_{x,x} &= - \sum_{y \in \mathbb{N}_0^{2K}: y \neq x} Q_{x,y} \\ Q_{x,y} &= 0 \quad y \notin \{x, x+e_i, x-e_i\} \end{aligned}$$

Unter den obigen Voraussetzungen ist der Markovprozess  $X$ , welcher das Orderbuchvolumen modelliert, ein komponentenweise unabhängiger  $2K$ -dimensionaler Prozess, wobei die einzelnen Komponenten  $(X_t^i)_{t \geq 0}$  eindimensionale *Birth and Death* Prozesse sind.

Das heißt für alle  $i \in \{-K, \dots, -1, 1, \dots, K\}$  ist  $(X_t^i)_{t \geq 0}$  ein zeitstetiger homogener Markovprozess auf dem Zustandsraum  $\mathbb{N}_0$  mit Intensitätsmatrix  $Q^i = (Q_{l,k}^i)_{(l,k) \in \mathbb{N}_0^2}$  definiert durch:

$$Q^i := \begin{pmatrix} -\lambda_i^L(0) & \lambda_i^L(0) & 0 & \dots & & & \\ \lambda_i^C(1) + \lambda_i^M(1) & -(\lambda_i^C(1) + \lambda_i^M(1) + \lambda_i^L(1)) & \lambda_i^L(1) & 0 & \dots & & \\ 0 & \lambda_i^C(2) + \lambda_i^M(2) & -(\lambda_i^C(2) + \lambda_i^M(2) + \lambda_i^L(2)) & \lambda_i^L(2) & 0 & \dots & \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & \\ & \vdots & & & & & \end{pmatrix} \tag{3.2}$$

Wie man erkennt, ist  $Q$  eine Tridiagonalmatrix. In der oberen Nebendiagonale stehen die Intensitäten für eine Erhöhung des Ordervolumens am  $i$ -ten Limit d.h. die Intensität vom Zustand  $n \in \mathbb{N}_0$  zum Zustand  $n+1$  zu wechseln (birth). Die Einträge in der unteren Nebendiagonale bezeichnen die Intensität vom Zustand  $n$  zum Zustand  $n-1$  zu wechseln (death). Die Einträge in der Hauptdiagonale spiegeln die Rate, mit der man im selben Zustand verweilt, wieder.

Für eine festes Preislimit  $P_i$  bezeichne nun  $J_n$  die Sprungzeitpunkte und  $S_n$  die Halteperioden des Markovprozesses  $(X_t^i)_{t \geq 0}$ . Sei weiters der dazugehörige Sprungprozess gegeben durch  $Z_n^i := X_{J_n}^i$ . Aus der Theorie von Markovprozessen wissen wir, dass die Halteperioden bedingt auf den letzten Zustand einer Exponentialverteilung folgen. Die Rate der Exponentialverteilung ist dabei durch die jeweiligen negativen Diagonaleinträge der Intensitätsmatrix  $Q$  gegeben:

$$S_n | (Z_{n-1}^i = k) \sim \text{Exp}(-Q_{k,k}) \quad , k \in \mathbb{N}_0$$

Bei der Modellierung des gesamten LOB  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$  können wir aufgrund

der Unabhängigkeit die einzelnen Komponenten getrennt voneinander betrachten.

## 3.2 Schätzung der Intensitäten

In diesem Abschnitt widmen wir uns der Schätzung der drei verschiedenen Intensitäten  $\lambda_i^L$ ,  $\lambda_i^C$  und  $\lambda_i^M$  zu gegebenen Daten. Wir können hierbei aufgrund der Annahme, dass die Ankunftsprozesse der drei verschiedenen Ordertypen unabhängig voneinander sind (siehe Annahme 5), diese Intensitäten einzeln schätzen. Dazu werden wir die Maximum Likelihood Methode heranziehen. Da die Intensitäten von einem Preislimit  $P_i$  abhängen, gehen wir folgendermaßen vor:

Wir fixieren ein festes Preislimit  $P_i, i \in \{-K, \dots, -1, 1, \dots, K\}$ . Im Folgenden bezeichnen wir mit  $(J_n)_{n \in \mathbb{N}_0}$  und  $(S_n)_{n \in \mathbb{N}}$  die Sprungzeitpunkte respektive Halteperioden des Markovprozesses  $(X_t^i)_{t \geq 0}$  (siehe Definition 2.8 und Definition 2.9). Die Zeitpunkte einer Order am Preislimit  $P_i$ , welche durch eine der drei verschiedenen Ordertypen hervorgerufen wird, sind gerade durch die Sprungzeitpunkte  $(J_n)_{n \in \mathbb{N}}$  gegeben. Die Zeit zwischen der  $k$ -ten und der  $(k+1)$ -ten Order ist somit gegeben durch  $S_{k+1}$ . Des Weiteren notieren wir zu jedem Sprungzeitpunkt den Ordertyp und bezeichnen diesen als  $\mathcal{T}(J_{k+1}) \in \{L, M, C\}$ . Als dritte Größe speichern wir die Anzahl der Orders am Limit  $P_i$  vor dem Sprung (in  $AES_i^2$ ), die gegeben ist als  $X_{J_k}^i \in \mathbb{N}$ .

In der Zeit eines konstanten Referenzpreises sammeln wir also zusammenfassend zum Limit  $P_i$  folgende Daten:

$$D := \{(S_{k+1}, X_{J_k}^i, \mathcal{T}(J_{k+1})) | k \in \{0, \dots, N\}\}$$

Aus Abschnitt 3.1 wissen wir bereits, dass die Halteperiode  $S_{k+1}$  des Markovprozesses  $(X_t^i)_{t \geq 0}$  bedingt auf die Anzahl der Orders davor:  $\{X_{J_k}^i = n\}$  exponentialverteilt sind mit Parameter  $\Lambda_i(n) := \lambda_i^C(n) + \lambda_i^M(n) + \lambda_i^L(n)$ . Um die ML Schätzer für die drei Intensitäten zu bestimmen, berechnen wir zuerst die auf  $\{X_{J_k}^i = n\}$  bedingte gemeinsame Verteilung von  $\mathcal{T}$  und  $S_{k+1}$ . Diese lässt sich im Falle einer Limit Order mittels Lemma A.3 folgendermaßen für  $k \in \mathbb{N}$  bestimmen:

$$\begin{aligned} F_i(t, L) &:= \mathbb{P}[S_{k+1} \leq t, \mathcal{T}(J_{k+1}) = L | X_{J_k}^i = n] = \\ &= \mathbb{P}[S_{k+1} \leq t, X_{J_{k+1}}^i = n + 1 | X_{J_k}^i = n] = \\ &\stackrel{A.3}{=} \mathbb{P}[S_{k+1} \leq t | X_{J_k}^i = n] \cdot \mathbb{P}[X_{J_{k+1}}^i = n + 1 | X_{J_k}^i = n] = \\ &= \left(1 - e^{-t \cdot \Lambda_i(n)}\right) \cdot \frac{\lambda_i^L(n)}{\Lambda_i(n)} \quad t \geq 0 \end{aligned}$$

wobei die letzte Gleichheit daraus resultiert, dass der assoziierte Sprungprozess  $Z_k^i := X_{J_k}^i$  des Markovprozesses  $(X_t^i)_{t \geq 0}$  eine diskrete Markovkette mit Übergangsmatrix  $\Pi^i$ , welche in Gleichung (3.12) definiert wird, ist.

Mittels obiger Umformung folgt weiters

$$\mathbb{P}[\mathcal{T}(J_{k+1}) = L | X_{J_k}^i = n] = \mathbb{P}[X_{J_{k+1}}^i = n + 1 | X_{J_k}^i = n] = \frac{\lambda_i^L(n)}{\Lambda_i(n)}$$

<sup>2</sup>AES bezeichnet hierbei Average Event Size am betrachteten Limit  $i$

Analog bekommen wir

$$\begin{aligned}
F_i(t, \{C, M\}) &= \mathbb{P} [S_{k+1} \leq t, \mathcal{T}(J_{k+1}) \in \{M, C\} | X_{J_k}^i = n] = \\
&= \mathbb{P} [S_{k+1} \leq t, X_{J_{k+1}}^i = n - 1 | X_{J_k}^i = n] = \\
&= \mathbb{P} [S_{k+1} \leq t | X_{J_k}^i = n] \cdot \mathbb{P} [X_{J_{k+1}}^i = n - 1 | X_{J_k}^i = n] = \\
&= \left(1 - e^{-t \cdot \Lambda_i(n)}\right) \cdot \frac{\lambda_i^M(n) + \lambda_i^C(n)}{\Lambda_i(n)} \quad t \geq 0
\end{aligned}$$

Des weiteren gilt aufgrund der Disjunktheit der Ereignisse:

$$\begin{aligned}
&\mathbb{P} [\mathcal{T}(J_{k+1}) = M | X_{J_k}^i = n] + \mathbb{P} [\mathcal{T}(J_{k+1}) = C | X_{J_k}^i = n] = \\
&= \mathbb{P} [X_{J_{k+1}}^i = n - 1 | X_{J_k}^i = n] = \frac{\lambda_i^M(n)}{\Lambda_i(n)} + \frac{\lambda_i^C(n)}{\Lambda_i(n)}
\end{aligned}$$

Unter der folgenden zusätzlichen Annahme hinsichtlich der Unterscheidung von Market und Cancellation Order erhalten wir schließlich die gemeinsame Dichte des Zufallsvektors:

$$(S_{k+1}, \mathcal{T}(J_{k+1})) \Big| \{X_{J_k}^i = n\}$$

**Annahme 5** *Unter obigen Voraussetzungen gelten folgende Beziehungen:*

$$\begin{aligned}
F_i(t, M) &= \left(1 - e^{-t \cdot \Lambda_i(n)}\right) \cdot \frac{\lambda_i^M(n)}{\Lambda_i(n)} \quad t \geq 0 \\
F_i(t, C) &= \left(1 - e^{-t \cdot \Lambda_i(n)}\right) \cdot \frac{\lambda_i^C(n)}{\Lambda_i(n)} \quad t \geq 0 \\
\mathbb{P} [\mathcal{T}(J_{k+1}) = M | X_{J_k}^i = n] &= \frac{\lambda_i^M(n)}{\Lambda_i(n)} \\
\mathbb{P} [\mathcal{T}(J_{k+1}) = C | X_{J_k}^i = n] &= \frac{\lambda_i^C(n)}{\Lambda_i(n)}
\end{aligned}$$

Daraus folgt unmittelbar, dass die beiden Zufallsvariablen  $S_{k+1}$  und  $\mathcal{T}(J_{k+1})$  bedingt auf die Menge  $\{X_{J_k}^i = n\}$  unabhängig sind.

Des Weiteren ergibt sich für die Dichte  $f_i(t, k)$  von  $(S_{k+1}, \mathcal{T}(J_{k+1})) | \{X_{J_k}^i = n\}$  bezüglich<sup>3</sup>  $d(\lambda \otimes \zeta)$  folgender Ausdruck:

$$f_i(t, k) = f_i^{S_{k+1}}(t) \cdot f_i^{\mathcal{T}(J_{k+1})}(k) = \lambda_i^k(n) \cdot e^{-t \cdot \Lambda_i(n)} \quad t \geq 0, k \in \{L, M, C\}$$

Um die Intensitäten für ein festes Preislevel  $n$  zu schätzen, filtern wir aus unseren gesamten Beobachtungen  $D$ , diejenigen Daten, bei denen die Anzahl der Orders vor dem Event gleich  $n$  war heraus und definieren:

$$D_n := \{(S_{k+1}, \mathcal{T}(J_k)) | (S_{k+1}, X_{J_k}^i, \mathcal{T}(J_k)) \in D, X_{J_k}^i = n\}$$

<sup>3</sup> $\lambda$  bezeichnet hierbei das Lebesguemaß und  $\zeta$  das Zählmaß

Des Weiteren definieren wir die Anzahl der verschiedenen Typen der Events zu gegebener Ordersize  $n$ :

$$\begin{aligned} N_n^L &:= |\{\mathcal{T}(J_{k+1}) = L | (S_{k+1}, \mathcal{T}(J_k)) \in D_n\}| \\ N_n^M &:= |\{\mathcal{T}(J_{k+1}) = M | (S_{k+1}, \mathcal{T}(J_k)) \in D_n\}| \\ N_n^C &:= |\{\mathcal{T}(J_{k+1}) = C | (S_{k+1}, \mathcal{T}(J_k)) \in D_n\}| \end{aligned}$$

Aus obiger Definition ergibt sich sofort  $N_n^L + N_n^M + N_n^C = |D_n|$ . Als nächstes bestimmen wir die Likelihoodfunktion  $L$  zu den gesammelten Daten  $D_n$ :

$$L(\lambda_i^L(n), \lambda_i^M(n), \lambda_i^C(n) | D_n) = \prod_{d_n \in D_n} f_i(d_n) = [\lambda_i^L(n)]^{N_n^L} \cdot [\lambda_i^M(n)]^{N_n^M} \cdot [\lambda_i^C(n)]^{N_n^C} \cdot e^{-S \cdot \Lambda_i(n)}$$

wobei  $S := \sum_{d_n \in D_n} (d_n)_1$ .

Hierbei bezeichnet  $(d_n)_1$  die erste Komponente des Tupels  $d_n \in D_n$  und beschreibt somit die Zeit zwischen zwei aufeinanderfolgenden Events  $S_{k+1}$  aus der Stichprobe  $D_n$ .

Die Loglikelihoodfunktion  $\ell(\cdot | D_n) := \ln(L(\cdot | D_n))$  ist gegeben durch:

$$\ell(\lambda_i^L(n), \lambda_i^M(n), \lambda_i^C(n) | D_n) = N_n^L \ln(\lambda_i^L(n)) + N_n^M \ln(\lambda_i^M(n)) + N_n^C \ln(\lambda_i^C(n)) - \Lambda_i(n)S$$

Maximierung bezüglich der Intensitäten führt auf folgende Gleichungen:

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda_i^L(n)} = \frac{N_n^L}{\lambda_i^L(n)} - S \stackrel{!}{=} 0 &\iff \widehat{\lambda_i^L(n)} = \frac{N_n^L}{S} \\ \frac{\partial \ell}{\partial \lambda_i^M(n)} = \frac{N_n^M}{\lambda_i^M(n)} - S \stackrel{!}{=} 0 &\iff \widehat{\lambda_i^M(n)} = \frac{N_n^M}{S} \\ \frac{\partial \ell}{\partial \lambda_i^C(n)} = \frac{N_n^C}{\lambda_i^C(n)} - S \stackrel{!}{=} 0 &\iff \widehat{\lambda_i^C(n)} = \frac{N_n^C}{S} \end{aligned}$$

Man kann sich im obigen Fall leicht davon überzeugen, dass die Hesse Matrix von  $\ell$  negativ definit ist und deshalb ein Maximum vorliegt. Durch Erweitern des Bruches erhält man schlussendlich folgenden Ausdruck der ML Schätzer für die Intensitäten am Preislimit  $i$  und für die Ordergröße  $n$ :

$$\widehat{\lambda_i^L(n)} = \frac{N_n^L}{S} = \frac{|D_n|}{S} \frac{N_n^L}{|D_n|} = \widehat{\Lambda_i(n)} \cdot \frac{N_n^L}{|D_n|} \quad (3.3)$$

$$\widehat{\lambda_i^M(n)} = \widehat{\Lambda_i(n)} \cdot \frac{N_n^M}{|D_n|} \quad (3.4)$$

$$\widehat{\lambda_i^C(n)} = \widehat{\Lambda_i(n)} \cdot \frac{N_n^C}{|D_n|} \quad (3.5)$$

wobei  $\widehat{\Lambda_i(n)} := \left(\frac{1}{|D_n|} S\right)^{-1} = \left(\frac{1}{|D_n|} \sum_{d_n \in D_n} (d_n)_1\right)^{-1}$  genau dem ML Schätzer der Zufallsvariablen  $S_{k+1} | \{X_{J_k}^i = n\} \sim \text{Exp}(\Lambda_i(n))$  entspricht.

### 3.3 Konfidenzintervalle für die Intensitäten

Wir berechnen in diesem Abschnitt Konfidenzintervalle für die Intensitäten  $\lambda_i^L(n)$ ,  $\lambda_i^M(n)$  und  $\lambda_i^C(n)$  für  $n \in \mathbb{N}_0$ . Für ein festes Preislimit  $P_i$  sei im folgenden  $(S_k)_{k \in \mathbb{N}}$  die Folge der Halteperioden des Markovprozesses  $(X_t^i)_{t \geq 0}$  und bezeichne mit  $(\mathcal{T}(J_k))_{k \in \mathbb{N}}$  die Folge der Ordertypen zu den jeweiligen Ordnern. Im weiteren Verlauf widmen wir uns der Berechnungen eines Konfidenzintervalles für die Limit Order Intensität  $\lambda_i^L(n)$ . Konfidenzintervalle für die beiden anderen Intensitäten berechnen sich analog.

Dazu betrachten wir für ein festes Preislevel  $n$  und  $k \in \mathbb{N}_0$  den Zufallsvektor:

$$(S_{k+1}, \mathbb{1}_{\{\mathcal{T}(J_{k+1})=L\}}) \Big| \{X_{J_k}^i = n\} \quad (3.6)$$

Aufgrund der Messbarkeit der Indikatorfunktion sind dessen Komponenten ebenso unabhängig (siehe Abschnitt 3.2). Bedingt auf  $\{X_{J_k}^i = n\}$  sind diese folgendermaßen verteilt:

$$\begin{aligned} S_{k+1} &\sim \text{Exp}(\Lambda_i(n)) \\ \mathbb{1}_{\{\mathcal{T}(J_{k+1})=L\}} &\sim \text{Ber}\left(\frac{\lambda_i^L(n)}{\Lambda_i(n)}\right) \end{aligned}$$

Die Maximum Likelihood Schätzer für die Parameter  $\rho_i^L(n) := \frac{\lambda_i^L(n)}{\Lambda_i(n)}$  beziehungsweise  $\Lambda_i(n)$  werden wir mit  $\widehat{\rho_i^L(n)}$  und  $\widehat{\Lambda_i(n)}$  bezeichnen. In Abschnitt 3.2 haben wir bereits gesehen, dass zu gegebener Stichprobe für das Preislimit  $i$  und Preislevel  $n$ :

$$D_n := \{(S_{k+1}, \mathcal{T}(J_k)) \mid (S_{k+1}, X_{J_k}^i, \mathcal{T}(J_k)) \in D, X_{J_k}^i = n\}$$

der Maximum Likelihood Schätzer für den Parameter  $\Lambda_i(n)$  gegeben ist durch:

$$\widehat{\Lambda_i(n)} = \frac{|D_n|}{\sum_{d_n \in D_n} (d_n)_1}$$

Überdies haben wir in diesem Abschnitt auch den Maximum Likelihood Schätzer für  $\lambda_i^L(n)$  berechnet. Dieser ist gegeben als:

$$\widehat{\lambda_i^L(n)} = \frac{N_n^L}{\sum_{d_n \in D_n} (d_n)_1}$$

Mittels Theorem B.1 folgt nun, dass  $\widehat{\rho_i^L(n)}$  gegeben ist als:

$$\widehat{\rho_i^L(n)} = \frac{\widehat{\lambda_i^L(n)}}{\widehat{\Lambda_i(n)}} = \frac{\lambda_i^L(n)}{\Lambda_i(n)} = \frac{N_n^L}{|D_n|}$$

Um ein Konfidenzintervall zu bilden, benötigen wir eine Aussage über das asymptotische Verhalten des Maximum Likelihood Schätzers:

$$\hat{\theta} := \begin{pmatrix} \widehat{\Lambda_i(n)} \\ \widehat{\rho_i^L(n)} \end{pmatrix}$$

Aus [7, Theorem 5.1] folgt dass der Maximum Likelihood Schätzer asymptotisch normalverteilt ist mit folgenden Parametern:

$$\hat{\theta} \xrightarrow[|D_n| \rightarrow \infty]{(d)} \mathcal{N} \left( \theta, \frac{1}{|D_n|} \mathcal{I}^{-1}(\theta) \right) \approx \mathcal{N} \left( \theta, \frac{1}{|D_n|} \mathcal{I}^{-1}(\hat{\theta}) \right)$$

wobei  $\mathcal{I}$  die Fisher Matrix zum Zufallsvektor aus (3.6) bezeichnet und  $\theta := \begin{pmatrix} \Lambda_i(n) \\ \rho_i^L(n) \end{pmatrix}$ .

Zunächst werden wir die Fisher Matrix  $\mathcal{I}(\theta)$  berechnen. Im Folgenden definieren wir die gemeinsame Dichte des Zufallsvektors aus (3.6) mit  $f(s, x|\theta) = f(s, x|\Lambda_i(n), \rho_i^L(n))$ . Diese ist aufgrund der Unabhängigkeit für  $s \geq 0$  und  $x \in \{0, 1\}$  gegeben als:

$$f(s, x | \Lambda_i(n), \rho_i^L(n)) = f(s | \Lambda_i(n)) \cdot f(x | \rho_i^L(n)) = \Lambda_i(n) e^{-s \cdot \Lambda_i(n)} \cdot (\rho_i^L(n))^x (1 - \rho_i^L(n))^{1-x}$$

Mittels Lemma B.2 folgt, dass die Fisher Matrix  $\mathcal{I}$  gegeben ist als:

$$\mathcal{I}(\theta) = -\mathbb{E} [H_\theta(\log(f(S_{k+1}, \mathbb{1}_{\{\mathcal{T}(J_{k+1})=L\}}|\theta)\}) | \{X_{J_k}^i = n\})]$$

wobei wir mit  $H_\theta(\log f(s, x, \theta))$  die Hesse Matrix der Funktion  $\log(f(s, x, \cdot))$  bezüglich des Parameters  $\theta$  bezeichnen. Eine einfache Rechnung führt auf folgenden Ausdruck:

$$H_\theta(\log f(s, x|\theta)) = \begin{pmatrix} -\frac{1}{\Lambda_i(n)^2} & 0 \\ 0 & -\left(\frac{x}{\rho_i^L(n)} + \frac{1-x}{(1-\rho_i^L(n))^2}\right) \end{pmatrix}$$

somit ergibt sich für die Fisher Matrix:

$$\mathcal{I}(\theta) = \begin{pmatrix} \frac{1}{\Lambda_i(n)^2} & 0 \\ 0 & \left(\frac{1}{\rho_i^L(n) \cdot (1-\rho_i^L(n))}\right) \end{pmatrix}$$

Schließlich folgt dass der Maximum Likelihood Schätzer  $\hat{\theta}$  asymptotisch normalverteilt ist:

$$\hat{\theta} \approx \mathcal{N} \left( \theta, \frac{1}{|D_n|} \begin{pmatrix} \widehat{\Lambda_i(n)}^2 & 0 \\ 0 & \left(\widehat{\rho_i^L(n)} \cdot (1 - \widehat{\rho_i^L(n)})\right) \end{pmatrix} \right)$$

Mittels Theorem B.3 angewendet auf die Funktion  $g(x, y) := x \cdot y$  folgt somit, dass

$$g(\hat{\theta}) = \widehat{\lambda_i^L(n)} \approx \mathcal{N} \left( \widehat{\lambda_i^L(n)}, \frac{2\widehat{\rho_i^L(n)}\widehat{\Lambda_i(n)}^2 - \widehat{\rho_i^L(n)}^2\widehat{\Lambda_i(n)}^2}{|D_n|} \right)$$

Daraus lässt sich für  $\alpha \in [0, 1]$  folgendes  $(1 - \alpha)$ -Konfidenzintervall für  $\lambda_i^L(n)$  konstruieren:

$$\left[ \widehat{\lambda_i^L(n)} - \frac{z_{1-\frac{\alpha}{2}} \cdot \widehat{\Lambda_i(n)} \sqrt{\widehat{\rho_i^L(n)} (2 - \widehat{\rho_i^L(n)})}}{\sqrt{|D_n|}}, \widehat{\lambda_i^L(n)} - \frac{z_{\frac{\alpha}{2}} \cdot \widehat{\Lambda_i(n)} \sqrt{\widehat{\rho_i^L(n)} (2 - \widehat{\rho_i^L(n)})}}{\sqrt{|D_n|}} \right] \quad (3.7)$$

wobei wir mit  $z_\alpha$  das  $\alpha$  - Quantil der Standardnormalverteilung bezeichnen. Konfidenzintervalle

für  $\lambda_i^M(n)$  und  $\lambda_i^C(n)$  lassen sich analog berechnen. Die Funktion  $Int()$  aus Kapitel 9, (6) berechnet mittels (3.3) - (3.5) den Maximum Likelihood Schätzer der Intensitäten und das Konfidenzintervall aus (3.7).

### 3.4 Berechnung der invarianten Verteilung

Aus Abschnitt 2.3 wissen wir bereits, dass unser Markovprozess  $X$  eine Invariante Verteilung  $\pi$  besitzt. Im Modell I wird es uns gelingen diese in geschlossener Form darzustellen. Die invariante Verteilung  $\pi$  hängt im betrachteten Modell I nur von den Intensitäten  $\lambda_i^L$ ,  $\lambda_i^C$  und  $\lambda_i^M$  ab. Bei der Herleitung der invarianten Verteilung  $\pi$  folgen wir im Wesentlichen [4, Abschnitt 2.1]. Zum Schluss besprechen wir den Zusammenhang zwischen den invarianten Verteilungen des assoziierten Sprungprozesses  $Z$  und dem zeitstetigen Markovprozess  $X$  selbst.

Wir betrachten weiterhin den Markovprozess  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$ , welcher das Orderbuchvolumen modelliert. Wie bereits oben erwähnt bezeichnen wir dessen invariante Verteilung mit  $\pi$  d.h.:

$$(\pi P(t))(x) = \pi(x) \quad \forall x \in \mathbb{N}_0^{2K}$$

wobei  $(P(t))_{t \geq 0}$  die Übergangsmatrix des Markovprozesses  $X$  beschreibt.

Da die Komponenten von  $X$  im Modell I unabhängig sind, zerfällt die invariante Verteilung  $\pi$  in das Produkt der Randverteilungen der Prozesse  $(X_t^i)_{t \geq 0}$ :

$$\pi(x) = \pi_{-K}(x_{-K}) \cdot \dots \cdot \pi_{-1}(x_{-1}) \cdot \pi_1(x_1) \cdot \dots \cdot \pi_K(x_K)$$

wobei  $\pi_i$  die invariante Verteilung des eindimensionalen birth and death Prozesses  $(X_t^i)_{t \geq 0}$  beschreibt:

$$(\pi_i P^i(t))(n) = \pi_i(n) \quad \forall n \in \mathbb{N}_0$$

Hier bezeichnet  $(P^i(t))_{t \geq 0}$  die Übergangsmatrix von  $(X_t^i)_{t \geq 0}$ , welche durch die Intensitätsmatrix  $Q^i$  aus (3.2) bestimmt ist.

**Lemma 3.1 (siehe [8], Theorem 3.5.5)** *Sei  $Q^i$  die Intensitätsmatrix des irreduziblen und (positiv) rekurrenten Markovprozesses  $(X_t^i)_{t \geq 0}$ . Bezeichne weiters  $\pi_i$  ein Maß auf  $\mathbb{N}_0$ , dann sind folgenden Aussagen äquivalent:*

1.  $\pi Q^i(\cdot) = \sum_{n \in \mathbb{N}_0} \pi(n) Q_{n,\cdot}^i \equiv 0$
2.  $\pi P^i(t)(\cdot) = \sum_{n \in \mathbb{N}_0} \pi(n) P^i(t)_{n,\cdot} = \pi(\cdot) \quad \forall t \geq 0$  (invariante Verteilung)

**Beweis.** siehe [8]. □

**Theorem 3.2** *Sei  $X$  der Markovprozess aus Modell I. Sei weiters  $\pi$  dessen invariante Verteilung. Dann lässt sich  $\pi$  folgendermaßen darstellen:*

$$\pi(x) = \pi_{-K}(x_{-K}) \cdot \dots \cdot \pi_{-1}(x_{-1}) \cdot \pi_1(x_1) \cdot \dots \cdot \pi_K(x_K) \quad , \forall x \in \mathbb{N}_0^{2K}$$

wobei  $\pi_i$  die invariante Verteilung des eindimensionalen birth and death Prozesses  $(X_t^i)_{t \geq 0}$  beschreibt und für  $i \in \{-K, \dots, -1, 1, \dots, K\}$  gegeben ist durch:

$$\pi_i(0) = \left( 1 + \sum_{n=1}^{\infty} \prod_{j=1}^n \rho_i(j-1) \right)^{-1} \quad (3.8)$$

$$\pi_i(n) = \pi_i(0) \cdot \prod_{j=1}^n \rho_i(j-1) \quad \forall n \in \mathbb{N} \quad (3.9)$$

und  $\rho_i$  das Verhältnis der Intensität vom Zustand  $n$  auf den Zustand  $n+1$  zu wechseln, zu der Intensität vom Zustand  $n+1$  auf den Zustand  $n$  zu wechseln, beschreibt (engl. arrival departure ratio) d.h.

$$\rho_i(n) := \frac{\lambda_i^L(n)}{\lambda_i^M(n+1) + \lambda_i^C(n+1)}$$

**Beweis.** Wir fixieren ein festes Limit  $i \in \{-K, \dots, -1, 1, \dots, K\}$  und definieren zuerst folgende Hilfsgrößen:

$$\begin{aligned} \lambda^+(\cdot) &:= \lambda_i^L(\cdot) \\ \lambda^-(\cdot) &:= \lambda_i^M(\cdot) + \lambda_i^C(\cdot) \\ \Lambda(\cdot) &:= \lambda^+(\cdot) + \lambda^-(\cdot) \end{aligned}$$

Aus Theorem 3.1 folgt unmittelbar, dass für die invariante Verteilung  $\pi_i$  gelten muss:

$$\pi_i Q^i(n) = 0 \quad \forall n \in \mathbb{N}_0 \quad (3.10)$$

Da  $Q^i$  eine Tridiagonalmatrix ist (siehe (3.2)) vereinfacht sich Gleichung (3.10) folgendermaßen:

$$\begin{aligned} n = 0 : \quad & 0 = -\pi_i(0)\lambda^+(0) + \pi_i(1)\lambda^-(1) \\ n \geq 1 : \quad & 0 = \pi_i(n-1)\lambda^+(n-1) - \pi_i(n)\Lambda(n) + \pi_i(n+1)\lambda^-(n+1) \end{aligned}$$

dies liefert durch algebraische Umformung:

$$\begin{aligned} n = 0 : \quad & \pi_i(1) = \pi_i(0) \cdot \frac{\lambda^+(0)}{\lambda^-(1)} \\ n \geq 1 : \quad & \pi_i(n+1) = \pi_i(n) \cdot \frac{\Lambda(n)}{\lambda^-(n+1)} - \pi_i(n-1) \cdot \frac{\lambda^+(n-1)}{\lambda^-(n+1)} \end{aligned}$$

betrachten wir nun exemplarisch den Fall  $n = 1$ , dann lässt sich  $\pi_i(2)$  folgendermaßen ausdrücken:

$$\begin{aligned} \pi_i(2) &= \pi_i(1) \cdot \frac{\Lambda(1)}{\lambda^-(2)} - \pi_i(0) \cdot \frac{\lambda^+(0)}{\lambda^-(2)} = \pi_i(0) \cdot \frac{\lambda^+(0)}{\lambda^-(1)} \cdot \frac{\Lambda(1)}{\lambda^-(2)} - \pi_i(0) \cdot \frac{\lambda^+(0)}{\lambda^-(2)} = \\ &= \pi_i(0) \cdot \frac{\lambda^+(0)\lambda^+(1)}{\lambda^-(1)\lambda^-(2)} \end{aligned}$$

Hieraus lässt sich schon ein gewisses Muster erkennen. Wir werden mittels einem Induktionsbeweis zeigen dass:

$$\pi_i(n) = \pi_i(0) \prod_{j=1}^n \frac{\lambda^+(j-1)}{\lambda^-(j)} \quad (3.11)$$

- $n = 0$ : dies folgt sofort aus der Tatsache, dass  $\prod_{j=1}^0 := 1$
- Des Weiteren wurde der Fall  $n = 1$  bereits oben gezeigt.
- Angenommen Gleichung (3.11) gelte für  $\{0, \dots, n\}$ , dann folgt:

$$\begin{aligned} \pi_i(n+1) &= \pi_i(n) \cdot \frac{\Lambda(n)}{\lambda^-(n+1)} - \pi_i(n-1) \cdot \frac{\lambda^+(n-1)}{\lambda^-(n+1)} = \\ &= \left( \pi_i(0) \prod_{j=1}^n \frac{\lambda^+(j-1)}{\lambda^-(j)} \right) \cdot \frac{\Lambda(n)}{\lambda^-(n+1)} - \left( \pi_i(0) \prod_{j=1}^{n-1} \frac{\lambda^+(j-1)}{\lambda^-(j)} \right) \cdot \frac{\lambda^+(n-1)}{\lambda^-(n+1)} = \\ &= \left( \pi_i(0) \prod_{j=1}^n \frac{\lambda^+(j-1)}{\lambda^-(j)} \right) \left[ \frac{\lambda(n) + \lambda^-(n)}{\lambda^-(n+1)} - \frac{\lambda^-(n)\lambda^+(n-1)}{\lambda^+(n-1)\lambda^-(n+1)} \right] = \pi_i(0) \prod_{j=1}^{n+1} \frac{\lambda^+(j-1)}{\lambda^-(j)} \end{aligned}$$

Daraus folgt schließlich (3.11). Schlussendlich folgt mittels Summation über  $\mathbb{N}_0$  in (3.11):

$$1 = \pi_i(0) \sum_{n=0}^{\infty} \prod_{j=1}^n \frac{\lambda^+(j-1)}{\lambda^-(j)} \iff \pi_i(0) = \left( 1 + \sum_{n=1}^{\infty} \prod_{j=1}^n \frac{\lambda^+(j-1)}{\lambda^-(j)} \right)^{-1}$$

Aus der Tatsache, dass  $\rho_i(j-1) = \frac{\lambda^+(j-1)}{\lambda^-(j)}$  folgt die Aussage des Theorems.  $\square$

Die Funktion  $InvDistrCont()$  aus Kapitel 9, (7) berechnet die invariante Verteilung  $\pi$  aus Theorem 3.2 zu gegebenen Intensitäten im Modell I.

