



## DIPLOMARBEIT

---

# Modellierung eines Limit Order Buches mittels Lévy-Prozessen

---

AUSGEFÜHRT AM INSTITUT FÜR

FINANZ- UND VERSICHERUNGSMATHEMATIK

DER

TECHNISCHEN UNIVERSITÄT WIEN

UNTER DER ANLEITUNG VON

UNIV.PROF. DIPL.-MATH. DR.RER.NAT.

THORSTEN RHEINLÄNDER

DURCH

NIKOLAUS ALTMANN BSc

HOLZWEG 22, 3430 TULLN

---

Datum

---

Unterschrift (Student)

# Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Die vorliegende Arbeit wurde bisher in gleicher oder ähnlicher Form noch nicht als Diplomarbeit eingereicht.

---

Datum

---

Unterschrift

Wenn die Börsenkurse fallen,  
regt sich Kummer fast bei allen,  
aber manche blühen auf:  
Ihr Rezept heißt Leerverkauf

Keck verhökern diese Knaben  
Dinge, die sie gar nicht haben,  
treten selbst den Absturz los,  
den sie brauchen- echt famos!

Und wenn Kurse wieder steigen,  
fängt von vorne an der Reigen-  
ist halt Umverteilung pur,  
stets in eine Richtung nur.

Aber sollten sich die Massen  
das mal nimmer bieten lassen,  
ist der Ausweg längst bedacht:  
Dann wird bisschen Krieg gemacht.

Kurt Tucholsky, 1930

# Danksagung

Mein tiefer Dank gilt meiner Familie und Studienkollegen, die mich in vielfältiger Art und Weise im Laufe meines Studiums unterstützt und mir dieses ermöglicht haben. Es war eine tolle Zeit!

Weiters möchte ich meinem Betreuer, Prof. Rheinländer, für seine stets freundliche, unterstützende und nachsichtige Betreuung danken.

Wien, im Mai 2017

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>6</b>
<b>2</b>	<b>Theoretische Grundlagen zum Limit Order Buch</b>	<b>10</b>
2.1	Ein einfaches stochastisches Modell . . . . .	10
2.1.1	Einführung . . . . .	11
2.1.2	Ein Markovmodell für das Limit Order Buch . . . . .	11
2.1.3	Resultate . . . . .	12
2.1.4	Diffusionsgrenzwert des Preisprozesses . . . . .	14
2.2	Order Flow Imbalance . . . . .	14
2.2.1	Die Order Flow Imbalance . . . . .	15
2.2.2	Price Impact Model . . . . .	15
2.2.3	Schätzen des Parameters und Resultate . . . . .	16
2.2.4	Anwendungen . . . . .	17
2.2.5	Intraday Volatility dynamics . . . . .	17
2.2.6	Conclusio . . . . .	18
2.3	First Passage Time Modell . . . . .	18
2.3.1	Abkürzungen . . . . .	18
2.3.2	First Passage Time . . . . .	19
2.3.3	Ein einfaches Modell für Time To Fill und Time To Cancel . . . . .	19
2.4	Ein allgemeineres LOB Modell . . . . .	20
2.4.1	Allgemeines . . . . .	20
2.4.2	Limit Orders und Stornierungen . . . . .	20
2.4.3	Markt Orders . . . . .	20
<b>3</b>	<b>Datenanalyse</b>	<b>23</b>
3.1	Beschreibung der Datenstruktur . . . . .	23
3.2	Qualitative Analyse . . . . .	25
3.3	Der Arrival Prozess . . . . .	36
<b>4</b>	<b>Model fitting</b>	<b>38</b>
4.1	Lévy-Prozess . . . . .	38
4.1.1	Definition und ausgewählte Eigenschaften . . . . .	38
4.1.2	Modellschätzungen . . . . .	40
4.2	Poisson-Prozess . . . . .	44
4.2.1	Theorie . . . . .	44
4.2.2	Berechnung . . . . .	45

5	Fazit	47
6	Anhang	48

# Kapitel 1

## Einleitung

An der Börse gibt es verschiedene Arten, eine Order, also einen Auftrag zum Handeln eines Wertpapiers, zu platzieren. Da ist zum einen natürlich die *Markt Order*. Hierbei wird ein Wertpapier sofort zum aktuellen Marktpreis ge- oder verkauft, sofern der Markt für diese Aktie liquide genug ist.

Im Gegensatz zur Markt Order zeichnet sich nun eine *Limit Order* dadurch aus, dass man einen Preis definieren kann, zu dem man handeln möchte. Im Falle eines Kaufes wird man hier logischerweise an einen Maximalbetrag denken, den man höchstens bereit ist zu zahlen. Im Falle eines Verkaufes möchte man einen gewissen Mindestpreis für das Papier bekommen.

Die Limit Orders an einer Börse werden im sogenannten *Limit Order Buch* (LOB) gesammelt. Das LOB teilt sich in zwei "Hälften". Auf der linken Seite, der *Bid* Seite, werden alle Kauf Limit Orders gesammelt. Hingegen werden auf der *Ask* Seite alle Verkauf Limit Orders eingetragen. Die Bezeichnungen mögen auf den ersten Blick etwas konträr zur Intuition erscheinen, jedoch sind diese Zuordnungen im Börsejargon so üblich.

Nun kann man sich vorstellen, dass das LOB aus mehreren Preis *Levels* besteht. Es wird also ein bestes Kaufangebot mit einer gewissen Stückzahl an Wertpapieren in Form eines Preises geben, und dahinter wird es normalerweise auf den "schlechteren" Preisniveaus auch Kaufangebote geben. Analog gilt dies für die Ask Seite. Sollte nun eine Markt-Verkaufs Order getätigt werden, so wird zuerst die Limit Order am *Best Bid* dazu verwendet, um diese Markt Order zu bedienen. Sollte dies nicht ausreichen, werden die dahinterliegenden Kaufangebote herangezogen. Den Mittelwert aus Best Bid und Best Ask bezeichnet man als den *Mid* Preis.

Für eine bestehende Limit Order gibt es nur zwei Möglichkeiten, was mit ihr geschehen kann. Entweder, sie wird (teilweise) ausgeführt (*execution*) oder sie wird storniert (*cancellation*). Zusammen mit den Markt Orders bilden diese drei Typen die *Order Book Events*.

Diese Diplomarbeit hat die Modellerierung des Mid Preises zum Thema. Zum einen werden theoretische Modelle diskutiert, zum anderen anhand von realen Börsedaten diver-

se stochastische Prozesse angepasst und kalibriert. Zu diesem Zweck wurden *NASDAQ* Börsedaten von der Internetseite *lobsterdata.com* verwendet. Untersucht werden 20 große Aktientitel von der NASDAQ am 16.11.2016. Die Aktien wurden der Größe nach ausgewählt, um sicherzustellen, dass liquide Titel analysiert werden. Das Datum ist ein rein willkürlich gewählter Handelstag.

Im nachfolgenden Kapitel werden theoretische Überlegungen und entwickelte Modelle vorgestellt. Die hierbei verwendeten Ansätze unterscheiden sich voneinander doch auf grundsätzlicher Ebene. Der darauffolgende Teil handelt schon von der Datenanalyse: Zunächst wird ausführlich auf die vorhandene Datenstruktur eingegangen, und auf welche Details man Acht geben muss.

Die Analyse und Auswertung der Daten wird anschließend vorgenommen. Wie sieht das Limit Order Buch qualitativ aus für die diversen Aktien? Verändert sich die Form über den Tag oder bleibt diese konstant? Und welchen Einfluss hat das LOB im Prozess von Angebot und Nachfrage und folgend in der Preisformation? Diese und weitere Fragen werden in Kapitel 3 behandelt. Der nächste Abschnitt, Kapitel 4, befasst sich mit der Kalibrierung von diversen Prozessen an die vorhandenen Daten. Dabei werden Lévy- und Poisson-Prozesse betrachtet.

Zum Schluss wird ein Fazit gegeben, in welchem noch einmal über den Inhalt dieser Arbeit reflektiert wird. Weitere Hinweise zu aktueller und weiterführender Forschung werden gegeben.



# Abstract

In a stock market several types of orders are usual. Surely, the most convenient orders are *market orders*. The current market price of a share is known by both the buyer and the seller and if they agree to trade the share, the order becomes valid.

But what if, say the buyer, is not very satisfied with the current market price because the share is too expensive in his opinion. So he places a *limit order*, where he can say that he wants to buy a certain number of a certain share to a certain maximum price which he defines. An analogue process is done by a seller of a share, who thinks that the price currently paid on the stock market is too low.

All limit orders are collected in the so-called *limit order book* (LOB). The LOB is divided into two halves, the buyer side and the seller side. The first one is actually called *bid side*. At every price offered by someone to buy a share, called *levels*, is noted how many shares and by whom the order was placed. The other side is called *ask side*.

Apparently, there are more levels than one on each side. How many there are exactly only depends on ask and demand. But the most important levels are the best levels of each side, i.e. the highest level of the bid side and the lowest level of the ask side, called *best bid* and *best ask*. The delta of them, i.e. how far are they away from each other, is called *spread*. In very liquid markets the spread is usually as big as the smallest possible difference (*tick size*). So usually the ticket size mounts up to one cent.

The *mid price* is the mean of best bid and best ask and will be further discussed in this thesis. For an existing limit order there are only three ways what can happen to them. Either it gets (partially) executed against a market order or it gets simply cancelled or it expires, respectively. These events are called *order book events*.

In this thesis, as mentioned above, we will deal with some approaches of fitting stochastic processes to the observed path of the mid price for a share. Theoretical models will be discussed as well as real stock data will be presented and analysed. For this purpose we use *NASDAQ* data from the web page *lobsterdata.com*. 20 big shares of the NASDAQ stock exchange will be analysed on the 16th November 2016, an arbitrarily chosen date by the author of this thesis.

In the consecutive chapter theoretical approaches and results will be presented as well as developed mathematical limit order book models. The approaches used are basically very different, though. After that we will already discuss our used data, what is the structure

of the data like and to which issues do we have to pay attention.

Finally, the analysis and evaluation of the data is done. The following questions will be answered: How does the qualitative structure of the limit order book look like and does it vary over time? If yes, what may be the reasons for that? Does the structure of the LOB have an influence in the process of supply and demand and, subsequently, to the price formation process? These and other questions will be discussed in chapter 3. In the consecutive chapter 4 we will calibrate some stochastic processes, namely Lévy and Poisson processes. After this chapter a conclusion will be given where we will reflect on the things we have done and what the learnings are.

# Kapitel 2

## Theoretische Grundlagen zum Limit Order Buch

Im Folgenden werden einige Ansätze zur Modellierung eines Limit Order Buches diskutiert. Zunächst wird ein einfaches stochastisches Modell vorgestellt, welches aus der Arbeit von CONT und DE LARRARD [7] entnommen wurde. Der Ansatz von diesen Autoren ist, dass der Preisprozess über die Order Buch *Tiefe* wesentlich determiniert wird. Daraus wird die Verteilung der Dauer bis zum nächsten Preissprung hergeleitet.

Das zweite Modell stützt sich auf die Beobachtung, dass der Preis, zumindest kurzfristig, von Überhängen auf der Bid oder auf der Ask Seite abhängig ist. Dieses Modell stammt aus einer Arbeit von CONT, KUKANOV und STOIKOV [3].

Im darauffolgenden Abschnitt 2.3 betrachten wir die Verteilung der Wartedauer bis zum nächsten Preissprung einer vorgegebenen Höhe. Das Stichwort heißt hier *First Passage Time* und stammt von EISLER, KERTESZ, LILLO und MANTEGNA [8].

Schlussendlich wird ein allgemeines Limit Order Buch Modell dargelegt, welches die drei vorhergegangenen Resultate und Modelle vereinigt. RICCARDI hat dies in seiner Thesis ausgearbeitet [11].

### 2.1 Ein einfaches stochastisches Modell

In diesem Abschnitt wird ein einfaches stochastisches Modell für das Verhalten eines Limit Order Buches vorgestellt. Wie in der Einleitung dieses Kapitels bereits erwähnt, wurde dieses aus [7] entnommen. Markt Orders, Limit Orders und Stornierungen werden im Kontext eines Markov'schen Warteschlangensystems beschrieben. Mit diesem Ansatz kann man dann Ausdrücke von verschiedenen interessanten Größen angeben, zum Beispiel:

- die Verteilung der Dauer zwischen Preissprüngen
- die Verteilung und Autokorrelation von Preissprüngen
- die Wahrscheinlichkeit eines Preisanstieges bedingt auf den Order Buch Status

### 2.1.1 Einführung

Die grundlegende Idee rührt von der Beobachtung her, dass der überwiegende Teil des Handelsvolumens am Level vom besten Bid und Ask Preis stattfindet und, dass das Ungleichgewicht der beiden Seiten der Hauptgrund von Preissprüngen ist. Daher wird das Limit Order Buch durch die *Anzahl* der Limit Orders  $(q_t^b, q_t^a)$  am Level vom besten Bid und Ask modelliert.

Da über Zeitspannen von 10 Sekunden tausende von Order Buch Ereignissen stattfinden, und für die Anwendung von Order Ausführungen der volumsgewichtete durchschnittliche Preis ausschlaggebend ist, werden asymptotische Methoden behandelt. Diese stellen durch Betrachtung des Diffusionslimits eine Verbindung zwischen der Volatilität des Preises und des Order Flusses her.

### 2.1.2 Ein Markovmodell für das Limit Order Buch

Das Limit Order Buch wird als zeitstetiger Prozess  $X_t = (s_t^b, q_t^b, q_t^a) \in \delta\mathbb{Z} \times \mathbb{N}^2$  aufgefasst. Hierbei bezeichnen  $s_t^b$  und  $s_t^a$  den (besten) Bid und Ask Preis. Sei oBdA  $s_t^a = s_t^b + \delta$  mit  $\delta > 0$ , was eine Reduktion der Dimension und damit eine Vereinfachung bewirkt.  $\delta$  kann als die sogenannte *Tick Size* aufgefasst werden, also jener Minimalabstand, der zwischen zwei möglichen Orders sein muss, um sie unterscheiden zu können. Bid und Ask Preis sind Vielfache von  $\delta$ . Die Größe  $s_t^a - s_t^b$  bezeichnet man als den *Bid-Ask spread*. Für große liquide Aktien ist der Bid-Ask Spread in 98% der Fälle nicht größer als einen Tick, wie die CONT und DE LARRARD in ihrer Arbeit festgestellt haben [7].

Falls entweder die Bid oder Ask Seite durch eine Markt Order oder durch eine Stornierung abgebaut wird, so bewegt sich der Preis hinauf oder hinunter zum nächsten Level der Order Buches. Die Prozesse  $s_t^b$  und  $s_t^a$  sind also stückweise konstant mit Übergängen dort, wo

$$q_t = (q_t^b, q_t^a) \in \{(0, y), y \in \mathbb{N}\} \cup \{(x, 0), x \in \mathbb{N}\} \quad (2.1)$$

Der Preisprozess hat stückweise konstante Pfade, Änderungen entsprechen Order Buch Ereignissen (Markt, Limit, Stornierung).  $(T_i^a, i \geq 1)$  sei die Zeit zwischen zwei aufeinanderfolgenden Ask Orders, analog  $T_i^b; V_i^a$  bzw.  $V_i^b$  die dazugehörigen Größenveränderungen in der Warteschlange.

Bezeichne weiters mit  $f(x, y) = P[(q_t^b, q_t^a) = (x, y)]$  nach einem Preisanstieg,  $\tilde{f}(x, y) = P[(q_t^b, q_t^a) = (x, y)]$  nachdem der Preis gefallen ist.

Sei  $(R_i)_{i \geq 1} = (R_i^b, R_i^a)_{i \geq 1}$  eine iid Folge von Zufallsgrößen mit (gemeinsamer) Verteilung  $f$ ,  $(\tilde{R}_i)_{i \geq 1}$  analog. Daraus folgt, falls eine (Markt) Order oder eine Stornierung auf der Ask Seite zum Zeitpunkt  $T$  ankommt, dass

$$(s_T^b, q_T^b, q_T^a) = (s_{T-}^b, q_{T-}^b, q_{T-}^a + V_i^a) \mathbb{1}_{q_{T-}^a > -V_i^a} + (s_{T-}^b + \delta, R_i^b, R_i^a) \mathbb{1}_{q_{T-}^a \leq -V_i^a} \quad (2.2)$$

Also: Werden weniger Stück verkauft, als im Order Buch momentan notiert sind ( $q_{T-}^a > -V_i^a \iff q_{T-}^a + V_i^a > 0$ ) verändert sich gerade nur die Stückanzahl. Andernfalls wurden alle verfügbaren Stück aufgebraucht. Der Bid Preis steigt um  $\delta$ , da der Spread immer

diese Größe haben muss. Die Warteschlangen ändern sich in zufälliger Art und Weise. Dies geschieht analog auf der Bid Seite.

Die Ereignisse  $T_i^a, T_i^b, V_i^a, V_i^b$  sollen nach einem Poisson-Prozess auftreten. Markt Orders mit Rate  $\mu$ , Limit Orders mit Rate  $\lambda$  und Stornierungen mit Rate  $\theta$ , unabhängig voneinander und oBdA von gleicher Order Größe = 1.