### 3.4.1 Invariante Verteilung des Sprungprozesses

Wir besprechen nun den Zusammenhang zwischen der invarianten Verteilung des zeitstetigen Markovprozess  $X$ , die wir wie bisher mit  $\pi$  bezeichnen, zu der invarianten Verteilung des mit  $X$  assoziierten Sprungprozess  $Z$ , welche wir im folgenden mit  $\mu$  bezeichnen werden. Wie wir sehen werden, stimmen  $\pi$  und  $\mu$  im Allgemeinen nicht überein. Diese Unterscheidung wird vor allem später wichtig sein, wenn wir die empirische Verteilung von simulierten Markovprozessen mit deren theoretisch invarianten Verteilung vergleichen. Aufgrund der Unabhängigkeitsannahme im Modell I zerfallen die invarianten Verteilungen  $\pi$  und  $\mu$  in das Produkt der Randverteilungen d.h. für  $x \in \mathbb{N}_0^{2K}$  gilt:

$$\begin{aligned} \pi(x) &= \pi_{-K}(x_{-k}) \cdot \dots \cdot \pi_{-1}(x_{-1}) \cdot \pi_1(x_1) \cdot \dots \cdot \pi_K(x_K) \\ \mu(x) &= \mu_{-K}(x_{-k}) \cdot \dots \cdot \mu_{-1}(x_{-1}) \cdot \mu_1(x_1) \cdot \dots \cdot \mu_K(x_K) \end{aligned}$$

Darum werden wir im Folgenden ein festes Preislimit  $P_i$  fixieren und einen Zusammenhang zwischen der invarianten Verteilung  $\pi_i$  des eindimensionalen zeitstetigen *birth and death* Prozesses  $(X_t^i)_{t \geq 0}$  und der invarianten Verteilung  $\mu_i$  des assoziierten Sprungprozesses herleiten. Dadurch folgt sofort aus obigen Gleichungen der Zusammenhang zwischen  $\pi$  und  $\mu$ .

Sei im Folgenden  $(J_n)_{n \in \mathbb{N}}$  die Folge der Sprungzeiten von  $X^i$  (siehe 2.8) und bezeichne mit  $Z_n^i := X_{J_n}^i$  dessen assoziierten Sprungprozess. Aus [8, Theorem 2.8.2] folgt, dass  $Z_n^i$  eine diskrete Markovkette mit Zustandsraum  $\mathbb{N}_0$  ist. Wir bezeichnen, sofern diese existiert, mit  $\mu_i$  deren invariante Verteilung sowie mit  $\Pi^i$  deren Übergangsmatrix d.h.  $Z^i \sim \text{Markov}(\mu_i, \Pi^i)$ . Die Übergangsmatrix  $\Pi^i$  kann mithilfe der Intensitätsmatrix des zeitstetigen Prozesses  $Q^i$  folgendermaßen berechnet werden:

Sei  $(k, j) \in \mathbb{N}_0^2$  dann gilt:

$$k \neq j : \quad (\Pi^i)_{k,j} = \begin{cases} \frac{Q_{k,j}^i}{-Q_{k,k}^i} & -Q_{k,k}^i > 0 \\ 0 & -Q_{k,k}^i = 0 \end{cases} \quad (3.12)$$

$$(\Pi^i)_{k,k} = \begin{cases} 0 & -Q_{k,k}^i > 0 \\ 1 & -Q_{k,k}^i = 0 \end{cases}$$

Diese ist, wie man leicht sieht, tatsächlich eine stochastische Matrix und hat, unter der Annahme, dass  $-Q_{k,k}^i > 0 \forall k \in \mathbb{N}_0$ , aufgrund der Tridiagonalgestalt von  $Q^i$  lediglich in den beiden Nebendiagonalen Einträge ungleich 0. In der Literatur wird  $\Pi^i$  auch als assoziierte Sprungmatrix (engl. jump matrix) des Markovprozesses  $X_i$  bezeichnet.

Wir werden nun einen Zusammenhang zwischen  $\mu_i$  und  $\pi_i$  herleiten. Aus Lemma 3.1 folgt, dass für die invariante Verteilung  $\pi_i$  von  $X^i$  folgende Gleichung erfüllt sein muss:

$$\pi_i Q^i(n) = \sum_{k \in \mathbb{N}_0} \pi_i(k) Q_{k,n}^i = 0 \quad \forall n \in \mathbb{N}_0$$

Dies ist äquivalent zu

$$\sum_{k \neq n} \pi_i(k) Q_{k,n}^i = -\pi_i(n) Q_{n,n}^i$$

woraus unter Annahme von  $-Q_{n,n}^i > 0, \forall n \in \mathbb{N}_0$  mittels 3.12 folgt

$$\sum_{k \neq n} [-\pi_i(k) Q_{k,k}^i] \Pi_{k,n}^i = -\pi_i(n) Q_{n,n}^i$$

da per Definition gilt  $\Pi_{n,n}^i = 0$  folgt weiters:

$$\sum_{k \in \mathbb{N}_0} [-\pi_i(k) Q_{k,k}^i] \Pi_{k,n}^i = -\pi_i(n) Q_{n,n}^i$$

somit ist  $[-\pi_i(\cdot) Q_{\cdot,\cdot}^i] : \mathbb{N}_0 \rightarrow [0, 1]$  ein invariantes Maß des assoziierten Sprungprozesses  $Z^i$ . Mittels Theorem A.6 und Theorem 2.3 folgt, dass  $Z^i$  die Eigenschaften der Rekurrenz und Irreduzibilität von dem zeitstetigen Markovprozess  $X^i$  erbt. Aus [8, Theorem 1.7.5 und Theorem 1.7.6] folgt, dass  $Z^i$  bis auf Multiplikation mit einer positiven Konstanten ein eindeutiges invariantes Maß besitzt. Nehmen wir nun an, dass  $\mu_i$  die invariante Verteilung von  $Z^i$  beschreibt. Insbesondere ist dann  $\mu_i$

auch ein invariantes Maß und erfüllt somit für alle  $n \in \mathbb{N}_0$ :

$$\mu_i(n) = c \cdot [-\pi_i(n)Q_{n,n}^i] \quad c > 0$$

womit sofort aus den Bedingungen  $\sum_{k \in \mathbb{N}_0} \mu_i(n) = 1$  beziehungsweise  $\sum_{k \in \mathbb{N}_0} \pi_i(n) = 1$  folgt:

$$\mu_i(n) = \frac{-\pi_i(n)Q_{n,n}^i}{\sum_{k \in \mathbb{N}_0} (-\pi_i(k)Q_{k,k}^i)} \quad \forall n \in \mathbb{N}_0 \quad (3.13)$$

$$\pi_i(n) = \frac{\frac{\mu_i(n)}{-Q_{n,n}^i}}{\sum_{k \in \mathbb{N}_0} \left( \frac{\mu_i(k)}{-Q_{k,k}^i} \right)} \quad \forall n \in \mathbb{N}_0 \quad (3.14)$$

Die Existenz von  $\pi_i$  und somit die Konvergenz der Summe im Nenner von (3.14) wurde bereits in 2.3 bewiesen. Dies folgte aus der positiven Rekurrenz des Markovprozesses. Falls die Summe im Nenner von (3.13) konvergiert, stellt dies auch die Existenz einer invarianten Verteilung des assoziierten Sprungprozesses sicher. Die Konvergenz dieser Summe ist somit äquivalent zu der positiven Rekurrenz des assoziierten Sprungprozesses  $Z^i$  (siehe [8, Theorem 1.7.7]). Jedoch kann im Allgemeinen nicht von der positiven Rekurrenz des zeitstetigen Prozesses  $X^i$  auf die des assoziierten Sprungprozesses  $Z^i$  gefolgert werden.

Um eine explizite Darstellung von  $\mu_i$  im Modell I zu bekommen, setzen wir nun die aus Theorem 3.2 erhaltene Darstellung von  $\pi_i$  in Gleichung (3.13) ein. Wir verwenden weiterhin die folgende Notation:

$$\begin{aligned} \lambda_i^+(n) &:= \lambda_i^L(n) \quad \forall n \in \mathbb{N}_0 \\ \lambda_i^-(n) &:= \lambda_i^C(n) + \lambda_i^M(n) \quad \forall n \in \mathbb{N} \quad (\lambda_i^-(0) := 0) \\ \Lambda_i(n) &:= \lambda_i^C(n) + \lambda_i^M(n) + \lambda_i^L(n) = -Q_{n,n}^i \quad \forall n \in \mathbb{N}_0 \end{aligned}$$

Damit folgt schließlich:

$$\begin{aligned} \mu_i(n) &= \frac{-\pi_i(n)Q_{n,n}^i}{\sum_{n \in \mathbb{N}_0} (-\pi_i(n)Q_{n,n}^i)} = \frac{\prod_{k=1}^n \rho_i(k-1) \cdot \Lambda_i(n)}{\sum_{n \in \mathbb{N}_0} \prod_{k=1}^n \rho_i(k-1) \cdot \Lambda_i(n)} = \frac{\prod_{k=1}^n \frac{\lambda_i^+(k-1)}{\lambda_i^-(k)} \cdot \Lambda_i(n)}{\sum_{n \in \mathbb{N}_0} \prod_{k=1}^n \frac{\lambda_i^+(k-1)}{\lambda_i^-(k)} \cdot \Lambda_i(n)} = \\ &= \frac{\prod_{k=1}^n \frac{\lambda_i^+(k-1)}{\Lambda_i(k-1)} \cdot \frac{\Lambda_i(k)}{\lambda_i^-(k)} \cdot \Lambda_i(0)}{\sum_{n \in \mathbb{N}_0} \prod_{k=1}^n \frac{\lambda_i^+(k-1)}{\Lambda_i(k-1)} \cdot \frac{\Lambda_i(k)}{\lambda_i^-(k)} \cdot \Lambda_i(0)} \end{aligned}$$

Aus der Definition der Sprungmatrix in Gleichung (3.12) ist unmittelbar ersichtlich, dass  $\forall k \in \mathbb{N}_0$  die Wahrscheinlichkeit eines *birth* beziehungsweise *deaths* am Level  $k$  gegeben ist durch

$$(\Pi^i)_{k,k+1} = \frac{\lambda_i^+(k)}{\Lambda_i(k)} \quad (\Pi^i)_{k,k-1} = \frac{\lambda_i^-(k)}{\Lambda_i(k)}$$

Setzen wir dies nun in obige Gleichung ein, so bekommen wir die explizite Darstellung der invari-

anten Verteilung  $\mu_i$  in Abhängigkeit von der Sprungmatrix  $\Pi^i$ :

$$\mu_i(n) = \frac{\prod_{k=1}^n \left[ \frac{(\Pi^i)_{k-1,k}}{(\Pi^i)_{k,k-1}} \right]}{\sum_{n \in \mathbb{N}_0} \prod_{k=1}^n \left[ \frac{(\Pi^i)_{k-1,k}}{(\Pi^i)_{k,k-1}} \right]} \quad \forall n \in \mathbb{N}_0$$

wobei  $\prod_{k=1}^0 := 1$ . Zusammenfassend haben wir somit folgendes Theorem gezeigt.

**Theorem 3.3** Sei  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$  der homogene zeitstetige Markovprozess, welcher das Orderbuchvolumen im Modell I beschreibt. Dann ist  $(X_t^i)_{t \geq 0}$  ein eindimensionaler birth and death Prozess mit Intensitätsmatrix  $Q^i$  gegeben durch Gleichung (3.2) und invarianter Verteilung  $\pi_i$  aus Theorem 3.2.

Sei weiters  $(Z_n^i)_{n \in \mathbb{N}_0}$  der zu  $X^i$  assoziierte Sprungprozess d.h.  $Z_n^i := X_{J_n}^i$  wobei  $J_n$  die Sprungzeiten von  $X^i$  bezeichnet. Dann ist  $Z^i$  eine diskrete Markovkette mit Übergangsmatrix  $\Pi^i$  definiert in 3.12. Angenommen  $Z^i$  besitzt eine invariante Verteilung  $\mu_i$ , dann ist diese gegeben durch:

$$\mu_i(0) = \left( 1 + \sum_{n \in \mathbb{N}} \prod_{k=1}^n \left[ \frac{(\Pi^i)_{k-1,k}}{(\Pi^i)_{k,k-1}} \right] \right)^{-1} \quad (3.15)$$

$$\mu_i(n) = \mu_i(0) \cdot \prod_{k=1}^n \left[ \frac{(\Pi^i)_{k-1,k}}{(\Pi^i)_{k,k-1}} \right] \quad \forall n \in \mathbb{N} \quad (3.16)$$

Hierbei ist ersichtlich, dass die Darstellung der invarianten Verteilung  $\mu_i$  der Formel für die invariante Verteilung des zeitstetigen Markovprozesses (3.8) - (3.9) entspricht, wenn man die *arrival departure ratio*  $\rho_i(k-1)$  durch den Term  $\frac{(\Pi^i)_{k-1,k}}{(\Pi^i)_{k,k-1}}$  ersetzt.

Die Funktion *InvDistrJump()* aus Kapitel 9, (8) berechnet die invariante Verteilung  $\mu$  aus Theorem 3.3 zu gegebenen Intensitäten im Modell I.

# Kapitel 4

## Schätzung des Referenzpreises

Wie bereits zu Beginn von Kapitel 2 erwähnt, spielt in unserem Modell der Referenzpreis  $p_{ref}$  eine zentrale Rolle. Dieser definiert das Zentrum unserer Preislimits  $P_i, i \in \{-K, \dots, -1, 1, \dots, K\}$  und bestimmt somit das Gitter auf dem sich der Markovprozess  $X$  bewegt.

Insbesondere beim Schätzen der Intensitäten und in weiterer Folge der invarianten Verteilung ist es wesentlich, die vorhanden Daten, nämlich Zeit zwischen zwei aufeinanderfolgenden Events, Typ und Anzahl der Orders vor dem Event dem richtigen Preislimit zuzuordnen.

In Zeiten eines konstanten Referenzpreises können wir unser allgemeines Modell und insbesondere Modell I zur Beschreibung des Orderbuchvolumens verwenden. Ändert sich der Referenzpreis, ändert sich das Gitter beziehungsweise die Preislimits  $P_i$ , auf dem sich der Markovprozess  $X$  bewegt. Das Kapitel basiert auf Abschnitt 2.2.2 von [6].

Im Folgenden bezeichnen wir mit  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$  weiterhin den Markovprozess aus Kapitel 2, wobei wir mit  $x_i \in \mathbb{N}_0$  die Anzahl der Orders am Preislimit  $P_i \in \Delta_{tick} \cdot \mathbb{N}$  für alle  $i \in \{-K, \dots, -1, 1, \dots, K\}$  bezeichnen.

Zu gegebenen Daten aus dem LOB bestimmen wir den Referenzpreis auf folgende Art:

Zu jedem Zeitpunkt betrachten wir einen Ausschnitt des LOB und bezeichnen mit  $p_{bestAsk}$  beziehungsweise  $p_{bestBid}$  den *best Ask Price* respektive den *best Bid Price* des aktuellen Zustandes des LOB.

- Falls der *Bid-Ask Spread*:  $p_{bestBid} - p_{bestAsk} \in \Delta_{tick} \cdot \mathbb{N}$  ungerade ist, so können wir einfach das arithmetische Mittel als Referenzpreis heranziehen d.h. den *Mid Price*:

$$p_{ref} := \frac{p_{bestAsk} + p_{bestBid}}{2} \quad (4.1)$$

- Im Falle eines geraden *Bid-Ask Spreads* ist der *Mid Price* nicht länger geeignet, da er selbst eine mögliche Position für die Preislimits  $P_1$  oder  $P_{-1}$  ist und somit neue Order an diese Positionen eintreffen können. Andernfalls könnte keine neue Order in den *Bid-Ask Spread*

gesetzt werden. Daher ziehen wir für den Referenzpreis folgenden zwei Möglichkeiten heran:

$$\frac{p_{bestAsk} + p_{bestBid}}{2} + \frac{\Delta_{tick}}{2} \quad (4.2)$$

$$\frac{p_{bestAsk} + p_{bestBid}}{2} - \frac{\Delta_{tick}}{2} \quad (4.3)$$

und wählen diejenige, welche den kleinsten Abstand zum vorigen Wert von  $p_{ref}$  aufweist. Bei der empirischen Berechnung definierten wir den Anfangswert von  $p_{ref}$  im Falle eines geraden *Bid-Ask Spreads* mittels (4.2).

Die Funktion *refPrice()* aus Kapitel 9, (1) berechnet mittels obiger Methode den Referenzpreis aus zugrunde liegenden Daten.

# Kapitel 5

## LOBSTER Datenbank

Für die empirische Studie von Modell I verwenden wir historische Daten von der Limit Orderbuch Datenbank LOBSTER<sup>1</sup>. Diese stellt ein rekonstruiertes LOB für an der NASDAQ gehandelte Aktien zur Verfügung. Im Folgenden werden wir den Aufbau, der von LOBSTER zur Verfügung gestellten Daten, näher erläutern und im Anschluss daran die Vorgehensweise im Bezug auf das Schätzen der Intensitäten aus Abschnitt 3.2 erklären.

### 5.1 Output Struktur

LOBSTER stellt für jeden Handelstag zwei Datensätze, ein *Message-* und ein *Orderbookfile*, für eine ausgewählte Aktie zur Verfügung. Um den Datensatz herunterzuladen, gibt man den gewünschten Ticker der Aktie und die Zahl der nichtleeren Preislimits ein. Das *Messagefile* beinhaltet zu jeder getätigten Order, die eine Veränderung des Orderbuchs für die betrachteten Preislimits zur Folge hat, folgende Informationen:

- Time: Sekunden nach Mitternacht.
- Type: Ordertyp. Es werden folgende unterschiedliche Typen von Ordnern betrachtet:
  1. Eingabe einer neuen Limit Order
  2. partielle Cancellation einer Limit Order (im Bezug auf die Ordergröße)
  3. vollständige Cancellation einer Limit Order
  4. Ausführung (Exekution) einer sichtbaren Limit Order
  5. Ausführung (Exekution) einer versteckten Limit Order
  6. Auktionshandel<sup>2</sup>
  7. Indikator einer Handelsunterbrechung
- Order ID
- Size: Anzahl der Aktien einer Order.

---

<sup>1</sup>LOBSTER: Limit Order Book System - The Efficient Reconstructor at Humboldt Universität zu Berlin, Germany. <http://LOBSTER.wiwi.hu-berlin.de>

<sup>2</sup>Kumulierte Repräsentation von mehreren sichtbaren oder versteckten Limit Order. Da das Orderbuch sowieso bezüglich all dieser Orders aktualisiert wurde, ist diese Meldung redundant und wird nicht näher betrachtet.

- Price: Zu welchem Preis die Order ausgeführt/eingetragen wurde (in USD·10000).
- Direction: -1 definiert eine Verkaufs Limit Order, 1 definiert eine Kauf Limit Order.<sup>3</sup>

Wir zeigen exemplarisch die Darstellung eines *Messagefiles*. Als Aktie wurde France Telekom (FTE) vom 27.04.2012 ausgewählt.<sup>4</sup>

	Time	Type	OrderID	Size	Price	TradeDirection
2	34201.03	2	6334433	100	22.32	1
3	34201.22	1	6460303	1100	22.36	-1
4	34201.22	1	6460314	700	22.36	-1
5	34201.30	3	6222770	800	22.37	-1
6	34201.30	3	6315100	700	22.37	-1
7	34201.51	4	6460303	150	22.36	-1

Tabelle 5.1: Ausschnitt des Messagefiles

Zu jedem *Messagefile* gehört ein *Orderbookfile*, welches den aktualisierten Zustand des LOB nach der jeweiligen Order angibt. Die Veränderung des *Orderbookfile* von der  $(i-1)$ -ten auf die  $i$ -te Zeile wird durch die  $i$ -te Zeile im *Messagefile* beschrieben. Die ausgewählten Preislimits bestimmen bis zu welchem nichtleeren Limit die Information im *Orderbookfile* enthalten sein soll. Hierbei ist eine wesentliche Unterscheidung im Bezug auf unsere allgemeine Modellbeschreibung zu treffen: Das von LOBSTER zur Verfügung gestellte *Orderbookfile* gibt lediglich die besten nichtleeren Preislimits an. Überdies muss die Differenz zwischen den jeweiligen Preislimits nicht der *ticksiz*<sup>5</sup> entsprechen. In unserem Modell sind leere Preislimits  $P_i$  möglich. Hat man zu einem konstanten Referenzpreis die Preislimits  $P_i$  in eindeutiger Weise identifiziert, dann definieren diese ein äquidistantes Gitter um den Referenzpreis mit den Abständen  $|P_{\pm 1} - p_{ref}| = \frac{\Delta_{tick}}{2}$  und  $|P_i - P_{i+1}| = \Delta_{tick}$  (siehe Kapitel 2). Daher ist darauf Acht zu geben, welchem Preislimit  $P_i$  das im *Orderbookfile* angezeigte Preislimit entspricht.

Das *Orderbookfile* enthält folgenden Informationen:

- Ask Price 1 (ASKp1): Der Preis des ersten nichtleeren Ask-Limits in USD·10000. Dies entspricht nicht notwendigerweise  $P_1$ . (*Best Ask Price*)
- Ask Size 1 (ASKs1): Die Anzahl der Orders am ersten Ask-Limit (in Aktien).
- Bid Price 1 (BIDp1): Der Preis des ersten nichtleeren Bid-Limits in USD·10000. Dies entspricht nicht notwendigerweise  $P_{-1}$ . (*Best Bid Price*)
- Bid Size 1 (BIDs1): Die Anzahl der Orders am ersten Bid-Limit (in Aktien).
- Ask Price 2 (ASKp2) : Der Preis des zweitbesten Ask-Limits in USD·10000. Die Differenz zum Ask Price 1 entspricht nicht unbedingt der *ticksiz*.
- Ask Size 2 (ASKs2): Die Anzahl der Orders am zweiten nichtleeren Ask-Limit.
- ...

<sup>3</sup>Eine Ausführung einer Kauf (Verkaufs) Limit Order ist ein von einem Verkäufer (Käufer) initiiertes Handel und wird deshalb mit -1 (1) ausgewiesen.

<sup>4</sup>Die Variable Price wurde bereits auf Dollar Preise normiert

<sup>5</sup>Bei Aktien die einen Wert von 1 USD übersteigen, beträgt die *ticksiz* üblicherweise:  $\Delta_{tick} = 0.01$  USD

Wir geben nun die zum *Messagefile* in Tabelle 5.1 korrespondierenden Zeilen des *Orderbookfiles* an:

Die Anzahl der ausgewählten Preislimits (in beide Richtungen) beträgt 3: <sup>6</sup>

	ASKp1	ASKs1	BIDp1	BIDs1	ASKp2	ASKs2	BIDp2	BIDs2	ASKp3	ASKs3	BIDp3	BIDs3
1	22.37	4400	22.33	20000	22.38	1100	22.32	300	22.85	100	22.30	1400
2	22.37	4400	22.33	20000	22.38	1100	22.32	200	22.85	100	22.30	1400
3	22.36	1100	22.33	20000	22.37	4400	22.32	200	22.38	1100	22.30	1400
4	22.36	1800	22.33	20000	22.37	4400	22.32	200	22.38	1100	22.30	1400
5	22.36	1800	22.33	20000	22.37	3600	22.32	200	22.38	1100	22.30	1400
6	22.36	1800	22.33	20000	22.37	2900	22.32	200	22.38	1100	22.30	1400
7	22.36	1650	22.33	20000	22.37	2900	22.32	200	22.38	1100	22.30	1400

Tabelle 5.2: Ausschnitt des Orderbookfiles

Betrachtet man nun exemplarisch das Event mit der Zeilennummer 3 des *Messagefiles*: Es wird eine Limit Order zum Preis von 22.36 USD abgegeben. Die Ordergröße beträgt 1100 Stück an Aktien. Eine *TradeDirection* von  $-1$  signalisiert, dass dies eine Verkauf Limit Order ist und somit auf die Ask Seite im LOB kommt. Im dazugehörigem *Orderbookfile* in Tabelle 5.1 wirkt diese Order auf die Zeilennummer 2 und resultiert in der Zeilennummer 3. Hierbei wird in den *Bid Ask Spread* hineingesetzt und es gibt ein neuen *best Ask Price* von 22.36 USD. Das Ask Limit zum Preis von 22.37 wandert vom besten Ask Limit zum Zweitbesten. Anhand dieses Beispiels sieht man auch, dass sich die Ordergröße an einem Limit (hier Limit 2) ohne eine zu diesem Limit getätigte Order verändern kann.

Wie bereits erwähnt, beinhalten die beiden output Dateien von LOBSTER detaillierte Information für die reguläre Handelsperiode an der NASDAQ d.h. von 09:30 - 16:00 Uhr. Hierbei kommen natürlich einige Fragen auf:

*Woraus setzt sich der Anfangszustand des LOB um 09:30 zusammen?* Zwischen 07:00 und 09:30 Uhr ist es Händler an der NASDAQ erlaubt in einer Art Vorauktion Limit Orders abzugeben. Diese werden dann gesammelt und formen die Basis für den Anfangszustand des LOB.

*Gibt es einen Zusammenhang mit den verbliebenen Limit Orders im Orderbuch am Ende der regulären Handelsperiode um 16:00 und dem Anfangszustand am folgenden Tag?* Das LOB beziehungsweise die korrespondierende *Trading Plattform* startet jeden Tag um 03:00 Uhr mit einem leeren Zustand. Vom Vortag verbliebene Orders, welche theoretisch noch eine Gültigkeit am nächsten Tag besitzen würden z.B. Orders vom Typ *Good-till-Date*, müssen erneut abgegeben werden. Dies geschieht typischerweise noch vor der Vorauktion d.h. vor 07:30 Uhr. Diese Eigenschaft wird bei der empirischen Studie der Aktie Microsoft Corporation (siehe 6.2) eine zentrale Rolle spielen.

### 5.1.1 Begriffsunterscheidung Trade - Exekution im Kontext von LOBSTER

Das rekonstruierte LOB von LOBSTER zeigt wie bereits besprochen jede Exekution einer Limit Order als eigene Zeile im *Messagefile* an. Da das *Orderbookfile* zu den angegebenen Ordnern im

<sup>6</sup>Die Preise der Limits wurden bereits auf Dollar Preise normiert

*Messagefile* aktualisiert wird, führt somit auch jede einzelne Exekution einer Limit Order zu einer Aktualisierung des *Orderbookfiles*. Dies führt dazu, dass der Output eines Trades, womit wir im Folgenden den Abschluss einer in bestimmter Höhe abgegebenen Market Order bezeichnen, gegebenenfalls von der Zusammensetzung der am betrachteten Preislimit stehenden Limit Order beziehungsweise deren Größe abhängt. Wir verdeutlichen diesen Sachverhalt anhand des folgenden Beispiels:

Nehmen wir an ein Händler A gibt eine *Market Order* in der Höhe von 800 Aktien am Preislimit  $P_i$  ab. Wir betrachten folgende zwei Szenarien der bereits abgegeben Limit Order an diesem Preislimit:

1. Szenario: Es befindet sich eine einzelne Limit Order in der Höhe von 1500 Aktien an diesem Preislimit (nicht vom Typ *Fill-or-Kill*).
2. Szenario: Es befinden sich 4 verschiedenen Limit Order zu je 200 Aktien an diesem Preislimit.

Der Output des *Messagefiles* sieht in den beiden Szenarien folgendermaßen aus:

1. Szenario: Es wird eine einzelne Order vom Typ 4 in der Höhe von 800 Aktien ausgewiesen. Hierbei stimmen die Begriffe Exekution und Trade überein.
2. Szenario: Es werden 4 Order in 4 Zeilen vom Typ 4 in der Höhe von jeweils 200 Aktien mit dem gleichen Zeitstempel im *Messagefile* ausgewiesen. Jede einzelne der 4 bereits im Orderbuch angegebenen Limit Order wird gegen die Market Order ausgeführt.

Zusammenfassend behandelt das rekonstruierte LOB von LOBSTER jede einzelne Limit Order Exekution individuell und gibt diese in einer eigenen Zeile aus. Daher werden wir im Folgenden Begriffe wie *Trade* oder *Tradesize* immer in diesem Kontext sehen und als Synonym für eine einzelne Limit Order Exekution beziehungsweise deren Größe in Aktien verstehen.

## 5.2 Methodik der Schätzung im Modell I

In diesem Abschnitt erklären wir, wie man aus dem von LOBSTER erhaltenen *Message-* und *Orderbookfile* die Intensitäten  $\lambda_i^L$ ,  $\lambda_i^C$  und  $\lambda_i^M$  für den Markovprozess  $X$  im Modell I schätzt. In Abschnitt 3.2 haben wir ausführlich erläutert, wie man zu einer gegebenen Stichprobe für das Preislimit  $P_i$  bestehend aus Zeitdifferenz zum vorhergehenden Event, Ordertyp und Ordergröße vor der Order:

$$\{(S_{k+1}, X_{J_k}^i, \mathcal{T}(J_{k+1})) | k \in \{0, \dots, N\}\}$$

diese Intensitäten mittels der Maximum Likelihood Methode schätzt.

Daher fokussieren wir uns in diesem Abschnitt darauf, wie man aus den beiden Outputfiles von LOBSTER Stichproben für die relevanten Preislimits  $P_i$  erhält.

1. Im ersten Schritt fügen wir das *Messagefile* und das *Orderbookfile* spaltenweise zusammen. Dabei können wir die erste Zeile des Messagefiles vergessen, da diese zu keiner sichtbaren Veränderung im *Orderbookfile* korrespondiert. Des Weiteren ist die Spalte *Order ID* für unsere Analyse nicht von Bedeutung und kann ebenso vernachlässigt werden.
2. Im nächsten Schritt berechnen wir für jede Zeile des *Orderbookfiles* den Referenzpreis  $p_{ref}$ . Hierbei verwenden wir die in Kapitel 4 vorgestellte Methode. Nach diesen beiden Schritten

erhalten wir als Ergebnis folgende Tabelle:

ASKp1	ASKs1	BIDp1	BIDs1	ASKp2	ASKs2	BIDp2	BIDs2	ASKp3	ASKs3	BIDp3	BIDs3	Time	Type	Size	Price	TDir	pRef
22.37	4400	22.33	20000	22.38	1100	22.32	300	22.85	100	22.30	1400	34201.03	2	100.00	22.32	1	22.355
22.37	4400	22.33	20000	22.38	1100	22.32	200	22.85	100	22.30	1400	34201.22	1	1100.00	22.36	-1	22.355
22.36	1100	22.33	20000	22.37	4400	22.32	200	22.38	1100	22.30	1400	34201.22	1	700.00	22.36	-1	22.345
22.36	1800	22.33	20000	22.37	4400	22.32	200	22.38	1100	22.30	1400	34201.30	3	800.00	22.37	-1	22.345
22.36	1800	22.33	20000	22.37	3600	22.32	200	22.38	1100	22.30	1400	34201.30	3	700.00	22.37	-1	22.345
22.36	1800	22.33	20000	22.37	2900	22.32	200	22.38	1100	22.30	1400	34201.51	4	150.00	22.36	-1	22.345

Tabelle 5.3: *Orderbook-* und *Messagefile* mit Referenzpreis

- Als nächstes bestimmen für jede Zeile, zu welchem Preislimit  $P_i$  aus unserem Modell die angegebene Order korrespondiert. Dies ist eindeutig durch den Abstand zum Referenzpreis bestimmt. Der Preis der ersten Order (22.32 USD) ist um 0.035 USD kleiner als der berechnete Referenzpreis (22.355 USD). Demnach korrespondiert diese Order zum Preislimit  $P_{-4}$ . Da wir uns bei unserer Analyse auf die ersten drei Preislimits beschränken, können wir diese Order vernachlässigen. Für die zweite Order gilt, dass deren Preis um eine halbe *ticks* größer ist als der momentane Referenzpreis. Daher wird sie dem Limit  $P_1$  zugeordnet. Wie bereits besprochen, wird durch diese Order eine Änderung des Referenzpreises hervorgerufen. Auf analoge Art und Weise lassen sich die restlichen Order den Preislimits aus unserem Modell zuordnen.
- Nachdem wir die Order den Preislimits zugeordnet haben, bestimmen wir die Anzahl der Orders vor dem jeweiligem Event  $X_{J_k}^i$ . Dabei gehen wir wie folgt vor:

Zuerst definieren wir als Hilfsgröße  $\tilde{A}_i$ , welche die tatsächliche Anzahl der Orders am Preislimit  $P_i$  im LOB von LOBSTER in Aktien beschreibt. Entspricht der Preis der Order einem nichtleeren angezeigten Limit im *Orderbookfile*, dann bestimmt sich die Anzahl der Orders in Aktien vor diesem Event als das zugehörige angezeigte Orderbuchvolumen  $\tilde{A}_i(J_k)$ . Falls der Preis der Order keinem angezeigten Preislimit im *Orderbookfile* gleicht, dann war das Limit leer und  $\tilde{A}_i(J_k) = 0$ . Der Preis der zweiten Order (22.36 USD) gleicht keinem aktuell angezeigten Preis des *Orderbookfiles*. Demnach ist  $\tilde{A}_1(J_1) = 0$  (in Aktien). Nachdem die Anzahl der Orders vor dem Event für alle Orders bestimmt wurde, wird diese Größe mittels  $AES_i$  skaliert.  $AES_i$  bezeichnet hierbei *Average Event Size* am Preislimit  $P_i$ . Dies wird als der empirische Mittelwert über alle Ordergrößen (Size) am jeweiligen Preislimit berechnet. Die Normierung erfolgt auf folgende Art und Weise <sup>7</sup>:

$$X_{J_k}^i = \left\lceil \frac{\tilde{A}_i(J_k)}{AES_i} \right\rceil \in \mathbb{N}_0$$

- Als letztes bestimmen wir den Ordertyp und die Zeitdifferenz zur vorhergehenden Order<sup>8</sup> an dem zugehörigen Preislimit. Diese beiden Kenngrößen lassen sich aus dem *Messagefile* direkt ablesen.

<sup>7</sup> $\lceil \cdot \rceil$  beschreibt hierbei das Aufrundungsfunktion und ist definiert als  $\lceil x \rceil := \min\{k \in \mathbb{N} : k \geq x\}$

<sup>8</sup>Für die erste Order wurde als Zeitdifferenz der Abstand zum Beginn der regulären Handelsperiode: 09:30 (34200 Sek.) gewählt.

Aufgrund der Symmetrieeigenschaft von Modell I werden die erhaltenen Stichproben der Preislimits  $P_i$  und  $P_{-i}$  vereinigt. Falls wir annehmen, dass  $AES_{\pm 1} = AES_{\pm 2} = AES_{\pm 3} = 200$  (in Aktien), erhalten wir zusammenfassend folgende drei Stichproben für die betrachteten Preislimits  $P_{\pm i}, i \in \{1, 2, 3\}$ .

Timediff	Lastsize	Type	Lastsize_AES
1.2263459	0	1	0

Tabelle 5.4: Preislimit:  $P_{\pm 1}$ 

Timediff	Lastsize	Type	Lastsize_AES
1.2263459	1100	1	6
0.2833028	1800	4	9

Tabelle 5.5: Preislimit:  $P_{\pm 2}$ 

Timediff	Lastsize	Type	Lastsize_AES
1.3058759	4400	3	22
0.0016620	3600	3	18

Tabelle 5.6: Preislimit:  $P_{\pm 3}$ 

Die Funktion *getSample1()* aus Kapitel 9, (2) berechnet zu gegebenen LOBSTER Daten die gewünschte Stichprobe nach der oben vorgestellten Methode.

Bei der oben erwähnten Methode für den Erhalt einer Stichprobe, hat eine Referenzpreisänderung keinen Einfluss. Wir betrachteten bei unserer Analyse noch zwei weitere Methoden für den Erhalt einer Stichprobe:

Bei der zweiten Methode wird in Anlehnung an [6] bei einer Referenzpreisänderung dieselbe Prozedur (Schritte 1 – 5) jedes mal neu gestartet. Dies hat zur Folge, dass wir die Daten in Blöcke mit konstantem Referenzpreis unterteilen und die Zeitdifferenz der Order, im Bezug auf den betrachteten Block, berechnet wird. Für die erste Order pro Preislimit wurde als Zeitdifferenz der empirische Mittelwert der Zeitdifferenzen des Blockes herangezogen. Falls ein Block nur aus einer Zeile besteht, wurde diese Order nicht betrachtet.

Die Funktion *getSample2()* aus Kapitel 9, (3) berechnet mit der zweiten Methode zu gegebenen LOBSTER Daten die gewünschte Stichprobe

Bei der dritten Methode unterteilen wir, analog wie in Methode 2, die Daten in Blöcke mit konstantem Referenzpreis. Die Variable Zeitdifferenz wird nun analog zur Methode 2 berechnet. Überdies löschen wir jedoch alle Order/Zeilen aus den Daten, welche eine Referenzpreisänderung hervorrufen, und kumulieren die Ordergröße von Ordnern, welche den gleichen Zeitstempel, den gleichen Ordertyp, den gleichen Orderpreis und die gleiche *Trade Direction* aufweisen (vgl. Abschnitt 5.1.1).

Wir erhielten für alle drei Methode sehr ähnliche Ergebnisse im Bezug auf die Intensitäten, weshalb wir uns bei der weiteren Analyse auf die erste Methode beschränken werden.

# Kapitel 6

## Empirische Studie Modell I - LOBSTER Daten

In diesem Kapitel präsentieren wir verschiedene Ergebnisse bezüglich der Modellierung des Orderbuches einer Aktie mit Modell I. Für das folgende Kapitel wählen wir für unsere Analyse zuerst in Anlehnung an [6] die Aktie France Telekom (FTE) aus. Des Weiteren betrachten wir auch die Aktie Microsoft Corporation (MSFT). Die Daten beziehen wir von der Limit Order Buch Datenbank LOBSTER, welche ein rekonstruiertes Orderbuch der NASDAQ zur Verfügung stellt (siehe Kapitel 5). Für beide Beispiele betrachten wir jeweils die ersten drei Preislimits d.h.  $K := 3$  beziehungsweise  $X_t = (X_t^{-3}, X_t^{-2}, X_t^{-1}, X_t^1, X_t^2, X_t^3) \in \mathbb{N}_0^6$ .

### 6.1 France Telekom - FTE

Für die Aktie France Telekom betrachten wir den Zeitraum: 27.04.2010 - 30.03.2012<sup>1</sup>. Aufgrund der fehlenden Information über die Ordergröße vor einem Event  $X_{J_k}$  wird für die folgende Analyse keine Order vom Typ 5 (Ausführung einer versteckten Limit Order) betrachtet. Des Weiteren werden aufgrund der bereits besprochenen Redundanz alle Order vom Typ 6 vernachlässigt.

Da die Daten von LOBSTER eine feinere Unterteilung der Ordertypen als im Modell I betrachtet aufweisen, treffen wir folgende Unterscheidung bezüglich der Ordertypen in den Daten im Bezug auf Modell I:

LOBSTER Datenbank	Modell I
1	Limit Order
2 und 3	Cancellation Order
4	Market Order

Als ersten Schritt extrahieren wir, wie in Abschnitt 5.2 ausführlich erläutert, aus dem gesamten Datensatz die relevanten Daten für die betrachteten Preislimits  $P_{\pm i}, i \in \{1, 2, 3\}$ . Wir erhielten folgende Kenngrößen zu den betrachteten Daten:

<sup>1</sup>Dieser Zeitraum wurde in Anlehnung an [6] gewählt. Da jedoch im Zeitraum von 1.1.2010 bis 26.04.2010 nur sehr wenige Order (< 100) pro Tag abgegeben wurden, haben wir uns dazu entschlossen diesen Zeitraum nicht zu betrachten.

- Average Event Size (AES):

Aktie	$AES_1$	$AES_{-1}$	$AES_2$	$AES_{-2}$	$AES_3$	$AES_{-3}$
FTE	214.34	198.47	445.56	421.58	595.23	573.95

Hierbei fällt auf, dass die erhaltenen Werte deutlich kleiner sind als in [6, Table 2.]. Dies kann einerseits damit zusammenhängen, dass in [6] Daten von der Börse *Euronext Paris* verwendet wurden, andererseits, dass sich die Struktur des Orderbuches von LOBSTER von der in [6] verwendeten *Chevreux LOB database* unterscheidet. Im Speziellen sei hierbei der Unterschied von Trades und der Exekution einer Limit Order erwähnt (siehe 5.1.1). Diese individuelle Betrachtungsweise eines *Trades* führt im Gegensatz zu einer kumulierten Darstellung zu einer Verkleinerung der *Average Event Size*. Da lediglich 0.5% der Order in den gesammelten Stichproben mit dem gleichen Zeitstempel ausgewiesen waren und somit von einem kumulierten Trade stammen, ist dieser Effekt jedoch vernachlässigbar.

- Sonstige Kenngrößen:

Aktie	$\Delta_{tick}$	$\varnothing$ Bid-Ask-Spread	$\varnothing$ Anzahl Orders/Tag	$\varnothing$ Anzahl Trades (Typ 4 Order) /Tag	$\varnothing$ Frequenz Referenzpreisänderungen
FTE	0.01 \$	1.40 Ticks	58448	959	61 Order

Der durchschnittliche *Bid-Ask-Spread* ist fast ident zu [6, Table 1]. In der obigen Tabelle ist zu erkennen, dass sich an der NASDAQ im Vergleich zur *Euronext Paris* weniger Orders beziehungsweise Trades pro Tag ereignet haben.

Die folgenden Grafiken zeigen exemplarisch die Entwicklung des Orderbuchvolumens aller 6 Preislimits im Zeitraum von 10.05.2010 - 24.05.2010<sup>2</sup>.

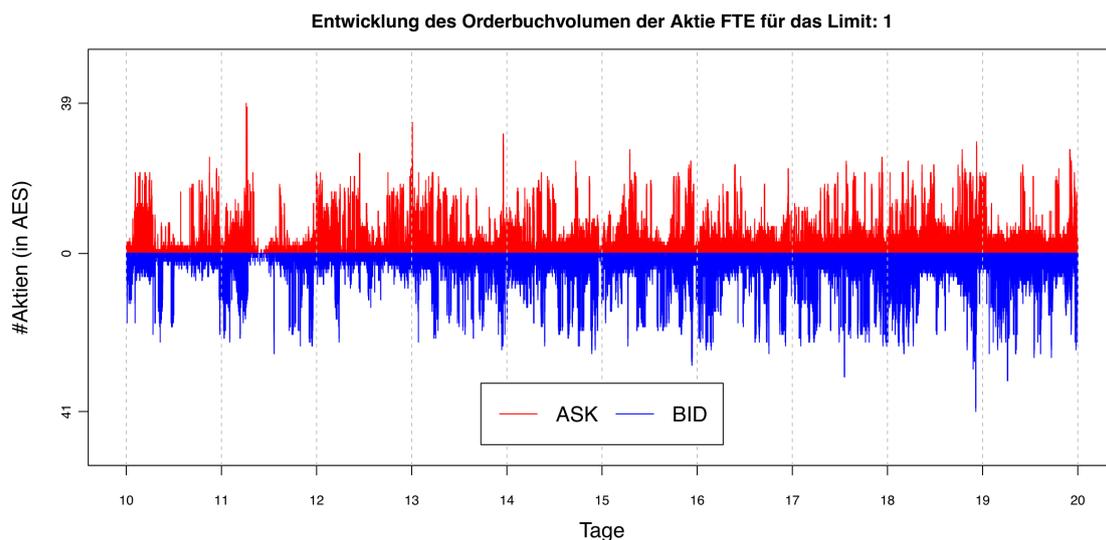


Abbildung 6.1: Orderbuchentwicklung FTE Limit 1: 10.05.2010 - 24.05.2010

<sup>2</sup>Bei diesen Grafiken wurde aufgrund der besseren Lesbarkeit Ausreißer über dem 99%-Quantil nicht berücksichtigt.

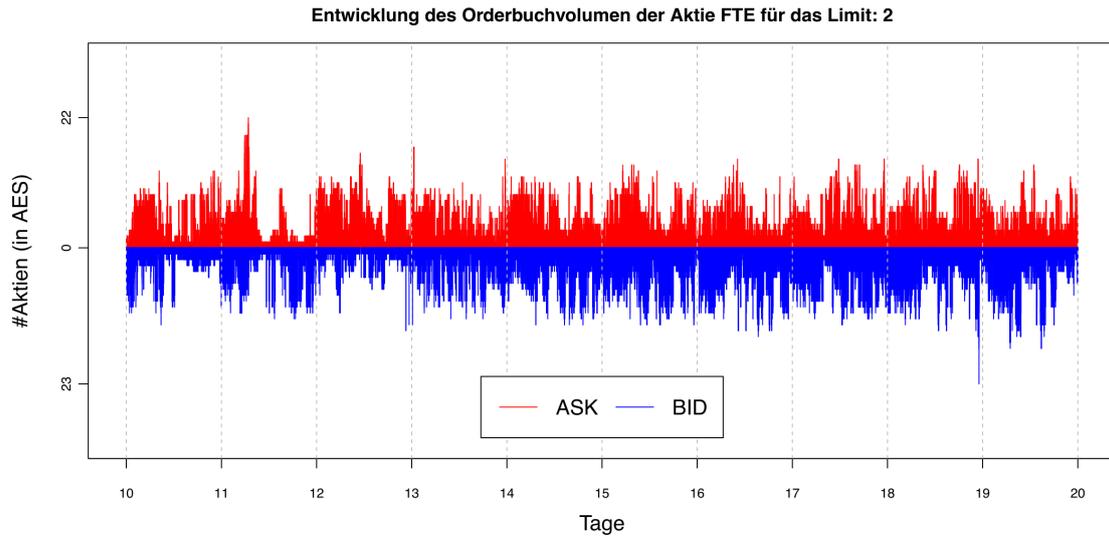


Abbildung 6.2: Orderbuchentwicklung FTE Limit 2: 10.05.2010 - 24.05.2010

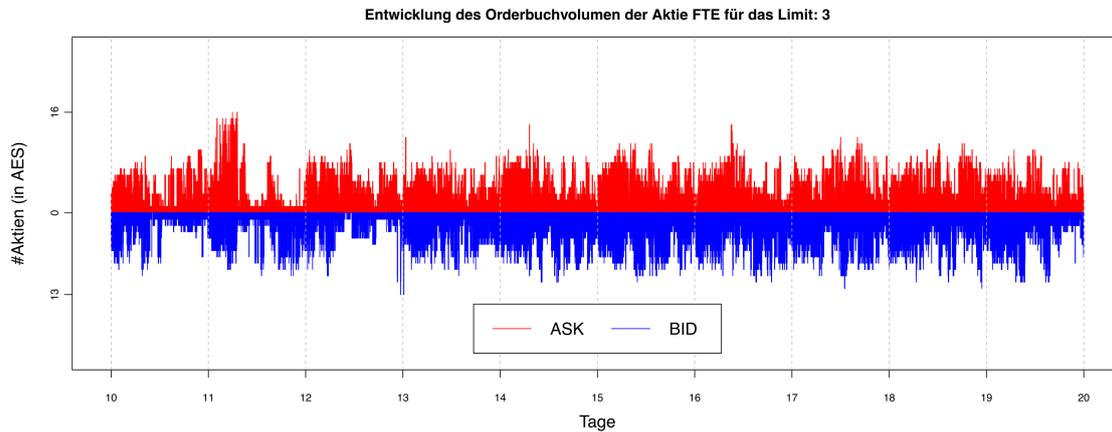


Abbildung 6.3: Orderbuchentwicklung FTE Limit 3: 10.05.2010 - 24.05.2010

Für die Schätzung der Intensitäten im Modell I  $\lambda_i^L(\cdot)$ ,  $\lambda_i^M(\cdot)$  und  $\lambda_i^C(\cdot)$  verwenden wir die in Abschnitt 3.2 beschriebenen Maximum Likelihood Schätzer (siehe (3.3) - (3.5)). Die Ergebnisse der Schätzung bis 35 AES pro Preislimit werden in Abbildung 6.4 präsentiert. In dieser Abbildung ist ersichtlich, dass die geschätzten Intensitäten für die Preislimits  $P_{\pm i}$  im Wesentlichen dieselbe Charakteristik aufweisen. Daher ist die in Modell I angenommene Symmetrieannahme durchaus plausibel. Wir werden somit im Folgenden die Stichproben für die Preislimits  $P_{\pm i}$  wie in [6] zusammenführen und bei der Schätzung der Intensitäten nicht in ASK und BID unterscheiden d.h.  $\lambda_i^k(\cdot) = \lambda_{-i}^k(\cdot)$ ,  $(k, i) \in \{L, C, M\} \times \{-3, -2, -1, 1, 2, 3\}$ . Das Ergebnis der Schätzung mit zusammengeführten Daten wird in Abbildung 6.5 dargestellt. Konfidenzintervalle wurden mittels (3.7) berechnet und werden als punktierte Linien dargestellt.

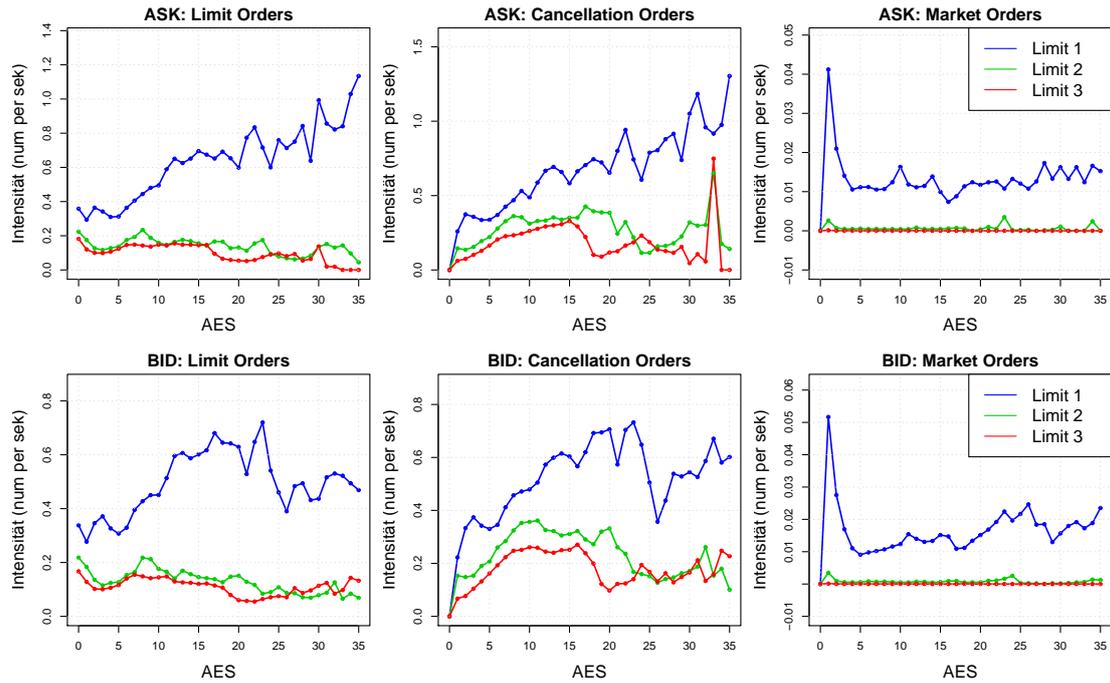


Abbildung 6.4: Intensitätsschätzung FTE: ASK/BID getrennt

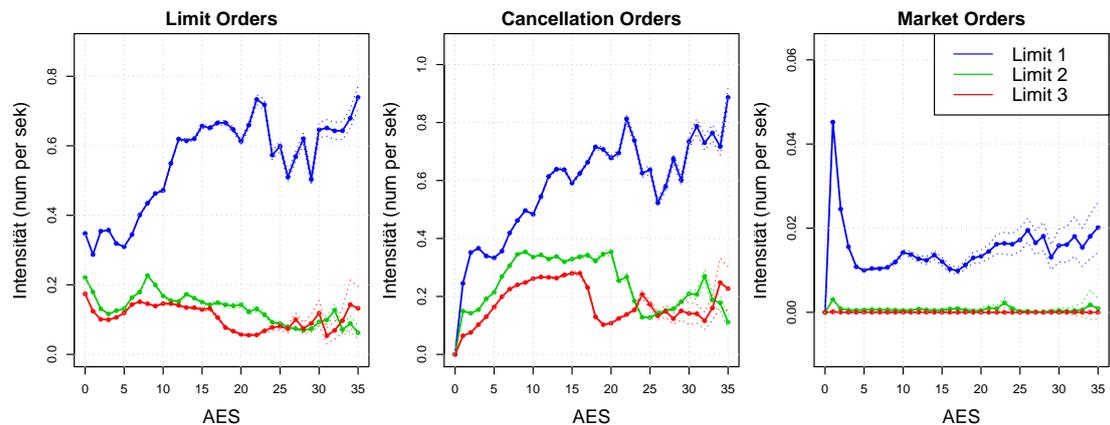


Abbildung 6.5: Intensitätsschätzung FTE

Wir kommentieren kurz die erhaltenen Ergebnisse von Abbildung 6.5 und deren Unterschiede zu [6, Figure 2.]. Für ökonomische Interpretationen sei auf [6] verwiesen:

- Limit Order:
  - Limit 1:  $\lambda_1^L(\cdot)$  ist ähnlich wie in [6] eine leicht steigende Funktion von dem vorhandenen Orderbuchvolumen in AES. Hierbei tritt im Gegensatz zu [6] kein signifikant geringerer Wert bei 0 AES auf.
  - Limit 2 / Limit 3:  $\lambda_{2,3}^L(\cdot)$  sind leicht fallende Funktionen im Bezug auf das Orderbuchvolumen. Diese weisen beide geringere Intensitäten als für das Limit 1 auf. Des Weiteren

haben diese im Gegensatz zu [6] keine eklatant größeren Werte bei einem vorhandenen Orderbuchvolumen von 0 AES.

- Cancellation Order:

- Limit 1:  $\lambda_1^C(\cdot)$  ist ähnlich wie in [6] bis 23 AES ungefähr eine konkav steigende Funktion vom Orderbuchvolumen. Auch der Wertebereich dieser stimmt mit [6] überein. Für AES größer als 23 ist im Gegensatz zu [6] kein klarer Trend erkennbar, was natürlich auch an zu wenig vorhandenen Daten zu diesem gegebenen Orderbuchvolumen liegen kann.
- Limit 2 / Limit 3:  $\lambda_{2,3}^C(\cdot)$  sind bis 17 AES ebenso konkav steigende Funktion wie die Cancellation Intensität für das erste Limit. Ab 17 AES ist bei beiden ein leicht fallender Trend zu erkennen. Die geschätzte Intensität für das Limit 3 ist im Gegensatz zu [6] weder eine linear steigende Funktion noch weist diese einen signifikant größeren Wert bei 1 AES auf.

- Market Order

- Limit 1:  $\lambda_1^M(\cdot)$  ist wie in [6] eine exponentiell fallende Funktion vom Orderbuchvolumen, deren Wertebereich jedoch kleiner ist als in [6].
- Limit 2 / Limit 3: Eine Market Order am Limit 2 oder 3 kann in der Praxis nur auftreten, falls die Preislimits, welche sich näher am Referenzpreis befinden, leer sind. Dies passiert, wie man an dem durchschnittlichen *Bid-Ask-Spread* von 1.40 sehen kann, eher selten. Deshalb sind die erhaltenen Werte sehr klein.

Mittels den geschätzten Intensitäten können nun auch Annahme 2 und 3 visuell überprüft werden. In Abbildung 6.6 und 6.7 ist ersichtlich, dass die beiden Regularitätsvoraussetzungen, welche eine invariante Verteilung garantieren, für die Aktie France Telekom erfüllt sind.

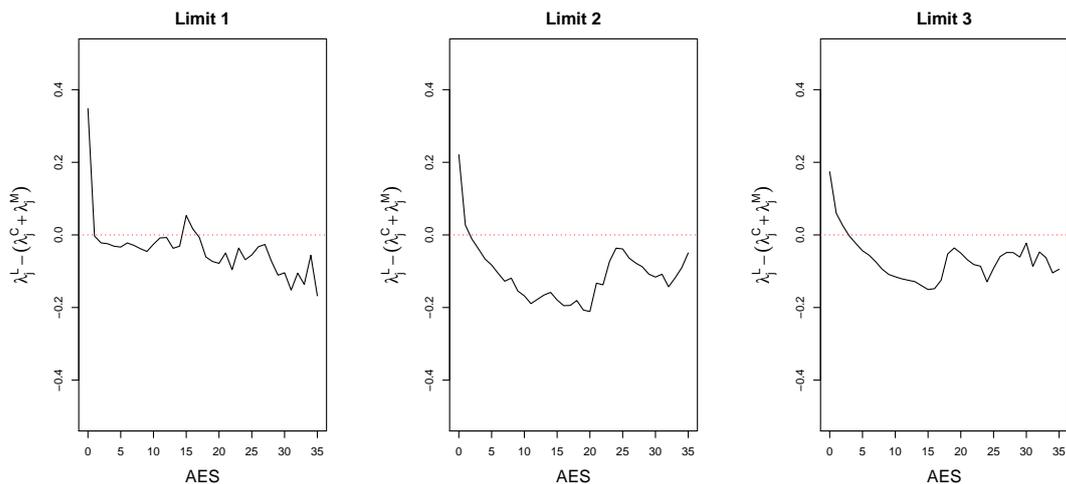


Abbildung 6.6: Überprüfung von Annahme 2 FTE

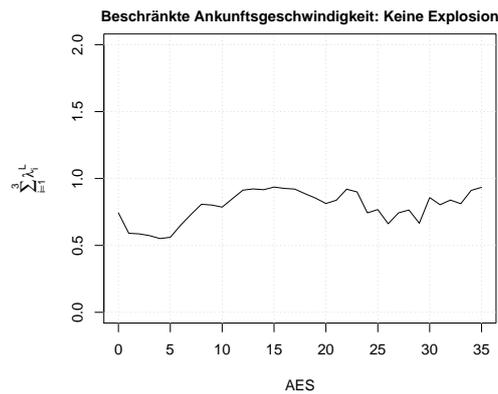
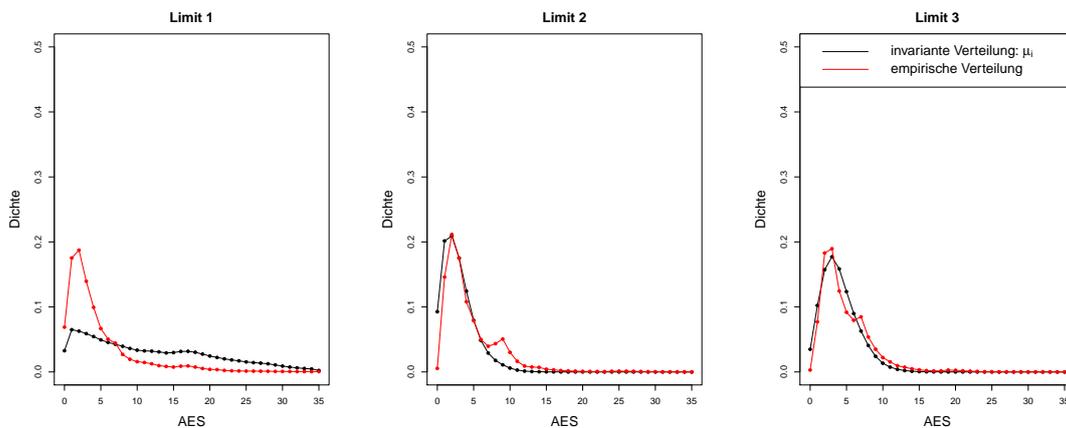


Abbildung 6.7: Überprüfung von Annahme 3 FTE

Als nächstes vergleichen wir die theoretische invariante Verteilung aus Modell I mit der empirischen Verteilung der Daten. Hierbei muss man aufpassen, dass man die richtigen Verteilungen miteinander vergleicht. Die Daten von LOBSTER zur Aktie France Telecom beinhalten für  $i \in \{1, 2, 3\}$  die Werte des mit  $X^i$  assoziierten Sprungprozesses  $Z^i$ . Schätzt man aus der nach Abschnitt 5.2 erhaltenen Stichprobe zu den LOBSTER Daten die empirischen Häufigkeiten, so muss man die erhaltene empirische Verteilung mit der invarianten Verteilung  $\mu_i$  des Sprungprozesses  $Z_i$  aus Theorem 3.3 vergleichen und nicht mit der invarianten Verteilung  $\pi_i$  des stetigen Markovprozesses  $X^i$ .