Daraufhin stellt sich heraus, dass  $(T_i^a)_{i \geq 0}$  und  $(T_i^b)_{i \geq 0}$  Folgen von unabhängigen exponentialverteilten Zufallsgrößen mit Parameter  $\lambda + \theta + \mu$  sind. Weiters lassen sich Ausdrücke wie  $P[V_i^a = 1]$  berechnen durch

$$P[V_i^a = 1] = \frac{\lambda}{\lambda + \theta + \mu}, \quad (2.3)$$

und  $q_t = (q_t^b, q_t^a)$  ist ein Markovprozess in  $\mathbb{N}^2$ . Mit Rate  $\lambda$  erreichen neue Limit Orders die jeweilige (Bid, Ask) Warteschlange, welche um eine Einheit erhöht wird.

$\lambda, \theta + \mu$  lassen sich mit einfachen Mitteln aus vorhandenen Daten schätzen, siehe [1]

### 2.1.3 Resultate

Wir starten mit einem Status  $(x, y)$  des Order Buches.

- $\sigma_a$  sei der erste Zeitpunkt, an dem  $(q_t^a, t \geq 0)$  verschwindet
- $\sigma_b$  sei der erste Zeitpunkt, an dem  $(q_t^b, t \geq 0)$  verschwindet

Da die Warteschlangen zwischen zwei Order Buch Ereignissen konstant sind, kann man diese *Stoppzeiten* auch folgendermaßen anschreiben:

- $\sigma_a = \inf\{T_1^a + \dots + T_i^a : q_{T_1^a + \dots + T_i^a}^a + V_i^a = 0\}$
- $\sigma_b = \inf\{T_1^b + \dots + T_i^b : q_{T_1^b + \dots + T_i^b}^b + V_i^b = 0\}$

Da der Preis sich ändert, falls  $q_t = (q_t^b, q_t^a)$  eine der Achsen berührt, ist die Dauer bis zum nächsten Preissprung gleich  $\tau = \sigma_a \wedge \sigma_b$ .

Die Verteilung von  $\tau$  bedingt auf den Order Buch Status ist gegeben durch:

$$P[\tau > t | q_0^b = x, q_0^a = y] = \sqrt{\left(\frac{\mu + \theta}{\lambda}\right)^{x+y}} \psi_{x,\lambda,\mu+\theta}(t) \psi_{y,\lambda,\mu+\theta}(t) \quad (2.4)$$

mit

$$\psi_{n,\lambda,\mu+\theta}(t) = \int_t^\infty \frac{n}{u} I_n(2\sqrt{\lambda(\theta + \mu)u}) e^{-u(\lambda+\theta+\mu)} du \quad (2.5)$$

und  $I_n$  ist die modifizierte Besselfunktion der ersten Art. Die bedingte Verteilung von  $\tau$  hat ein regulär variierendes Ende

- mit tail exponent 2 falls  $\lambda < \mu + \theta$

- mit tail exponent 1 falls  $\lambda = \mu + \theta$ , d.h.  $\mathbb{E}[\tau|_0^b = x, q_0^a = y] = \infty$  falls  $x, y > 0$ .

Der Beweis arbeitet mit Laplace Transformation.

Die Wahrscheinlichkeit  $p_1^{up}(n, p), (n, p) \in \mathbb{N}^2$ , dass die nächste Preisbewegung ein Anstieg ist, gegeben, dass  $n$  Orders auf der Bid Seite und  $p$  Orders auf der Ask Seite sind, ist:

$$p_1^{up}(n, p) = \frac{1}{\pi} \int_0^\pi \left(2 - \cos(t) - \sqrt{(2 - \cos(t))^2 - 1}\right)^p \frac{\sin(nt) \cos(\frac{t}{2})}{\sin(\frac{t}{2})} dt \quad (2.6)$$

Dieser Ausdruck ist nicht abhängig von Order Flow Parametern im Falle eines gleichgewichteten Order Buches (die Intensität des Flusses von Limit Orders ist gleich groß wie die des Flusses von Markt Orders und Stornierungen).

Bei einem asymmetrischen Order Fluss wird die Symmetrie von Bid und Ask Seite abgeschwächt, d.h. die Intensitäten von Limit- und Markt Orders auf der Bid und Ask Seite dürfen verschieden sein. Der Ausdruck für  $p_1^{up}(n, p)$  ist dann deutlich komplexer.

Zu Vereinfachungszwecken wird ab nun angenommen, dass  $\tilde{f}(x, y) = f(y, x)$ .

Modelliere die Folge von Preissprüngen mit  $(X_1, X_2, \dots)$ , wobei  $X_i \in \{-\delta, \delta\}$  für  $i \geq 1$ . Wir interessieren uns für die  $n$ -Vorwärtsschrittverteilung

$$p_n^{up}(x, y) = P[X_n = \delta | (q_0^b, q_0^a) = (x, y)]. \quad (2.7)$$

Um uns diesem Ausdruck anzunähern, betrachten wir die Wahrscheinlichkeit von zwei Preissprüngen in die selbe Richtung zwei Mal hintereinander:

$$p_{cont} = P[X_{k+1} = \delta | X_k = \delta] = P[X_{k+1} = -\delta | X_k = -\delta] \quad (2.8)$$

Man erhält

- $p_{cont} = \sum_{i=1}^\infty \sum_{j=1}^\infty f(i, j) p_1^{up}(i, j)$
- $\mathbb{E}[X_k | q_0^b = x, q_0^a = y] = (2p_1^{up}(x, y) - 1)(2p_{cont} - 1)^{k-1}, k \geq 1$
- $\text{Cov}(X_1, X_2 | q_0^b = x, q_0^a = y) = \delta^2(2p_{cont} - 1)(1 - (2p_1^{up}(x, y) - 1)^2)$
- $p_n^{up}(x, y) = \frac{1 + (2p_{cont} - 1)^{n-1}(2p_1^{up}(x, y) - 1)}{2}$

Da, wie man oben sieht,  $\text{Cov}(X_1, X_2 | q_0^b = x, q_0^a = y)$  nicht von  $(q_0^b, q_0^a)$  abhängt, gilt  $\text{Cov}(X_k, X_{k+1}) = \text{Cov}(X_1, X_2) < 0$  genau dann wenn

$$p_{cont} < 1/2 \quad (2.9)$$

Nach einem Preisanstieg ist die neue Bid Warteschlange im Allgemeinen kleiner als die Ask Warteschlange, da die Ask Schlange nun auf dem ehemaligen zweitbesten Limit liegt. Die Bid Schlange hingegen resultiert aus der Akkumulation der Orders über die sehr kleine Zeitspanne seit dem letzten Preissprung. Unter diesen Umständen korrelieren die Hochfrequenzinkremente des Preises negativ, d.h. auf einen Preisanstieg folgt eher ein Sinken des Preises im nächsten Schritt.

Die Folge  $(X_1, X_2, \dots)$  ist unkorreliert genau dann, wenn  $p_{cont} = 1/2$ .

## 2.1.4 Diffusionsgrenzwert des Preisprozesses

Über Zeiträume, die viel länger sind als die zwischen zwei Order Buch Ereignissen, werden diffusive Dynamiken des Preises beobachtet und auch so modelliert. Um nun eine Verbindung zwischen hochfrequenten Preisdynamiken und diesem diffusiven Verhalten auf längeren Zeitskalen herzustellen, wird ein skaliertes Preisprozess

$$(s_t^n := \frac{s_{t_n}}{\sqrt{n}}, t \geq 0)_{n \geq 1} \quad (2.10)$$

betrachtet, der dann für  $n \rightarrow \infty$  gegen einen nicht-degenerierten Prozess  $(p_t, t \geq 0)$  in Verteilung konvergiert. Es wird gezeigt, dass unter einer Symmetrieannahme dieser Grenzwert als Diffusionsprozess identifiziert werden kann. Der Diffusionskoeffizient kann über den Order Flow des Limit Order Buches bestimmt werden.

Wir nehmen an, dass die gemeinsame Verteilung  $f$  der Warteschlangengröße nach einem Preissprung folgende Bedingung erfüllt:

$$D(f) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} ijf(i, j) < \infty \quad (2.11)$$

$D(f)$  ist ein Maß für die Markttiefe, bzw. ist  $\sqrt{D(f)}$  das geometrische Mittel der Bid und Ask Seite nach einem Preissprung.

*Theorem:* Für  $\lambda = \mu + \theta$  gilt

$$\left( \frac{s_{tn \log n}}{\sqrt{n}}, t \geq 0 \right) \Rightarrow \left( \delta \sqrt{\frac{\pi \lambda}{D(f)}} W_t, t \geq 0 \right), n \rightarrow \infty. \quad (2.12)$$

Die Aussage von Theorem 1 verbindet die grobkörnige Volatilität von Tagesreturns mit geringeren Frequenzen mit hochfrequenten Orders. Der Preis hat ein diffusives Verhalten mit einem Diffusionskoeffizienten, der gegeben ist durch

$$\sigma_n = \delta \sqrt{\frac{n \pi \lambda}{D(f)}}, \quad (2.13)$$

wobei  $n$  die durchschnittliche Anzahl an Orders während eines ca. 10-minütigen Zeitintervalles ist.

Wenn man den Fall betrachtet, wo Markt Orders und Stornierungen dominieren, so wird hier ebenfalls ein asymptotischer Ansatz für den Preisprozess gewählt. Benutzt wird hierbei dann die erwartete Trefferzeit an den Achsen für ein Markov'sches Warteschlangensystem  $q$  mit Parameter  $(\lambda, \theta + \mu)$ .

## 2.2 Order Flow Imbalance

Dieses Modell betrachtet die Auswirkungen von Order Buch Ereignissen auf den Preis. Order Buch Ereignisse sind wie immer Limit Orders, Markt Orders und Stornierungen

[3]. In diesem Abschnitt wird die These behandelt, dass auf kurzen Zeitintervallen Preisveränderungen hauptsächlich durch die *Order Flow Imbalance* (OFI) ausgelöst werden. Es wird ein linearer Zusammenhang zwischen der OFI und Preisveränderungen gezeigt, und zwar mit einer zur Markttiefe, i.e. den ausstehenden Limit Orders, proportionalen Steigung. Es reicht also nur eine einzige Variable aus, um den Impact auf die Preise zu modellieren.

Dieses Modell ist robust und impliziert altbekannte Beziehungen zwischen dem Handelsvolumen und den Ausschlägen von Preisbewegungen. Es wird eine potenzielle Anwendung bei Limit Order Aufträgen des Modells vorgestellt und es wird demonstriert, wie es zur Analyse von Intraday Volatilitätsdynamiken verwendet werden kann.

### 2.2.1 Die Order Flow Imbalance

Die OFI ist das Ungleichgewicht zwischen dem besten Angebot und der besten Nachfrage, also das Netto zwischen Bid und Ask Orders. Sie bildet Veränderungen in den Bid und Ask Warteschlangen ab, indem sie ansteigt, falls die Bid Seite ansteigt oder fällt, falls die Ask Seite fällt.

Die OFI wertet einen Marktverkauf und eine Stornierung von der selben Größe als äquivalent, da beide den selben Effekt auf die Größe der besten Bid Warteschlange haben (sie wird kleiner). Für das lineare Modell, welches hier vorgestellt wird, muss man nur einen Koeffizienten schätzen, und zwar den *Price Impact Coefficient*.

Die OFI hat auch noch weitere nützliche Eigenschaften.

### 2.2.2 Price Impact Model

Folgende Bezeichnungen werden im weiteren Verlauf verwendet. Diese werden in analoger Weise auch für die Ask Seite definiert.

- $P_k^b$  ... bester Bid Preis zum Zeitpunkt  $t_k$
- $\Delta P_k^b := (P_k^b - P_{k-1}^b)$
- $L_k^b, C_k^b$  ... Anzahl der Kauf Orders, die zum besten Preis hinzukommen bzw. storniert werden
- $M_k^b$  ... Anzahl der verfügbaren Markt Kauf Orders, die beim aktuell besten Ask ankommen
- $D$  ... maximale Anzahl an Aktien, die bei einem Level im Order Buch eingetragen sein können
- $\delta$  ... tick size, die Einheit im Order Buch, z.B. 1 Cent
- $P_k = \frac{P_k^b + P_k^s}{2\delta}$  ... Mittelpreis (*Mid Price*) aus Bid und Ask



Folgende lineare Beziehung existiert in diesem simplen Order Buch Modell (analog für Ask Seite; Beweis durch Induktion über die Anzahl der Preissprünge in  $[t_{k-1}, t_k]$ ):

$$\Delta P_k^b = \delta \left[ \frac{L_k^b - C_k^b - M_k^b}{D} \right] \quad (2.14)$$

Hierbei werden also alle möglichen Ereignisse aufaddiert, und der Preis ändert sich um das aufgerundete Netto aus Markt Orders, Stornierungen und Limit Orders.

Nun kommen wir zur Definition der Order Flow Imbalance in diesem Modell:

$$OFI_k = L_k^b - C_k^b - M_k^s - L_k^s + C_k^s + M_k^b \quad (2.15)$$

$$= M_k^b - M_k^s + (L_k^b - C_k^b) - (L_k^s - C_k^s) \quad (2.16)$$

Die  $OFI_k$  ist also zum Zeitpunkt  $t_k$  die Differenz aus Markt Kauf und Verkaufs Ordnern ("Wurde mehr ver- oder mehr gekauft?") plus das Netto aus Limit Orders (Bid orders minus Ask orders), bereinigt um die Stornierungen. Also: gibt es generell eher einen Angebots- oder eher einen Nachfrageüberschuss?

Weitere Modellparameter:

$$\Delta P_k = \frac{OFI_k}{2D} + \eta_k, \quad (2.17)$$

$$TI_k = M_k^b - M_k^s, \quad (2.18)$$

wobei  $TI_k$  die *Handelsimbalance* ist und  $\eta_k$  den Rundungsfehler bezeichnet. Man sieht, dass die Preisänderung primär von  $OFI_k$  abhängt, welches wiederum nur von den Bewegungen an den besten Ask und Bid Stellen im Order Buch abhängt.

Falls Limit Order Aktivitäten gegenüber den Marktaktivitäten dominieren, so ist die Korrelation der Preisänderungen mit  $TI_k$  schwächer als mit  $OFI_k$ , da Limit Orders und Stornierungen nur als Rauschen in (2.17) vorkommen.

Man nimmt deshalb die folgende Beziehung an, die für kurze Zeitintervalle  $[t_{k-1,i}, t_{k,i}] \subset [T_{i-1}, T_i]$  gilt:

$$\Delta P_{k,i} = \beta_i OFI_{k,i} + \varepsilon_{k,i} \quad (2.19)$$

$\beta_i$  ist der in der Einleitung erwähnte *Price Impact Coefficient* für das  $i$ -te Zeitintervall. Man nimmt weiters an, dass

$$\beta_i = \frac{c}{D_i^\lambda} + \nu_i \quad (2.20)$$

wobei  $c, \lambda = const.$  und  $\nu_i$  ein Rauschen darstellt. In unserem Beispiel sind  $c = \frac{1}{2}, \lambda = 1$ .

## 2.2.3 Schätzen des Parameters und Resultate

Angewandt auf empirischen Daten muss man die Definitionen von  $OFI, \Delta P, TI$  und  $VOL$  an diese anpassen. Die besten empirischen Resultate gäbe es freilich mit den ursprünglichen Definitionen. Die *Tiefe* wird durch Mittelung von den Bid/Ask Warteschlangen-Größen vor und nach eines Preissprunges modelliert.

Für jede halbe Stunde werden die Parameter aus (2.21) neu regressiert mittels der Methode der kleinsten Quadrate:

$$\Delta P_{k,i} = \hat{\alpha}_i + \hat{\beta}_i OFI_{k,i} + \hat{\varepsilon}_{k,i} \quad (2.21)$$

Für diese Beziehung wurde eine hohe statistische Signifikanz gefunden. Beispielhafte Werte für die Schlumberger Aktie wären  $\hat{\alpha}_i = -0.002$ ,  $\hat{\beta}_i = 0.009$ ,  $R^2 = 76\%$ . Für höhere Abhängigkeiten (zB quadratische Abhängigkeit) gibt es keine Signifikanz. Auch Mittelung über alle verfügbaren Aktien gab ähnlich gute Signifikanzen [3].

Man erhält weiters die Einsicht, dass *OFI* Preisbewegungen viel besser erklären kann als die Trade Imbalance, wobei *OFI* als ein viel allgemeineres Maß für Angebot und Nachfrage dient, da es die Trade Imbalance angemessener abbildet.

Beziehung (2.20) wurde ebenfalls geschätzt und es wurde eine signifikante Abhängigkeit gefunden. Also korreliert die Markttiefe  $D$  mit den Koeffizienten, die auch Preisänderungen verursachen.  $\frac{1}{2\hat{\beta}_i}$  ist die *implizite Limit Order Buch Tiefe*.

## 2.2.4 Anwendungen

In einer ersten Anwendung wird die *OFI* als Maß für Negativselektion bei Orders verwendet. Wenn eine Limit Order ausgeführt wird und vor bzw. während dieser Ausführung hat es ein positives *OFI* gegeben, dann ist es wahrscheinlich, dass ein Preisanstieg stattfinden wird.

Im Falle einer Verkaufs Limit Order bedeutet ein positiver Preisanstieg nach der Ausführung der Order, dass diese verlustreich verkauft worden ist, d.h. Negativselektion hat stattgefunden.