Andererseits kann man aus den Sprungdaten von LOBSTER zu einer gegebenen festen Beobachtungsfrequenz (engl. sampling frequency) z.B. 30 Sekunden jeweils den Wert des Markovprozesses  $X^i$  vor diesem festen Zeitpunkt als Grundlage für die empirischen Häufigkeiten heranziehen. Dies liefert eine empirische Verteilung des zeitstetigen Markovprozesses  $X^i$  und kann mit der invarianten Verteilung  $\pi_i$  aus Theorem 3.2 verglichen werden. Die Rechtfertigung dieser Schätzmethoden findet sich im Anhang im Abschnitt A.4.

Im Folgenden werden die erhaltenen Ergebnisse zu beiden Methoden in den Abbildungen 6.8 und 6.9 verglichen. Bei der Letzteren wurde eine Beobachtungsfrequenz von 30 Sekunden gewählt.

Abbildung 6.8: FTE Vergleich: empirische und invariante Verteilung  $\mu_i$  von  $Z$

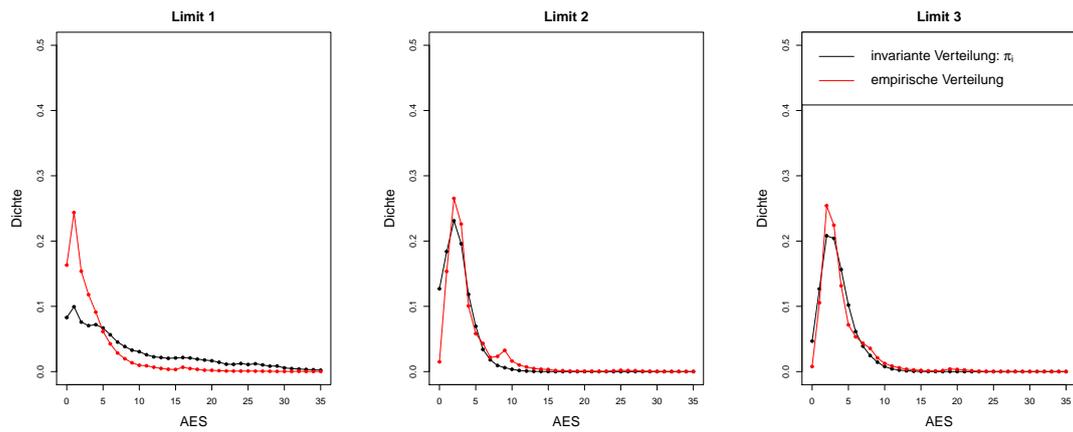


Abbildung 6.9: FTE Vergleich: empirische und invariante Verteilung  $\pi_i$  von  $X$

Hierbei wird deutlich, dass die invarianten Verteilungen  $\pi$  und  $\mu$  aus Modell I für die Preislimits 2 und 3 eine sehr gute Approximation liefern. Lediglich bei Limit 1 weichen diese von der empirischen Verteilung für AES Werte zwischen 1 und 4 von der empirischen Verteilung ab. Zusammenfassend lässt sich die Dynamik des Orderbuchvolumens der Aktie France Telekom in Zeiten eines konstanten Referenzpreises mittels Modell I relativ gut beschreiben. Als nächstes wenden wir uns der Aktie *Microsoft Corporation* zu.

## 6.2 Microsoft Corporation - MSFT

Für die Aktie Microsoft Corporation (MSFT) wählten wir den Zeitraum von 01.09.2017 bis 05.05.2017 aus. Die Analyse der Aktie MSFT geschieht auf analoge Art und Weise wie zuvor.

Wir erhielten folgende Kenngrößen:

- Average Event Size (AES):

Aktie	$AES_1$	$AES_{-1}$	$AES_2$	$AES_{-2}$	$AES_3$	$AES_{-3}$
MSFT	203.54	195.50	208.69	199.04	247.31	255.21

Hierbei wird ersichtlich, dass die erhaltenen Werte für die Preislimits  $P_{\pm 2}$  und  $P_{\pm 3}$  deutlich kleiner sind als wie bei der Aktie France Telekom.

- Sonstige Kenngrößen:

Aktie	$\Delta_{tick}$	$\varnothing Bid-Ask-Spread$	$\varnothing$ Anzahl Orders/Tag	$\varnothing$ Anzahl Trades (Typ 4 Order) /Tag	$\varnothing$ Frequenz Referenzpreisänderungen
MSFT	0.01 \$	1.12 Ticks	582260	21775	232 Order

Anhand der Werte ( $\approx 580000$  Orders/Tag und  $\approx 21000$  Trades/Tag) wird deutlich, dass die Aktivität im LOB für die Aktie MSFT viel größer ist als bei FTE.

Die Intensitätsschätzung<sup>3</sup> lieferte für  $\lambda_i^L(n)$ ,  $\lambda_i^C(n)$  und  $\lambda_i^M(n)$  mit  $i \in \{1, 2, 3\}$  und  $n \in \{0, \dots, 50\}$  folgende Werte:

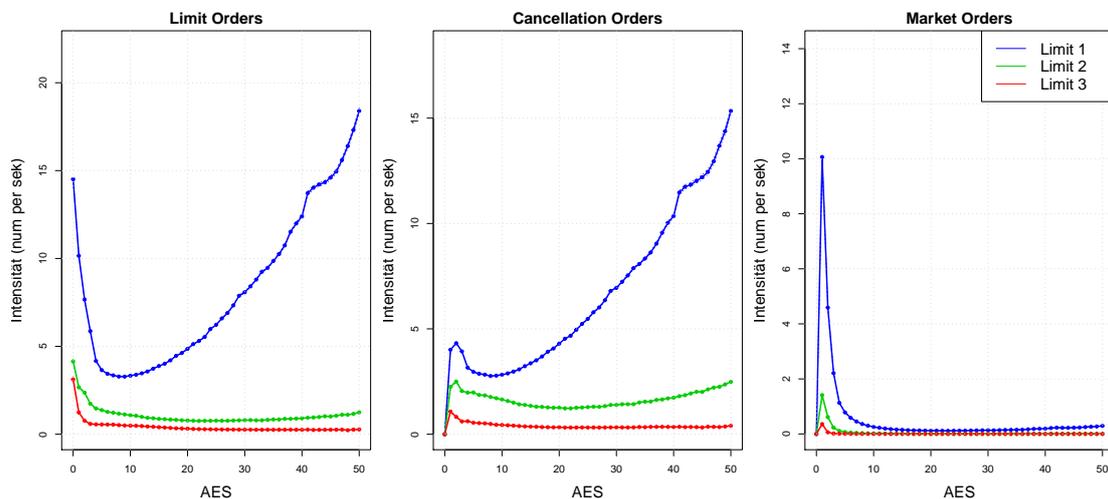


Abbildung 6.10: Intensitätsschätzung MSFT

In Abbildung 6.10 sieht man, dass die Intensitäten generell größer sind als jene der Aktie France Telekom. Dies lässt sich durch die bereits angesprochene größere Anzahl von Order und Trades pro Tag erklären. Des Weiteren fällt auf, dass die Intensitäten teilweise eine deutlich andere Charakteristik als jene der Aktie France Telekom aufweisen. Während die Intensitäten einer Market

<sup>3</sup>Die Daten für die Preislimits  $P_{\pm i}$  wurden aufgrund der Symmetrieannahme im Modell I bereits zusammengeführt

Order ähnlich wie bei FTE exponentiell fallend (Limit 1) beziehungsweise fast ident 0 (Limit 2 und 3) sind, unterscheiden sich die Intensitäten von Limit und Cancellation Order im Vergleich zu FTE stark. Die Intensität einer Limit Order am Preislimit  $P_{\pm 1}$  ist ungefähr eine quadratische Funktion, die ihr Minimum in der Nähe von 10 AES annimmt. Das heißt Marktteilnehmer geben insbesondere bei kleinem Orderbuchvolumen (0-3 AES) und bei sehr großen Volummen ( $\geq 30$  AES) vermehrt Limit Orders ab. Ebenso wie die Limit Order Intensität ist die Intensität einer Cancellation Order am Preislimit  $P_{\pm 1}$  ab circa 10 AES wachsend, jedoch vom Wertebereich her kleiner als die korrespondierende Limit Order Intensität.

Als nächstes betrachten wir für die erhaltenen geschätzten Intensitäten die beiden Regularitätsvoraussetzungen (Annahme 2 und 3). In Abbildung 6.11 ist klar ersichtlich, dass Annahme 2 für das Preislimit  $P_1$

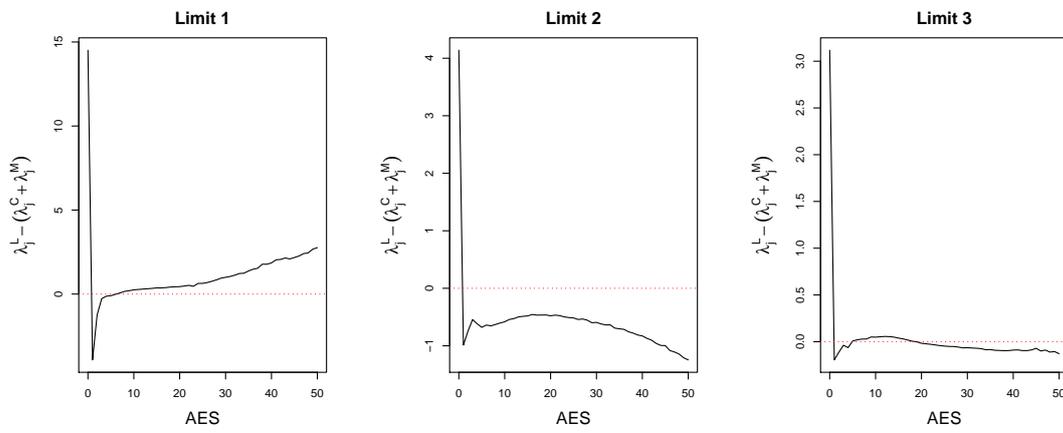


Abbildung 6.11: Überprüfung von Annahme 2 MSFT

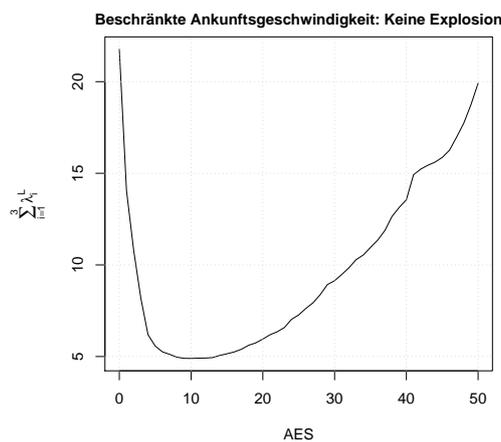


Abbildung 6.12: Überprüfung von Annahme 3 MSFT

nicht erfüllt ist. Hierbei wird deutlich, was wir bei den geschätzten Intensitäten bereits beobachtet haben:

Die Intensität eines Sprunges der Höhe +1 (Limit Order Intensität) wächst stärker als die Intensität eines Sprunges der Höhe -1 (Cancellation Order + Market Order Intensität). Somit tendiert das

Orderbuchvolumen dazu, stets zu wachsen. Dies hat zur Folge, dass der Markovprozess  $(X_t^{\pm 1})_{t \geq 0}$  keine invariante Verteilung  $\pi_1$  hat, beziehungsweise diese, sofern man genügend Daten für große Werte zur Verfügung hat und annimmt, dass die Intensitäten sich für AES größer als 50 analog verhalten, degeneriert ist.

Berechnen wir dennoch die invariante Verteilung  $\pi_i$  für  $i \in \{1, 2, 3\}$  mit Formel (3.8) - (3.9) aus Theorem 3.2 und vergleichen diese mit der empirischen Verteilung aus den Daten,<sup>4</sup> so erhalten wir folgendes Ergebnis:

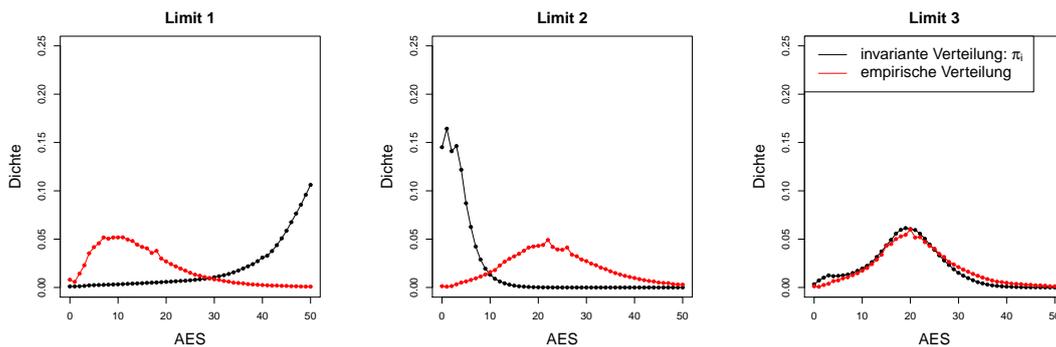


Abbildung 6.13: FTE Vergleich: empirische und invariante Verteilung  $\pi_i$  von  $X$

Der Vergleich der empirischen Verteilung aus den Sprungdaten zu der invarianten Verteilung  $\mu_i$  aus Theorem 3.3 des mit  $X$  assoziierten Sprungprozesses  $Z$  liefert folgendes Ergebnis:

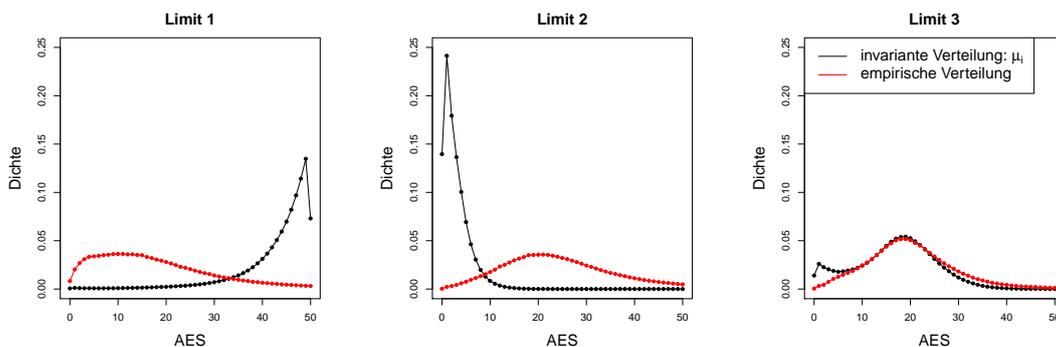


Abbildung 6.14: FTE Vergleich: empirische und invariante Verteilung  $\mu_i$  von  $Z$

Anhand von Abbildung 6.13 und Abbildung 6.14 ist deutlich zu erkennen, dass unser Modell keine zufriedenstellende Ergebnisse für die Preislimits  $P_{\pm 1}$  und  $P_{\pm 2}$  liefert.

Wir werden nun die Problematik für diese beiden Preislimits näher analysieren. Wir beginnen mit den Preislimits  $P_{\pm 1}$ :

<sup>4</sup>Es wurde eine Beobachtungsfrequenz von 30 Sekunden gewählt

Zuerst zeigen wir in Abbildung 6.15 eine exemplarische Darstellung des Orderbuchvolumen der Preislimits  $P_{\pm 1}$  für die ersten 5 betrachteten Tage (9.01.2017-13.01.2017). Dabei haben wir die Größen bereits mit den  $AES_i$  Werten skaliert und Ausreißer ( $> 99\%$  Quantil) zur besseren visuellen Darstellung eliminiert.

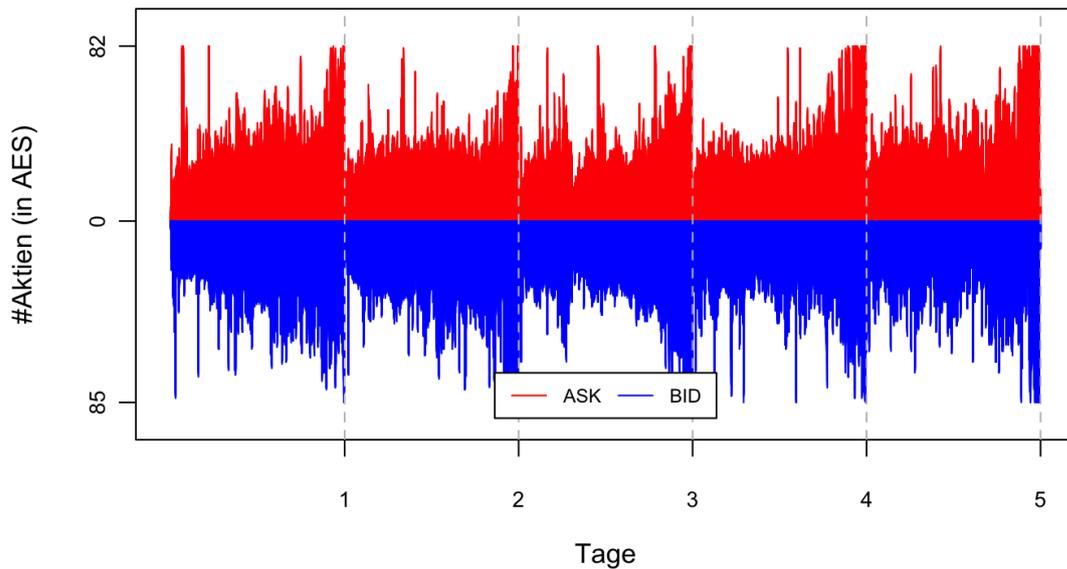


Abbildung 6.15: Orderbuchentwicklung MSFT Limit 1: 10.05.2010 - 24.05.2010

Daran sieht man, dass die Daten einen zyklischen Verlauf widerspiegeln. Das Orderbuchvolumen startet niedrig und nimmt über den Tag betrachtet zu, bis es schließlich wieder zurückgesetzt wird. Dieses Zurücksetzen des LOB wurde bereits in Kapitel 5 diskutiert. Das LOB an der NASDAQ startet jeden Tag mit einem leeren Zustand, ehe es schließlich um 9:30 mit einem in der Vorauktion ermittelten Zustand für den regulären Handel geöffnet wird. Diesen exogenen Einfluss beinhaltet unser Modell rund um dem Markovprozesses  $X^{\pm 1}$  nicht. Dies kann ein möglicher Grund sein, warum es in Abbildung 6.13 und Abbildung 6.14 beim Vergleich von empirischer und invarianter Verteilung zu eklatant falschen Ergebnissen kommt.

Das durchschnittliche Startvolumen und Endvolumen des LOB der Aktie MSFT am Preislimit 1 ist gegeben durch:

	ASK1	BID1
$\varnothing$ Vol. Start (AES)	9.5	1.64
$\varnothing$ Vol. Ende (AES)	73.38	83.65

Die erhaltenen Werte stützen unsere These eines zyklischen Verlaufs.

Im Folgenden simulieren wir nun mithilfe von Methode I aus Abschnitt 7.1.1 beziehungsweise dem Pseudocode aus Abschnitt 7.2 den Markovprozess  $X^1$  bis zu einer maximalen Größe von 100 AES und setzen diesen nach einer gewissen Anzahl an Order  $\Delta_0$  wieder zurück auf  $X_0$ . Als Inten-

sitäten verwenden wir die geschätzten Intensitäten aus den Daten bis 100 AES<sup>5</sup>. Wir haben zwei Varianten betrachtet, bei denen wir jeweils insgesamt 8 Millionen Orders simuliert haben:

- Variante 1:  $\Delta_0 = 10^5$  und  $X_0 \sim U\{0, \dots, 10\}$
- Variante 2:  $\Delta_0 = 500$  und  $X_0 \sim U\{5, \dots, 15\}$

Die Ergebnisse werden in der folgenden Abbildung mit der empirischen und invarianten Verteilung aus Abbildung 6.14 verglichen.

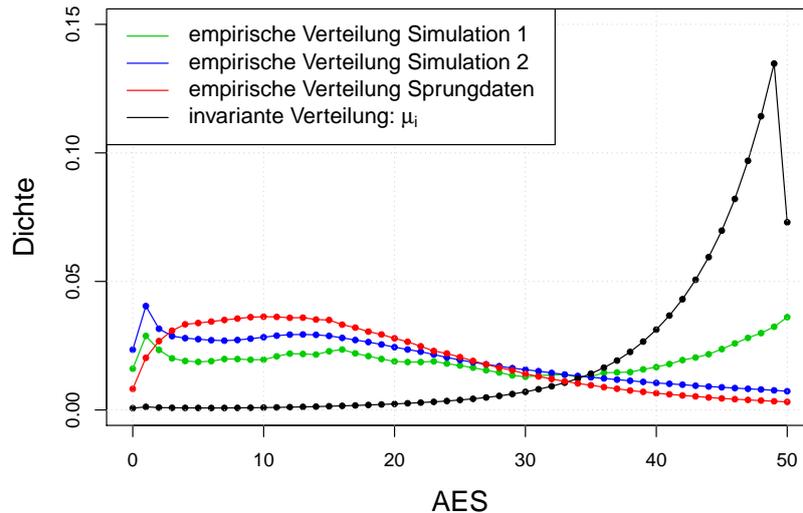


Abbildung 6.16: Nachbildung der empirischen Verteilung des Preislimits 1 mittels Simulation

Hierbei ist ersichtlich, dass wir durch das künstliche zurücksetzen des Markovprozesses eine akzeptable Approximation der empirischen Verteilung bekommen. Variante 1 (alle 100000 Orders zurücksetzen) entspricht circa einem Tag an der NASDAQ in dem betrachteten Zeitraum. Dabei sieht man jedoch, dass alleine durch ein Zurücksetzen am Beginn jeden Tages die empirische Verteilung nicht gänzlich erklärt werden kann, da diese sich für AES größer als 35 nicht analog verhält. Daraus lässt sich schließen, dass wir mehrmals pro Tag den Prozess zurücksetzen müssen, um die empirische Verteilung der Daten erklären zu können. Dies geschah in Variante 2 (alle 500 Orders), welche eine bessere Approximation der zugrunde liegenden empirischen Verteilung aus den Daten liefert.

Diese Analyse verdeutlicht einen möglichen Problemfaktor bezüglich der Aktie MSFT und den Preislimits  $P_{\pm 1}$ . Unsere Vermutung, warum das Modell für dieses Preislimit nicht gut funktioniert, ist zusammenfassend, dass das Orderbuchvolumen mehrmals über den Tag hinweg abrupt zurückgesetzt wird beziehungsweise plötzlich zu kleineren Werten springt. Dies kann nun einerseits durch eine große Cancellation Order oder Market Order passieren oder andererseits, indem eine neue kleinere Order in den *Bid Ask Spread* gesetzt wird. Dieses Zurücksetzen verhindert, dass das Orderbuchvolumen in der Realität explodiert, was mit unserem geschätzten Modell eigentlich der

<sup>5</sup>Diese verhalten sich bis 80 AES analog wie die präsentierten Intensitäten bis 50 AES aus Abbildung 6.10 und fangen danach, aufgrund von zu wenig Beobachtungen an großen AES Werten, an stark zu oszillieren.

Fall sein müsste.

Wir kommen nun zur Analyse der Preislimits  $P_{\pm 2}$ . In den Abbildungen 6.13 und 6.14 sieht man, dass es auch bei diesen Preislimits zu falschen Ergebnissen bezüglich der invarianten Verteilungen kommt. In Abbildung 6.12 ist ersichtlich, dass die Annahme 2 theoretisch erfüllt ist. Hierbei erkennt man, dass der *negative individual Drift* stark sinkt und nicht wie bei den ersten Limits stark ansteigt. Dies bedeutet, dass der Markovprozess  $(X_t^2)_{t \geq 0}$  für große Werte zu fallen tendiert. Da der Zustandsraum in unserem Fall jedoch mit 0 nach unten beschränkt ist, gibt es hierbei keine Probleme bezüglich der Existenz der invarianten Verteilung (Annahme 3 ist aufgrund der Intensitäten des ersten Preislimits verletzt).

Betrachtet man exemplarisch wieder einen Ausschnitt aus der Orderbuchentwicklung für die ersten 5 Tage, so ergibt sich folgendes Bild. Anhand von Abbildung 6.17 ist zu erkennen, dass das

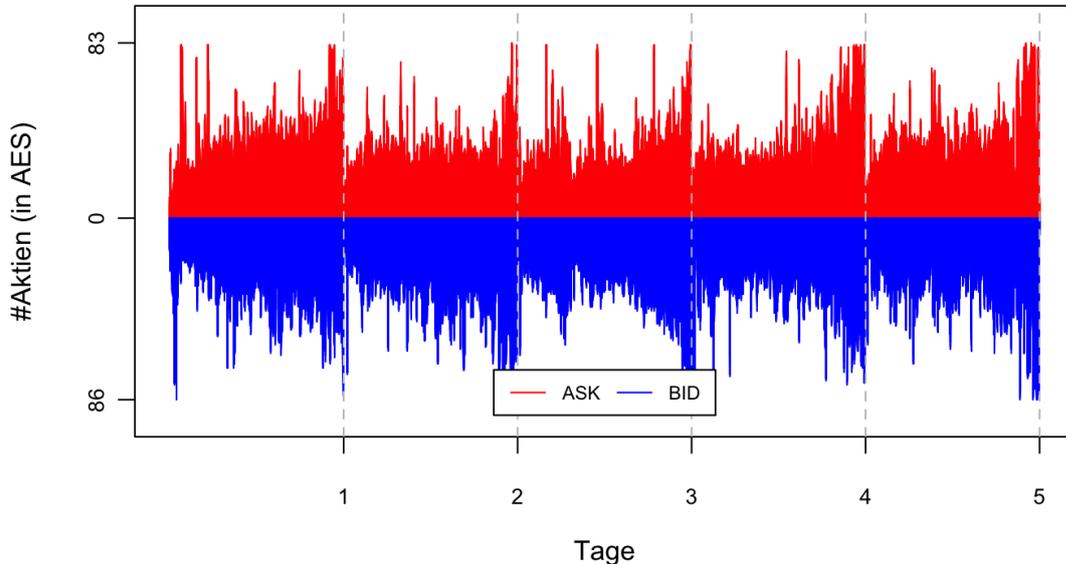


Abbildung 6.17: Orderbuchentwicklung MSFT Limit 2: 10.05.2010 - 24.05.2010

Orderbuchvolumen entgegen der Erwartung ein ähnlich zyklisches Verhalten wie beim Preislimit 1 aufweist. Hierbei ist unsere Vermutung, dass das Volumen mehrmals pro Tag im Zuge von Referenzpreisänderungen große Volumina der ersten Preislimits erbt und somit plötzlich hinaufgesetzt wird. Diese These werden wir nun wieder mit einer Simulation untersuchen. Exemplarisch betrachteten wir folgende Varianten, wobei wir jeweils insgesamt 1 Million Orders simulierten:

- Variante 1:  $\Delta_0 = 500$  und  $X_0 \sim U\{15, \dots, 45\}$
- Variante 2:  $\Delta_0 = 50$  und  $X_0 \sim U\{15, \dots, 45\}$

Die Ergebnisse werden in Abbildung 6.18 präsentiert. Je kleiner wir die Frequenz des Hochsetzens wählen, desto besser wird die empirische Verteilung aus den Daten erklärt.

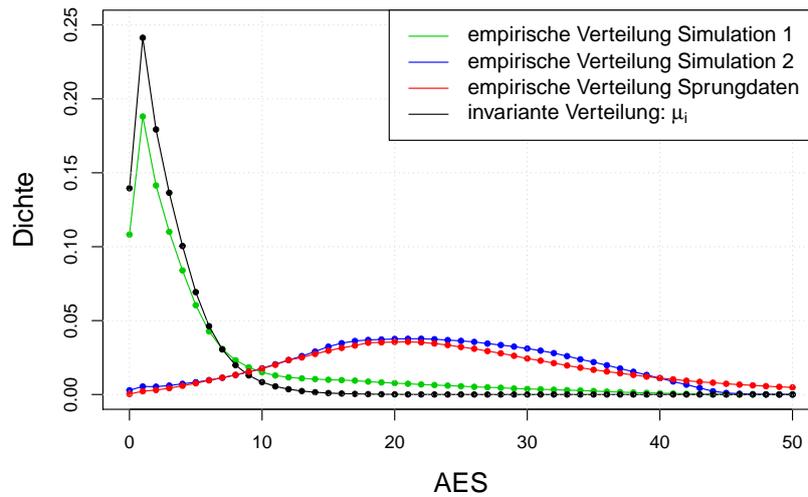


Abbildung 6.18: Nachbildung der empirischen Verteilung des Preislimits 2 mittels Simulation

In Abbildung 6.19 zeigen wir die empirische Verteilung der Differenz des Zustandes der verschiedenen Preislimits vor einer Referenzpreisänderung im Vergleich zu deren Zustand danach<sup>6</sup>. Wie besprochen würden wir nun für das erste Preislimit eher positive Differenzen erwarten, da es nach unserer These mehrmals pro Tag zurückgesetzt wird. Für das Limit 2 hingegen würden wir nach unserer These negative Differenzen erwarten.

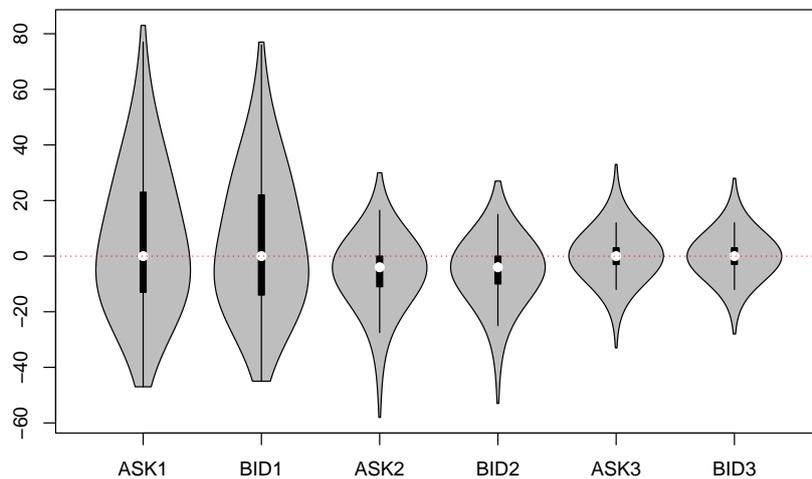


Abbildung 6.19: Darstellung der empirischen Verteilung der Sprünge bei gegebener Referenzpreisänderung MSFT

Zusammenfassend sieht man, dass unser Modell für die Aktie MSFT und die Preislimits 1 und 2 keine guten Ergebnisse im Bezug auf den Vergleich der empirischen mit der invarianten Ver-

<sup>6</sup>Ausreißer (> 99% Quantil wurden hierbei wieder zur besseren visuellen Darstellung eliminiert)

teilung liefert. Mögliche Ursachen dafür könnte das Zusammenspiel der hohen Anzahl an Referenzpreisänderungen und die unterschiedliche Charakteristik der Preislimits 1 und 2 sein (siehe Abbildung 6.11). Bei der Analyse von anderen Aktien z.B. Facebook (1.05.2017 - 31.05.2017) wurde generell beobachtet, dass ein stabiler *negativer individual Drift* mit Werten in der Nähe von 0 ein Indiz dafür ist, dass die invarianten Verteilung aus den geschätzten Intensitäten die empirische Verteilung gut approximiert.

# Kapitel 7

## Simulation des Orderbuches - Modell I

In diesem Kapitel werden wir eine Methode vorstellen, wie man das Orderbuchvolumen beziehungsweise den Markovprozess  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$  im Modell I zu gegebenen Intensitäten simulieren kann. Am Ende des Kapitels werden wir schließlich aus dem simulierten Orderbuch die Intensitäten wie in Abschnitt 3.2 schätzen und mit den gegebenen Intensitäten vergleichen. Da die Komponenten von  $X$  im Modell I als unabhängig angenommen werden, fokussieren wir uns hierbei auf ein festes Preislimit  $P_i$  und definieren für die ersten beiden Abschnitte 7.1 und 7.2 dieses Kapitels  $X_t := X_t^i$ . Aufgrund der Symmetrieannahme ist überdies auch keine Unterscheidung von  $X_t^{\pm i}$  notwendig.

### 7.1 Simulation eines einzelnen Preislimits

Wie wir bereits in Abschnitt 3.1 gesehen haben ist  $X_t$  ein eindimensionaler *Birth and Death* Prozess. Wir bezeichnen mit  $Q := Q^i$  und  $\pi := \pi^i$  dessen Intensitätsmatrix respektive invariante Verteilung d.h.  $X \sim \text{Markov}(\pi, Q)$ . Die Intensitätsmatrix  $Q$  kann mittels Definition von  $\lambda^+(\cdot) := \lambda_i^L(\cdot)$  und  $\lambda^-(\cdot) := \lambda_i^C(\cdot) + \lambda_i^M(\cdot)$  geschrieben werden als (siehe 3.2):

$$Q := \begin{pmatrix} -\lambda_i^+(0) & \lambda_i^+(0) & 0 & 0 & 0 & \dots \\ \lambda^-(1) & -(\lambda^-(1) + \lambda^+(1)) & \lambda_i^+(1) & 0 & 0 & \dots \\ 0 & \lambda^-(2) & -(\lambda^-(2) + \lambda_i^+(2)) & \lambda_i^+(2) & 0 & \dots \\ 0 & 0 & \lambda^-(3) & -(\lambda^-(3) + \lambda_i^+(3)) & \lambda_i^+(3) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Sei  $(J_n)_{n \in \mathbb{N}}$  die Folge der Sprungzeiten von  $X$  (siehe 2.8) und bezeichne weiters mit  $Z_n := X_{J_n}$  den assoziierten Sprungprozess des Markovprozesses  $(X_t)_{t \geq 0}$ . Wie wir bereits in Abschnitt 3.4.1 gesehen haben, folgt, dass  $Z_n$  eine diskrete Markovkette mit Übergangsmatrix  $\Pi$  ist. Falls  $Z$  eine invariante Verteilung  $\mu$  besitzt, kann diese mittels Theorem 3.3 von der invarianten Verteilung  $\pi$  des zeitstetigen Prozesses  $X$  berechnet werden. Wie wir in Abschnitt 3.4.1 bereits gesehen haben,

ist deren Übergangsmatrix  $\Pi$  gegeben durch:

$$i \neq j : \quad (\Pi)_{i,j} = \begin{cases} \frac{Q_{i,j}}{-Q_{ii}} & -Q_{i,i} > 0 \\ 0 & -Q_{i,i} = 0 \end{cases}$$

$$(\Pi)_{i,i} = \begin{cases} 0 & -Q_{i,i} > 0 \\ 1 & -Q_{i,i} = 0 \end{cases}$$

Diese hat aufgrund der Tridiagonalgestalt von  $Q$  lediglich in den beiden Nebendiagonalen Einträge ungleich 0.

Um den zeitstetigen homogenen Markovprozess  $X_t \sim \text{Markov}(\pi, Q)$  zu konstruieren, kann man folgendermaßen vorgehen [8, Abschnitt 2.6]:

### 7.1.1 Simulation Markovprozess Methode I

Bezeichne  $\Pi$  die zu  $Q$  gehörige Sprungmatrix. Sei  $(Z_n)_{n \in \mathbb{N}}$  eine diskrete Markovkette mit Übergangsmatrix  $\Pi$  und Anfangsverteilung gleich der invarianten Verteilung  $\pi$  aus Theorem 3.2 d.h.  $Z \sim \text{Markov}(\pi, \Pi)$ . Seien weiters  $T_1, T_2, \dots$  i.i.d,  $T_1 \sim \text{Exp}(1)$  und unabhängig von  $(Z_n)_{n \in \mathbb{N}}$ . Definiere

$$S_n := \frac{T_n}{-Q_{Z_{n-1}Z_{n-1}}} \quad (\text{Halteperioden})$$

$$J_n := S_1 + S_2 + \dots + S_n \quad (J_0 := 0)(\text{Sprungzeiten})$$

$$X_t := \begin{cases} Z_n & J_n \leq t < J_{n+1} \\ \infty & \text{sonst} \end{cases}$$

dann gilt, dass  $Z$  der assoziierte Sprungprozess zu  $X$  ist und dass  $S_1, \dots, S_n$  bedingt auf  $Z_0, \dots, Z_{n-1}$  unabhängig exponentialverteilt sind mit  $S_n | \{Z_{n-1} = n\} \sim \text{Exp}(-Q_{n,n})$ . Laut Theorem 2.8.2 [8] ist somit  $X \sim \text{Markov}(\pi, Q)$ .

## 7.2 Algorithmus zur Simulation eines Preislimits

Wir besprechen nun, wie man beispielsweise ein einzelnes Preislimit  $P_i$  beziehungsweise den ein-dimensionalen *Birth and Death* Prozess  $X_t$  und die dazugehörigen Parameter Zeit der Order und Ordertyp, simulieren kann. Da der Prozess  $X_t$  laut Annahme nur Sprünge der Höhe  $\pm 1$  aufweist, ist die Variable Ordergröße obsolet. Hat man zu gegebenem Preislimit  $P_i$  die durchschnittliche Ordergröße  $AES_i$  (siehe Annahme 1) definiert, so bedeutet ein Sprung der Höhe  $\pm 1$  eine Veränderung des Orderbuchvolumens  $X$  in der Höhe von  $\pm AES_i$  Aktien.

Zu gegebener Intensitätsmatrix  $Q$  und invarianter Verteilung  $\pi$  kann wie folgt vorgegangen werden:

*Pseudocode zur Simulation von  $(X_t)_{t \leq T}$ :*

1. Schritt: Berechne die zu der Intensitätsmatrix  $Q$  gehörige Sprungmatrix  $\Pi$  wie in 3.12.

2. Schritt: Setze  $n := 0$  und  $t = t_0 := 0$ . Ziehe den Startwert  $i_0 \in \mathbb{N}$  von der Verteilung  $\pi$  und setze  $X_0 = Z_0 := i_0$
3. Schritt: Ziehe  $S_{i_0} \sim \text{Exp}(-Q_{i_0, i_0})$  und setze  $t = t_1 := S_{i_0}(\omega)$
4. Schritt: Falls  $t \leq T$ 
  - (a) Setze  $i := Z_n$  und ziehe von  $Y \sim (\Pi)_i$ . (i-te Zeile) wobei  $\mathbb{P}[Y = j] = (\Pi)_{i,j}$ .<sup>1</sup>
  - (b) Setze  $n := n + 1$  und  $Z_n := Y(\omega)$ .
  - (c) (Bestimme Typ der Order)<sup>2</sup>  
Falls  $Y(\omega) = i + 1$  setze  $\mathcal{T}_n := L$   
Falls  $Y(\omega) = i - 1$  ziehe  $\mathcal{T}_n \sim \mathbb{P}_i$  wobei  $\mathbb{P}_i$  definiert ist durch:
$$\mathbb{P}_i[\mathcal{T}_n = C] = \frac{\lambda^C(i)}{\lambda^C(i) + \lambda^M(i)}$$

$$\mathbb{P}_i[\mathcal{T}_n = M] = \frac{\lambda^M(i)}{\lambda^C(i) + \lambda^M(i)}$$
  - (d) Ziehe  $S_{Z_n} \sim \text{Exp}(-Q_{Z_n, Z_n})$  und setze  $t := t + S_{Z_n}(\omega)$ ,  $t_{n+1} := t$ .
  - (e) Wiederhole Schritt 4.

Für  $N_T := \max\{n \in \mathbb{N} : t_n \leq T\}$  liefert uns dieses Verfahren alle Sprungzeitpunkte  $t_1, \dots, t_n$  im Intervall  $[0, T]$ , die dazugehörigen Werte der Markovkette  $X_{t_n} = Z_n$ ,  $0 \leq n \leq N_T$  und die Ordertypen  $\mathcal{T}_1, \dots, \mathcal{T}_{N_T}$ , die zu der Veränderung des Orderbuchvolumens geführt haben.

Die Funktion *simulMod1()* aus Kapitel 9, (12) simuliert mittels dem vorgestellten Pseudocode ein einzelnes Preislimit bis zu einer maximalen Anzahl an AES.

### 7.3 Empirische Studie zur Orderbuchmodellierung

In diesem Abschnitt werden wir die Ergebnisse einer Simulation und der auf den simulierten Daten basierenden Intensitätsschätzung präsentieren.

Im Folgenden bezeichnen wir mit  $X_t$  wieder das komplette Orderbuch beziehungsweise den 6-dimensionalen Markovprozess  $X_t = (X_t^{-3}, X_t^{-2}, X_t^{-1}, X_t^1, X_t^2, X_t^3) \in \mathbb{N}_0^6$  d.h. für die folgende Analyse betrachten wir die ersten drei Preislimits und wählen somit  $K:=3$ .

In Anlehnung an [6, Abschnitt 2.3], wählen wir für  $n \in \{0, \dots, 50\}$  folgende Intensitäten:

- Limit Order Intensitäten:

$$\lambda_1^L(n) := \begin{cases} 0.25 & n = 0 \\ -0.00028 \cdot n^2 + 0.016 \cdot n + 0.6 & n \in \{1, \dots, 50\} \end{cases}$$

$$\lambda_2^L(n) := 2.3 \cdot \exp\left(-\frac{n}{4}\right) + 0.5$$

$$\lambda_3^L(n) := 2.5 \cdot \exp(-n) + 0.5$$

<sup>1</sup>Aufgrund der besonderen Gestalt der Matrix  $\Pi$  gilt  $\mathbb{P}[Y \in \{i+1, i-1\}] = 1$ .

<sup>2</sup>L=Limit Order, M=Market Order, C=Cancellation Order

- Cancellation Order Intensitäten:

$$\lambda_1^C(n) := \begin{cases} -\frac{0.9}{n+1} + 0.9 & n \leq 25 \\ 0.865 - 0.005 \cdot (n - 26) & n \in \{26, \dots, 50\} \end{cases}$$

$$\lambda_2^C(n) := -\frac{0.6}{5n+1} + 0.6$$

$$\lambda_3^C(n) := \begin{cases} 0 & n = 0 \\ 0.25 & n = 1 \\ \frac{n}{160} + 0.2875 & n \in \{2, \dots, 50\} \end{cases}$$

- Market Order Intensitäten:

$$\lambda_1^M(n) := \begin{cases} 0 & n = 0 \\ 0.25 & n = 1 \\ 0.288 \cdot \exp\left(-\frac{n}{2}\right) + 0.0755 & n \in \{4, \dots, 50\} \end{cases}$$

$$\lambda_2^M(n) := \begin{cases} 0 & n = 0 \\ 0.025 & n = 1 \\ 0.01 & n = 2 \\ 0.005 & n = 3 \\ 0 & n \in \{4, \dots, 50\} \end{cases}$$

$$\lambda_3^M(n) := \begin{cases} 0 & n = 0 \\ 0.005 & n = 1 \\ 0 & n \in \{2, \dots, 50\} \end{cases}$$

In der folgenden Grafik werden diese generierten Intensitäten illustriert.

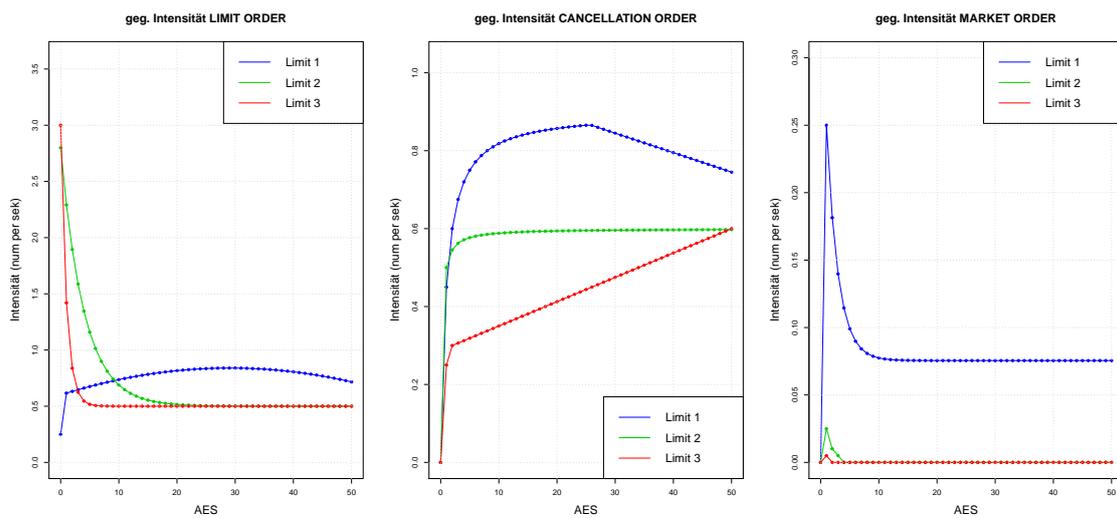


Abbildung 7.1: gegebene Intensitäten

Die daraus resultierenden Intensitätsmatrizen  $Q^i$ ,  $i \in \{1, 2, 3\}$  und deren assoziierten Sprungma-

trizen  $\Pi^i, i \in \{1, 2, 3\}$  sind Elemente von  $\mathbb{R}^{51 \times 51}$ . Insbesondere nehmen wir hierbei an, dass das Volumen der Preislimits nach oben mit 50  $AES_i$  beschränkt ist. Um die Schätzmethode der Intensitäten aus Abschnitt 3.2 zu analysieren, haben wir 30 Millionen Orders pro Preislimit mittels dem in Abschnitt 7.2 beschriebenen Algorithmus zu den oben definierten Intensitäten simuliert. Aus diesen simulierten Daten wurden dann die Intensitäten mit der im Abschnitt 3.2 beschriebenen Methode bis 50  $AES_i$  geschätzt.

Die Ergebnisse für die Schätzung werden in den folgenden Grafiken präsentiert:

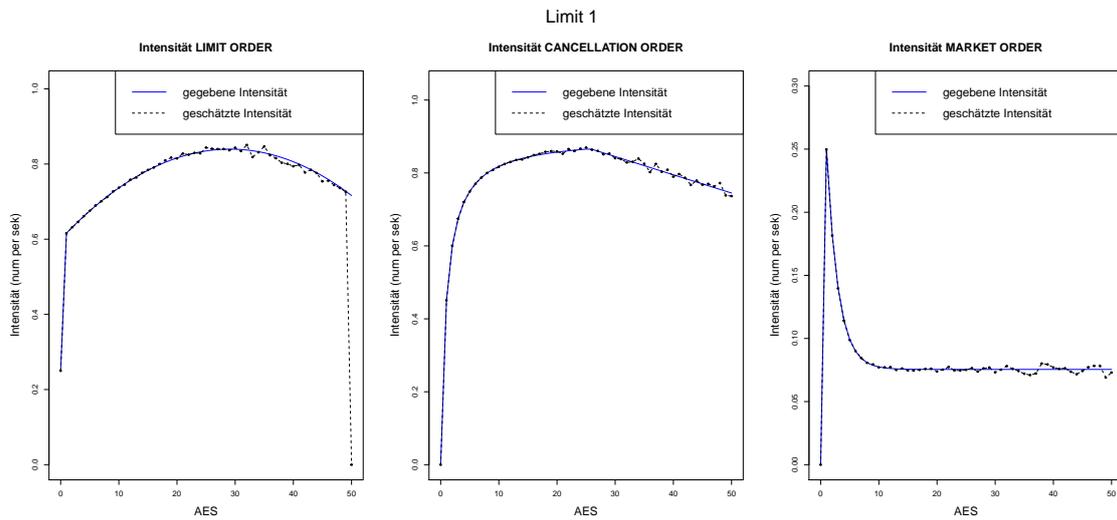


Abbildung 7.2: Aus simulierten Daten geschätzte Intensitäten für das erste Preislimit

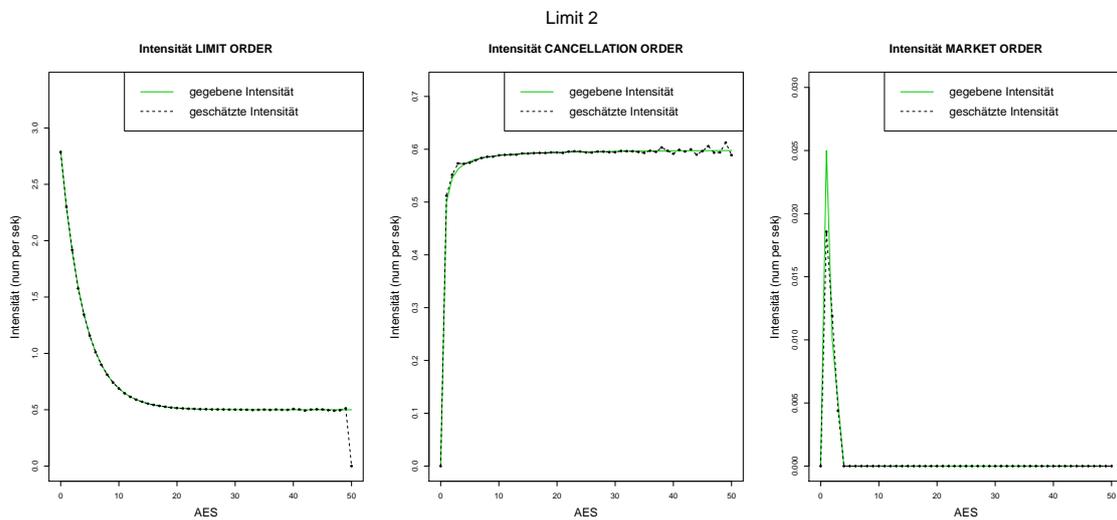


Abbildung 7.3: Aus simulierten Daten geschätzte Intensitäten für das zweite Preislimit

Hierbei ist ersichtlich, dass die in Abschnitt 3.2 vorgestellte Maximum Likelihood Methode bei simulierten Daten eine sehr gute Approximation der zugrunde liegenden Intensitäten liefert.

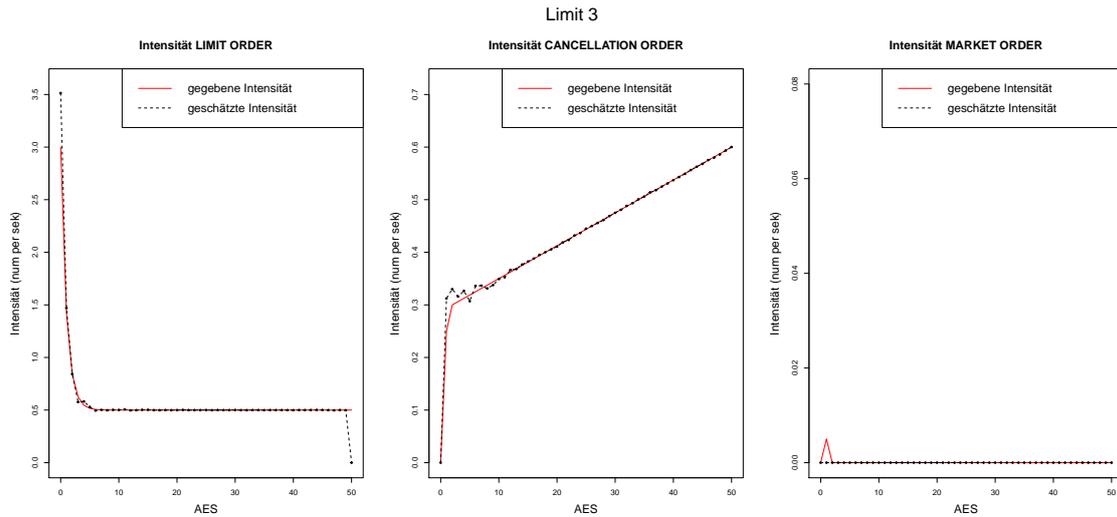


Abbildung 7.4: Aus simulierten Daten geschätzte Intensitäten für das dritte Preislimit

Anhand der Schätzung wird ersichtlich, dass der geschätzte Wert für die Intensität bei Limit Orders bei einer Ordergröße von 50 AES gleich null beträgt und somit auf den ersten Blick stark von den gegebenen Intensitäten abweicht. Die lässt sich jedoch aufgrund der Endlichkeitsannahme der Intensitätsmatrix  $Q \in \mathbb{R}^{51 \times 51}$  beziehungsweise wegen der angenommenen Beschränktheit des simulierten Markovprozesses erklären.

Zum Abschluss dieses Kapitels vergleichen wir für  $i \in \{1, 2, 3\}$  die theoretische invariante Verteilung des Markovprozesses  $X^i$  zu den gegebenen Intensitäten mit der empirischen Verteilung.

Hierbei ist wie bereits zuvor erwähnt besonders darauf Acht zu geben, dass man die richtigen Verteilungen miteinander vergleicht. Die Simulation aus Abschnitt 7.2 liefert die Werte des zu  $X^i$  assoziierten Sprungprozesses  $Z^i$ . Schätzt man nun aus den simulierten Daten die empirischen Häufigkeiten dieser Werte, so muss man die erhaltene empirische Verteilung mit der theoretisch invarianten Verteilung  $\mu_i$  des Sprungprozesses  $Z^i$  aus Theorem 3.3 vergleichen und nicht mit der invarianten Verteilung  $\pi_i$  des stetigen Markovprozesses  $X^i$ .

Eine zweite Möglichkeit ist es aus den simulierten Sprungdaten zu einer gegebenen festen Beobachtungsfrequenz (engl. sampling frequency) z.B. 10 Sekunden jeweils den Wert des Markovprozesses  $X^i$  davor als Grundlage für die empirischen Häufigkeiten heranzuziehen. Dies liefert die empirische Verteilung des zeitstetigen Markovprozesses  $X^i$  und kann somit mit der invarianten Verteilung  $\pi_i$  verglichen werden. Im Folgenden verwenden wir erstere Methode, um die empirische Verteilung aus der Simulation mit der theoretisch invarianten Verteilung zu vergleichen.

Wir überprüfen zuerst die Regularitätsvoraussetzungen (Annahme 2 und 3), welche die Existenz der invarianten Verteilung sicherstellen. Die Ergebnisse werden in Abbildung 7.5 und 7.6 dargestellt.

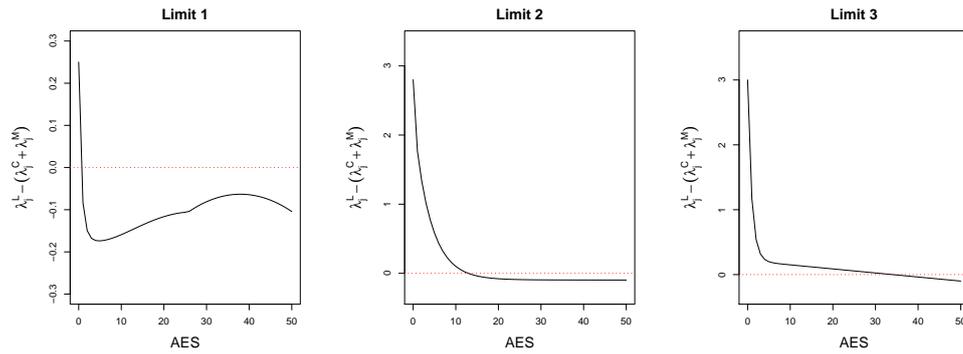


Abbildung 7.5: Regularitätsvoraussetzung: Annahme 2

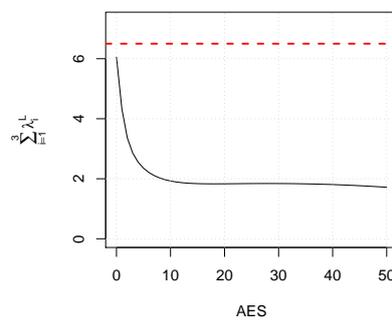
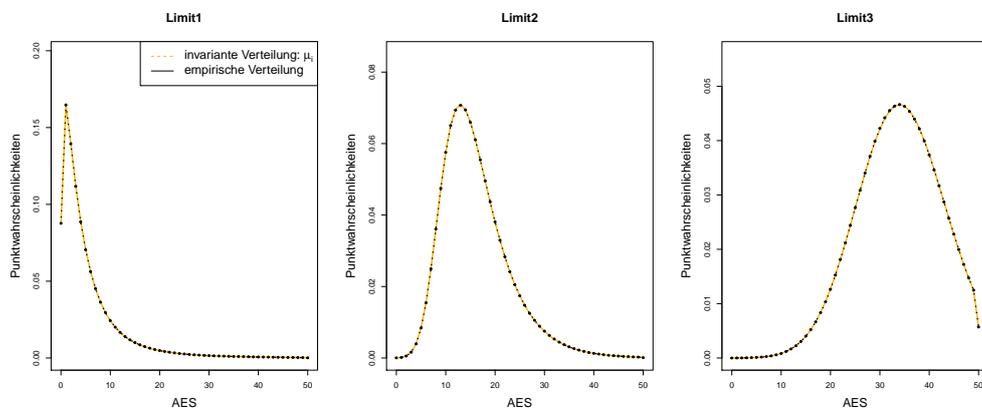


Abbildung 7.6: Regularitätsvoraussetzung: Annahme 3

Anhand der beiden Abbildungen ist klar ersichtlich, dass beide Bedingungen für die an [6] angelehnten Intensitäten erfüllt sind und somit eine invariante Verteilung des Markovprozesses zu diesen Intensitäten existiert. Bei der empirischen Verteilung wurden wie bereits besprochen die relativen Häufigkeiten des simulierten Orderbuches als Schätzer für die Punktwahrscheinlichkeiten verwendet und mit der invarianten Verteilung  $\mu_i$  des assoziierten Sprungprozesses  $Z^i$  verglichen. Anhand des Ergebnisses in Abbildung 7.7 sieht man, dass die mittels Theorem 3.3 berechnete invariante Verteilung mit der empirischen Verteilung der simulierten Daten übereinstimmt.

Abbildung 7.7: Vergleich der empirischen Verteilung mit der invarianten Verteilung  $\mu_i$  zu simulierten Daten

## 7.4 Anwendungsbeispiel - Wahrscheinlichkeit der Exekution einer Limit Order

In diesem Abschnitt besprechen wir ein konkretes Anwendungsbeispiel von Modell I. Mithilfe der Simulationstechnik von Abschnitt 7.2 können wir die Wahrscheinlichkeit einer Orderausführung, bevor sich der *Midprice* bewegt, mittels Monte Carlo Simulation berechnen. Dieser Abschnitt basiert auf [6, Abschnitt 2.5].

Wir nehmen an, dass ein Händler A zum Zeitpunkt  $t = 0$  eine Kauf oder Verkauf Limit Order an die Preislimits  $P_{-1}$  respektive  $P_1$  in der Höhe von  $n_0 \cdot AES_{\pm 1}$  Aktien abgibt. Des Weiteren fordern wir, dass das bestehende Ordervolumen an diesen Preislimits nicht leer ist d.h.  $X_0^{-1} \neq 0, X_0^1 \neq 0$ . Der Händler wartet nun solange, bis eines der beiden folgenden Ereignisse eintritt:

- Die Order wird vollständig ausgeführt.
- Das gegenüberliegende Preislimit:  $P_{\mp 1}$  ist leer. Dieses Ereignis ist äquivalent zu einer Bewegung des *Mid Price*. In diesem Falle wurde die Order nicht rechtzeitig vollständig ausgeführt.

Ziel ist es die Wahrscheinlichkeit der Orderausführung numerisch zu berechnen. Aufgrund der angenommenen Symmetrieeigenschaft der Intensitäten des Markovprozesses  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$  im Modell I ist die Wahrscheinlichkeit einer Orderausführung für das obige Szenario im Falle einer Kauf Limit Order gleich wie im Falle einer Verkauf Limit Order. Daher betrachten wir im weiteren Verlauf dieses Abschnittes o.B.d.A. den Fall einer Kauf Limit Order beziehungsweise einer Orderabgabe an das Preislimit  $P_{-1}$ . Des Weiteren können wir uns bei der Simulation aufgrund der Unabhängigkeit der Komponenten von  $X$  auf die ersten beiden Preislimits beschränken und definieren für den restlichen Teil dieses Abschnitts  $X_t := (X_t^{-1}, X_t^1)$ .

Es gibt zwei Arten von Limit Orders am Preislimit  $P_{-1}$ . Jene die bereits vor  $t = 0$  abgegeben wurden und eine höhere Priorität als die vom Händler A Abgegebene aufweisen und jene, die nach  $t = 0$  abgegeben werden und somit eine geringere Priorität aufweisen (vgl. Preis - Zeit - Priorität in 1.2.2). Wir bezeichnen im Folgenden mit  $N_b(t)$  und  $N_a(t)$  die Anzahl der Orders zum Zeitpunkt  $t$  mit einer höheren (below) respektive geringeren (above) Priorität als die abgegebene Order vom Händler A. Das heißt ein Teil der vom Händler A abgegebenen Order wird zum ersten Mal ausgeführt, wenn alle Orders mit einer höheren Priorität bereits ausgeführt oder gelöscht wurden d.h. zum Zeitpunkt  $\inf\{t > 0 : N_b(t) = 0\}$ .

Im Falle einer *Market Order* am Preislimit  $P_{-1}$  wird wie im normalen Setting mittels der Preis - Zeit - Priorität ermittelt, welche stehenden Limit Order ausgeführt werden. Im Falle einer *Cancellation Order* treffen wir zusätzlich folgende Annahmen:

**Annahme 6** *Die vom Händler A abgegebene Order kann niemals gelöscht werden. Falls eine Cancellation Order bei  $P_{-1}$  eintrifft, haben alle Order, bis auf die vom Händler A abgegebene, die gleiche Wahrscheinlichkeit gelöscht zu werden. Für einen festen Zeitpunkt  $t \geq 0$  ist somit die Wahrscheinlichkeit, dass eine Order mit höherer beziehungsweise mit niedriger Priorität gelöscht wird, gegeben durch  $\frac{N_b(t)}{N_a(t)+N_b(t)}$  respektive  $\frac{N_a(t)}{N_b(t)+N_a(t)}$ .*

**Annahme 7** *Die Cancellation Intensität am Preislimit  $P_{-1}$  ist gegeben durch  $\lambda_{-1}^C(x_{-1}) := \lambda_{-1}^C(x_{-1}) \cdot \frac{x_{-1} - n_0(t)}{x_{-1}}$ , wobei  $n_0(t)$  die verbliebene Größe der abgegebenen Order zum Zeitpunkt  $t$  bezeichnet.*

*Pseudocode: Monte Carlo Simulation*

1. Bestimme Intensitätsmatrix  $Q^1$  und Sprungmatrix  $\Pi^1$  zu gegebenen Intensitäten  $\lambda_1^L, \lambda_1^C$  und  $\lambda_1^M$ .
2. Setze Startwerte für Bid und Ask Limits  $X_0^1 := j$  und  $X_0^{-1} := i$  zu definiertem Gitter  $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$  mit  $m, n \in \mathbb{N}$ .
3. Einzelne Simulation: Sei  $n_0$  die Größe der abgegebenen Order von Händler A.
  - Bestimme Anzahl der Orders mit höherer Priorität  $N_b(0) := i$ , Anzahl der Orders mit niedrigerer Priorität  $N_a(0) := 0$  und setze  $T := 0, n := n_0$ .
  - Simuliere mithilfe von  $Q^1$  und  $\Pi^1$  den ersten Zeitpunkt an dem sich der *Mid Price* bewegt d.h. bestimme  $\tau := \inf\{t \geq 0 : X_t^1 = 0\}$  (vgl. Abschnitt 7.2).
  - Solange  $T \leq \tau$ :
    - Manipuliere Cancellation Intensität  $\lambda_{-1}^C$  wie in Annahme 7 und erhalte  $\tilde{\lambda}_{-1}^C, \tilde{Q}^{-1}$  und  $\tilde{\Pi}^{-1}$ .
    - Simulieren mithilfe von  $\tilde{Q}^{-1}$  und  $\tilde{\Pi}^{-1}$  eine Order zum Preislimit  $P_{-1}$ , deren Zeitpunkt  $t$  und deren Typ  $\mathcal{T}$  (vgl. Abschnitt 7.2). Setze  $T := T + t$ .
    - Falls  $\mathcal{T} = L$ :  $N_a(t) := N_a(t) + 1$
    - Falls  $\mathcal{T} = C$ : Bestimme wie in Annahme 6 beschrieben ob eine Limit Order mit höherer oder niedrigerer Priorität gelöscht wurde und setze anschließend:
      - \* Falls Limit Order mit höherer Priorität gelöscht wurde:  $N_b(t) := N_b(t) - 1$ .
      - \* Falls Limit Order mit niedrigerer Priorität gelöscht wurde:  $N_a(t) := N_a(t) - 1$ .
    - Falls  $\mathcal{T} = M$  dann gehe wie folgt vor:
      - \* Falls  $N_b(t) > 0$ : setze  $N_b(t) := N_b(t) - 1$ .
      - \* Falls  $N_b(t) = 0$ : setze  $n := n - 1$ . Falls  $n = 0$ : Order erfolgreich vor  $\tau$  ausgeführt.