## 2.2.5 Intraday Volatility dynamics

Durch diverse zeitliche und über die Aktien hinweg durchgeführte Mittelwertbildungen wurden 13  $\hat{\lambda}_i$ ,  $\hat{c}_i$  für jede Aktie ermittelt [3]. Diese Price Impact Koeffizienten zeigten, dass die Markttiefe zwischen 9:30 und 10:00 Uhr zweimal so klein ist wie durchschnittlich. Es gibt zu dieser Zeit also einen flachen Markt, das heißt, eingehende Orders können leicht zu Preisveränderungen (Mid Preis) führen. Gegen Ende des Handelstages hin sinken diese Koeffizienten weiter deutlich. Mit diesem Wissen kann man nun beobachtete Intraday Muster erklären.

Weiters besteht die Vermutung, dass Preisveränderung und Informationsasymmetrie ein und dasselbe sein könnten. Wenn es am Morgen viel private Information gibt, und den Limit Order Händlern dies bewusst ist, werden sie vernünftigerweise am Morgen eher weniger am Marktgeschehen teilnehmen.

Die Beobachtung, dass Stärke von Preisänderungen und Handelsvolumen positiv miteinander korrelieren, wurde bestätigt [1]. Jedoch ist weiterhin unklar, ob Handelsvolumina

tatsächlich die Stärke von Preisprüngen beeinflussen und ob es ein gutes Maß dafür ist. Es gibt aber Hinweise darauf, dass dies auch auf täglicher Datenbasis der Fall sein könnte.

## 2.2.6 Conclusio

Das lineare Modell zwischen hochfrequenten Preisänderungen und der eingeführten Maßzahl *OFI* wurde theoretisch und praktisch validiert. Das Modell passt auf die meisten Aktien und Zeitskalen. Der Price Impact Coefficient ist umgekehrt proportional zur Markttiefe. Daraus folgt, dass Preise auf Änderungen in Angebot und Nachfrage bei der besten Limit Order reagieren. *OFI* ist überdies ein allgemeineres Maß für die Angebots- und Nachfrageänderungen als Handelsüberschüsse bzw. -defizite. Außerdem kann es zu Analyse- und Aufsichtszwecken verwendet werden.

## 2.3 First Passage Time Modell

### 2.3.1 Abkürzungen

- *FPT = First Passage Time*. Das ist jene Zeit, die man abwarten muss, um einen gegebenen Preissprung  $\Delta > 0$  zu beobachten.
- *TTF = Time To Fill*. Die Zeit von der Platzierung der Order bis zur kompletten Ausführung.
- *TTC = Time To Cancel*. Zeit von der Platzierung einer Order bis zur Stornierung.
- *TFFF = Time To First Fill*. Zeit bis zur ersten Transaktion, wo die Order "teilnimmt" (Aufträge werden oft nicht in einer einzigen Transaktion ausgeführt).

Es wurde beobachtet, dass die Verteilungen von *FPT*, *TTF* und *TTC* asymptotisch als Potenzgesetze fallen, wobei die First Passage Time ein deutlich schwereres Ende hat [8]. Der Grund für diesen Unterschied liegt darin, dass Stornierungen durchgeführt werden. Für diesen Sachverhalt wird nun im Folgenden ein geeignetes, einfaches Modell vorgestellt, welches die beobachteten Daten hinreichend gut erklären kann.

Die getroffenen Annahmen, auf denen dieses Modell beruht, lauten:

1. Preise werden durch die empirisch beobachtete *FPT* bestimmt
2. Aufträge werden in zufälliger Art und Weise storniert. Die Lebensdauer fällt mit einem Potenzgesetz.

Mit diesem Modell ist es nun möglich, die Qualität von Stornierungsstrategien mittels empirischer Daten zu schätzen.

### 2.3.2 First Passage Time

Die First Passage Time eines Preisprozesses  $S(t)$  zum Level  $S_0 + \Delta$ ,  $\Delta > 0$ , ist definiert als der erste Zeitpunkt, zu dem eine Transaktion zum Preis  $S(t) \geq S_0 + \Delta$  durchgeführt wird. Notiere die Verteilung von  $t$  bzgl.  $\Delta$  mit  $P_{\text{FPT};\Delta}(t)$ .  $\Delta > 0$  bedeutet, dass der Limit Order Preis schlechter ist als der aktuell gehandelte, sprich der Normalfall einer Limit Order.

Da die diskreten Aspekte vernachlässigbar sind für unsere Zwecke, werden die Preisprozesse als stetig angenommen. Der Grund dafür liegt in den betrachteten Zeitintervallen.

Die Verteilung der FPT kann man gut mit einer deterministischen Funktion anpassen, da sich die Verteilungen der FPT von Preis, Bid und Ask nicht wesentlich voneinander unterscheiden. Die folgende, zwei-parametrische Form eignet sich dafür gut:

$$P_{\text{FPT};\Delta}(t) = \frac{Ct^{-\lambda_{\text{FPT}}}}{1 + \left[\frac{t}{T_{\text{FPT}}(\Delta)}\right]^{-\lambda_{\text{FPT}} + \lambda'_{\text{FPT}}}} \quad (2.22)$$

Die Annahme, dass man Stornierungen aus der Analyse des Limit Order Buches ausnehmen kann, führt zu einem Widerspruch mit den beobachteten Daten. Das hat den Grund, da Aufträge, die zu lange im Order Buch stehen, einfach storniert werden und daher aus der Statistik fallen. Die zwei Parameter  $\lambda_{\text{FPT}}$  und  $\lambda'_{\text{FPT}}$  werden aus den Daten geschätzt.

### 2.3.3 Ein einfaches Modell für Time To Fill und Time To Cancel

Das Modell betrachtet oBdA folgende Situation:

1. Es wird bei  $t = 0$  ein einziger Kaufauftrag mit einem Abstand von  $\Delta > 0$  vom aktuell besten Preis platziert.
2. Schon bei der Platzierung wird ein Zeitpunkt  $t'$  definiert, bis zu dem die Order höchstens im Order Buch verbleiben soll. Danach wird storniert.  $t'$  wird zufällig aus ihrer Lifetime Distribution  $P_{\text{LT};\Delta}(t')$  ausgewählt.
3. Der Markt ist liquide und die Tick Size ist klein.

Unter diesen Annahmen kann man nun eine zusammengesetzte Dichtefunktion anschreiben, die sowohl den Preisdiffusionsprozess als auch Stornierungen beschreibt. Die Wahrscheinlichkeit  $P_{\Delta}(t, t')$ , dass der Preis im Zeitintervall  $(t, t')$  eine Order erreicht, die  $\Delta$ -weit weg ist vom aktuell besten Preis, und, dass die Order bei  $t'$  storniert wird, lässt sich zerlegen in

$$P_{\Delta}(t, t') = P_{\text{FPT};\Delta}(t)P_{\text{LT};\Delta}(t') \quad (2.23)$$

Dies entspricht der Verteilung, wie lange der Investor gewillt ist zu warten, bis seine Order ausgeführt wird. Diese Beziehung beschreibt aber nicht, wie lange der Investor die Aktie hält.

Mit diesem Modell kann man nun gut die Differenz zwischen dem asymptotischen Abflachen der FPT und der TTF beschreiben. Weiters kann man die Verteilung der Lebensdauer beachten.

Eine Order mit  $\Delta \leq 0$  bedeutet, dass es sich entweder um eine Markt Order (Fall 0) oder sogar um eine noch bessere Order (Fall  $< 0$ ) handelt. Beides ergibt Sinn, betrachtet zu werden. Warum aber sollten diese Orders nicht sofort ausgeführt werden? Es könnte nämlich sein, dass, bevor unser Auftrag ausgeführt wird, ein noch besseres Angebot hereinkommt, sodass wiederum  $\Delta > 0$  effektiv gilt, und unser Modell somit wieder Anwendung findet.

## 2.4 Ein allgemeineres LOB Modell

### 2.4.1 Allgemeines

Ein Limit Order Buch bezüglich einer gehandelten Aktie sei ein Preisgitter  $\{k\tau\}_{k \in \mathbb{Z}}$  mit der Schrittgröße (a.k.a tick size)  $\tau > 0$ . Der Spread ist die Differenz zwischen dem besten Ask und dem besten Bid Angebot. Man nimmt meistens an, dass der Spread konstant ist (zB  $2\tau$ ), da dies in der Praxis nichts Ungewöhnliches ist. Wenn die Liquidität auf einem Level verbraucht ist, wird eine neue Order zwischen dem besten Angeboten durch einen Händler oder einen Marktmacher platziert, um Liquidität zu gewährleisten. Die Dynamiken, die bei diesem Vorgang passieren, sind Gegenstand dieses Abschnittes, welcher aus [11] entnommen wurde.

Wir befinden uns stets in einem filtrierten Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}, P)$  und es gebe nur folgende Order Buch Ereignisse: Limit Orders, Markt Orders und Stornierungen.

### 2.4.2 Limit Orders und Stornierungen

Limit Orders werden anhand eines Poisson-Prozesses ausgewählt. Jedes Mal, wenn dieser Prozess einen Sprung macht, wird eine Order platziert. D.h. seien  $\{T_i\}_{i \geq 1}$  diese Sprungzeiten, dann ist die Dauer zwischen zwei Sprüngen exponentialverteilt:  $T_{i+1} - T_i \sim \exp(\lambda(T_{i+1} - T_i))$ .

Zuerst wird die Seite ausgewählt, wobei Ask oder Bid Seite beide gleichwahrscheinlich sind. Danach wird die Order nach einem empirisch bestätigten Potenzgesetz ([1],[2],[3],[4])  $\mu$  ticks vom aktuellen Mid preis (Mittelwert aus bestem Bid und Ask) entfernt platziert, wobei für  $\mu$

$$P[\mu = j] = \frac{k}{j^{1+\alpha}}, \quad j = 1, \dots, 100 \quad (2.24)$$

gilt mit  $\alpha > 0$  und  $k = \left[ \sum_{i=1}^{100} \frac{1}{i^{1+\alpha}} \right]^{-1}$ . Die Zahl 100 ist willkürlich gewählt und völlig ausreichend. Stornierungen werden analog modelliert.

### 2.4.3 Markt Orders

Sei  $\sigma > 0$  und  $B$  eine Brownsche Bewegung bezüglich ihrer erzeugten Filtration. Der mittlere Preisprozess ist gegeben durch  $S_t = \sigma B_t, t \geq 0$ .

Approximiere nun den (stetigen) Preisprozess durch einen konstruierten, diskreten Prozess mithilfe des Satzes von *Donsker*. Die Approximation ist bezüglich Konvergenz in Verteilung aufzufassen.

$$S_t = \sigma B_t \approx \sigma X_t^{(n)} = \frac{\sigma}{\sqrt{n}} Y_{nt} \quad (2.25)$$

$Y$  wird dabei aus einer iid Folge von Zufallsvariablen  $\{\xi_j\}_{j \geq 1}$  konstruiert, wobei  $\mathbb{E}[\xi_1] = 0$  und  $\text{Var}[\xi_1] = s^2, 0 < s^2 < \infty$ .

Man definiert nun den acht-stündigen Handelstag als eine Zeiteinheit und betrachtet als diskretes Inkrement die Sekunden  $h = 1/n$ , wenn  $n$  die Anzahl der Sekunden  $n = 60 \cdot 60 \cdot 8$  bezeichnet. Für  $n$  gibt es jedoch eine Mindestgröße aus Approximationsgründen. Man schaut nun nach, ob auf jeder Sekunde der Preisprozess eine Barriere (beste Bid oder Ask Offer) berührt. Da  $n$  sehr groß ist, ist es nicht tragisch, nicht zwischen den Sekunden zu prüfen.

Wenn man mit  $p_A(t)$  den aktuell besten Ask Preis bezeichnet und der Preisprozess größer oder gleich diesem Level ist, so werden alle Orders auf diesem Level ausgeführt und der Mid Preis geht einen Tick nach oben und die besten Levels werden aktualisiert:

$$p_A^{new} = p_A(t) + \tau, \quad (2.26)$$

und analog für die Bid Seite. Dieses Vorgehen kann man nun aber noch verbessern.

Da auf kurzen Zeitintervallen die bekannte Verteilung der First Passage Time für eine Brownsche Bewegung sehr stark davon abweicht, was tatsächlich beobachtet wird, wird auf folgenden, oben bereits vorgestellten Ansatz zurückgegriffen, wo die Dichte als

$$h(t) = \frac{Ct^{-\lambda_F}}{1 + [t/T_F(\Delta)]^{-\lambda_F + \lambda'_F}} \quad (2.27)$$

modelliert wird.  $T_F, \lambda_F, \lambda'_F$  können leicht aus den Daten geschätzt werden.  $\lambda_F$  hat zum Beispiel typischerweise einen Wert von  $3/2$ .

Nun arbeitet man heraus, wie man zu einer gegebenen Verteilung  $H$  auf  $\mathbb{R}_+$  einen càdlàg stochastischen Prozess  $Y = \{Y_t\}_{t \geq 0}$  findet, sodass die First Passage Time unter Null,

$$T_0^Y := \inf\{t \geq 0 : Y_t < 0\}, \quad (2.28)$$

nach  $H$  verteilt ist. Dies kann man dazu benutzen, die Verteilung der First Passage Time eines beliebigen Levels  $\Delta$  im Preisgitter anzugeben:

$$P\left[ \underbrace{T_\Delta^Y}_{\inf\{t \geq 0 : Y_t > \Delta\}} \leq t \right] = \int_0^t \frac{Cs^{-\lambda_F}}{1 + [s/T_F(\Delta)]^{-\lambda_F + \lambda'_F}} ds \quad (2.29)$$

Es werden nun quantitative Größen untersucht, die nützlich sind, um einerseits das Limit Order Buch besser zu verstehen, und andererseits, um Handelsstrategien daraus abzuleiten. Zu diesem Zweck betrachte die Stoppzeiten, wenn die ganze Liquidität zum ersten

Mal verbraucht wird, sowohl auf der Bid als auch auf der Ask Seite:

$$T_A = \inf\{t \geq 0 : Y_t > \tau\} \quad (2.30)$$

$$T_B = \inf\{t \geq 0 : Y_t < -\tau\}$$

1. Die Wahrscheinlichkeit, dass ein Preisanstieg vor einem Preisverfall stattfindet, ist aus Symmetriegründen gleich

$$P[T_A < T_B] = \frac{1}{2} \quad (2.31)$$

2. Daraus lässt sich die Dichtefunktion der Dauer bis zur nächsten Preisbewegung des Mittelpreises herleiten:

$$P[T_A \wedge T_B \in dt] = f_1(I_t)\sigma_t^2 dt \quad (2.32)$$

wobei  $f_1(t)$  eine deterministische Funktion ist und  $I_t = \int_0^t \sigma_s^2 ds < \infty$ .

3. Die Dichtefunktion des Ereignisses, eine (zB Verkaufs) Order ausgeführt zu bekommen bis zur nächsten Bewegung des Mid Preises, ist gleich

$$P[T_A < T_B, T_A \in dt] = f_2(I_t)\sigma_t^2 dt, \quad (2.33)$$

wobei  $f_2(t) = (2t)^{-1}f_1(t)$ .

4. Definiere die *Ausführungen* als das Verhältnis der Anzahl an Trades und Anzahl an Orders im Order Buch. Die Wahrscheinlichkeit einer Ausführung lässt sich nun bestimmen.

# Kapitel 3

## Datenanalyse

### 3.1 Beschreibung der Datenstruktur

Die Datengrundlage dieser Arbeit stellt ein Datensatz von der *NASDAQ* Börse dar. Grundsätzlich werden in dieser Arbeit 20 große NASDAQ Aktien betrachtet. Es sind dies (in Klammern das NASDAQ Symbol): Apple (AAPL), Adobe (ADBE), Alexion (ALXN), Amgen (AMGN), Amazon (AMZN), Baidu (BIDU), Biogen (BIIB), Costco (COST), Alphabet (GOOGL), Intuitive Surgical (ISRG), Lam Research (LRCX), Microsoft (MSFT), Netflix (NFLX), Netease (NTES), NVIDIA (NVDA), O'Reilly (REGN), Pricelinecom (PCLN), Regeneron (REGN), Tesla (TSLA) und Ulta Salon (ULTA). Diese Titel wurden danach ausgewählt, ob der Preis über einem gewissen Niveau ist und nach Bekanntheit, grundsätzlich aber rein willkürlich. So sollten auch die zugehörigen Limit Order Bücher reichhaltig sein.

Beim Bezug der Daten ist die gewünschte Anzahl der Levels des Order Buches anzugeben. "1 Level" würde zum Beispiel bedeuten, dass man nur Informationen über den besten Bid und besten Ask Preis haben möchte, und nicht, was sich einen tick vor dem besten Bid oder einen tick nach dem besten Ask abspielt. Bei "2 Levels" würde man folgerichtig auf jeder Seite Informationen zu den je zwei "besten" Levels erhalten usw. Natürlich bzw. leider steigt mit zunehmenden Levels auch die Datengröße massiv an. In dieser Arbeit wurde stets mit 3 Levels gearbeitet. Dies bietet einerseits genügen Informationen über das LOB, andererseits schont dies auch den verfügbaren Speicherplatz.

Sobald man die oben erwähnten Parameter definiert hat, bekommt man einen Datensatz geliefert. Dieser besteht stets aus zwei separaten Dateien. Die erste Datei, das sogenannte *message file*, stellt eine Art Journal bzw. Historie dar über die chronologische Entwicklung des LOB. Es enthält Indikatoren über die Art des Ereignisses, welches den Status des LOB verändert hat. Jedes Ereignis ist zeitlich markiert in Sekunden nach Mitternacht, wobei zumindest Millisekunden angegeben werden und bis zu Nanosekundengenauigkeit möglich ist. Nachfolgend eine schematische Darstellung des Aufbaus des message files.