Speichere, ob bei dieser Simulation die Orderausführung erfolgreich war oder nicht, und wiederhole Schritt 3, sofern die Anzahl der gewünschten Simulationen pro Startwert noch nicht erreicht wurde.
4. Berechne die relative Häufigkeit der Simulationen mit erfolgreicher Orderausführung und gehe zu Schritt 2, falls noch nicht alle Kombinationen an Startpaaren erschöpft sind.

Dieser Algorithmus liefert schlussendlich zu jedem Startwert  $(i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$  einen Schätzer für die Wahrscheinlichkeit der Exekution der abgegebenen Limit Order.

Die folgende Grafik zeigt die erhaltenen Ergebnisse für die Wahl von  $n_0 = 1, m = n = 30$  und 10000 Simulationen pro Startwertpaar. Als gegebene Intensitäten wurden die bereits in Abschnitt 7.3 eingeführten Funktionen verwendet (siehe Abbildung 7.1).

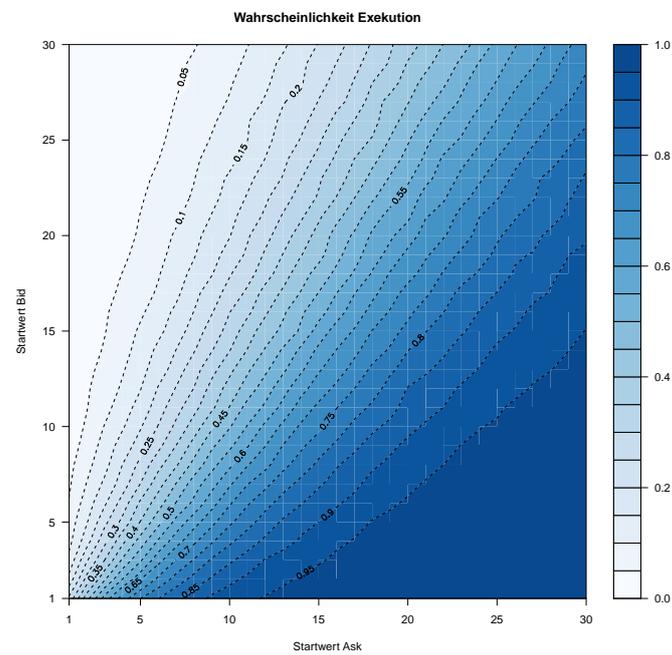


Abbildung 7.8: Wahrscheinlichkeit einer Exekution einer Buy Limit Order mit Ordergröße 1, gesendet an das Preislimit  $Q_{-1}$ .

# Kapitel 8

## Modell III - dynamischer Referenzpreis

In diesem Kapitel präsentieren wir eine Erweiterung von Modell I. Da im Modell I das Gitter  $\{P_{-K}, \dots, P_{-1}, P_1, \dots, P_K\}$ , auf dem sich der Markovprozess  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$  bewegt, durch den Referenzpreis  $p_{ref}$  bestimmt ist, ist dieses Modell in Zeiten eines konstanten Referenzpreises anwendbar, beinhaltet jedoch keinerlei dynamische Beschreibung des Referenzpreises. Mit Modell III wird es möglich sein das Orderbuchvolumen  $X$  für beliebige Zeitperioden, in denen sich der Referenzpreis möglicherweise mehrmals ändert, zu beschreiben. Dies wird durch eine dynamische Modellierung des Referenzpreises erreicht, wobei in Zeiten eines konstanten Referenzpreises das Modell I herangezogen wird. Dieses Kapitel folgt im Wesentlichen Abschnitt 3.1 von [6].<sup>1</sup>

### 8.1 Modellbeschreibung - Modell III

Im Folgenden bezeichnen wir weiterhin mit  $\Delta_{tick}$  die *ticks*ize der betrachteten Aktie. Der *Midprice* des LOB wird mit  $p_{mid}$  bezeichnet. Dieser ist definiert als:

$$p_{mid} := \frac{p_{bestAsk} + p_{bestBid}}{2} \quad (8.1)$$

wobei wir mit  $p_{bestAsk}$  beziehungsweise  $p_{bestBid}$  den *best Ask Price* respektive den *best Bid Price* des aktuellen Zustandes des LOB bezeichnen. Wie bereits besprochen wird nun der Referenzpreis zu jedem Zeitpunkt als stochastischer Sprungprozess mit den Sprunghöhen  $\pm \Delta_{tick}$  beschrieben:

$$p_{ref} : \Omega \times \mathbb{R}_+ \longrightarrow \frac{\Delta_{tick}}{2} \cdot \mathbb{N} \\ (\omega, t) \mapsto p_{ref}(\omega, t)$$

In diesem Modell nehmen wir an, dass sich der Referenzpreis  $p_{ref}$  mit Wahrscheinlichkeit  $\theta \in [0, 1]$  ändert, falls es zu einer Veränderung des *Midprice*  $p_{mid}$  kommt. Genauer gesagt gilt, falls  $p_{mid}$  steigt (sinkt), steigt (sinkt)  $p_{ref}$  um  $\Delta_{tick}$  mit Wahrscheinlichkeit  $\theta$ , sofern für die Ordergröße am

---

<sup>1</sup>Die Bezeichnung des Modells mit dynamischem Referenzpreis als Modell III ist an das zugrunde liegende Paper angelehnt. In diesem wurde überdies auch ein Modell II vorgestellt, welches jedoch in dieser Arbeit nicht besprochen wird.

ersten Preislimit gilt  $x_1 = 0$  ( $x_{-1} = 0$ ).

Im Folgenden bezeichnen wir mit  $J_n^Y$  die Sprungzeitpunkte eines stochastischen Prozesses  $Y$ . Weiters definieren wir mit  $\Delta_n(Y) := Y_{J_n^Y} - Y_{J_{n-1}^Y}$  die Sprunghöhe des  $n$ -ten Sprunges.

Mathematisch formalisiert gilt somit für den stochastischen Prozess  $p_{ref}(t)$  und  $n \geq 1$ :

$$\begin{aligned} \mathbb{P} [p_{ref}(J_n) = p_{ref}(J_{n-1}) + \Delta_{tick} | \Delta_n(p_{mid}) > 0, X_{J_n}^1 = 0] &= \theta \\ \mathbb{P} [p_{ref}(J_n) = p_{ref}(J_{n-1}) - \Delta_{tick} | \Delta_n(p_{mid}) < 0, X_{J_n}^{-1} = 0] &= \theta \end{aligned}$$

In unserem Modell können somit drei Szenarien unterschieden werden, aufgrund derer sich der Referenzpreis ändert:

1. Abgabe einer neuen Kauf Limit Order in den *Bid Ask Spread*, wobei die Ordergröße zur Zeit der Abgabe am ersten Ask Limit gleich null beträgt d.h.  $x_1 = 0$ . Analog führt eine Abgabe einer neuen Verkauf Limit Order in den *Bid Ask Spread* zu einer Veränderung des Referenzpreises, sofern  $x_{-1} = 0$ .
2. Durch eine Cancellation einer Limit Order am *best Bid Price* (oder *best Ask Price*), sodass die Ordergröße am betrachteten Preislimit nach der Cancellation gleich null beträgt. Hierbei ist die Voraussetzung  $x_{-1} = 0$  ( $x_1 = 0$ ) immer erfüllt.
3. Eine Exekution einer Limit Order durch eine Market Order am *best Bid Price* (oder *best Ask Price*), sodass die Ordergröße nach dieser Market Order gleich null beträgt. Wie im zweiten Punkt ist hierbei die Voraussetzung  $x_{-1} = 0$  ( $x_1 = 0$ ) automatisch erfüllt.

Falls sich der Referenzpreis  $p_{ref}(t)$  zum Zeitpunkt  $T$  durch eine der oben genannten Möglichkeiten ändert, wechseln die jeweiligen Ordergrößen  $x_i$  zum Zeitpunkt  $T$  zu ihren Nachbarn. Da die Preislimits  $P_i$  womöglich unterschiedliche  $AES_i$  haben, müssen diese Größen richtig skaliert werden. Erhöht sich der Referenzpreis aufgrund von (1) einer Kauf Limit Order Abgabe in den *Bid Ask Spread* oder (2) einer Cancellation einer Verkauf Limit Order am *best Ask Price*, am Preislimit  $i^*$  um einen Tick, dann gilt: <sup>2</sup>

$$\begin{aligned} X_T^i &:= \left[ \frac{X_{T-}^{i+1}}{AES_i} \cdot AES_{i+1} \right], \quad \forall i \in \{-K, \dots, -2, 1, \dots, K-1\} \setminus \{i^*\} \\ X_T^{-1} &:= 0 \\ (1) X_T^{i^*} &:= +1 \quad (2) X_T^{i^*} := 0 \\ X_T^K &\text{ wird von der invarianten Verteilung } \pi_K \text{ gezogen.} \end{aligned}$$

wobei  $i^*$  als das nächst kleinere Preislimit zu  $i^*$  definiert ist.

Analog gilt bei einer Verminderung der Referenzpreises aufgrund von (1) einer Verkauf Limit Order Abgabe in den *Bid Ask Spread* oder (2) einer Cancellation einer Kauf Limit Order am *best Bid*

<sup>2</sup>Mit  $X_{T-}^i$  bezeichnen wir den linksseitigen Grenzwert d.h.  $X_{T-}^i := \lim_{t \uparrow T} X_t^i$

Price) am Preislimit  $i^*$ :

$$X_T^i := \left\lfloor \frac{X_{T-}^{i-1}}{AES_{i-1}} \cdot AES_i \right\rfloor, \forall i \in \{-K+1, \dots, -2, -1, 2, \dots, K\} \setminus \{i^*\}$$

$$X_T^1 := 0$$

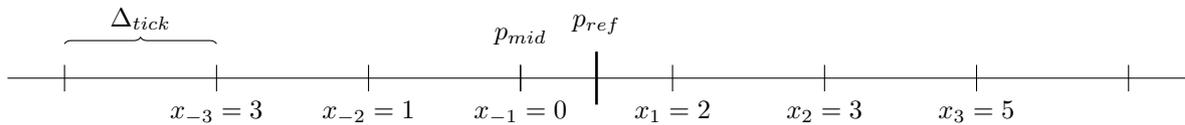
$$(1) X_T^{i^*} := +1 \quad (2) X_T^{i^*} := 0$$

$X_T^{-K}$  wird von der invarianten Verteilung  $\pi_{-K} = \pi_K$  gezogen.<sup>2</sup>

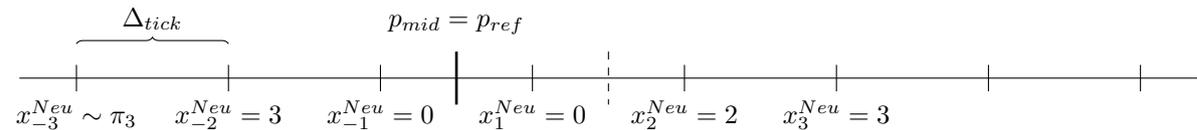
wobei  $i_+^*$  als das nächst größere Preislimit zu  $i^*$  definiert ist.

Illustrativ stellen wir nun drei mögliche Szenarien einer Referenzpreisänderung vor. In allen drei Beispielen gehen wir zur Vereinfachung von einheitlichen  $AES_i$  für alle  $i$  aus. Des Weiteren setzen wir  $\theta := 1$  und  $K := 3$ .

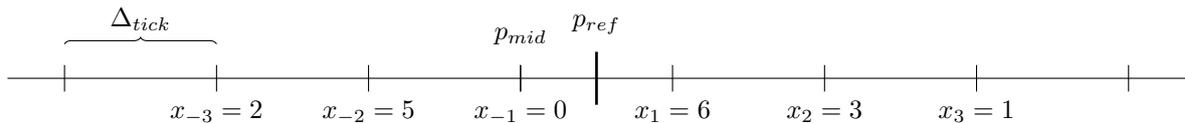
Als erstes Beispiel besprechen wir den Fall 2 und 3 d.h. Cancellation oder Market Order. Zum Zeitpunkt  $t = 0$  nehmen wir an, dass der Zustand des LOB gegeben ist durch  $X_0 = (x_{-3}, x_{-2}, x_{-1}, x_1, x_2, x_3)$  mit  $x_i$  definiert als:



Zu einem späteren Zeitpunkt  $T > 0$  wird die verbliebene Limit Order am Preislimit  $P_{-2}$  gelöscht oder durch eine Market Order ausgeführt. Der Referenzpreis vermindert sich um eine *ticksize*. In der folgenden Grafik wird das veränderte LOB  $X_T = (x_{-3}^{Neu}, x_{-2}^{Neu}, x_{-1}^{Neu}, x_1^{Neu}, x_2^{Neu}, x_3^{Neu})$  dargestellt, wobei  $x_i^{Neu}$  gegeben ist als:



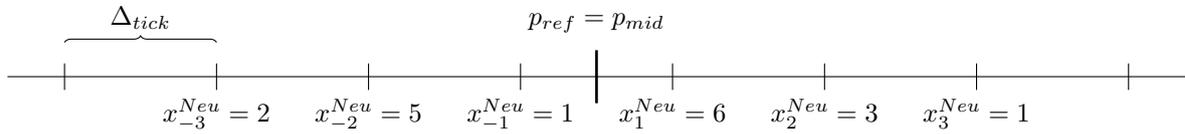
Als zweites Beispiel widmen wir uns dem Fall 1 d.h. einer neuen Kauf Limit Order in den *Bid Ask Spread*. Zum Zeitpunkt  $t = 0$  nehmen wir an, dass der Zustand des LOB gegeben ist durch  $X_0 = (x_{-3}, x_{-2}, x_{-1}, x_1, x_2, x_3)$  mit  $x_i$  definiert in der folgenden Grafik:



Hierbei ist ersichtlich, dass der *Bid Ask Spread*  $2 \cdot \Delta_{tick}$  beträgt.

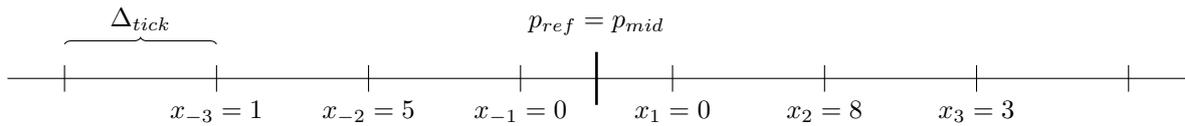
<sup>2</sup>Die Gleichheit der invarianten Verteilungen gilt aufgrund der weiterhin bestehenden Symmetrieannahme aus Modell I.

Zu einem späteren Zeitpunkt  $T > 0$  wird eine neue Kauf Limit Order zum Preislimit  $P_{-1}$  in der Höhe von  $1 AES_{-1}$  abgegeben. Da jedoch  $x_1 = 6 > 0$  erhöht sich der Referenzpreis nicht. In der folgenden Grafik wird das LOB  $X_T = (x_{-3}^{Neu}, x_{-2}^{Neu}, x_{-1}^{Neu}, x_1^{Neu}, x_2^{Neu}, x_3^{Neu})$  dargestellt, wobei  $x_i^{Neu}$  gegeben ist als:

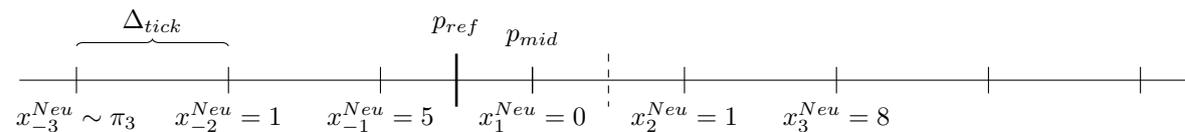


Es kommt in diesem Szenario somit zu keiner Verschiebung der Preislimits.

Im letzten Beispiel diskutieren wir für den Fall 1 ein Szenario, bei dem sich der Referenzpreis tatsächlich ändert. Dies geschieht bei einem ungeraden *Bid Ask Spread*. Hierbei nehmen wir an, dass der Zustand des LOB zum Zeitpunkt  $t = 0$  gegeben ist durch  $X_0 = (x_{-3}, x_{-2}, x_{-1}, x_1, x_2, x_3)$  mit  $x_i$  definiert als:



Der *Bid Ask Spread* beträgt in diesem Szenario  $3 \cdot \Delta_{tick}$  Ticks. Zu einem späteren Zeitpunkt  $T > 0$  wird eine neue Verkauf Limit Order zum Preislimit  $P_1$  in der Höhe von  $1 AES_1$  abgegeben. Da in diesem Fall gilt  $x_{-1} = 0$ , sinkt der Referenzpreis um  $\Delta_{tick}$ . Der Zustand des LOB nach der Order  $X_T = (x_{-3}^{Neu}, x_{-2}^{Neu}, x_{-1}^{Neu}, x_1^{Neu}, x_2^{Neu}, x_3^{Neu})$  wird in der folgenden Grafik veranschaulicht, wobei  $x_i^{Neu}$  gegeben ist als:



Um exogene Information, wie zum Beispiel Nachrichten, in unserem Modell zu berücksichtigen, nehmen wir an, dass im Falle einer vorliegenden Referenzpreisänderung der neue Zustand des LOB um den neuen Referenzpreis mit Wahrscheinlichkeit  $\theta^{reinit} \in [0, 1]$  von seiner invarianten Verteilung  $\pi$  gezogen wird. Dies widerspiegelt das Szenario, dass alle Marktteilnehmer sehr schnell ihre Positionen im LOB im Bezug auf die exogene Information ändern d.h. der gesamte Zustand des LOB wird von der invarianten Verteilung neu gezogen.

Das Modell III ist somit durch das Paar  $(X, p_{ref})$  bestimmt. Die wichtigsten Bestandteile des Modell III sind nochmals zusammenfassend:

1. Der Markovprozess  $X_t = (X_t^{-K}, \dots, X_t^{-1}, X_t^1, \dots, X_t^K)$ . Dieser wird jeweils in Zeiten eines konstanten Referenzpreises durch Modell I beschrieben.

2. Der stochastische Sprungprozess  $p_{ref}(t)$ .
3. Der Parameter  $\theta$ , welcher die Wahrscheinlichkeit einer Änderung des Referenzpreises bei gegebener Änderung des *Mid Prices* und unter der Voraussetzung, dass zu diesem Zeitpunkt das Ordervolumen des ersten Preislimits leer ist d.h.  $x_{\pm 1} = 0$ , bezeichnet.
4. Der Parameter  $\theta_{reinit}$ , welcher die Wahrscheinlichkeit eines neu gezogenen LOB Zustandes bei gegebener Referenzpreisänderung widerspiegelt.
5. Die Skalierungsfaktoren  $\{AES_{-K}, \dots, AES_{-1}, AES_1, \dots, AES_K\}$  mit  $AES_i \in \mathbb{R}_+, \forall i \in \{-K, \dots, -1, 1, \dots, K\}$ .

## 8.2 Simulation des Orderbuches - Modell III

In diesem Abschnitt präsentieren wir einen Algorithmus zur Simulation des gesamten Orderbuches und des dazugehörigen Referenzpreises im Modell III  $(X, p_{ref})$ . Dieser fasst im Wesentlichen die allgemeine Beschreibung des vorangegangenen Kapitels zusammen.

Neben dem in Abschnitt 7.1 vorgestellten Algorithmus zur Simulation eines zeitstetigen Markov Sprungprozesses werden wir im Folgenden eine zweite Möglichkeiten angeben, um einen solchen Prozess zu simulieren. Dieser dient dazu die Dynamik des Orderbuchvolumens  $X$  im Modell III während Zeiten eines konstanten Referenzpreises zu simulieren. Der Algorithmus beruht auf [8, Abschnitt 2.6].

Bezeichne im Folgenden weiterhin mit  $X$  den zeitstetigen  $2K$ -dimensionalen Markov Sprungprozess, welcher das Orderbuchvolumen modelliert d.h.  $X \sim \text{Markov}(\pi, Q)$ , wobei die Intensitätsmatrix  $Q$  definiert ist in Formel (3.1). Bezeichne weiters mit  $(J_n)_{n \in \mathbb{N}_0}$  die Sprungzeitpunkte, mit  $(S_n)_{n \in \mathbb{N}}$  die Halteperioden und mit  $Z_n := X_{J_n}$  die assoziierte diskrete Markovkette zu  $X$  mit der zu  $Q$  korrespondierenden Sprungmatrix  $\Pi$ . Um  $X$  beziehungsweise  $Z$  zu modellieren, kann man wie folgt vorgehen:

### 8.2.1 Simulation Markovprozess Methode II

Ziehe einen Anfangszustand  $x \in \mathbb{N}_0^{2K}$  von der invarianten Verteilung  $\pi$  und setze:  $\tilde{X}_0 = Z_0 := x$   
 $\tilde{J}_0 := 0$ .

Für  $n \geq 0$  und  $I := \{-K, \dots, -1, 1, \dots, K\}$ , falls  $\tilde{X}_{J_n} = \tilde{Z}_n = x = (x_{-K}, \dots, x_{-1}, x_1, \dots, x_K) \in \mathbb{N}_0^{2K}$  erzeuge  $4K$  unabhängig exponentialverteilte Zufallsvariablen  $\{(S_{n+1,j}^+, j \in I), (S_{n+1,j}^-, j \in I)\}$  mit folgenden Parametern

$$\begin{aligned} S_{n+1,j}^+ &\sim \text{Exp}(\lambda^L(x_j)) \\ S_{n+1,j}^- &\sim \text{Exp}(\lambda^M(x_j) + \lambda^C(x_j)) \end{aligned}$$

Die Zufallsvariable  $S_{n+1,j}^+ (S_{n+1,j}^-)$  symbolisiert hierbei die Zeit bis zu einem hypothetischen Sprung der Höhe  $+1 (-1)$  am Preislimit  $P_j$  falls der aktuelle Zustand des Orderbuches gegeben ist durch  $x \in \mathbb{N}_0^{2K}$ .

Definiere weiters:

$$\begin{aligned}\tilde{S}_{n+1} &:= \min_{j \in I} \{S_{n+1,j}^+, S_{n+1,j}^-\} \\ \tilde{J}_{n+1} &:= \tilde{J}_n + \tilde{S}_{n+1} \\ \tilde{Z}_{n+1} &:= \begin{cases} x + e_j & \tilde{S}_{n+1} = S_{n+1,j}^+ \\ x - e_j & \tilde{S}_{n+1} = S_{n+1,j}^- \\ x & \tilde{S}_{n+1} = \infty \end{cases} \\ \tilde{X}_{\tilde{J}_{n+1}} &:= \tilde{Z}_{n+1}\end{aligned}$$

Laut Konstruktion ist  $\tilde{Z}_{n+1}$  und  $\tilde{S}_{n+1}$  bedingt auf  $\{\tilde{Z}_n = x\}$  unabhängig von  $\sigma(\tilde{Z}_k, k \leq n-1)$  respektive  $\sigma(\tilde{S}_k, k \leq n)$ . Mittels Lemma C.1 folgt weiters, dass  $\tilde{S}_{n+1} | \{\tilde{Z}_n = x\} \sim \text{Exp}(Q_{x,x})$ . und die Verteilung von  $\tilde{Z}_{n+1}$  gegeben ist durch  $\Pi_{x,\cdot}$ .<sup>3</sup> Mittels [8, Theorem 2.8.4] folgt nun, dass  $\tilde{X} \stackrel{(d)}{=} X$ .

Um das gesamte Orderbuch im Modell III zu simulieren gehen wir wie folgt vor:

*Pseudocode zur Simulation im Modell III:*

1. Schritt: Definiere einen Startwert für den Referenzpreis  $p_{ref}(0) \in \frac{\Delta_{tick}}{2} \cdot \mathbb{N}$  und ziehe um diesen Referenzpreis von der invarianten Verteilung  $\pi$  einen Anfangszustand  $x^{Start} \in \mathbb{N}_0^{2K}$ . Setze  $X_0 = Z_0 := x^{Start}$  und berechne den *Mid Price*  $p_{mid}(0)$  aus dem aktuellen Zustand des Orderbuches  $X_0$ .
2. Für  $n \geq 1$  verwende Algorithmus 8.2.1, um einen einzelnen Sprung beziehungsweise eine einzelne Order zu simulieren. Speichere dabei die Zeit bis zur Order  $S_n \in \mathbb{R}_+$  und den neuen Zustand  $X_{J_n} = Z_n =: x \in \mathbb{N}_0^{2K}$ . Um den Typ der Order  $\mathcal{T}_n$  zu simulieren, gehe analog wie in Abschnitt 7.2 [4.Schritt (c)] vor. Berechne schließlich aus dem aktuellen Zustand  $x \in \mathbb{N}_0^{2K}$  den *Mid Price*  $p_{mid}(J_n)$ .
3. Falls  $p_{mid}(J_n) \neq p_{mid}(J_{n-1})$ : Ziehe  $U \sim U[0, 1]$ .
  - Falls  $U > \theta$  setze  $p_{ref}(J_n) := p_{ref}(J_{n-1})$  und gehe zu Schritt 2.
  - Falls  $U < \theta$ ,  $p_{mid}(J_n) > p_{mid}(J_{n-1})$  und  $x_1 = 0$ : Ziehe  $V \sim U[0, 1]$ :
    - Falls  $V > \theta_{reinit}$ : Setze  $p_{ref}(J_n) := p_{ref}(J_{n-1}) + \Delta_{tick}$  und bestimme den neuen Zustand des Orderbuches  $x^{neu}$  mithilfe von  $x$ ,  $\{AES_j, j \in \{-K, \dots, -1, 1, \dots, K\}\}$  und der invarianten Verteilung  $\pi_K$  wie in Abschnitt 8.1 beschrieben und setze  $X_{J_n} := x^{neu}$ .
    - Falls  $V < \theta_{reinit}$ : Setze  $p_{ref}(J_n) := p_{ref}(J_{n-1}) + \Delta_{tick}$  und ziehe das gesamte Orderbuch um  $p_{ref}(J_n)$  neu von der invarianten Verteilung d.h. ziehe  $X_{J_n} \sim \pi$
  - Falls  $U < \theta$  und  $p_{mid}(J_n) < p_{mid}(J_{n-1})$  und  $x_{-1} = 0$ : Ziehe  $V \sim U[0, 1]$ :
    - Falls  $V > \theta_{reinit}$ : Setze  $p_{ref}(J_n) := p_{ref}(J_{n-1}) - \Delta_{tick}$  und bestimme den neuen Zustand des Orderbuches  $x^{neu}$  mithilfe von  $x$ ,  $\{AES_j, j \in \{-K, \dots, -1, 1, \dots, K\}\}$

<sup>3</sup>Lemma C.1 liefert uns überdies auch die Unabhängigkeit der Zufallsvariablen  $\tilde{S}_{n+1} | \{\tilde{Z}_n = x\}$  und  $\tilde{Z}_{n+1} | \{\tilde{Z}_n = x\}$ .

und der invarianten Verteilung  $\pi_{-K}$  wie in Abschnitt 8.1 beschrieben und setze  $X_{J_n} := x^{neu}$ .

- Falls  $V < \theta_{reinit}$ : Setze  $p_{ref}(J_n) := p_{ref}(J_{n-1}) - \Delta_{tick}$  und ziehe das gesamte Orderbuch um  $p_{ref}(J_n)$  neu von der invarianten Verteilung d.h. ziehe  $X_{J_n} \sim \pi$

4. Gehe zu Schritt 2, sofern die gewünschte Anzahl an simulierten Orders noch nicht erreicht wurde.

Dieser Algorithmus generiert für eine gewünschte Anzahl an Orders  $N$ , die Zeiten der Order  $(S_n)_{n \leq N}$ , die Typen  $(\mathcal{T}_n)_{n \leq N}$ , den Zustand des Orderbuches  $(X_{J_n})_{n \leq N}$  und den Wert des Referenzpreises  $(p_{ref}(J_n))_{n \leq N}$ .

Die Funktion `simulMod3()` aus Kapitel 9, (14) realisiert diesen Pseudocode.

### 8.2.2 Kalibrierung der Parameter

In diesem Abschnitt besprechen wir kurz eine mögliche Variante, um die Parameter  $\theta$  und  $\theta_{reinit}$  von Modell III an die vorhandenen Daten anzupassen. Wir verwenden hierbei die Methode von [6, Abschnitt 3.1.1].

Um  $\theta$  und  $\theta_{reinit}$  für die Periode  $[0, t]$  zu kalibrieren verwenden wir folgende Kenngrößen:

- Die empirische Standardabweichung der diskreten Returns des *Mid Price*:  $p_{mid}(s)$  für  $s \in [0, t]$  zu einer festen Beobachtungsfrequenz. Wir werden diese im Folgenden mit  $SR$  bezeichnen und die verwendete Beobachtungsfrequenz immer explizit erwähnen.
- Die *Mean Reversion Ratio*  $\eta_t$ , welche definiert ist durch:

$$\eta_t := \frac{1}{2} \cdot \left( \frac{N^c(t)}{N^a(t)} \right) \quad N^a(t) \neq 0 \quad (8.2)$$

wobei wir mit  $N^c(t)$  die Anzahl jener Referenzpreisänderungen im betrachteten Intervall  $[0, t]$  bezeichnen, bei denen die vorhergehende Referenzpreisänderung die gleiche Richtung wie die Betrachtete aufweist (engl. *continuations*). Analog bezeichnet  $N^a(t)$  die Anzahl jener Referenzpreisänderungen, bei denen die vorhergehende Änderung die entgegengesetzte Richtung als die Betrachtete aufweist (engl. *alternations*).

Für nähere Details zum Parameter  $\eta$  verweisen wir auf [10]. Dort werden die *continuations* beziehungsweise *alternations* auf Basis eines Sprunges des *transaction price* von genau einem *Tick* berechnet. Da sich im Modell III der Referenzpreis bei einem Zustandsübergang lediglich um  $\pm \Delta_{tick}$  ändern kann, entspricht somit die Definition von [6] für simulierte Daten der von [10], wobei anstelle des *transaction price* der Referenzpreis betrachtet wird. In den Daten von LOBSTER kann es jedoch zu größeren Sprüngen als  $\Delta_{tick}$  kommen. Diese treten jedoch extrem selten auf, weshalb wir diese Ungenauigkeit vernachlässigen werden.

Im Folgenden werden wir die Parameter  $\theta$  und  $\theta_{reinit}$  für die Aktie FTE kalibrieren. Dazu berechnen wir zuerst die beiden Kenngrößen aus den LOBSTER Daten. Hierbei wählen wir eine Beobachtungsfrequenz von 3 Minuten. Wir berechnen diese beiden Größen für jeden der betrachteten Tage (488) und bilden danach den empirischen Mittelwert. Das Ergebnis wird in der folgenden Tabelle dargestellt.

Aktie	$\bar{\eta}_{LOB}$	$\overline{SR}_{LOB}$	$\overline{SR}_{LOB}$ in Basispunkten (bp)
FTE	0.257219	0.0010087	10.087

Als nächstes simulieren wir für verschiedene Werte von  $\theta$  und  $\theta_{reinit}$  mit der in Abschnitt 8.2.1 beschriebenen Methode für insgesamt 20 Tage je 60.000 Order pro Tag (dies entspricht in etwa einem durchschnittlichen Tag für die Aktie FTE). Bei der Simulation verwenden wir die geschätzten Parameter aus Abschnitt 6.1 (Intensitäten,  $AES_i, \dots$ ). Anschließend berechnen wir für jeden Tag die Kenngrößen  $\eta_M$  und  $SR_M$  und bilden den empirischen Mittelwert dieser. Als Startwert für den Referenzpreis wählen wir den durchschnittlichen Startwert aus den Daten, welcher gegeben ist durch 19.0350 USD. Abbildung 8.1 zeigt die bei der Simulation erhaltenen Werte von  $\overline{SR}_M$  (links) und  $\bar{\eta}_M$  (rechts).

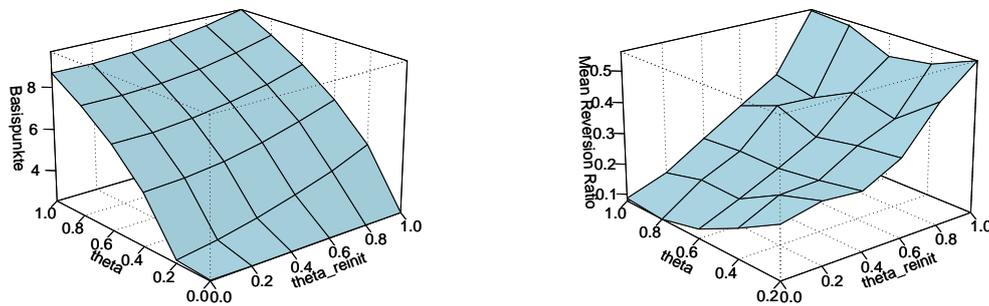


Abbildung 8.1: Kalibrierung der Parameter  $\theta$  und  $\theta_{reinit}$

Ausgewählte Werte werden in der folgenden Tabelle gezeigt:

$\theta$	$\theta_{reinit}$	$\bar{\eta}_M$	$\overline{SR}_M$
0	0	NA	0.00025
0.8	0.4	0.22	0.00078
1	0	0.08	0.00087
1	0.6	0.28	0.00090
1	1	0.55	0.00097

Hierbei ist ersichtlich, dass die Wahl von  $\theta := 1$  und  $\theta_{reinit} := 0.6$  eine brauchbare Approximation der empirischen Werte liefert, welche man natürlich durch ein feineres Gitter weiter verbessern könnte. Die *maximal mechanical volatility*, d.h. jener Wert für  $\overline{SR}_M$ , der ohne exogenen Einfluss ( $\theta := 1$  und  $\theta_{reinit} := 0$ ) erreicht wird beträgt 8.7 bp (vgl. [6, Abschnitt 3.2.1]).

Verwenden wir bei der Modellierung die in Abschnitt 7.3 an [6] angelehnten Intensitäten, sowie die in [6, Table 2.] präsentierten  $AES_i$  Größen, so erhalten wir folgende Abbildung:

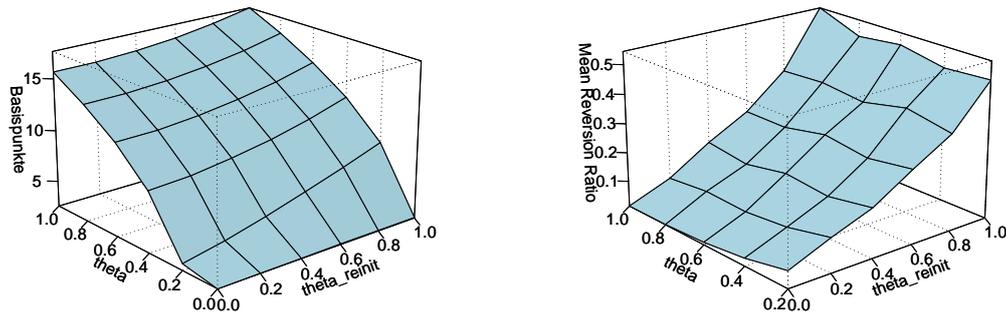


Abbildung 8.2: An [6] angelehnt: Kalibrierung der Parameter  $\theta$  und  $\theta_{reinit}$  (vgl. [6], Figure 10.)

Zum Schluss vergleichen wir für die Aktie FTE die empirische Sprungverteilung des mit Modell III zu den kalibrierten Parametern  $\theta = 1$  und  $\theta_{reinit} = 0.6$  simulierten Orderbuches mit der invarianten Verteilung  $\mu_i$  des Sprungprozesses  $Z$  und der empirischen Verteilung aus den Sprungdaten von LOBSTER.

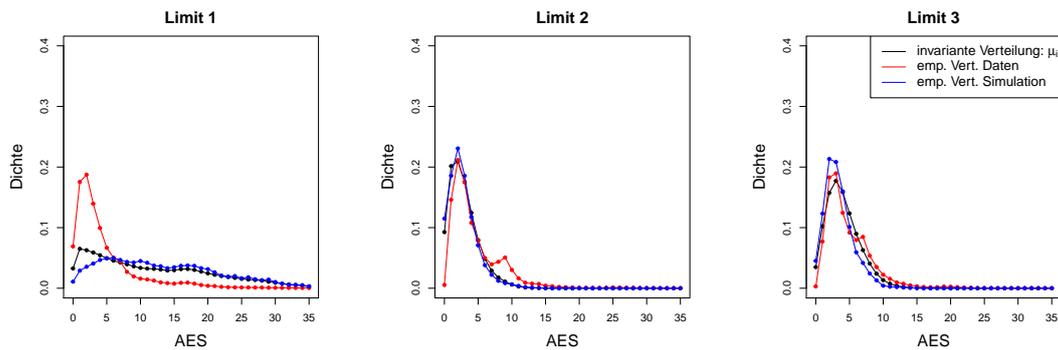


Abbildung 8.3: Vergleich der Verteilungen im Modell III

In Abbildung 8.3 sieht man, dass das Orderbuch im Modell III im Wesentlichen gleich verteilt ist, wie die invariante Verteilung  $\mu_i$  aus Modell I.

Zusammenfassend erhalten wir nach geeigneter Kalibrierung der Parameter  $\theta$  und  $\theta_{reinit}$  ein Modell, welches für die Simulation des LOB und des Referenzpreises der Aktie France Telekom für den Zeitraum von einem Tag geeignet ist.

# Kapitel 9

## Quellcode - R

In diesem Kapitel präsentieren wir einen Ausschnitt der wichtigsten Funktionen in R, die im Zuge der Arbeit verwendet wurden. Die Anwendung derer, ist im jeweiligen Abschnitt der vorigen Kapitel, immer explizit erwähnt. Die Funktionen werden aufeinander aufbauend aufgezählt.

Es wurde folgende Version von R verwendet: *R Version 3.3.2*

1. Die Funktion *refPrice()* berechnet, wie in Kapitel 4 beschrieben, zu gegebenem *Orderbookfile* von LOBSTER den Referenzpreis.

```
1 #Description of Arguments:
2 #OB...Orderbookfile obtained from LOBSTER
3
4 refPrice<-function(OB){
5
6     BAP<-OB[,1] #best ask price
7     BBP<-OB[,3] #best bid price
8     N<- length(BAP)
9     PR<- rep(0,N) # reference price
10    PR[1]<- (BBP[1]+BAP[1])/2+ifelse((BAP[1]-BBP[1])%%200==0,50,0)
11
12    for(I in 2:N){
13        PM<- (BBP[I]+BAP[I])/2
14        if((BAP[I]-BBP[I])%%200!=0){
15            PR[I]<- PM
16        } else {
17            PR1<- PM+50
18            PR2<- PM-50
19            PR[I]<- ifelse(abs(PR1-PR[I-1])<abs(PR2-PR[I-1]),PR1,PR2)
20        }
21    }
22    return(PR)
23 }
```

2. Die Funktion *getSample1()* berechnet aus den Daten von LOBSTER die für die Intensitätsschätzung benötigte Stichprobe für ein festes Preislimit (vgl. Abschnitt 5.2, Variante 1).

```

1 #Description of Arguments:
2 #OB...Orderbookfile obtained from LOBSTER
3 #MB...Messagebookfile obtained from LOBSTER, first row must be already
   deleted
4 #j...considered price limit
5
6 getSample1<-function(OB,MB,j){
7
8   N<-(nrow(OB)-1)
9   AP<- OB[1:N,c(1,5,9)] # best nonempty ask prices
10  AS<- OB[1:N,c(2,6,10)] # best nonempty ask sizes
11  BP<- OB[1:N,c(3,7,11)] # best nonempty bid prices
12  BS<- OB[1:N,c(4,8,12)] # best nonempty bid sizes
13
14  T<- MB[,1] # time
15  Y<- MB[,2] # type
16  S<- MB[,4] # size
17  P<- MB[,5] # price
18  TD<- MB[,6] # tradeDirection
19
20  REFP<-refPrice(OB[1:N,]) # calculate reference price using the
   function refPrice()
21
22  Z<-abs(P-REFP) #calculate the distance from the orderprice to the
   reference price
23  W<- which( (Z-(50+(j-1)*100)) == 0 ) # indices with ask_j or bid_j
   events
24  WA<- W[TD[W]==-1] # indices with Ask_j events
25  WB<- W[TD[W]==1] # indices with Bid_j events
26
27  DT<- rep(NA,length(W)) # timediff
28  DT[TD[W]==-1]<- c((T[WA])[1]-T[1],diff(T[WA]))
29  DT[TD[W]==1]<- c((T[WB])[1]-T[1],diff(T[WB]))
30
31  L<- AS[W,1]*(AP[W,1]==P[W])+BS[W,1]*(BP[W,1]==P[W])+AS[W,2]*(AP[W
   ,2]==P[W])+BS[W,2]*(BP[W,2]==P[W])+AS[W,3]*(AP[W,3]==P[W])+BS[W
   ,3]*(BP[W,3]==P[W]) # last size in AES
32
33  return(cbind("Timediff"=DT,"Lastsize"=L,"Type"=Y[W],"Eventsize"=S[W
   ],"Trade Direction"=TD[W]))
34 }

```

3. Die Funktion *getSample2()* berechnet aus den Daten von LOBSTER die für die Intensitätsschätzung benötigte Stichprobe für ein festes Preislimit (vgl. Abschnitt 5.2, Variante 2).

```

1 #Description of Arguments:
2 #OB...Orderbookfile obtained from LOBSTER

```

```

3 #MB...Messagebookfile obtained from LOBSTER, first row must be already
  deleted
4 #j...considered price limit
5
6 getSample2<-function(OB,MB,j){
7
8   RES<-NULL #result
9   REFP<-refPrice(OB[1:(nrow(OB)-1),]) # calculate reference price
    using the function refPrice()
10  I<-c(0,which(diff(REFP)!=0),nrow(OB)-1) # indices where reference
    price changes
11
12  #loop over blocks with constant reference price
13  for (k in 1:(length(I)-1)){
14    J<-(I[k]+1):(I[k+1]) # current block
15
16    if(length(J)==1){next} # do not consider blocks consisting of a
      single row
17
18    AP<- OB[J,c(1,5,9)] # best nonempty ask prices
19    AS<- OB[J,c(2,6,10)] # best nonempty ask sizes
20    BP<- OB[J,c(3,7,11)] # best nonempty bid prices
21    BS<- OB[J,c(4,8,12)] # best nonempty bid sizes
22
23    T<- MB[J,1] # time
24    Y<- MB[J,2] # type
25    S<- MB[J,4] # size
26    P<- MB[J,5] # price
27    TD<- MB[J,6] # tradeDirection
28
29    Z<-abs(P-REFP[J]) #calculate the distance from the orderprice to
      the reference price
30    W<- which( (Z-(50+(j-1)*100)) == 0 ) # indices with ask_j or bid_j
      events
31
32    if(length(W)==0){next} # no existing orders for this limit
33
34    WA<- W[TD[W]==-1] # indices with ask_j events
35    WB<- W[TD[W]==1] # indices with bid_j events
36
37    DT<- rep(NA,length(W)) # timediff
38    STARTASK<-ifelse( (WA[1]==1)|(is.na(WA[1])) , mean(diff(T)) , (T[
      WA][1]-T[1]) ) # timediff for the first askorder of current
      block
39    STARTBID<-ifelse( (WB[1]==1)|(is.na(WB[1])) , mean(diff(T)) , (T[
      WB][1]-T[1]) ) # timediff for the first bidorder of current
      block
40    DT[TD[W]==-1]<- c( STARTASK , diff(T[WA]) )
41    DT[TD[W]==1]<- c( STARTBID , diff(T[WB]) )

```

```

42
43   L<- AS[W,1]*(AP[W,1]==P[W])+BS[W,1]*(BP[W,1]==P[W])+AS[W,2]*(AP[W
,2]==P[W])+BS[W,2]*(BP[W,2]==P[W])+AS[W,3]*(AP[W,3]==P[W])+BS[W
,3]*(BP[W,3]==P[W]) # last size in AES
44
45   RES<-rbind(RES,cbind(DT,L,Y[W],S[W],TD[W])) #RESULT
46
47 }
48 colnames(RES)<-c("Timediff","Lastsize","Type","Eventsize","Trade
Direction")
49 return(RES)
50 }

```

4. Die Funktion *cumOrder()* kumuliert die Ordersize von Ordnern mit gleichem Zeitstempel, Typ, Preis und gleicher *Trade Direction* (vgl. Abschnitt 5.2, Variante 3). Diese ist lediglich eine Hilfsfunktion für die Funktion *getSample3()* (siehe unten).

```

1 #Description of Arguments:
2 #OB...Orderbookfile obtained from LOBSTER
3 #MB...Messagebookfile obtained from LOBSTER, first row must be already
  deleted
4
5 cumOrder<-function(OB,MB){
6
7   OB<-OB[1:(nrow(OB)-1),]
8   I<-c(0,(diff(MB[,1])==0) & (diff(MB[,2])==0) & (diff(MB[,5])==0) & (
  diff(MB[,6])==0)) # same time&type&price&tradeDirection
9   J<-which(diff(I)==1)
10  for(j in J){
11    i<-j+1
12    while(I[i]==1){
13      MB[j,4]<-MB[j,4]+MB[i,4] # cumulate ordersize
14      ifelse(i<nrow(OB),i<-i+1,break)
15    }
16  }
17  return(list(OB[-which(I==1),],MB[-which(I==1),]))
18 }

```

5. Die Funktion *getSample3()* berechnet aus den Daten von LOBSTER die für die Intensitätsschätzung benötigte Stichprobe für ein festes Preislimit (vgl. Abschnitt 5.2, Variante 3).

```

1 #Description of Arguments:
2 #OB...Orderbookfile obtained from LOBSTER
3 #MB...Messagebookfile obtained from LOBSTER, first row must be already
  deleted
4 #j...considered price limit
5
6 getSample3<-function(OB,MB,j){
7

```

```

8   tmp<-cumOrder(OB,MB)
9   OB<-tmp[[1]] # OB with cumulated values
10  MB<-tmp[[2]] # MB with cumulated values
11
12  RES<-NULL # result
13  REFP<-refPrice(OB) # calculate reference price
14  I<-c(0,which(diff(REFP)!=0),nrow(OB)-1) # indices where reference
    price changes
15
16  #loop over blocks
17  for (k in 1:(length(I)-1)){
18
19      J<-(I[k+1]+1):(I[k+1]-1) # current block, do not consider order that
    caused a reference price change
20
21      if(length(J)==1|(J[1]>J[2])){next} # do not consider blocks
    consisting of a single row
22
23      AP<- OB[J,c(1,5,9)] # best nonempty ask prices
24      AS<- OB[J,c(2,6,10)] # best nonempty ask sizes
25      BP<- OB[J,c(3,7,11)] # best nonempty bid prices
26      BS<- OB[J,c(4,8,12)] # best nonempty bid sizes
27
28      T<- MB[J,1] # time
29      Y<- MB[J,2] # type
30      S<- MB[J,4] # size
31      P<- MB[J,5] # price
32      TD<- MB[J,6] # tradeDirection
33
34      Z<-abs(P-REFP[J]) # calculate the distance from the orderprice to
    the reference price
35      W<- which( (Z-(50+(j-1)*100)) == 0 ) # indices with Ask_j or Bid_j
    events
36
37      if(length(W)==0){next} # no existing orders for this limit
38
39      WA<- W[TD[W]==-1] # indices with ask_j events
40      WB<- W[TD[W]==1] # indices with bid_j events
41
42      DT<- rep(NA,length(W)) # timediff
43      STARTASK<-ifelse( (WA[1]==1)|(is.na(WA[1])) , mean(diff(T)) , (T[
    WA])[1]-T[1] ) # timediff for the first askorder of current
    block
44      STARTBID<-ifelse( (WB[1]==1)|(is.na(WB[1])) , mean(diff(T)) , (T[
    WB])[1]-T[1] ) # timediff for the first bidorder of current
    block
45      DT[TD[W]==-1]<- c( STARTASK , diff(T[WA]) )
46      DT[TD[W]==1]<- c( STARTBID , diff(T[WB]) )
47

```

```

48   L<- AS[W,1]*(AP[W,1]==P[W])+BS[W,1]*(BP[W,1]==P[W])+AS[W,2]*(AP[W
      ,2]==P[W])+BS[W,2]*(BP[W,2]==P[W])+AS[W,3]*(AP[W,3]==P[W])+BS[W
      ,3]*(BP[W,3]==P[W]) # last size in AES
49
50   RES<-rbind(RES,cbind(DT,L,Y[W],S[W],TD[W])) #RESULT
51 }
52 colnames(RES)<-c("Timediff","Lastsize","Type","Eventsize","Trade
      Direction")
53 return(RES)
54 }

```

6. Die Funktion *Int()* berechnet zu gegebener Stichprobe die Maximum Likelihood Schätzer der Intensitäten (vgl. (3.3) - (3.5)) für alle betrachteten Preislimits und bildet das Konfidenzintervall aus Gleichung (3.7).

```

1 #Description of Arguments:
2 #M...list with sample obtained from function get_sample1(),get_sample2
      () or get_sample3()
3 #nmax...maximal considered AES for estimation
4 #alpha...parameter for (1-alpha)-confidence interval
5
6 Int<-function(M,nmax,alpha){
7
8   laL<-matrix(0,nrow=nmax+1,ncol=3) # Limit Order Intensity
9   laM<-matrix(0,nrow=nmax+1,ncol=3) # Market Order Intensity
10  laC<-matrix(0,nrow=nmax+1,ncol=3) # Cancellation Order Intensity
11  CID_L<-matrix(0,nrow=nmax+1,ncol=3) # upper confidence band for LO
12  CIU_L<-matrix(0,nrow=nmax+1,ncol=3) # lower confidence band for LO
13  CID_M<-matrix(0,nrow=nmax+1,ncol=3) # upper confidence band for MO
14  CIU_M<-matrix(0,nrow=nmax+1,ncol=3) # lower confidence band for MO
15  CID_C<-matrix(0,nrow=nmax+1,ncol=3) # upper confidence band for CO
16  CIU_C<-matrix(0,nrow=nmax+1,ncol=3) # lower confidence band for CO
17
18  z<-qnorm(1-alpha/2) # quantile for confidence interval
19  nlimits<-length(M)
20
21  # loop over different limits
22  for (j in 1:nlimits){
23    X<-M[[j]] # select matrix for limit j of list M
24    DT<-X[,1] # timediff
25    L<-X[,2] # lastsize (not yet in AES)
26    Y<-X[,3] # type
27    S<-X[,4] # eventsize (not yet in AES)
28    D<-X[,5] # trade direction
29    I<-which(D==-1) # indices of ask events
30    Q<-ceiling(L/mean(S[I]))*(D==-1)+ceiling(L/mean(S[-I]))*(D==1) #
      calculate lastsize in AES
31    rm(X,L,S,D,I)
32
33    #ML estimation

```

```

34   for(n in 0:nmax){
35     NQ<- sum(Q==n)
36     sq_NQ<-sqrt(NQ)
37     if(NQ>0){
38       La<- 1/mean(DT[Q==n])
39       laL[n+1,j] <- La*sum((Q==n)&(Y==1))/NQ
40       laM[n+1,j] <- La*sum((Q==n)&((Y==4)))/NQ
41       laC[n+1,j] <- La*sum((Q==n)&((Y==2)|(Y==3)))/NQ
42       CID_L[n+1,j] <- laL[n+1,j] - (z*La*sqrt(laL[n+1,j]/La*(2-laL[n+1,j]
43         )/La))/sq_NQ # lower confidence band for L0
44       CIU_L[n+1,j] <- laL[n+1,j] + (z*La*sqrt(laL[n+1,j]/La*(2-laL[n+1,j]
45         )/La))/sq_NQ # upper confidence band for L0
46       CID_M[n+1,j] <- laM[n+1,j] - (z*La*sqrt(laM[n+1,j]/La*(2-laM[n+1,j]
47         )/La))/sq_NQ # lower confidence band for M0
48       CIU_M[n+1,j] <- laM[n+1,j] + (z*La*sqrt(laM[n+1,j]/La*(2-laM[n+1,j]
49         )/La))/sq_NQ # upper confidence band for M0
50       CID_C[n+1,j] <- laC[n+1,j] - (z*La*sqrt(laC[n+1,j]/La*(2-laC[n+1,j]
51         )/La))/sq_NQ # lower confidence band for C0
52       CIU_C[n+1,j] <- laC[n+1,j] + (z*La*sqrt(laC[n+1,j]/La*(2-laC[n+1,j]
53         )/La))/sq_NQ # upper confidence band for C0
54     }
55   }
56   rm(Q,DT,Y)
57 }
58 return(list(laL,laM,laC,CID_L,CIU_L,CID_M,CIU_M,CID_C,CIU_C))
59 }

```

7. Die Funktion *InvDistrCont()* berechnet die invariante Verteilung  $\pi$  des zeitstetigen Markovprozesses  $X$  aus Theorem 3.2.

```

1 #Description of Arguments:
2 #laL,laM,laC...Maximum likelihood estimator of intensities for all
3   considered price limits obtained from the function Int().
4 #laL,laM,laC are matrices with ncol(laL) == considered limits and nrow
5   (laL) == maximal considered volume in AES.
6
7 InvDistrCont<-function(laL,laM,laC){
8
9   nmax<-nrow(laL)-1
10  nlimits<-ncol(laL)
11  pi<-matrix(0,nrow=nmax+1,ncol=3) # invariant distribution pi
12
13  #loop over limits
14  for (j in 1:nlimits){
15    RHO<-laL[(1:nmax),j]/( laM[(2:(nmax+1)),j] + laC[(2:(nmax+1)),j] )
16    # arrival departure vector
17    TEMP<-cumprod(RHO)
18    pi[1,j] <- 1/(1+sum(TEMP))
19    pi[(2:(nmax+1)),j] <- pi[1,j]*TEMP
20  }
21 }

```

```

18   return(pi)
19 }

```

8. Die Funktion *InvDistrJump()* berechnet die invariante Verteilung  $\mu$  der zeitdiskreten Markovkette  $Z$  aus Theorem 3.3.

```

1 #Description of Arguments:
2 #laL,laM,laC...Maximum likelihood estimator of intensities for all
   considered price limits obtained from function Int()
3 #laL,laM,laC are matrices with ncol(laL) == considered limits and nrow
   (laL)== maximal considered volume in AES.
4
5 InvDistrJump<-function(laL,laM,laC){
6
7   nmax<-nrow(laL)-1
8   nlimits<-ncol(laL)
9   mu<-matrix(0,nrow=nmax+1,ncol=3) # invariant distribution mu
10
11   LAMBDA<-(laL+laM+laC)
12
13   UP<-laL/LAMBDA # probability of a jump of size +1
14   UP[1,]<-1
15   UP[nmax+1,]<-0
16
17   DOWN<-(laM+laC)/LAMBDA # probability of a jump of size -1
18   DOWN[nmax+1,]<-1
19
20   RATIO<-UP[1:nmax,]/DOWN[-1,]
21   TEMP<-apply(rbind(1,RATIO),2,FUN="cumprod")
22   mu<-TEMP/rep(colSums(TEMP),each=nmax+1)
23
24   return(mu)
25 }

```

9. Die Funktion *triDiag()* erstellt zu gegebenen Haupt und Nebendiagonalen eine Tridiagonalmatrix. Diese ist eine Hilfsfunktion für die Funktion *Qmat()* (siehe unten).

```

1 #Description of Arguments:
2 #upper...desired upper diagonal
3 #lower...desired lower diagonal
4 #main...desired main diagonal, it has to hold that length(upper/lower)
   ==length(main) -1
5
6 triDiag <- function(upper, lower, main){
7   mat <- matrix(0,length(main),length(main))
8   diag(mat) <- main
9   indx <- 1:length(upper)
10  mat[cbind(indx+1,indx)] <- lower
11  mat[cbind(indx,indx+1)] <- upper
12  return(mat)

```

```
13 }
```

10. Die Funktion  $Qmat()$  erstellt zu gegebenen Intensitäten eines festen Preislimits für Limit Orders, Market Orders und Cancellation Orders die Intensitätsmatrix  $Q^i$  aus Modell I (siehe (3.2)).

```
1 #Description of Arguments:
2 #laL,laM,laC...vector (not matrix) of intensities for considered price
   limit, obtained via the function Int(), take only a single column
   from the output i.e. Int()[[1-3]][,i]
3
4 Qmat<-function(laL,laM,laC){
5
6   N<-length(laL)-1
7   lam_plus<-laL
8   lam_minus<-laC+laM
9
10  main<-c(-lam_plus[1],-(lam_plus[2:N]+lam_minus[2:N]),-lam_minus[N
   +1]) # main diagonal
11  upper<-lam_plus[1:N] # upper diagonal
12  lower<-lam_minus[2:(N+1)] # lower diagonal
13
14  Q<-triDiag(upper,lower,main) # create (finite dimensional) intensity
   matrix
15
16  return(Q)
17 }
```

11. Die Funktion  $PImat()$  erstellt zu gegebener Intensitätsmatrix  $Q^i$  für ein festes Preislimit  $i$  die korrespondierende Sprungmatrix  $\Pi^i$  aus Modell I (siehe (3.12)).

```
1 #Description of Arguments:
2 #Q...intensity matrix obtained from the function Qmat()
3
4 PImat<-function(Q){
5   PI<-Q/-diag(Q)
6   diag(PI)<-0
7   return(PI)
8 }
```

12. Die Funktion  $simulMod1()$  simuliert ein einzelnes Preislimit  $i$  im Modell I (vgl. Abschnitt 7.2) bis zu einer maximalen betrachteten Größe in AES.

```
1 #Description of Arguments:
2 #Q...intensity matrix for a single price limit, obtained from the
   function Qmat()
3 #PI...corresponding jump matrix for a single price limit, obtained
   from the function PImat()
4 #pi...invariant distribution of  $X^i$  obtained via the function
   InvDistrCont()[,i] or InvDistrJump()[,i] (only i-th column of
   result)
```

```

5 #0...number of desired orders to be simulated
6 #laC,laM...corresponding intensities for this price limit, obtained
   via the function Int()[[1-3]] [,i](only ith-column of result)
7
8 simulMod1<-function(Q,PI,pi,0,laC,laM){
9
10  X<-numeric(0+1)
11  TIMEDIFF<-numeric(0+1)
12  TYP<-numeric(0+1)
13  N<-length(laC)-1    # number of maximal considered AES for
   simulation
14  q<-diag(Q)          # diagonal entries of the intensity matrix Q
15  ud<-c(diag(PI[, -1]),0) # probability of a jump of size +1
16  ld<-c(0,diag(PI[-1,])) # probability of a jump of size -1
17
18  TYP[1]<-0           # initial type
19  TIMEDIFF[1]<-0     # initial time
20  X[1]<-sample((0:N),1,prob=pi) # initial value
21  i<-X[1]
22
23  for(j in 2:0){
24    TIMEDIFF[j]<-rexp(1,rate=-q[i+1]) # next time of a jump
25    X[j]<-sample(c(i+1,i-1),1,prob=c(ud[i+1],ld[i+1])) # next state
26    if(X[j]==i+1){ TYP[j]<-1 } # determine
   type of the order: LO
27    if(X[j]==i-1){ TYP[j]<-sample((2:3),1,prob= c(laC[i+1]/(laM[i+1]+
   laC[i+1]),laM[i+1]/(laM[i+1]+laC[i+1])) ) } #determine type of
   the order (CO=2 | MO=3)
28    i<-X[j]
29  }
30  return(cbind(X, TIMEDIFF, "TIME"=cumsum(TIMEDIFF), TYP))
31 }

```

13. Die Funktion *fillOB()* dient als Hilfsfunktion für *simulMod3()* und befüllt zu gegebenem Referenzpreis und Volumen das Orderbuch.

```

1 #Description of Arguments:
2 #pref...considered reference price, defines center of the LOB
3 #values...volumes for different limits values=(ask1,bid1,ask2,bid2
   ,...)
4
5 fillOB<-function(pref, values){
6
7  n<-length(values) # volumes for the orderbook
8  OB<-numeric(n*2)
9
10  p<-seq(from=1,by=2,length.out = n) # prices
11  s<-seq(from=2,by=2,length.out = n) # sizes
12  OB[p]<-pref+rep(seq(from=50,by=100,length.out = n/2),each=2)*c
   (+1,-1) # fill prices

```

```

13  OB[s]<-values # fill volume
14
15  # calculate mid price for given reference price and values
16  ask<-seq(2,by=4,length.out = n/2)
17  bid<-seq(4,by=4,length.out = n/2)
18  IA<-ifelse(any(OB[ask]!=0),min(which(OB[ask]!=0)),n/2+1)
19  IB<-ifelse(any(OB[bid]!=0),min(which(OB[bid]!=0)),n/2+1)
20
21  # if there are only empty limits on one or both sides (Ask/Bid) we
    assume that the next limit is not empty
22  if((IA==(n/2+1))&!(IB==(n/2+1))){
23    pmid<-((OB[ask[IA-1]-1]+100)+OB[bid[IB]-1])/2
24  }else if(!(IA==(n/2+1))&(IB==(n/2+1))){
25    pmid<-((OB[ask[IA]-1]+(OB[bid[IB-1]-1]-100))/2
26  }else if((IA==(n/2+1))&(IB==(n/2+1))){
27    pmid<-((OB[ask[IA-1]-1]+100)+(OB[bid[IB-1]-1]-100))/2
28  }else{pmid<-((OB[ask[IA]-1]+OB[bid[IB]-1])/2
29  }
30  return(c(OB,pref,pmid))
31 }