Time (sec)	Event Type	Order ID	Size	Price	Direction
⋮	⋮	⋮	⋮	⋮	⋮
34713.685155243	1	206833312	100	118600	-1
34714.133632201	3	206833312	100	118600	-1
⋮	⋮	⋮	⋮	⋮	⋮

Was bedeuten die einzelnen Spalten im message file?

- Time: Sekunden nach Mitternacht mit mindestens Millisekunden Präzision (fallweise bis zu Nanosekunden Genauigkeit möglich)
- Event Type: Es gibt 6 verschiedene Arten eines LOB Ereignisses:
  - 1: Aufgabe einer neuen Limit Order
  - 2: Partielle Stornierung
  - 3: Komplette Stornierung
  - 4: Ausführung einer sichtbaren Limit Order
  - 5: Ausführung einer verdeckten Limit Order
  - 7: Indikator für Handelsstop
- Order ID: Eindeutige Kennzeichnung der Order
- Size: Anzahl der Aktien
- Preis: Preis in US \$ mal 10.000 (d.h. 118600 = \$ 11,86)
- Direction:
  - -1: Verkaufs Order
  - 1: Kauf Order

Neben dem message file wird das *order book file* geliefert. Dieses stellt die Entwicklung des Limit Order Buches mit der definierten Anzahl an Levels dar. Es hat folgenden schematischen Aufbau (am Beispiel von 2 Levels):

Ask Price 1	Ask Size 1	Bid Price 1	Bid Size 1	Ask Price 2	Ask Size 2	Bid Price 2	Bid Size 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
118600	9484	118500	8800	118700	22700	22700	14930
118600	9384	118500	8800	118700	22700	22700	14930
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Erläuterung:

- Ask Price 1: Level 1 Ask Preis (bester Ask Preis)
- Ask Size 1: Level 1 Ask Volumen (Volumen am besten Ask Level)
- Bid Price 1: Level 1 Bid Preis (bester Bid Preis)
- Bid Size 1: Level 1 Bid Volumen (Volumen am besten Bid Level)

Das message und order book file können auch als Matrizen der Größe  $N \times 6$  bzw.  $N \times (4 \cdot \text{Anzahl Levels})$  aufgefasst werden.  $N$  steht hierbei dann für die Anzahl der Ereignisse.

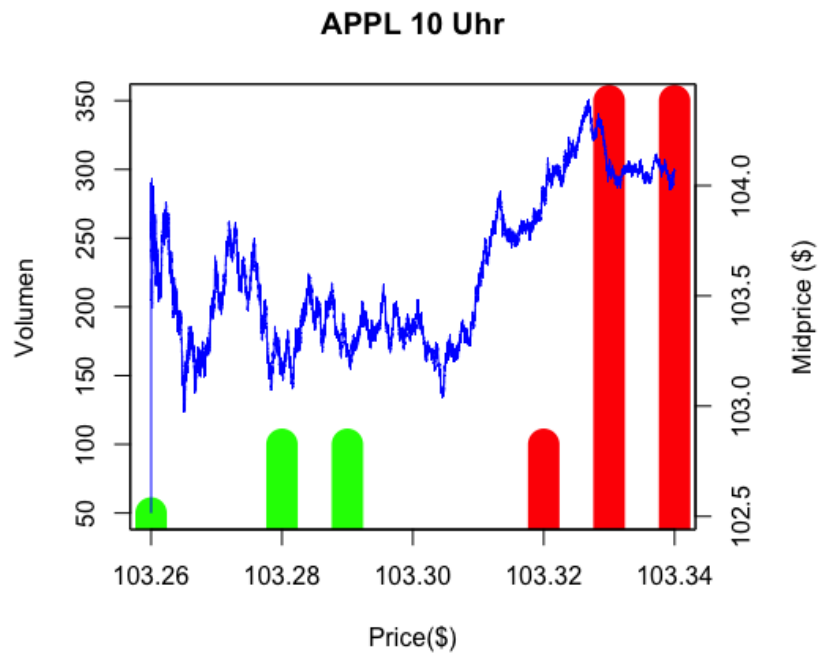
Die  $k$ -te Zeile im message file korrespondiert mit jenem Limit Order Ereignis, welches die  $k - 1$ -te Zeile im order book file in die  $k$ -te Zeile überführt. Betrachten wir als Beispiel die zweite Zeile in der schematischen Darstellung des message file oben. Wir haben ein Ereignis vom Typ 3, also eine (komplette) Stornierung von 100 Stück zum Preis von 11,68 Dollar. Im order book file sieht man dies dann in der zweiten Spalte. Auf dem besten Ask Preis werden statt 9484 Stück Aktien nur noch 9384 angeführt, also um genau jene 100 Stück weniger.

## 3.2 Qualitative Analyse

Betrachten wir nun die ersten Auswertungen aus den vorhandenen Daten. Diskutiert wird, wie oben bereits erwähnt, das Limit Order Buch von 20 großen Aktien am 16. November 2016 von der *NASDAQ* Börse. Die Anzahl der Levels beträgt drei, da dies sowohl speicherplatzschonend als auch schon aussagekräftig ist.

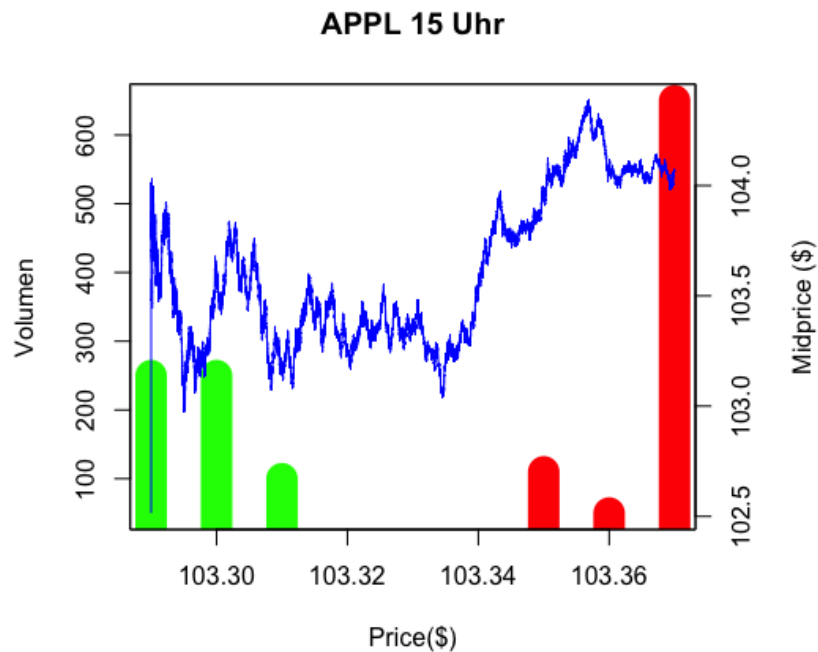
Im Folgenden betrachten wir Plots wie den nachfolgenden. Diese sind wie folgt zu lesen. In der Überschrift findet man den Ticker, in diesem Fall AAPL für die Apple Aktie. Auf der  $x$ -Achse sind die Preislevels aufgetragen, auf der linken (relevanten)  $y$ -Achse das Volumen. Die dritte grüne Säule von links kennzeichnet den aktuell besten Bid Preis (der Plot ist eine Momentaufnahme von 10 Uhr am 16.11.2016), links davon das zweite und das dritte Level. Analog gilt dies für die roten Balken, die Ask Seite. Man sieht also, dass der Spread (best Ask minus best Bid) wie in den meisten Fällen einen Cent beträgt.

Qualitativ sieht man, dass auf der Ask Seite beim innersten Level deutlich weniger Volumen sitzt als auf den äußeren Levels.



Der blaue Graph zeigt den Verlauf des Mid Preises (arithmetisches Mittel aus bestem Bid und bestem Ask Preis) für den kompletten Handelstag. Die rechte  $y$ -Achse ist hierbei die relevante Ordinate. Ein Handelstag geht von 9:30 los und endet um 16 Uhr. An diesem Tag ist der Mid Preis der Apple Aktie bei rund 106 Dollar gestartet und hat bei rund 110 Dollar den Handelstag abgeschlossen.

Um 15 Uhr sieht die Situation folgendermaßen aus:

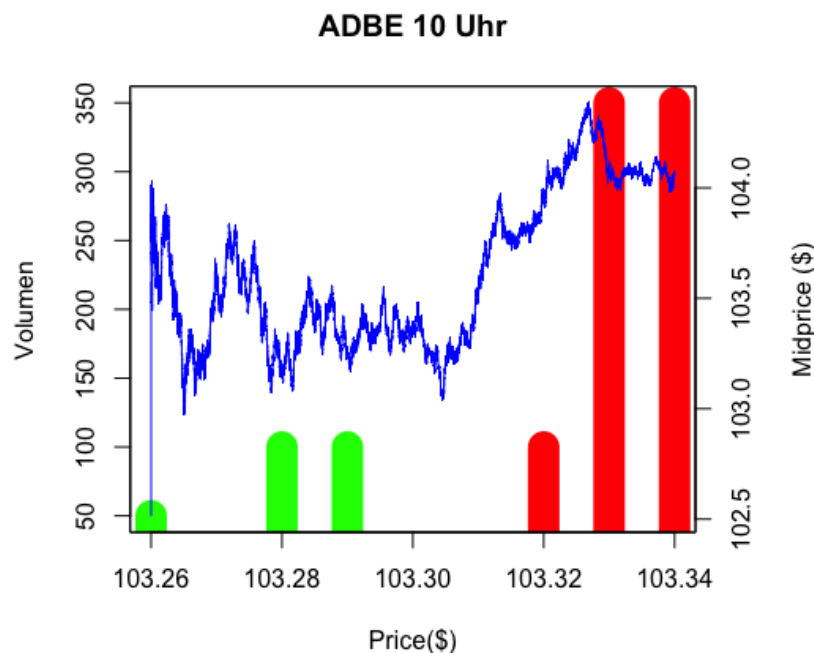


Die Summe der Volumina von Bid und Ask halten sich ungefähr die Waage und man sieht, dass um diese Tageszeit der Verlauf des Mid Preises tendenziell eher einen seitwärtsgerichteten Graphen zeigt. Von der qualitativen Interpretation her sieht man auch hier, dass auf den besten Bid und Ask Levels weniger Volumen liegt als auf den hinteren Levels.

Zu erwähnen ist, dass es zur Bestimmung des Mid Preises natürlich auch andere Ansätze als das reine arithmetische Mittel gibt. Eine naheliegende Verbesserung stellt die Verwendung eines gewichteten Mittelwertes dar, wobei die Gewichte die Volumina auf den beiden Seiten des Limit Order Buches darstellen. Somit werden die vorkommenden Asymmetrien viel besser berücksichtigt. Ebenfalls ist anzumerken, dass allfällige Rundungen aufgrund der Einschränkung durch die Tick Size auftreten können.

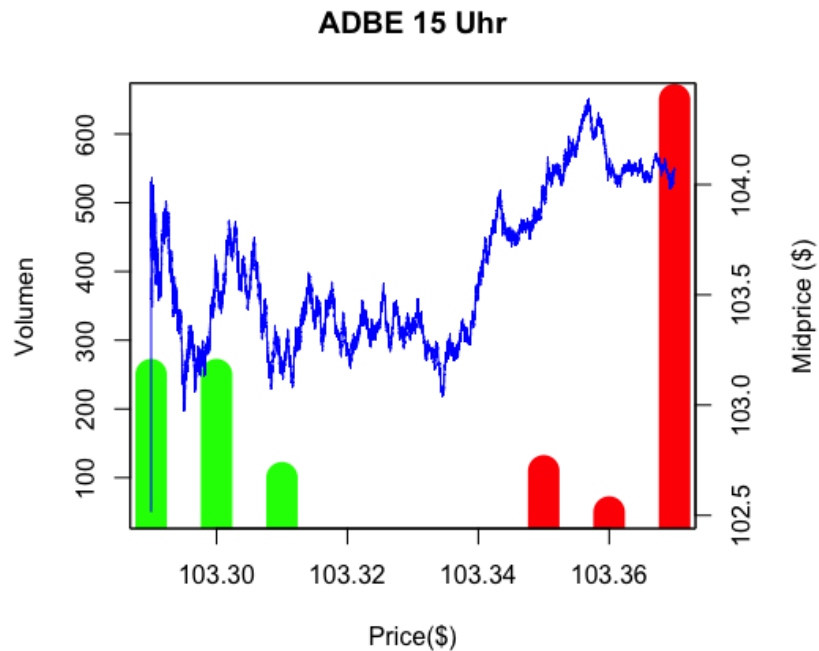
Die nächste Aktie, die wir betrachten, ist jene von Adobe (ADBE). Man erkennt auf den ersten Blick, dass zu diesem Zeitpunkt eindeutig ein Verkaufsdruck vorhanden ist, da auf den Ask Levels viel mehr Volumen liegt als auf der Kaufseite. Wie bei der Apple Aktie ist auch auf der Verkaufsseite am besten Ask weniger Volumen vorhanden als auf den hinteren Levels.

Den Verkaufsdruck erkennt man auch am Verlauf des Mid Preises, der um diese Tageszeit nach unten geht von knapp 140 Dollar bis hinunter nach ca. 130 Dollar.

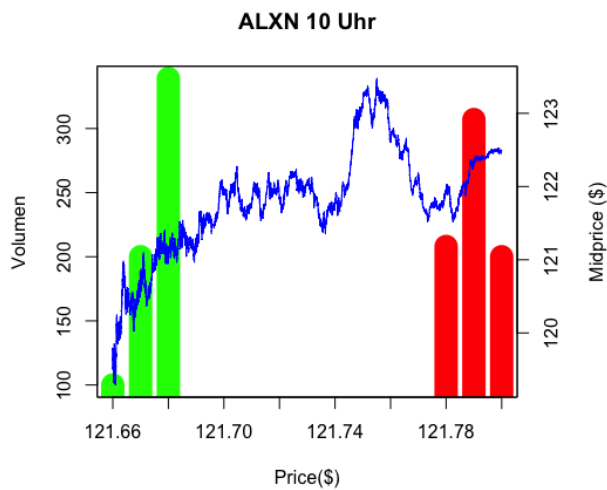


Interessant ist auch der fulminante Start in den Handelstag der Aktie Adobe. Der Mid Preis schießt förmlich von 102,5 Dollar nach oben zu knapp 104 Dollar. Hier war möglicherweise viel private Information vorhanden am Morgen, von welcher schon im Kapitel 2.2.5 die Rede war.

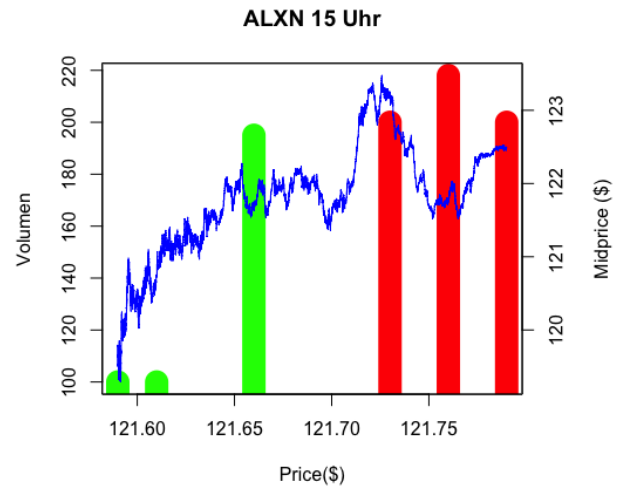
Um 15 Uhr zeigt sich das Limit Order Buch dann folgendermaßen, nachdem bereits ein Preisanstieg stattgefunden hat, trotz des Überhangs auf der Ask Seite. Anschließend werden die restlichen analysierten Aktientitel präsentiert.



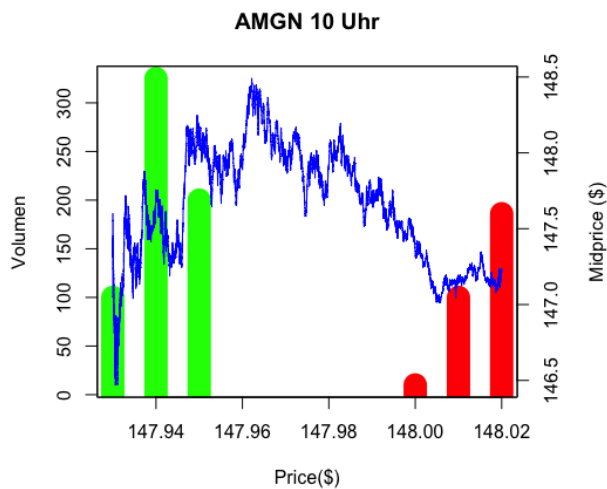
Im Folgenden werden die weiteren Plots der untersuchten Aktientitel dargestellt. Zu diesen gibt es Daten für 10 Uhr und für 15 Uhr. Die restlichen Titel werden anschließend behandelt, da hier am 16.11.2016 aus diversen Gründen (z.B. Handelsstop) keinen Handel stattgefunden hat zu diesen Zeiten.