```

14. Die Funktion *simulMod3()* simuliert das gesamte Orderbuch im Modell III (vgl. Abschnitt 8.2).

```

1  #Description of Arguments:
2  #laL,laC,laM...corresponding intensities (matrices) obtained via the
    function Int()[[1-3]]
3  #theta...probability of a reference price change
4  #thetaR...probability of a reinitialisation
5  #AES...vector with estimated AES_i per price limit i
6  #n0...number of orders to be simulated
7  #pref...initial reference price
8  #info...boolean variable for printing detailed info about reference
    change
9
10 simulMod3<-function(laL,laC,laM,theta,thetaR,AES,n0,pref,info){
11
12  laL[nrow(laL),]<-0 # intensity for a jump of +1 at the maximal
    considered size in AES is zero
13  N<-nrow(laL)-1 # maximal considered size in AES for simulation
14  nlimits<-ncol(laL) # considered limits
15
16  OB<-matrix(0,nrow=n0,ncol=(4*nlimits+6)) # resulting output, and
    formatting
17  columns <- c("ASKp1", "ASKs1", "BIDp1", "BIDs1")
18  for (i in 2:ncol(laL)){columns <- c (columns,paste("ASKp",i,sep=""),
    paste("ASKs",i,sep=""),paste("BIDp",i,sep=""),paste("BIDs",i,sep
    =""))}
19  colnames(OB)<-c(columns,"PREF","PMID","REINIT","TIME","TYP","LIMIT")
20

```

```

21 pi<-InvDistrCont(laL,laM,laC) # calculate the invariant distribution
    pi
22 X<-numeric(2*nlimits)
23
24 for(j in 1:nlimits){          # initialize orderbook, draw from
    invariant distribution pi
25   X[(1:2)+2*(j-1)]<-sample(0:N,2,prob=pi[,j])
26 }
27 t<-0
28 TYP<-0
29 limit<-0
30 OB[1,]<-c(fillOB(pref,X),0,t,TYP,limit) # fill the first row of the
    orderbook
31 q<-X          # current volume
32
33 # loop over simulations
34 for (i in 1:(n0-1)){
35
36   I<-cbind(q+1,rep(1:3,each=2)) # evaluate right indices for
    intensities
37   UP<-laL[I]          # intensities of a jump of size +1
38   DOWN<-(laM+laC)[I] # intensities of a jump of size -1
39
40   e<-suppressWarnings(rexp(12,rate=c(UP,DOWN))) # draw exponential
    random variables
41   t<-t+min(e,na.rm=TRUE) # calculate time of the next jump
42   l<-which.min(e) # evaluate which limit is changed 1-6
    corresponds to a jump of size +1 (7-12 to a jump of size -1)
43
44   # formatting for output
45   if(1%%6==1){limit<-1} # ASK1
46   if(1%%6==2){limit<--1} # BID1
47   if(1%%6==3){limit<-2} # ASK2
48   if(1%%6==4){limit<--2} # BID2
49   if(1%%6==5){limit<-3} # ASK3
50   if(1%%6==0){limit<--3} # BID3
51
52   # change volume at the evaluated limit
53   if(l<=6){
54     q[l]<-q[l]+1
55     TYP<-1 # evaluate typ of order = L0
56   }
57   if(l>6){
58     q[(1-6)]<-q[(1-6)]-1
59     K<-matrix(I[(1-6)],,ncol=2)
60     TYP<-sample(c(2,3),1,prob=c(laC[K]/(laM[K]+laC[K]),laM[K]/(laM[K]
    ]+laC[K]))) # evaluate typ of order (2=C0 | 3=M0)
61   }
62   tempOB<-fillOB(pref,q) # create temporary orderbook

```

```

63 U1<-(runif(1)<=theta) # draw theta
64
65 # [1] szenario of a decreasing mid price, where the reference
    price can also decrease
66 if( (U1)&(tempOB[4*nlimits+2]<OB[i,4*nlimits+2])&(q[2]==0)){
67   pref<-pref-100 # reference price decreases, 1Tick=100
68   U2<-(runif(1)<=thetaR) # draw theta_Reinit
69
70 # [1][1] szenario of a reinitialisation
71 if(U2){
72   if(info){print(paste("PREF gesunken nach Order",i,"+ REINIT"))
    }
73   X<-numeric(2*nlimits)
74   for(j in 1:nlimits){
75     X[(1:2)+2*(j-1)]<-sample(0:N,2,prob=pi[,j])
76   }
77   OB[(i+1),]<-c(fillOB(pref,X),1,t,TYP,limit) # fill the
    orderbook with the drawn reinitialisation
78   q<-X #current state
79 }
80 # [1][2] szenario of no reinitialisation
81 if(!U2){
82   if(info){print(paste("PREF gesunken nach Order",i))}
83   X<-numeric(2*nlimits)
84   X[1]<-0 # per definition
85   X[2*nlimits]<-sample(0:N,1,prob=pi[,nlimits]) # draw X_
    {-3} from the invariant distribution pi
86   bid<-seq(from=2,by=2,length.out=(nlimits-1)) # indices for
    bid side
87   ask<-seq(from=3,by=2,length.out = (nlimits-1)) # indices for
    ask side
88   X[bid]<-pmin(ceiling( q[bid+2]*(AES[bid+2]/AES[bid]) ),N)
    # scaling
89   X[ask]<-pmin(ceiling( q[ask-2]*(AES[ask-2]/AES[ask]) ),N)
    # scaling
90   OB[(i+1),]<-c(fillOB(pref,X),0,t,TYP,limit) # new
    orderbook
91   q<-X # current state
92 }
93
94 # [2] szenario of an increasing mid price, where the reference
    price can also increase
95 }else if( (U1)&(tempOB[4*nlimits+2]>OB[i,4*nlimits+2])&(q[1]==0) )
    {
96   pref<-pref+100 # reference price increases, 1Tick=100
97   U2<-(runif(1)<=thetaR) # draw theta_reinit
98
99 # [2][1] szenario of a reinitialisation
100 if(U2){

```

```

101     if(info){print(paste("PREF gestiegen nach Order",i,"+ REINIT")
102         )}
103     X<-numeric(2*nlimits)
104     for(j in 1:nlimits){
105         X[(1:2)+2*(j-1)]<-sample(0:N,2,prob=pi[,j])
106     }
107     OB[(i+1),]<-c(fillOB(pref,X),1,t,TYP,limit) # fill the
108         orderbook with the drawn reinitialisation
109     q<-X # current state
110 }
111 # [2][2] szenario of no reinitialisation
112 if(!U2){
113     if(info){print(paste("PREF gestiegen nach Order",i))}
114     X<-numeric(2*nlimits)
115     X[2]<-0 #per definition
116     X[(2*nlimits-1)]<-sample(0:N,1,prob=pi[,nlimits]) # draw X_
117         {3} from the invariant distribution pi
118     bid<-seq(from=4,by=2,length.out=(nlimits-1)) # indices for
119         bid side
120     ask<-seq(from=1,by=2,length.out = (nlimits-1)) # indices for
121         ask side
122     X[bid]<-pmin(ceiling( q[bid-2]*(AES[bid-2]/AES[bid]) ),N)
123         # scaling
124     X[ask]<-pmin(ceiling( q[ask+2]*(AES[ask+2]/AES[ask]) ),N)
125         # scaling
126     OB[(i+1),]<-c(fillOB(pref,X),0,t,TYP,limit) # new
127         orderbook
128     q<-X # current state
129 }
130 }
131 # [3] szenario no change of the reference price possible
132 }else{
133     OB[(i+1),]<-c(fillOB(pref,q),0,t,TYP,limit) #new orderbook
134 }
135 }
136 return(OB)
137 }

```

# Anhang A

## Markovprozesse mit abzählbarem Zustandsraum

Im Folgenden Kapitel betrachten wir einen zeitstetigen homogenen und rechtsstetigen Markovprozess  $(X_t)_{t \geq 0}$  auf einem abzählbarem Zustandsraum  $(E, \mathcal{E} := \mathcal{P}(E))$  mit Intensitätsmatrix  $Q$ , beliebiger Anfangsverteilung  $\lambda$  und Übergangshalbgruppe  $(P(t))_{t \geq 0} : P_{e_1, e_2}(t) = \mathbb{P}[X_t = e_2 | X_0 = e_1] =: \mathbb{P}_{e_1}[X_t = e_2]$ .

**Definition A.1** (*invariantes Maß*) Ein sigma endliches Maß  $\pi$  auf  $(E, \mathcal{E})$  mit der Eigenschaft:

$$\pi(e) = (\pi P(t))(e) \quad \forall t \geq 0, \forall e \in \mathcal{E} \quad (\text{A.1})$$

heißt *invariantes Maß* des Markovprozesses  $X$ .

Im Folgenden werden wir mit  $\pi$  immer, sofern es existiert, das invariante Maß des Markovprozesses  $X$  benennen.

Für das nächste Lemma benötigen wir folgende Definition.

**Definition A.2** Sei  $t \in \mathbb{R}_+$ . Dann bezeichne mit  $\tau(t)$  die verbleibende Zeit ab dem Zeitpunkt  $t$  bis zum nächsten Sprung des Markovprozesses  $X$  d.h.:

$$\tau(t) := \inf\{s > 0 : X_{t+s} \neq X_t\} \quad (\text{A.2})$$

Aufgrund der Rechtsstetigkeit des Markovprozesses  $X$  ist dies für alle  $t \in \mathbb{R}_+$  eine Stoppzeit.

Von nun an bezeichnen wir mit  $(J_n)_{n \in \mathbb{N}}$  die Folge der Sprungzeitpunkte und mit  $(S_n)_{n \in \mathbb{N}}$  die Folge der Halteperioden des Markovprozesses  $X$  (siehe Definition 2.8 und Definition 2.9). Das folgende Lemma zeigt, dass bedingt auf einen Zustand des Markovprozesses die Halteperioden bis zum nächsten Sprung und der Zustand, zu dem er springt, unabhängig voneinander sind.

**Lemma A.3** Sei  $X$  nicht explosiv und  $J_{k+1} < \infty$   $\mathbb{P}$ -f.s. für  $k \in \mathbb{N}$ . Dann gilt für  $e_1 \in E$ , dass die Zufallsvariablen  $S_{k+1}$  und  $X_{J_{k+1}}$  bedingt auf  $\{X_{J_k} = e_1\}$  unabhängig voneinander sind d.h.:

$$\mathbb{P}[S_{k+1} \leq t, X_{J_{k+1}} = e_2 | X_{J_k} = e_1] = \mathbb{P}[S_{k+1} \leq t | X_{J_k} = e_1] \cdot \mathbb{P}[X_{J_{k+1}} = e_2 | X_{J_k} = e_1] \quad (\text{A.3})$$

für alle  $t \in \mathbb{R}_+$  und  $e_2 \in E$ .

**Beweis.** Die linke Seite von (A.3) kann mittels der Bayes Formel folgendermaßen umgeformt werden:

$$\mathbb{P}[S_{k+1} > t, X_{J_{k+1}} = e_2 | X_{J_k} = e_1] = \mathbb{P}[S_{k+1} > t | X_{J_k} = e_1] \cdot \mathbb{P}[X_{J_{k+1}} = e_2 | S_{k+1} > t, X_{J_k} = e_1]$$

Da auf der Menge  $\{X_{J_k+h} = e_1, 0 \leq h \leq t\}$  gilt, dass  $J_{k+1} = J_k + t + \tau(t + J_k)$ , kann der zweite Ausdruck der obigen Multiplikation mithilfe der Markoveigenschaft\* beziehungsweise der Homogenität des Markovprozesses\*\* umgeformt werden zu:

$$\begin{aligned} \mathbb{P}[X_{J_{k+1}} = e_2 | S_{k+1} > t, X_{J_k} = e_1] &= \mathbb{P}[X_{J_{k+1}} = e_2, X_s = e_1, J_k \leq s < J_{k+1} | S_{k+1} > t, X_{J_k} = e_1] = \\ &= \mathbb{P}[X_{J_k+t+\tau(t+J_k)} = e_2, X_s = e_1, J_k \leq s < J_k + t + \tau(t + J_k) | X_{J_k+h} = e_1, 0 \leq h \leq t] = \\ &\stackrel{*}{=} \mathbb{P}[X_{J_k+t+\tau(t+J_k)} = e_2, X_s = e_1, J_k \leq s < J_k + t + \tau(t + J_k) | X_{J_k+t} = e_1] = \\ &\stackrel{**}{=} \mathbb{P}[X_{J_k+\tau(J_k+t)} = e_2, X_s = e_1, J_k \leq s < J_k + \tau(t + J_k) | X_{J_k} = e_1] = \\ &= \mathbb{P}[X_{J_{k+1}} = e_2, | X_{J_k} = e_1]. \end{aligned}$$

□

## A.1 Konvergenz gegen ein Äquilibrium

**Theorem A.4** (siehe [8], Theorem 3.6.2) Sei  $(X_t)_{t \geq 0}$  nicht explosiv und irreduzibel. Weiters sei  $\pi$  ein invariantes Maß. Dann gilt:

$$\lim_{t \rightarrow \infty} P_{e_1, e_2}(t) = \pi(e_2) \quad \forall e_1, e_2 \in E.$$

**Beweis.** Siehe [8]

□

## A.2 Explosivitätskriterium

**Theorem A.5** (siehe [8], Korollar 2.7.3) Sei  $X \sim \text{Markov}(\lambda, Q)$  und sei  $V : E \mapsto \mathbb{R}$ . Dann sind für beliebige  $\theta > 0$  folgende Aussagen äquivalent:

1.  $Q$  ist nicht explosiv
2. aus  $(QV)(\cdot) = \theta V(\cdot)$  und  $|V| \leq 1 \Rightarrow V \equiv 0$

**Beweis.** Beweis siehe [8].

□

## A.3 Rekurrenz und Transienz

**Theorem A.6** (siehe [8], Theorem 3.4.1) Folgende Aussagen gelten:

1. Falls  $e \in E$  rekurrent ist für den assoziierten Sprungprozess  $Z_n := X_{J_n}$ , dann ist  $e$  rekurrent für  $(X_t)_{t \geq 0}$ .

2. Falls  $e \in E$  transient ist für den assoziierten Sprungprozess  $Z_n := X_{J_n}$ , dann ist  $e$  transient für  $(X_t)_{t \geq 0}$ .
3. Jeder Zustand  $e \in E$  ist entweder transient oder rekurrent.
4. Rekurrenz und Transienz sind Klasseigenschaften (bezüglich der Äquivalenzrelation:  $\longleftrightarrow$  siehe 2.4)

**Beweis.** Beweis siehe [8]. □

Bezeichne  $T_e := \inf\{t \geq J_1 : X_t = e\}$  die erste Ankunftszeit (engl. first passage time) des Zustandes  $e \in E$ .

**Theorem A.7 (siehe [8], Theorem 3.4.2)** Sei  $e \in E$ . Es gelten folgende Aussagen:

1. Falls  $-Q_{e,e} = \sum_{l \in E: l \neq e} Q_{e,l} = 0$  oder  $\mathbb{P}_e[T_e < \infty] = 1$  gelten, dann folgt das der Zustand  $e$  rekurrent ist.
2. Falls  $-Q_{e,e} = \sum_{l \in E: l \neq e} Q_{e,l} > 0$  und  $\mathbb{P}_e[T_e < \infty] < 1$  gelten, dann folgt das der Zustand  $e$  transient ist.

**Beweis.** Beweis siehe [8]. □

**Definition A.8 (kompakt)** Wir nennen eine Funktion  $V : E \rightarrow \mathbb{R}_+$  kompakt, falls für alle  $d > 0$  gilt:  $|\{e \in E : V(e) < d\}| < \infty$

**Theorem A.9 (siehe [3], Theorem 1.18)** Sei  $(X_t)_{t \geq 0}$  irreduzibel mit Intensitätsmatrix  $Q = (Q_{e_1, e_2})_{(e_1, e_2) \in E \times E}$ . Seien weiters  $b > 0$ ,  $c > 0$  und  $V : E \rightarrow \mathbb{R}_+$  kompakt. Falls gilt:

$$\sum_{l \in E} Q_{e,l}(V(l) - V(e)) < -bV(e) + c$$

Dann ist der Markovprozess  $X$  positiv rekurrent.

**Beweis.** Beweis siehe [3]. □

**Theorem A.10 (siehe [8], Theorem 3.5.3)** Sei  $(X_t)_{t \geq 0}$  irreduzibel mit Intensitätsmatrix  $Q$ . Dann sind folgende Aussagen äquivalent:

1. Jeder Zustand  $e \in E$  ist positiv rekurrent.
2. Ein Zustand  $e \in E$  ist positiv rekurrent.
3.  $Q$  ist nicht explosiv und hat eine invariante Verteilung.

**Beweis.** Beweis siehe [8]. □

## A.4 Ergodensätze

Im Folgenden werden wir in aller Kürze die Rechtfertigung für die in Kapitel 6 und Abschnitt 7.3 verwendete Schätzmethode der invarianten Verteilung liefern.

Für eine diskrete homogene Markovkette  $(Z_n)_{n \in \mathbb{N}_0}$  auf  $E$  gibt uns folgendes Theorem Aufschluss über die Existenz einer invarianten Verteilung. Dies ist in gewisser Weise das zeitdiskrete Pendant zu Theorem A.10.

**Theorem A.11 (siehe [8], Theorem 1.7.7)** Sei  $(Z_n)_{n \in \mathbb{N}_0}$  irreduzibel mit Übergangsmatrix  $P$ . Dann sind folgende Aussagen äquivalent:

1. Jeder Zustand  $e \in E$  ist positiv rekurrent.
2. Ein Zustand  $e \in E$  ist positiv rekurrent.
3.  $Z$  hat eine invariante Verteilung  $\mu$ .

Diese ist für  $e \in E$  gegeben durch  $\mu(e) = \frac{1}{m(e)}$  wobei  $m$  die erwartete Ankunftszeit des Zustandes  $e$  bezeichnet d.h.  $m(e) := \mathbb{E}_e[\tilde{T}_e]$  mit  $\tilde{T}_e := \inf\{n \geq 1 : Z_n = e\}$ .

**Beweis.** Beweis siehe [8]. □

Das nächste Theorem rechtfertigt nun unsere Schätzmethode der invarianten Verteilung  $\mu_i$  des zeitdiskreten Sprungprozesses  $Z_n^i := X_{J_n}^i$  mit Übergangsmatrix  $\Pi^i$ .

**Theorem A.12 (siehe [8], Theorem 1.10.2)** Sei  $(Z_n)_{n \in \mathbb{N}_0}$  irreduzibel mit Übergangsmatrix  $P$  und beliebiger Anfangsverteilung  $\lambda$  auf  $E$ . Dann gilt:

$$\mathbb{P} \left[ \frac{1}{n} \sum_{k=1}^{n-1} \mathbb{1}_{\{Z_k=e\}} \xrightarrow{n \rightarrow \infty} \frac{1}{m(e)} \right] = 1$$

Falls  $(Z_n)_{n \in \mathbb{N}_0}$  überdies positiv rekurrent ist dann gilt für jede beschränkte Funktion  $f : E \mapsto \mathbb{R}$ :

$$\mathbb{P} \left[ \frac{1}{n} \sum_{k=1}^{n-1} f(Z_k) \xrightarrow{n \rightarrow \infty} \sum_{e \in E} \mu(e) f(e) \right] = 1$$

wobei  $\mu$  die eindeutige invariante Verteilung jener Markovkette bezeichnet.

**Beweis.** Beweis siehe [8]. □

Nun folgt entweder mittels Theorem A.11 und der ersten Aussage von Theorem A.12 oder mit Wahl von  $f(z) := \mathbb{1}_{\{z=e\}}$  und der zweiten Aussage von Theorem A.12, dass im Falle der positiven Rekurrenz für die eindeutige invariante Verteilung  $\mu$  von  $Z$  gilt:

$$\mathbb{P} \left[ \frac{1}{n} \sum_{k=1}^{n-1} \mathbb{1}_{\{Z_k=e\}} \xrightarrow{n \rightarrow \infty} \mu(e) \right] = 1 \quad \forall e \in E.$$

Angenommen  $(X_t)_{t \geq 0}$  sei positiv rekurrent mit eindeutiger invarianter Verteilung  $\pi$ , Intensitätsmatrix  $Q$  und korrespondierender Übergangsmatrix  $P$ . Dann gilt mittels Lemma 3.1 folgende Aussage:

$$\pi Q = 0 \iff \pi P(t) = \pi \quad \forall t \geq 0 \implies \pi P(\Delta) = \pi$$

Betrachtet man nun den zeitdiskreten Markovprozess  $(X_{n\Delta})_{n \in \mathbb{N}_0}$  mit Übergangsmatrix  $\tilde{P} := P(\Delta)$ . Dann folgt aus obigen Überlegungen, dass die invariante Verteilung  $\pi$  des zeitstetigen Prozesses  $(X_t)_{t \geq 0}$  auch eine invariante Verteilung des Prozesses  $(X_{n\Delta})_{n \in \mathbb{N}_0}$  darstellt. Zieht man nun zu einer festen Beobachtungsfrequenz  $\Delta$  Werte des zeitstetigen Prozesses  $(X_t)_{t \geq 0}$ , so ergibt jenes Verfahren den zeitdiskreten Prozess  $(X_{n\Delta})_{n \in \mathbb{N}_0}$ , welcher dieselbe invariante Verteilung besitzt. Berechnet man nun die empirischen Häufigkeiten so erhält man für  $n$  groß genug mittels Theorem A.12 einen (konsistenten) Schätzer für die invariante Verteilung  $\pi$ .

## Anhang B

# Parameter Punktschätzung und Asymptotisches Verhalten

**Theorem B.1** (siehe [1], Theorem 2.1) Sei  $\hat{\theta}$  der Maximum Likelihood Schätzer eines Parameters  $\theta \in \mathbb{R}^n$  und  $g : \mathbb{R}^n \mapsto \mathbb{R}^m$ . Dann ist der Maximum Likelihood Schätzer von  $\xi := g(\theta)$  gegeben durch  $\hat{\xi} := g(\hat{\theta})$ .

*Beweis.* Siehe [1, Theorem 2.1] □

**Theorem B.2** Unter gegebenen Regularitätsvoraussetzungen (siehe [7, Abschnitt 6 im speziellen Formel (6.11)]) gilt für die Fisher Matrix  $\mathcal{I}$  einer Zufallsvariablen  $X : \Omega \rightarrow \mathbb{R}^n$  mit Dichte  $f(\cdot, \theta) : \mathbb{R}^n \rightarrow \mathbb{R}_+^n$  und  $\theta \in \mathbb{R}^m$  folgende Beziehung:

$$\mathcal{I}(\theta) := \left( \mathbb{E} \left[ \frac{\partial}{\partial \theta_i} \log(f(X, \theta)) \cdot \frac{\partial}{\partial \theta_j} \log(f(X, \theta)) \right] \right)_{i,j \in \{1, \dots, m\}} = -\mathbb{E} [H_\theta(\log f(x, \cdot))(X, \theta)]$$

wobei  $H$  die Hesse Matrix bezeichnet.

*Beweis.* siehe [7]. □

**Theorem B.3 (Delta Methode, siehe [7], Theorem 8.22)** Angenommen für  $\hat{\theta}_n : \Omega \rightarrow \mathbb{R}^m$  mit  $m, n \in \mathbb{N}$ ,  $\theta \in \mathbb{R}^m$  und  $\Sigma \in \mathbb{R}^{m \times m}$  gelte, dass:

$$\sqrt{n} \cdot (\hat{\theta}_n - \theta) \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}(0, \Sigma)$$

Sei weiters für  $k \in \mathbb{N}$  die Funktion  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  in einer offenen Umgebung  $U_\theta \subset \mathbb{R}^m$  von  $\theta$  definiert und stetig partiell differenzierbar sodass dessen Jacobi-Matrix  $J_g : \mathbb{R}^m \rightarrow \mathbb{R}^k$  auf  $U_\theta$  invertierbar ist. Dann gilt:

$$\sqrt{n} \cdot (g(\hat{\theta}) - g(\theta)) \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{N}(0, J_g(\theta) \cdot \Sigma \cdot J_g(\theta)^T)$$

*Beweis.* siehe [7, Theorem 8.22] □

## Anhang C

# Wahrscheinlichkeitstheorie

**Lemma C.1** (siehe [8], Theorem 2.3.3) Sei  $I$  eine abzählbare Menge und  $(T_k)_{k \in I}$  unabhängig exponentialverteilt  $T_k \sim \text{Exp}(q_k)$  mit  $0 < q := \sum_{k \in I} q_k < \infty$ . Definiere  $T := \inf_{k \in I} T_k$ . Dann wird dieses Infimum bei einer Zufallsvariablen  $K : \Omega \rightarrow I$   $\mathbb{P}$ -f.s. angenommen. Des Weiteren sind die Zufallsvariablen  $K$  und  $T$  unabhängig und deren Verteilung ist gegeben durch  $\mathbb{P}[K = k] = \frac{q_k}{q}$ ,  $k \in I$  respektive  $T \sim \text{Exp}(q)$ .

**Beweis.** Setze  $K := k$  falls  $T_k < T_j$ ,  $j \neq k$ , andernfalls lasse  $K$  undefiniert <sup>1</sup>.

Nun gilt mithilfe des Satzes von der vollständigen Erwartung für  $t \geq 0$ :

$$\begin{aligned} \mathbb{P}[K = k, T \geq t] &= \mathbb{P}[T_k \geq t, T_j > T_k, j \neq k] = \int_0^\infty q_k e^{-q_k \cdot s} \mathbb{P}[s \geq t, T_j > s, j \neq k | T_k = s] ds = \\ &= \int_t^\infty q_k e^{-q_k \cdot s} \mathbb{P}[T_j > s, j \neq k | T_k = s] ds = \int_t^\infty q_k e^{-q_k \cdot s} \mathbb{P}[T_j > s, j \neq k] ds = \int_t^\infty q_k e^{-q_k \cdot s} \prod_{j \neq k} e^{-q_j \cdot s} ds = \\ &= \frac{q_k}{q} \cdot e^{-q \cdot t} \end{aligned}$$

Daraus folgt sofort für  $k \in I$  und  $t \geq 0$ :

$$\begin{aligned} \mathbb{P}[K = k] &= \mathbb{P}[K = k, T \geq 0] = \frac{q_k}{q} \\ \mathbb{P}[T \geq t] &= \sum_{k \in I} \mathbb{P}[K = k, T \geq t] = \sum_{k \in I} q_k \cdot \frac{e^{-q \cdot t}}{q} = e^{-q \cdot t} \end{aligned}$$

Damit folgt nun auch die Unabhängigkeit der beiden Zufallsvariablen  $T$  und  $K$ . Schlussendlich folgt aus obigen Überlegungen, dass  $\mathbb{P}[K = k, k \in I] = 1$  und somit wird das Infimum  $\mathbb{P}$ -f.s. angenommen.  $\square$

---

<sup>1</sup> $\mathbb{P}[\#k \in I : T_k < T_j, j \neq k] = 0$

# Literaturverzeichnis

- [1] Felix Abramovich and Ya'acov Ritov. *Statistical theory: a concise introduction*. CRC Press, 2013.
- [2] Álvaro Cartea, Sebastian Jaimungal, and José Penalva. *Algorithmic and high-frequency trading*. Cambridge University Press, 2015.
- [3] Mu-Fa Chen. On three classical problems for Markov chains with continuous time parameters. *Journal of Applied Probability*, 28(2):305–320, 1991.
- [4] Donald Gross, J. F. Shortle, J. M. Thompson, and C. M. Harris. *Fundamentals of queueing theory*. John Wiley & Sons, Inc., 2008.
- [5] Nikolaus Hautsch. *Econometrics of financial high-frequency data*. Springer Science & Business Media, 2011.
- [6] Weibing Huang, Charles-Albert Lehalle, and Mathieu Rosenbaum. Simulating and analyzing order book data: The queue-reactive model. *Journal of the American Statistical Association*, 110(509):107–122, 2015.
- [7] Erich Leo Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 1998.
- [8] James R Norris. *Markov chains*. Cambridge University Press, 1998.
- [9] Martin Predota. *Finanzmärkte und Finanzintermediation*. Facultas. WUV, 2011.
- [10] Christian Y Robert and Mathieu Rosenbaum. A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics*, 9(2):344–366, 2010.

# Abbildungsverzeichnis

1.1	LOB Nasdaq . . . . .	5
1.2	Illustration einer neu hinzukommende buy Limit Order . . . . .	6
1.3	Illustration einer neu hinzukommenden Market Order . . . . .	7
2.1	Limit Order Book: schematische Darstellung . . . . .	9
6.1	Orderbuchentwicklung FTE Limit 1 . . . . .	41
6.2	Orderbuchentwicklung FTE Limit 2 . . . . .	42
6.3	Orderbuchentwicklung FTE Limit 3 . . . . .	42
6.4	Intensitätsschätzung FTE: ASK/BID getrennt . . . . .	43
6.5	Intensitätsschätzung FTE . . . . .	43
6.6	Überprüfung von Annahme 2 FTE . . . . .	44
6.7	Überprüfung von Annahme 3 FTE . . . . .	45
6.8	FTE Vergleich: empirische und invariante Verteilung $\mu_i$ von $Z$ . . . . .	45
6.9	FTE Vergleich: empirische und invariante Verteilung $\pi_i$ von $X$ . . . . .	46
6.10	Intensitätsschätzung MSFT . . . . .	47
6.11	Überprüfung von Annahme 2 MSFT . . . . .	48
6.12	Überprüfung von Annahme 3 MSFT . . . . .	48
6.13	MSFT Vergleich: empirische und invariante Verteilung $\pi_i$ von $X$ . . . . .	49
6.14	MSFT Vergleich: empirische und invariante Verteilung $\mu_i$ von $Z$ . . . . .	49
6.15	Orderbuchentwicklung MSFT Limit 1 . . . . .	50
6.16	Nachbildung der empirischen Verteilung des Preislimits 1 mittels Simulation . . . . .	51
6.17	Orderbuchentwicklung MSFT Limit 2 . . . . .	52
6.18	Nachbildung der empirischen Verteilung des Preislimits 2 mittels Simulation . . . . .	53
6.19	Darstellung der empirischen Verteilung der Sprünge bei gegebener Referenzpreisänderung MSFT . . . . .	53
7.1	Illustration der gegebenen Intensitäten zur empirischen Studie im Modell I . . . . .	58
7.2	Aus simulierten Daten geschätzte Intensitäten für das Limit 1 . . . . .	59
7.3	Aus simulierten Daten geschätzte Intensitäten für das Limit 2 . . . . .	59
7.4	Aus simulierten Daten geschätzte Intensitäten für das Limit 3 . . . . .	60
7.5	Regularitätsvoraussetzung: Annahme 2 an die an [6] angelehnten Intensitäten . . . . .	61
7.6	Regularitätsvoraussetzung: Annahme 3 an die an [6] angelehnten Intensitäten . . . . .	61
7.7	Vergleich der empirischen Verteilung mit der invarianter Verteilung $mu_i$ zu simulierten Daten . . . . .	61
7.8	Wahrscheinlichkeit einer Exekution einer Buy Limit Order im Modell I . . . . .	64

8.1	FTE: Kalibrierung der Parameter $\theta$ und $\theta_{reinit}$ . . . . .	72
8.2	An [6] angelehnt: Kalibrierung der Parameter $\theta$ und $\theta_{reinit}$ . . . . .	73
8.3	Vergleich der Verteilungen im Modell III . . . . .	73

# Tabellenverzeichnis

5.1	Darstellung des <i>Messagefiles</i> von LOBSTER zur Aktie FTE . . . . .	35
5.2	Darstellung des <i>Orderbookfiles</i> von LOBSTER zur Aktie FTE . . . . .	36
5.3	Exemplarische Darstellung von zusammengeführtem <i>Orderbook</i> - und <i>Messagefile</i> zur Aktie FTE mit Referenzpreis . . . . .	38
5.4	Erhaltene Stichprobe zum Preislimit $P_{\pm 1}$ . . . . .	39
5.5	Erhaltene Stichprobe zum Preislimit $P_{\pm 2}$ . . . . .	39
5.6	Erhaltene Stichprobe zum Preislimit $P_{\pm 3}$ . . . . .	39