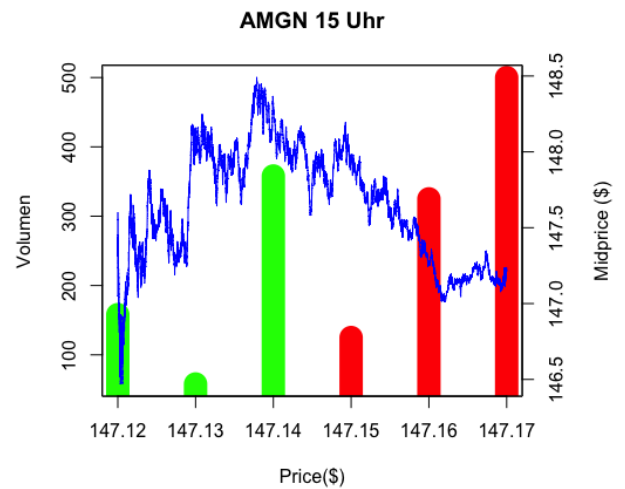
Auch Alexion verzeichnete an diesem Tag einen Anstieg des Mid Preises. Am besten Bid liegt die meiste Masse.



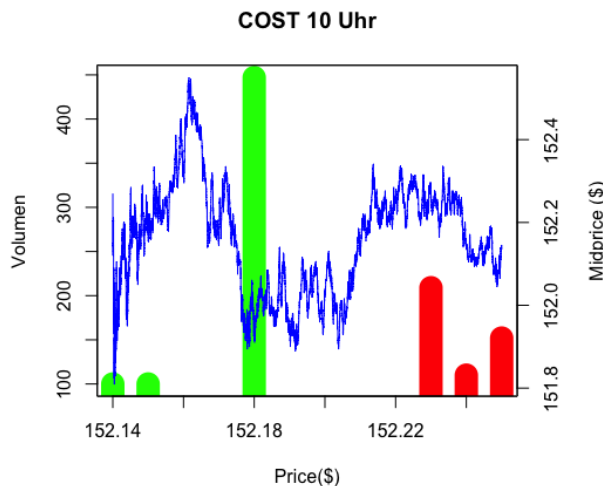
Um 15 Uhr herrschte ein Verkaufsdruck, was man auf der Ask Seite sieht und welcher sich in Niedergang des "kleinen Berges" widerspiegelt.



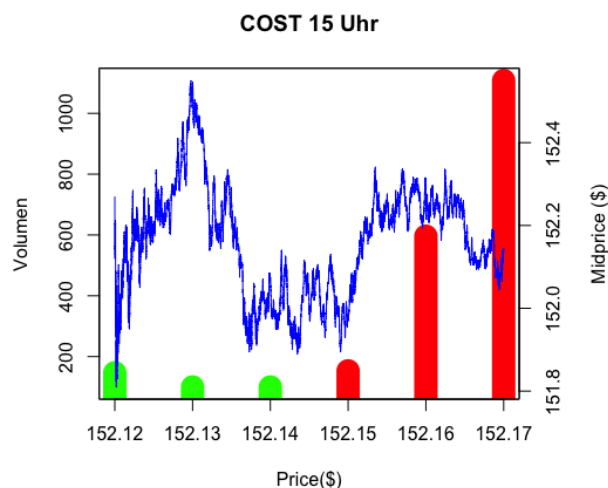
Das LOB von Amgen ist unsynchron. Auf der Bid Seite liegt die meiste Masse auf dem zweitbesten Level. Die Ask Seite ist stufenförmig.



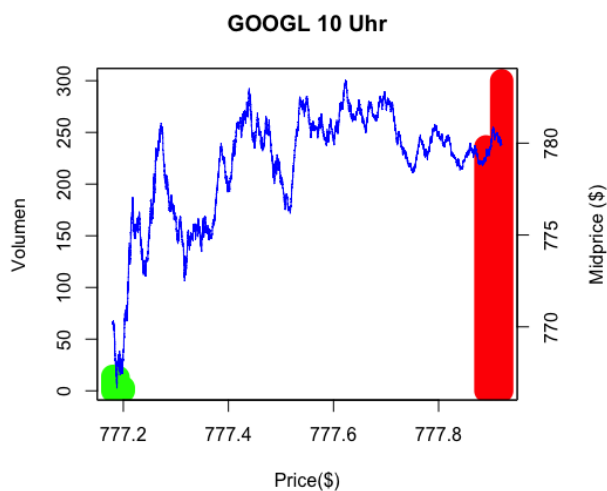
Auch um 15 Uhr hat die Ask Seite die selbe Form, hingegen hat sich auf der Bid Seite die Masse auf das beste Niveau begeben. Dies erklärt vermutlich auch den Verlauf des Mid Preis.



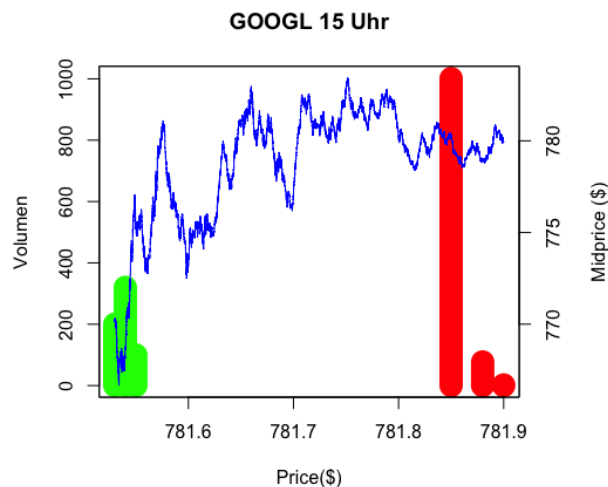
Um 10 Uhr liegt beim Titel Costco die gesamte Masse auf dem besten Bid und dementsprechend zieht der Mid Preis nach oben an...



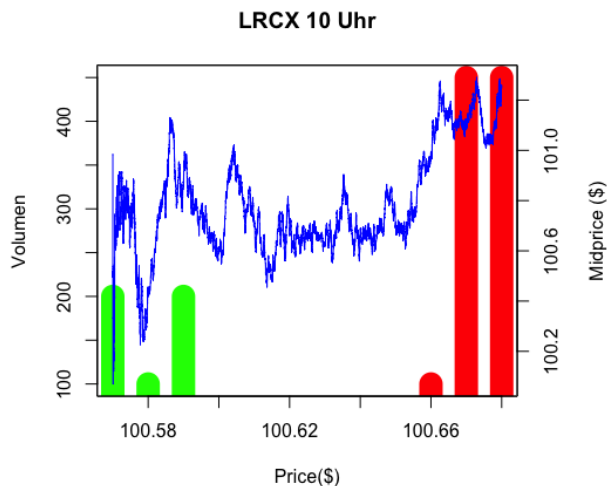
... hingegen ist am Nachmittag auf der Bid Seite keine Aktivität zu sehen, anders als auf der Ask Seite.



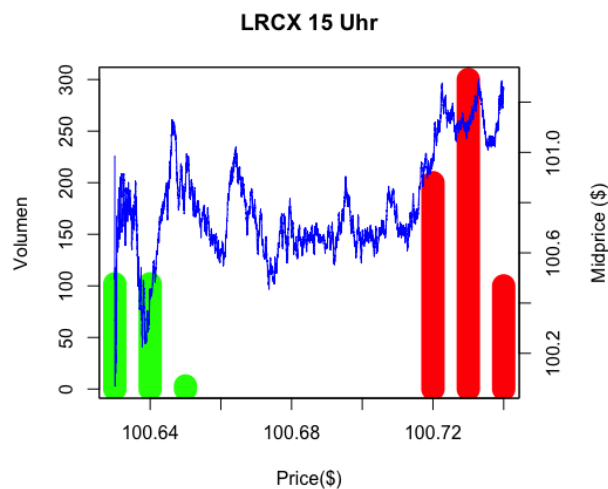
Das LOB von Google sieht sehr dynamisch aus. Zum einen ist der Spread ungewöhnlich groß, auf der Bid Seite liegt wenig Masse, hingegen auf der Ask Seite sehr viel.



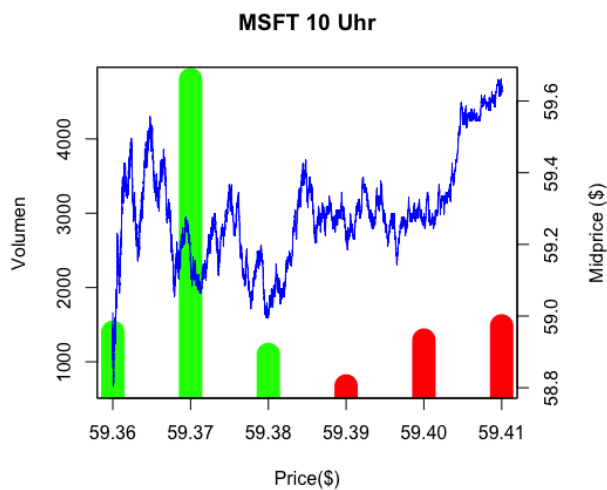
Um 15 Uhr ist die Bid Seite auch noch dicht gedrängt, und auf dem besten Ask Level liegt der Gutteil der Masse.



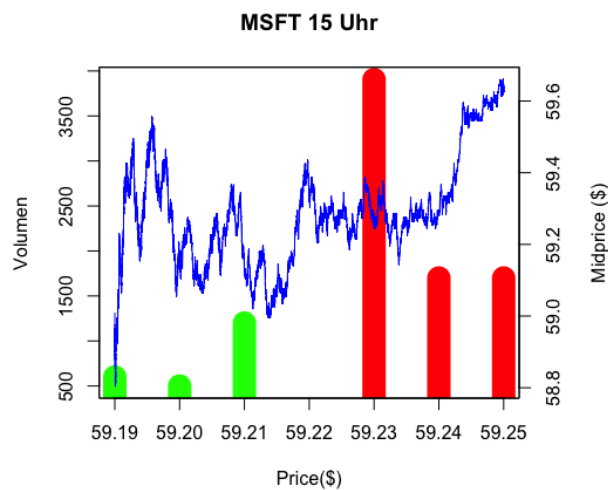
Dieser Status des LOBs von Lam Research stammt vom Zeitpunkt des mittleren ersichtlichen Einbruchs des Mid Preises. Viele Aktientitel stehen zum Verkauf und der Spread ist ebenfalls größer als ein Cent.



Um 15 Uhr ist es etwas ausgeglichener.

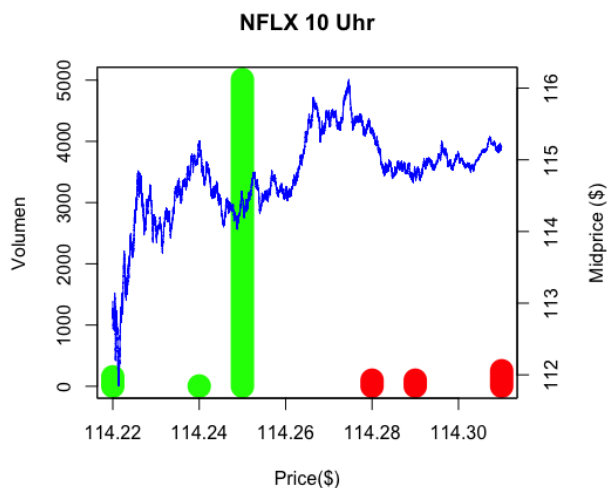


Bei Microsoft hat sich das Volumen von der Bid ...

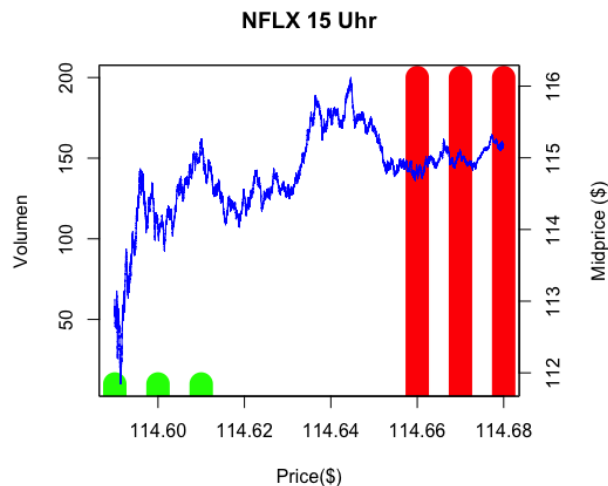


... auf die Ask Seite verschoben im Tagesverlauf.

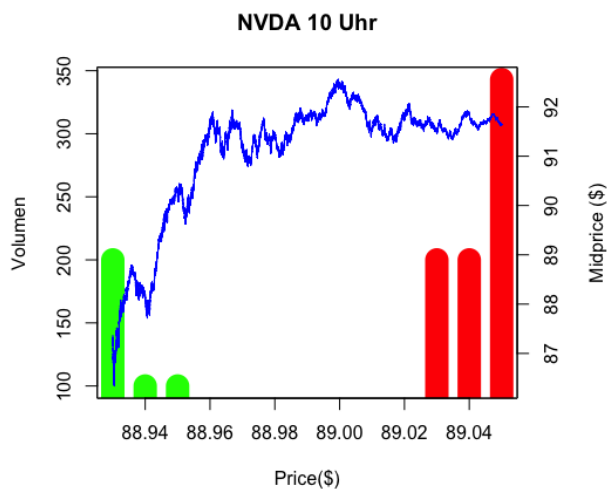




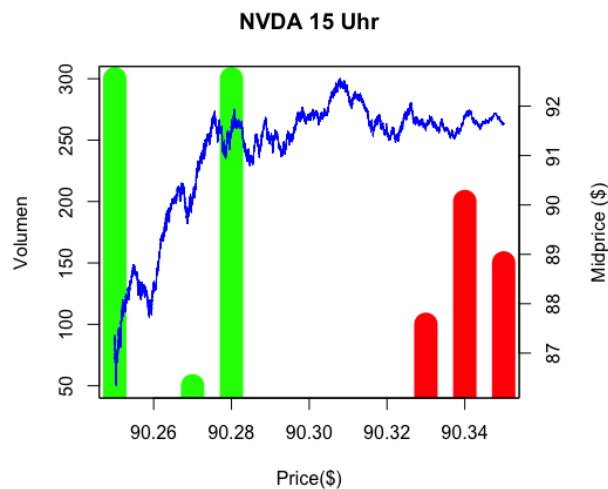
Die Form des LOB von Netflix ist eindeutig, es gibt einen Kaufdruck.



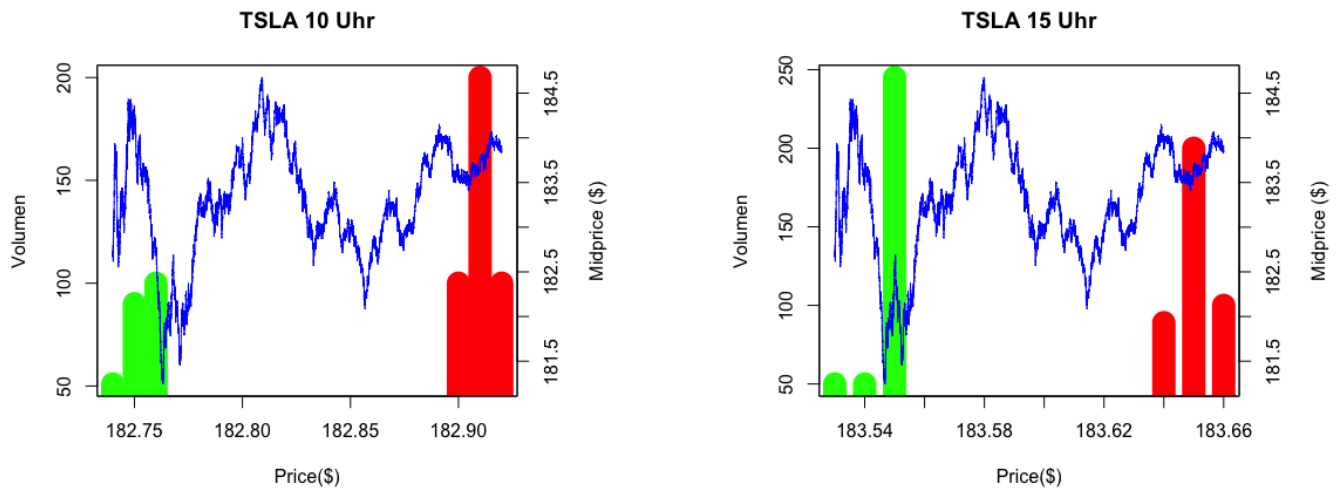
Hingegen präsentiert sich die Form am Nachmittag völlig anders, indem die Anleger die Aktie eindeutig wieder verkaufen möchten.



Nvidia konnte an diesem Tag ebenfalls einen Gewinn verzeichnen, der Spread des LOBs ist auch überdurchschnittlich groß.



Das LOB lässt dies jedoch am Vormittag nicht erahnen. Am Nachmittag ist der Mid Preis schon konstanter, obwohl im LOB deutlich mehr Kauf- als Verkaufsangebote verzeichnet sind.

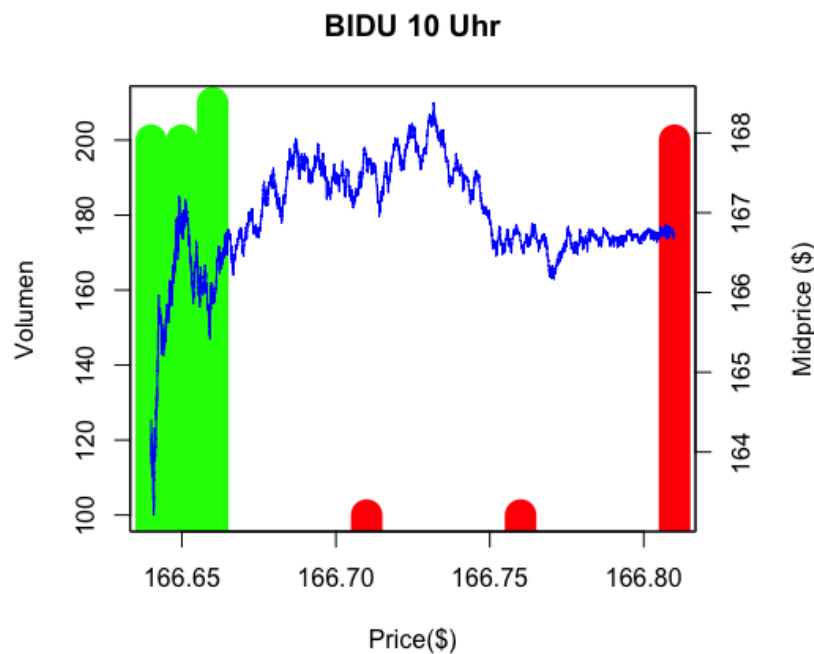


Die Form des Limit Order Buches von Tesla bleibt über den Tagesverlauf ...

... grundsätzlich konstant. Der Verlauf des Mid Preises ist jedoch etwas "wilder".

Nun folgen jene Titel, wo es keine Daten für 10 und 15 Uhr gibt. Dies kann unterschiedliche Gründe haben, in den Daten ist dies jedenfalls als "Handelsstop" gekennzeichnet.

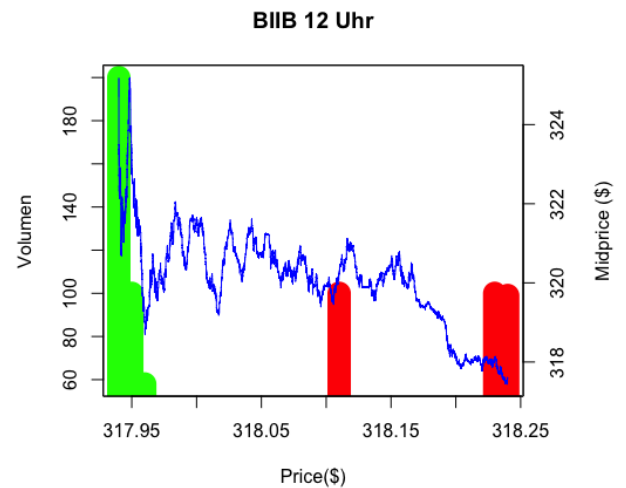
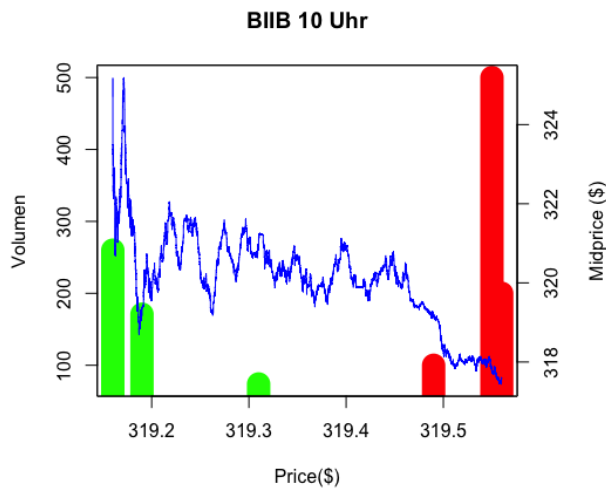
Für die Aktie Baidu liegen nur Daten für 10 Uhr vor:



Die Form des Limit Order Buches ist ungewöhnlich. Auf der Bid Seite liegt die Masse dicht gedrängt, auch hängt auf der Bid Seite viel mehr Masse als auf der Ask Seite. Letztere ist etwas zersetzt. Der beste Ask Level ist nicht der mit der meisten Masse. Diese

liegt auf dem drittbesten Level. Hier liegt offenbar viel Kaufdruck vor, was sich auch im Verlauf des Mid Preises niederschlägt wie man sehen kann.

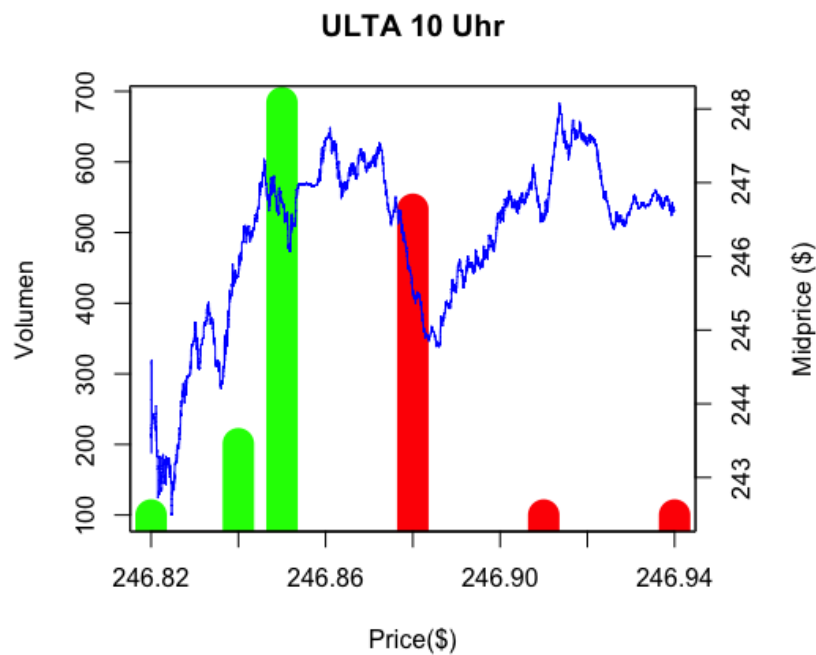
Als nächstes betrachten wir den Titel von Biogen. Hier haben wir Momentaufnahmen des LOBs von 10 und von 12 Uhr zur Verfügung.



Der Mid Preis von Biogen hat im Gegensatz zu den meisten anderen Titeln des NASDAQs an diesem Tag eine Talfahrt hingelegt. Man sieht dies am Überhang der Ask Seite.

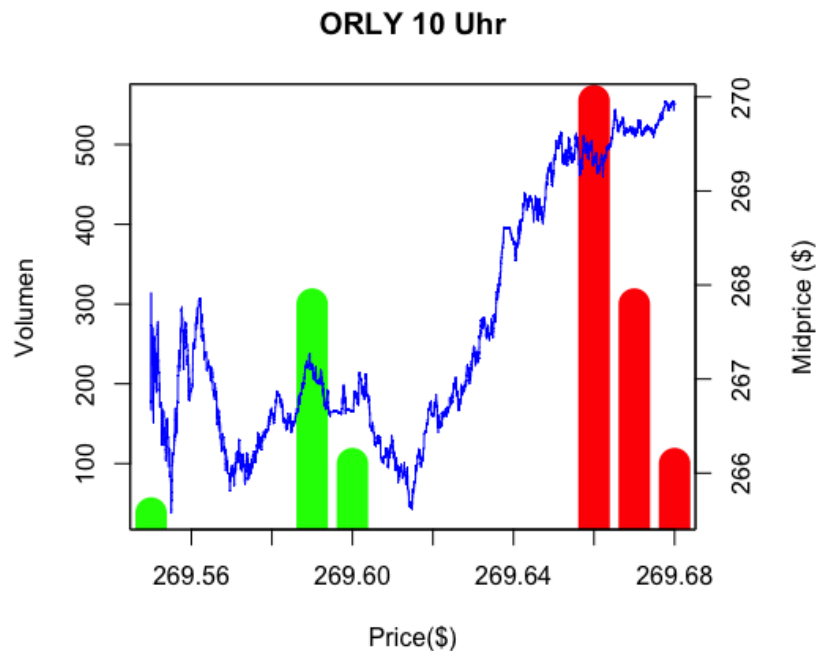
Zu Mittag wären schon mehr Anleger bereit gewesen, zu gewissen Preisen zu kaufen, der Spread war aber anscheinend zu groß.

Als nächstes betrachten wir die Aktie von Ulta Salon, einem Hersteller von Kosmetika. Hier können wir ebenfalls nur einen Plot von 10 Uhr liefern.



Hier ist die Form des Limit Order Buches sehr schön erkennbar, auf den besten Kauf- und Verkaufsseiten liegt die meiste Masse an Aufträgen. Hier wäre es interessant gewesen, ob diese Form am Nachmittag auch noch so zu sehen gewesen wäre.

Der Plot von O'Reilly präsentiert sich folgendermaßen:

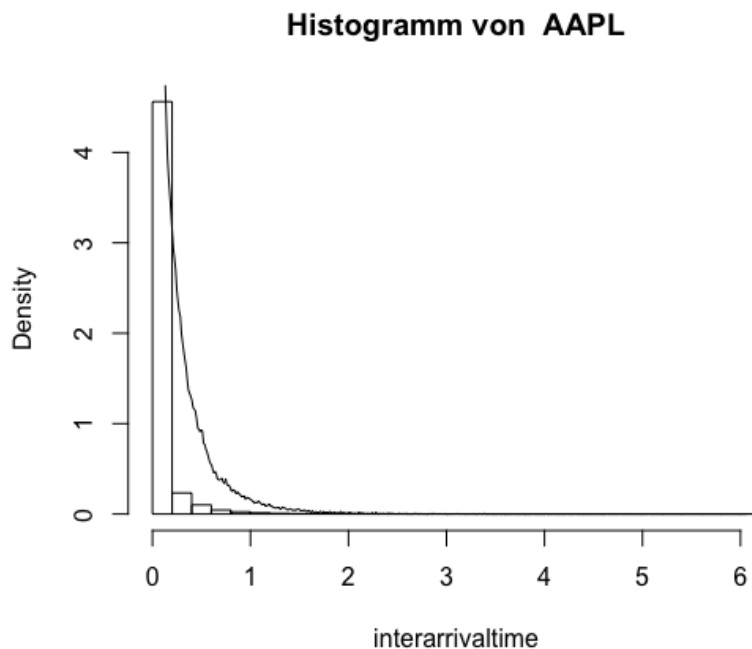


Zu den fehlenden drei Titeln (ISRG, PCLN, REGN) können wir keinen Plot des LOBs liefern, da die Daten sehr unvollständig sind.

### 3.3 Der Arrival Prozess

Betrachten wir nun den Arrival Prozess, also die Zeitpunkte  $T_i$  bzw. Ereignisse, wo Limit Orders ins Order Buch gelangen. Von Interesse sind hierbei natürlich die Abstände  $\Delta_i := T_{i+1} - T_i$  zwischen zwei Orders. Wir wollen uns hier wirklich nur auf Orders beschränken, und lassen Stornierungen und Ausführungen aus. Um genau zu sein, betrachten wir nur jene Zeilen im message file, deren Event Type gleich 1 ist, siehe dazu den Abschnitt 3.1.

Wir nehmen zunächst wieder das Beispiel Apple (AAPL). Hier sehen wir im Zuge unseres Handelstages insgesamt 418.636 Limit Orders, wobei 91,2% der  $\Delta_i$ , also der Großteil, unter 0,2 Sekunden liegt. Im Folgenden sehen wir das Histogramm der Verteilung der  $\Delta_i$  und die angepasste Dichtefunktion. Die Breite eines Balkens entspricht 0,2 Sekunden.



Offensichtlich und ganz nach Intuition handelt es sich hierbei um eine Exponentialverteilung mit Parameter  $\lambda = 17,9$ . Da die Limit Orders laut Modell nach den Sprüngen eines Poisson-Prozesses mit Intensität  $\lambda$  platziert werden, haben wir folgendes Modell:

$$T_{i+1} - T_i \sim \text{Exp}(\lambda(T_{i+1} - T_i)) \quad (3.1)$$

Für die entsprechenden Parameter der anderen Aktien erhalten wir folgende Werte:

Aktie	$\lambda$
AAPL	17,9
ADBE	3,1
ALXN	1,3
AMGN	1,2
BIDU	0,9
BIIB	1,0
COST	1,1
GOOGL	1,7
ISRG	0,6
LRCX	1,6
MSFT	19,0
NFLX	2,6
NTES	0,7
NVDA	6,1
ORLY	0,8
PCLN	0,5
REGN	0,7
TSLA	1,1
ULTA	0,8

Man sieht, es gibt zwei Ausreißer mit der Apple und Microsoft Aktie, welche nach oben herausstechen mit einem  $\lambda$  in Richtung 20. Bei den anderen Titeln bewegt sich  $\lambda$  im Bereich um 1 bzw. bei manchen im Bereich um 3. Nvidia liegt bei 6,1, einem Bereich, wo sonst keine Aktie zu finden ist. Der Mittelwert beträgt 3,3.

# Kapitel 4

## Model fitting

In diesem Kapitel soll es nun um den Mid Preis gehen. Wir wollen die Parameter eines Lévy- und Poisson-Prozesses bestmöglich an unsere Daten anpassen.

Zuerst beschäftigen wir uns mit einer sehr allgemeinen Klasse von Prozessen, und zwar mit Lévy-Prozessen mit Dimension eins. Wir geben die Definition und wichtige Resultate ohne Beweis an. Diese sind in [10] nachzulesen bzw. in der TU Vorlesung "Poisson'sche Punktprozesse" nachzuhören.

### 4.1 Lévy-Prozess

#### 4.1.1 Definition und ausgewählte Eigenschaften

*Definition:* Ein Lévy-Prozess ist ein an eine Filtration  $\mathbb{F} = (\mathcal{F}_t)$  adaptierter Prozess  $X$  mit folgenden Eigenschaften:

- (i) unabhängige Inkremente:  $X_t - X_s$  ist u.a. von  $\mathcal{F}_s$  für alle  $s \leq t$
- (ii) stationäre Inkremente:  $X_t - X_s \sim X_{t-s}$  für alle  $s \leq t$
- (iii) stetig in Wahrscheinlichkeit, also  $\lim_{s \rightarrow t} \mathbb{P}(|X_t - X_s| > \varepsilon) = 0$  für alle  $\varepsilon > 0$

Beispiele für Lévy-Prozesse sind der Poisson-Prozess und die Brownsche Bewegung selber. Ein triviales deterministisches Beispiel wäre der lineare Drift. Lévy-Prozesse gelten auch als die linearen Funktionen der Wahrscheinlichkeitswelt, warum, werden wir später sehen.

Die Charakteristische Funktion ist definiert als  $f_t(u) = \mathbb{E}[e^{iuX_t}]$ . Sie hat folgende Eigenschaften:

- $f_0(u) = 1$ ,
- $f_{t+s}(u) = f_t(u)f_s(u)$  und,
- da  $X$  stetig in Wahrscheinlichkeit ist, existiert eine positive Funktion  $\psi$  mit  $f_t(u) = e^{-t\psi(u)}$ .

*Satz:* Jeder Lévy-Prozess besitzt eine Modifikation, die càdlàg ist und selbst auch wieder ein Lévy-Prozess ist.

Definiere die *Stopzeit  $\sigma$ -Algebra* als  $\mathcal{F}_T = \{F \in \mathcal{F} : F \cap \{T \geq t\} \in \mathcal{F}_t \text{ für alle } t \geq 0\}, T \geq 0$ .

Lévy-Prozesse regenerieren sich nach Stopzeiten:

*Satz:* Sei  $X$  Lévy und  $T$  eine Stopzeit. Dann ist  $Y_t := X_{T+t} - X_T$  auch Lévy, adaptiert an  $\mathcal{H}_t := \mathcal{F}_{T+t}$  für alle  $t \geq 0$ . Weiters ist  $Y$  unabhängig von  $\mathcal{F}_T$  und  $Y$  hat die gleiche Verteilung wie  $X$ .

*Bemerkung:*  $\mathcal{H}_t$  ist wohldefiniert, da zur Stopzeit  $T$  nur eine deterministische Zeit  $t$  dazuaddiert wird.

*Satz:* Sei  $X$  Lévy mit beschränkten Sprüngen. Dann ist  $\mathbb{E}[|X^n|] < \infty$  für  $n \in \mathbb{N}$ .

Nun kommen wir zur Definition des Lévy-Maßes für einen Lévy-Prozess  $X$ . Sei dazu  $\Lambda$  eine Borelmenge mit  $0 \notin \bar{\Lambda}$ . Definiere dann induktiv eine Folge von Stopzeiten:

- $T_\Lambda^1 = \inf\{t > 0 : \Delta X_t \in \Lambda\}$
- $\vdots$
- $T_\Lambda^{n+1} = \inf\{t \geq T_\Lambda^n : \Delta X_t \in \Lambda\}$ .

Aus  $0 \notin \Lambda \Rightarrow T_\Lambda^1 > 0$  und, da die Pfade càdlàg sind, folgt  $\lim_{n \rightarrow \infty} T_\Lambda^n = \infty$ , d.h. es gibt nie eine Explosion.

Definiere nun

$$N_t^\Lambda := \sum_{0 < s \leq t} \mathbb{1}_\Lambda(\Delta X_s) = \sum_{n=1}^{\infty} \mathbb{1}_{T_\Lambda^n \leq t}.$$

Daher ist  $N^\Lambda$  ein Zählprozess. Er zählt, wieviele Sprünge in die Menge  $\Lambda$  fallen. Da  $X$  unabhängige und stationäre Inkremente besitzt, gilt dies auch für  $N^\Lambda$ . Wir folgern, dass  $N^\Lambda$  ein Poisson-Prozess ist.

Definiere nun

$$\nu(\Lambda) := \mathbb{E}[N_1^\Lambda] = \mathbb{E}\left[\sum_{0 < s \leq 1} \mathbb{1}_\Lambda(\Delta X_s)\right].$$

$\nu$  definiert ein  $\sigma$ -endliches Maß auf den Borelmengen  $\Lambda$  mit  $0 \notin \Lambda$  (mit der Konvention  $\nu(\emptyset) = 0$ ).  $\nu$  heißt *Lévy-Maß für  $X$* . Man kann zeigen, dass

$$\int_{|x| \leq 1} x^2 \nu(dx) < \infty$$



*Bemerkung:*  $\nu$  ist kein Wahrscheinlichkeitsmaß!

*Beispiel:* Sei  $X$  ein Compound-Poisson-Prozess. Die Intensität vom Poisson-Ankunftsprozess sei  $\lambda > 0$ . Seien weiters  $Y_i \sim F$ ,  $X_t = \sum_{i=1}^{N_t} Y_i$ . Dann ist

$$\nu(\Lambda) = \mathbb{E}\left[\sum_{0 < s \leq 1} \mathbf{1}_\Lambda(\Delta X_i)\right] = 1 \cdot \mathbb{E}[N_t] \mathbb{E}\left[\mathbf{1}_{\underbrace{\{X_s - X_{s-}\}_{Y_i} \in \Lambda}}\right] = \lambda \mathbb{P}[Y_i \in \Lambda]$$

Definiere die *Kumulantenfunktion*  $\kappa$  für  $t > 0$  implizit als

$$\begin{aligned} \kappa : \mathcal{D} := \{z \in \mathbb{C} : \mathbb{E}[e^{\operatorname{Re}(z)X_1}] < \infty\} &\rightarrow \mathbb{C} \\ e^{\kappa(z)t} &= \mathbb{E}[e^{zX_t}] \end{aligned}$$

Die Kumulante hat oftmals eine einfachere Gestalt als die Verteilung von  $X_1$ .

*Definition:* Die *Charakteristische Funktion* ist definiert als  $\psi(u) := \log \mathbb{E}[e^{iuX_1}]$ ,  $u \in \mathbb{R}$ .

Nun kommen wir zu einer wichtigen und viel gefeierten Aussage, nämlich der Lévy-Khinchine-Zerlegung. Diese sagt aus, dass sich ein Lévy-Prozess durch ein Trippel von Parametern eindeutig charakterisieren lässt.

*Theorem:* Sei  $h$  eine lokal beschränkte, messbare Funktion, für die lokal um 0 gilt, dass  $h(x) = 0$ , z.B.  $h(x) = x\mathbf{1}_{|x| \leq 1}$  ("Abschneidefunktion"). Dann existiert ein Trippel  $(\gamma, \sigma^2, \nu)$  sodass

$$\kappa(z) = \gamma z + \frac{1}{2}\sigma^2 z^2 + \int_{-\infty}^{+\infty} (e^{zx} - 1 - zh(x)) \nu(dx).$$

$\gamma = \gamma(h)$  nennt man *Lokation*,  $\sigma^2$  ist die *Diffusionskonstante* und  $\nu$  das sogenannte *Lévy-Maß*. An dieser Formel kann man erkennen, dass die Lévy-Prozesse eine Prozessklasse von großer Vielfalt ist. Ferner sieht man, dass die Verteilung eines  $X_t, t \neq 0$ , schon die Verteilung des ganzen Lévy-Prozesses eindeutig bestimmt.

Zuletzt definieren wir die sogenannten Subordinatoren, welche einem Zeitwechsel entsprechen.

*Definition:* Ein *Subordinator* ist ein nicht-fallender Lévy-Prozess, dessen Kumulantenfunktion sich stets folgendermaßen schreiben lässt:

$$\begin{aligned} \kappa(z) &= \gamma z + \int_0^\infty (e^{zx} - 1)\nu(dx) \\ \nu(-\infty, 0) &= 0 \end{aligned}$$

## 4.1.2 Modellschätzungen

Die Methoden in diesem Abschnitt stammen stets aus MASUDA [9].

## Beta - Stabile Lévy Prozesse

Wir werden uns in weiterer Folge auf die sogenannten  $\beta$ -stabilen Prozesse konzentrieren. Die charakteristische Funktion von  $\beta$ -stabilen Lévy-Prozessen hat folgende Form:

$$\psi(u) = \begin{cases} -(t^{\frac{1}{\beta}}\sigma)^\beta |u|^\beta \left(1 - i\rho \operatorname{sign}(u) \tan\left(\frac{\beta\pi}{2}\right) + it\gamma u\right) & \beta \neq 1 \\ -t\sigma |u| \left(1 + i\frac{2\rho}{\pi} \operatorname{sign}(u) \log |u|\right) + it\gamma u & \beta = 1 \end{cases}$$

mit dem Stabilitätsindex  $\beta \in (0, 2]$ , der Skalierung  $\sigma > 0$ , dem Grad der Schiefe  $\rho \in [-1, 1]$  sowie dem deterministischen Trend  $\gamma \in \mathbb{R}$ . Wir schreiben dann  $\mathcal{L}(X_t) = S_\beta(t^{\frac{1}{\beta}}\sigma, \rho, t\gamma)$  bzw. als Abkürzung  $S_\beta(\sigma) := S_\beta(\sigma, 0, 0)$ . Wir beschränken uns weiters auf den nicht-Gauss'schen Fall  $\beta \in (0, 2)$ , da hier das Lévy-Maß bekannt und von einfacher Struktur ist.

Wir sagen, dass ein stochastischer Prozess  $Y$  *selbstähnlich* bzw. dass er die *Skalierungseigenschaft* hat, falls  $\exists a, H > 0$  mit  $(Y_t) = (a^{-H}Y_{at})$  in Verteilung. Es ist bekannt, dass nur der stabile Lévy-Prozess selbstähnlich ist von allen Lévy-Prozessen [9].

Wir werden nun im Folgenden die Methode von MASUDA [9, Abschnitt 3.3.2] durchführen. Sei dazu  $X = (X_{t_j^n})_{j=1\dots n}$  der beobachtete Verlauf des Mid Preises. Wenn notwendig, wird die letzte Beobachtung  $X_{t_n^n}$  entfernt und der Prozess nur bis  $X_{t_{n-1}^n}$  betrachtet (siehe unten), um mit einem ungeraden  $n$  zu arbeiten.

Mit folgenden Größen werden wir im weiteren Verlauf arbeiten:

$$\begin{aligned} n &= 2k + 1, \quad k \in \mathbb{N} \\ \Delta_j^n X &= X_{t_{j+1}^n} - X_{t_j^n} \\ Y_{nj}(\beta, \gamma) &= h_n^{-\frac{1}{\beta}} (\Delta_j^n X - \gamma h_n), \quad j = 1 \dots n \\ h_n &= \max_{j=1\dots n} \Delta t_j^n := \max_{j=1\dots n} t_{j+1}^n - t_j^n \end{aligned}$$

$Y_{nj}(\beta, \gamma)$  ist nach MASUDA i.i.d.  $S_\beta(\sigma)$  verteilt. Wir betrachten die geordnete Statistik

$$Y_{n1}(\beta, \gamma) < \dots < Y_{nn}(\beta, \gamma),$$

und sei  $m_k$  der Stichprobenmedian von  $(\Delta_j^n X)_{j=1}^n$  und definiere

$$\hat{\gamma}_n := \frac{m_k}{h_n}.$$

MASUDA zeigt, dass diese Statistik ein bezüglich Konvergenz optimaler Schätzer ist sowie ein asymptotisch normalverteilter Schätzer vom tatsächlichen  $\gamma$  ist.

Wir untersuchen zwei Fälle, wo wir für die Anwendung der Konvergenztheoreme von MASUDA messbare Funktionen  $g_l$ , die um den Ursprung symmetrisch sind, verwenden. Wir wählen dabei

1.  $g_l(y) = (\log |y|)^l$ ,  $l = 1, 2$

$$2. g_l(y) = |y|^{p_l}, l = 1, 2, p \in (0, \frac{\beta}{6})$$

Bei 1. erhält MASUDA für die Schätzer  $\hat{\beta}$  und  $\hat{\gamma}$  auf folgende Ausdrücke, wobei  $x_{n_1} < \dots < x_{n_n}$  für die geordneten  $(\Delta_j^n X)$  stehen und  $\mathcal{C} = 0,5772$  für die Euler Konstante:

$$\begin{aligned} \bullet \hat{\beta}_{\log, n} &= \left( \frac{6}{2k\pi^2} \sum_{j \leq n, j \neq k+1} \left( \log |x_{nj} - \hat{\gamma}_n h_n| - \frac{1}{2k} \sum_{j \leq n, j \neq k+1} \log |x_{nj} - \hat{\gamma}_n h_n| \right)^2 - \frac{1}{2} \right)^{-\frac{1}{2}} \\ \bullet \hat{\sigma}_{\log, n} &= \exp \left( \frac{1}{\hat{\beta}_{\log, n}} \log \left( \frac{1}{h_n} \right) + \frac{1}{2k} \sum_{j \leq n, j \neq k+1} \log |x_{nj} - \hat{\gamma}_n h_n| - \mathcal{C} \left( \frac{1}{\hat{\beta}_{\log, n}} - 1 \right) \right) \end{aligned}$$

Bei der zweiten erhalten wir einen impliziten Ausdruck, welchen wir mit einem Standard-Nullstellenverfahren (z.B. mit dem Newton-Verfahren) lösen:

$$\frac{H_{1n}^2}{H_{2n}} = \frac{C(\hat{\beta}_{p, n}, p)^2}{C(\hat{\beta}_{p, n}, 2p)}, \quad \hat{\sigma}_{p, n} = \left( \frac{h_n^{-p/\hat{\beta}_{p, n}} H_{1n}}{C(\hat{\beta}_{p, n}, p)} \right)^{\frac{1}{p}},$$

wobei

$$C(\beta, p) := \frac{2^p \Gamma((p+1)/2) \Gamma(1-p/\beta)}{\sqrt{\pi} \Gamma(1-p/2)}$$

und

$$H_{ln} := \frac{1}{n} \sum_{j=1}^n |\Delta_j^n X - \hat{\gamma}_n h_n|^{p_l}, l = 1, 2$$

### Schiefes Lévy-Maß

Hierbei greifen wir auf eine weitere Parametrisierung einer strikt positiven  $\beta$ -stabilen Verteilung zurück. Schreibe für  $\beta \neq 1$ :

$$\mathcal{L}(S) = S'_\beta(\mathbf{p}, \sigma)$$

falls die charakteristische Funktion folgende Gestalt hat:

$$\phi_S(u) = \exp \left( -\sigma |u|^\beta \left( 1 - i \operatorname{sgn}(u) \tan(\beta\pi(\mathbf{p} - 1/2)) \right) \right), u \in \mathbb{R},$$

wobei  $\mathbf{p} := \mathbb{P}(S > 0)$ , welches sich für unsere Zwecke im Bereich  $[1 - 1/\beta, 1/\beta]$  bewegt.

Für unser Modell in diesem Abschnitt setzen wir einen positiven und  $\beta$ -stabilen Prozess  $Z$  mit  $Z_t \sim S'_\beta(\mathbf{p}, t)$ ,  $t \in [0, 1]$ , an und definieren  $X_t := \sigma Z_t$  mit einem  $\sigma > 0$ . Unser Ziel ist es wieder, den Parameter  $\theta = (\mathbf{p}, \beta, \sigma)$  anhand eines Pfades  $(X_{t_j^n})$  zu schätzen. Der Schätzer wird asymptotisch normalverteilt mit Konvergenzrate  $\sqrt{n}$  sein.

Hierbei wird zuerst ein Schätzer für  $(\mathbf{p}, \beta)$  angegeben und anhand dessen ein Schätzer für  $\sigma$  bestimmt.

- $\hat{\mathbf{p}} = \frac{1}{2}(1 + \frac{1}{n} \sum_{j=1}^n \text{sgn}(\Delta_j^n X))$
- $\hat{\beta}$  ist die Nullstelle der auf  $(1, 2)$  strikt monoton wachsenden Funktion

$$g(\beta) = \frac{(\Gamma(1 - \frac{1}{4\beta}))^2}{\Gamma(1 - \frac{1}{2\beta})}$$

- Als Schätzer für  $\sigma$  dient

$$\hat{\sigma} = \left( \frac{n^{\frac{1}{2\beta}-1}}{\hat{\mu}} \sum_{j=1}^n \sqrt{|\Delta_j^n X|} \right)^2, \text{ wobei}$$

$$\hat{\mu} = \frac{\Gamma(1 - \frac{1}{2\beta})}{\Gamma(\frac{1}{2})} \frac{\cos(\frac{\xi}{2\beta})}{\cos(\frac{\pi}{4}) |\cos(\xi)|^{\frac{\beta}{2}}} \text{ und}$$

$$\xi = \beta\pi(\mathbf{p} - \frac{1}{2})$$

## Ergebnisse

Bei der praktischen Anwendung der Methoden ergeben sich natürlich einige Probleme. Großteils sind es undefiniertheiten von reellen Größen, da Wurzeln aus negativen Zahlen gezogen werden sollen oder durch 0 geteilt werden soll. Im speziellen wurden folgende Anpassungen vorgenommen:

Bei der Methode mit  $g_l(y) = (\log |y|)^l$  werden in den Summen nur diejenigen  $\log |x_{nj} - \hat{\gamma}_n h_n|$  aufsummiert, welche  $> 0$  sind. Ebenso wird  $\hat{\beta}_{\log, n} = 0$  gesetzt, falls der Ausdruck in den großen Klammern nicht positiv ist. Es folgt, dass in diesem Falle auch  $\hat{\sigma}_{\log, n}$  auf 0 gesetzt wird. Diejenigen Aktien, bei denen dies der Fall war, sind in der Tabelle als 0-er Einträge ersichtlich. Falls  $n$  gerade ist, wird wie oben bereits erwähnt der letzte Eintrag aus den Daten entfernt.

Bei der nächsten Methode mit  $g_l(y) = |y|^{pl}$  wird  $p$  auf Empfehlung von MASUDA [9] auf  $p = 0,05$  gesetzt. Zur numerischen Ermittlung der Nullstellen wurde das R-Paket `nleqslv` verwendet, welches eine robuste und leicht zu implementierende numerische Berechnung garantiert. Für den genauen Code sei auf den Anhang verwiesen.

Die folgende Tabelle zeigt die Ergebnisse der angewandten Methoden auf die vorhandenen Daten.

Aktie	$\beta_{\log,n}$	$\sigma_{\log,n}$	$\beta_{p,n}$	$\sigma_{p,n}$	$\hat{\mathbf{p}}$	$\hat{\beta}$	$\hat{\sigma}$
AAPL	0	0	0.1	0	0.5	-0.1	0
ADBE	0	0	0.1	0	0.5	-0.14	0
ALXN	30376	0.34	0.11	0	0.5	0.567	4778
AMGN	25569	0.08	0.11	0	0.5	0.58	3585
AMZN	0	0	0.11	0	0.49	0.58	17937
BIDU	46753	0.04	-0.03	1.21E+35	0.51	0.59	3298
BIIB	18660	0.17	0.11	0	0.5	0.59	3767
COST	0	0	0.11	0	0.5	-0.16	0
GOOGL	0	0	0.11	0	0.49	0.58	19013
ISRG	44562	0.01	-0.03	2.19E+50	0.52	0.586	2425
LRCX	0	0	0.1	0	0.5	-0.15	0
MSFT	0	0	-0.01	230000	0.5	-0.09	0
NFLX	0	0	0.11	0	0.5	-0.14	0
NTES	25569	0.06	-0.03	6.84E+41	0.5	0.6	3500
NVDA	0	0	-0.03	1.74E+25	0.5	-0.15	0
ORLY	0	0	0.11	0	0.5	0.57	1879
PCLN	16834	0.12	-0.04	7.84E+49	0.5	0.63	5341
REGN	27760	0.7	-0.03	1.99E+44	0.51	0.61	3283
TSLA	43831	0.02	-0.04	2.24E+36	0.49	0.59	6884
ULTA	43042	0.24	0.11	0	0.5	0.57	1842

Man sieht, dass die Ansätze der ersten beiden Methoden bei rund der Hälfte der untersuchten Aktientitel nicht gegriffen hat, d.h. hier war die Anwendung der Methode aufgrund von nicht zulässigen mathematischen Operationen nicht möglich. Auch unsinnig große Ausdrücke bei  $\sigma_{p,n}$  kommen vor.

Die Ähnlichkeit bei manchen Schätzern, z.B. bei  $\beta_{p,n}$  oder  $\hat{\mathbf{p}}$  rührt sicherlich daher, dass bei den Methoden die  $\Delta_j^n X$  untersucht werden, welche wiederum unter den unterschiedlichen Titeln doch recht ähnlich aussehen (normalverteilt um 0).

## 4.2 Poisson-Prozess

### 4.2.1 Theorie

Die hier vorgestellte kurze Einführung in Poisson-Prozesse stammt aus der TU Vorlesung "Poisson'sche Punktprozesse".

Sei  $(\Omega, \mathcal{F}, \mathbb{P})$  ein filtrierter Wahrscheinlichkeitsraum, wobei  $\mathcal{F}$  die üblichen Bedingungen erfülle (vollständig, rechtsstetig). Wir betrachten eine streng monoton wachsende Folge von Zufallszeiten:

$$T_0 = 0, \quad (T_n)_{n \geq 1} \quad (T_n \geq 0 \quad \forall n \in \mathbb{N})$$

Seien weiters  $\mathbb{1}_{t \geq T}(\omega) = 1$ , falls  $t \geq T(\omega)$  und 0 andernfalls. Es folgt, dass dies eine rechts- aber nicht linksstetige Funktion in  $t$  ist.

*Definition* Der Prozess  $N = N(t, \omega) := (N_t)_{0 \leq t \leq \infty}$ , wobei

$$N_t := \sum_{n \geq 1} \mathbf{1}_{t \geq T_n},$$

heißt *Zählprozess assoziiert mit  $T_n$* .  $N_t$  zählt die Anzahl der Ereignisse, die bis  $t$  eingetreten sind. Definiere die *Explosionszeit* von  $N$  durch  $T := \sup_n T_n$ . Falls  $T = \infty$ , dann hat  $N$  keine Explosion. Weiters ist  $N_t - N_s = \sum_{n \geq 1} \mathbf{1}_{s < T_n \leq t}$  die Anzahl der Ereignisse, die zwischen  $s$  und  $t$  passieren.

Sei  $X$  ein an  $\mathcal{F}$  adaptierter Prozess. Falls  $X$  ein gewisses Level  $u$  trifft, wissen wir, ob dies zum ersten Mal oder nicht zum ersten Mal passiert. Wir können uns aber nicht sicher sein, ob dies auch zum letzten Mal der Fall war. Also sind Eintrittszeiten Stoppzeiten, aber Austrittszeiten nicht.

*Proposition* Ein Zählprozess  $N$  ist adaptiert genau dann, wenn die  $T_n$  Stoppzeiten sind.

*Beweis* "  $\Rightarrow$  "  $\{T_n \leq t\} = \{N_t \geq n\} \in \mathcal{F}_t \quad \forall t$

"  $\Leftarrow$  "  $\{N_t = n\} = \{T_n \leq t < T_{n+1}\} = \underbrace{\{T_n \leq t\}}_{\in \mathcal{F}_t} \cap \underbrace{\{t < T_{n+1}\}}_{\mathcal{F}_t} \in \mathcal{F}_t \quad \blacksquare$

*Definition* Ein Zählprozess heißt *Poisson-Prozess*, wenn er adaptiert ist sowie

- (i)  $\forall s \leq t : N_t - N_s$  ist unabhängig von  $\mathcal{F}_s$  ("unabhängige Inkremente")
- (ii)  $\forall s \leq t, u \leq v$  sodass  $t - s = v - u : N_t - N_s \sim N_v - N_u$  ("Stationarität")

$\Rightarrow N_v - N_u = N_{v-u} - N_0 = N_{v-u}$ , jeweils in Verteilung.

*Theorem* Sei  $N$  ein Poisson-Prozess. Dann existiert ein  $\lambda > 0$  sodass  $N_t \sim P(\lambda t)$ . Weiters ist  $N$  stetig in Wahrscheinlichkeit und hat keine Explosion.  $\lambda$  heißt die *Intensität* des Poisson-Prozesses und es folgt als Korollar, dass

$$\mathbb{E}[N_t] = \lambda t, \quad \text{Var}[N_t] = \lambda t$$

Weiters kann man aus der Unabhängigkeit der Inkremente und der Stationarität leicht folgern, dass  $N_t - \lambda t$  und  $(N_t - \lambda t)^2 - \lambda t$  Martingale sind.

## 4.2.2 Berechnung

Zur konkreten Berechnung wurde das R-Paket `poisson` verwendet, insbesondere die Funktion `hpp.mle`. Diese bestimmt anhand von vorgegeben Event-Zeiten die Intensität des Poisson-Prozesses mittels Maximum-Likelihood Ansatz. Für Details sei wieder auf den Anhang verwiesen.

Es ergab sich folgendes Resultat:

Aktie	$\hat{\lambda}$
AAPL	418636
ADBE	72660
ALXN	29959
AMGN	29185
AMZN	32912
BIDU	20298
BIIB	22453
COST	26687
GOOGL	39214
ISRG	14896
LRCX	38079
MSFT	445461
NFLX	61802
NTES	15540
NVDA	143273
ORLY	19357
PCLN	11592
REGN	16729
TSLA	26788
ULTA	18387

Man sieht, der Schätzer für die Dichte bewegt sich durchgehend im fünf- oder sechsstelligen Bereich. Dies rührt wahrscheinlich daher, dass wir in unserem betrachteten Zeitintervall sehr viele Sprünge erleben (im Millisekundenbereich) und dies dementsprechend in der Intensität Ausdruck findet.

# Kapitel 5

## Fazit

Das Ziel dieser Arbeit war es, Theorie und Praxis von der mathematischen Modellierung eines Limit Order Buches zu verbinden. Bereits entwickelte, mit unterschiedlichem Ansatz an die Arbeit herangehende Modelle wurden vorgestellt sowie praktische Methoden zur parametrischen Schätzung angewandt, nachdem zuvor ebenfalls erläutert wurde, wie diese aufgebaut sind.

Die graphische Visualisierung des LOB von diversen Aktientitel an zwei Zeitpunkten des Handelstages waren ebenfalls Bestandteil und haben gezeigt, wie unterschiedlich das LOB von an der selben Börse gehandelten Wertpapieren sein kann. Sicherlich könnte man aus diesen Daten geeignete Handelsstrategien ableiten, was eine sehr spannende Aufgabe wäre

Die erhaltene Datenstruktur eignet sich optimal für viele Arten von Analysen und zusammen mit R ergibt sich hier ein mächtiges Tool für den weiteren Erkenntnisgewinn.

Ich hoffe, mit dieser Arbeit einen kleinen Teil zur Forschung in diesem Bereich beigetragen zu haben.



# Kapitel 6

## Anhang

Im "Grundcode" werden die Daten aus den gelieferten .csv files in R eingespielt und anschließend noch etwas bearbeitet und zugleich erklärt. Man kann zum Beispiel die verwendete Zeitspanne einstellen, aus der die Daten stammen sollen und es werden diverse Prüfungen auf allfällige Handelsunterbrechnungen vorgenommen. Der Grundcode sieht folgendermaßen aus:

```
library(graphics)

#TICKER
ticker <- "AMZN"

# DATE for which data is downloaded, the file name you downloaded
# contains this string, say if you downloaded from 1st july 2009 ,
# type here 2009-07-01
demodate = "2012-06-21"
starttime <- 34200000
endtime <- 57600000

# Levels
nlevels      = 10;

# Name of Orderbook
ORDERBOOK <- paste(paste(ticker , demodate ,starttime,
                        endtime,"orderbook" ,
                        nlevels ,sep = "_"),"csv",sep = ".")

# Name of MSGbook
MSGBOOK <- paste(paste(ticker , demodate ,starttime,endtime,
                        "message" ,nlevels ,sep = "_"),"csv",sep = ".")

# Load data (message file)
dataM <- read.csv ( MSGBOOK )
```

```

# Name the columns
columns <- c ( "Time" , "Type" , "OrderID" , "Size" ,
              "Price" , "TradeDirection" )
colnames ( dataM ) <- columns

# Trading hours (start & end)
startTrad  = 9.5*60*60;      # 9:30:00.000 in ms after midnight
endTrad    = 16*60*60;      # 16:00:00.000 in ms after midnight

# dataM_part is data frame containing messages during the
# continue trading period
dataM_part = dataM[dataM$Time>=startTrad & dataM$Time<=endTrad,]

# Check for trading halts
# -----

tradehaltIdx = which(dataM[,2] == 7 & dataM[,5] == -1 );
tradequoteIdx = which(dataM[,2] == 7 & dataM[,5] == 0 );
traderesumeIdx = which(dataM[,2] == 7 & dataM[,5] == 1 );

if(length(tradehaltIdx)==0 & length(tradequoteIdx)==0
    & length(traderesumeIdx)==0 )
  print("No trading halts detected.")

if(length(tradehaltIdx) !=0)
  cat("Data contains trading halt! at time stamp(s)",
      dataM[tradehaltIdx,1],"\n" )

if(length(tradequoteIdx) !=0)
  cat(" Data contains quoting message! at time stamp(s)",
      dataM[tradequoteIdx,1], "\n")

if(length(traderesumeIdx) !=0)
  cat(" Data resumes trading! at time stamp(s) ",
      dataM[traderesumeIdx,1],"\n")

# Set Bounds for Intraday Intervals
# -----

```

```

# Define interval length
freq = 5*60; # Interval length in ms 5 minutes

# Number of intervals from 9:30 to 4:00
noint= (endTrad-startTrad)/freq
dataM_part$index = seq(from=1,to=dim(dataM_part)[1])

# Variables for 'for' loop
j= 0
l =0
bound =0
visible_count = 0
hidden_count = 0
visible_size = 0
hidden_size = 0

# Set Bounds for Intraday Intervals
for(j in c(1:noint)) {

  bound[j+1] = startTrad + j * freq
}
bound[1] = startTrad

# Logic to calculate number of visible/hidden
# trades and their volume
for(l in c(1:noint))
{
  visible_count[l] = nrow(dataM_part[dataM_part$Time > bound[l]
                                & dataM_part$Time < bound[l+1]
                                & dataM_part$Type == 4,])
  visible_size[l] = sum(dataM_part[dataM_part$Time > bound[l]
                                & dataM_part$Time < bound[l+1]
                                & dataM_part$Type == 4,4])/100

  hidden_count[l] = nrow(dataM_part[dataM_part$Time > bound[l]
                                & dataM_part$Time < bound[l+1]
                                & dataM_part$Type == 5,])
  hidden_size[l] = sum(dataM_part[dataM_part$Time > bound[l]
                                & dataM_part$Time < bound[l+1]
                                & dataM_part$Type == 5,4])/100

}

# Load data (order book)
dataOB <- read.csv( ORDERBOOK )

```

```

# Note: The file contains more than 250 000 entries.
# It takes a few seconds to load.

columns2 <- c("ASKp1" , "ASKs1" , "BIDp1", "BIDs1")

# naming the columns of data frame
if (nlevels > 1)
{
  for ( i in 2:nlevels )
  {
    columns2 <- c (columns2,paste("ASKp",i,sep=""),
                  paste("ASKs",i,sep=""),paste("BIDp",i,sep=""),
                  paste("BIDs",i,sep=""))
  }
}

colnames ( dataOB ) <- columns2

# Trading hours (start & end) 16:00:00.000 in ms after midnight
timeindex <-dataM$Time>=startTrad & dataM$Time<=endTrad

dataOB_part = dataOB[timeindex,]
# Convert prices into dollars
# Note: LOBSTER stores prices in dollar price times 10000

for ( i in c(seq(from = 1, length=2*nlevels, by = 2)) )
{
  dataOB_part[,i ] = dataOB_part[ ,i]/10000
}

```

Das Anpassen des Arrival-Prozesses aus Kapitel 3.3 wurde anschließend mit folgenden Codezeilen durchgeführt:

```

limitorders=dataM_part$Time[dataM_part$Type==1]
interarrivaltime=limitorders[-c(1)]-limitorders[-c(length(limitorders))]
h=hist(interarrivaltime,freq=FALSE,breaks=40, main=paste("Histogramm von ",ticker))
lines(density(interarrivaltime))
h
fitdistr(interarrivaltime,densfun="exponential")

```

Im Kapitel 4.1 wurde mit folgenden Codes gearbeitet, wo wir zunächst den Mid Preis berechnen sowie diverse andere Größen aus der Theorie definieren:

```

midpreis=(dataOB_part$ASKp1+dataOB_part$BIDp1)*0.5#Midpreis
dmid = midpreis[-c(1)]-midpreis[-c(length(midpreis))]#Differenzenprozess
hn=max(dataM_part$Time[-c(1)]-dataM_part$Time[-c(length(dataM_part$Time))])#maximaler

```

```

n=length(dmid)

if (n%%2==0)
{
  n=length(dmid)-1
  midpreis=midpreis[-c(length(midpreis))]
  dmid=dmid[-c(length(dmid))]
}
mk=median(dmid)
gamman=mk/hn
c=0.5772

```

Die erste Variante mit  $g_l(y) = (\log |y|)^l, l = 1, 2$  wurde folgendermaßen implementiert:

```

s1=sort(dmid)-gamman*hn
s2=log(abs(s1[s1>0]))#>0 wegen log
k=0.5*(n-1)
s3=1/(2*k)*sum(s2[-c(k)])

s4=(s1-s3)^2
s5=6/(2*k*pi)*sum(s4[-c(k)])-1/2
beta_logn=0
if(s5>0) beta_logn=s5^(-1/2)

sigma_logn=0
if(beta_logn>0) sigma_logn=exp(log(hn^(-1))/beta_logn+s3-c*(beta_logn^(-1)-1))

#Ungerades n eventuell wieder rückgängig machen
if (n%%2==0)
{
  n=n+1
  midpreis=(dataOB_part$ASKp1+dataOB_part$BIDp1)*0.5#Midpreis
  dmid = midpreis[-c(1)]-midpreis[-c(length(midpreis))>#Differenzenprozess
}

```

Nun die zweite Variante,  $g_l(y) = |y|^{pl}, l = 1, 2, p \in (0, \frac{\beta}{6})$ :

```

p=0.05
s6=dmid-gamman*hn
s71=abs(s6)^p
s72=s71^2
H1n=sum(s71)/n
H2n=sum(s72)/n

C=function(b,p){
  return((2^p*gamma(0.5*(p+1))*gamma(1-p/b))/(sqrt(pi)*gamma(1-p/2)))
}

```

```

#Nullstellenfunktion
g=function(b)
{
  return(H1n^2/H2n-C(b,p)^2/C(b,2*p))
}

beta_pn=nleqslv(1.1,g)$x
sigma_pn=((hn^(-p/beta_pn)*H1n)/(C(beta_pn,p)))^(1/p)

```

Die letzte Variante (schiefes Lévy-Maß) wurde in dieser Art und Weise umgesetzt:

```

pdach=0.5*(1+sum(sign(dmid))/n)
q=p/2

g2=function(b)
{
  h1a=sum(abs(dmid[-c(length(dmid))])^q*abs(dmid[-c(1)])^q)
  h1b=sum(abs(dmid)^(2*q))
  h1=h1a/h1b

  h2=gamma(1-2*q)*cos(q*pi)/(gamma(1-q)*cos(q*pi*0.5))^2
  h3=cos(q*pi*(pdach-0.5))^2/cos(2*q*pi*(pdach-0.5))
  h4=(gamma(1-1/(4*b)))^2/(gamma(1-1/(2*b)))
  return(h1-h2*h3*h4)
}
beta_k=nleqslv(1.5,g2)$x

xi=beta_k*pi*(pdach-0.5)
mu=(gamma(1-1/(2*beta_k))*cos(xi/(2*beta_k)))/(gamma(0.5)*cos(pi/4)*abs(cos(xi))^(bet
sigma_k=(n^(1/(2*beta_k)-1)*sum(sqrt(abs(dmid)))/mu)^2

```

# Literaturverzeichnis

- [1] Bouchaud, J. P.; Mezard, M. and Potters, M.: *Statistical Properties of Stock Order Books: Empirical Results and Models*. (2002) Science and Finance. Capital Fund Management.
- [2] Bouchaud, J. P. and Potters, M.: *More Statistical Properties of Order Books and Price Impact*. (2002) Working paper series. Available at SSRN: <http://ssrn.com/abstract=50735338>
- [3] Cont, R.; Kukanov, A. and Stoikov, S.: *The Price Impact of Order Book Events*. (2010) Working paper series. Available at SSRN: <http://ssrn.com/abstract=1712822>
- [4] Cont, R.; Stoikov, S. and Talreja, R.: *A Stochastic Model for Order Book Dynamics* (2008) Operations Research, Volume 58, 549-563
- [5] Karpoff, J.; *The relation between price changes and trading volume: A survey* (1987) Journal of Financial and Quantitative Analysis, 22 , p. 109
- [6] Cont, R., Stoikov, S. and Talreja, R.: *A stochastic model for order book dynamics* (2010b.)Operations Research 58, 549-563
- [7] Cont, R. and De Larrard, A.: *Price dynamics in a Markovian limit order market* (2010, revised 2012)
- [8] Eisler, Z.; Kertesz, J.; Lillo, F. and Mantegna, R.: *Diffusive behavior and the modeling of characteristic times in limit order executions* (2008) Available at *arXiv:physics/0701335v3 [physics.soc-ph]*
- [9] Masuda, H.: *Parametric estimation of Levy processes* (2015) 179 -286, MR3364258
- [10] Kallsen, J.: *Levy-Prozesse anschaulich erklärt*
- [11] Riccardi, F.: *Stochastic Models for the Limit Order Book* (2011), M Phil - Thesis, LSE