



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

Effizienter Preis

Ausgeführt am Institut für

Wirtschaftsmathematik

Forschungsgruppe: Finanz- und Versicherungsmathematik
der Technischen Universität Wien

unter der Anleitung von

Univ.Prof. Dipl.-Math. Dr.rer.nat. Thorsten Rheinländer

durch

David Christoph Hinterkörner

Kienmayergasse 19/13

1140 Wien

Wien, am 16. September 2014

Inhaltsverzeichnis

1	Einleitung	1
2	Begriffe und Definitionen	3
2.1	Limit Order Book	3
2.2	Mathematische Grundlagen	5
3	Schätzung des effizienten Preises aus dem Order Flow	8
3.1	Das Modell nach Delattre, Robert, Rosenbaum	8
3.1.1	Hintergrund	8
3.1.2	Die Idee hinter dem Schätzer	9
3.2	Beschreibung des Modells	9
3.3	Eigenschaften von Y_t	11
3.3.1	Markov-Eigenschaft	11
3.3.2	Erneuerungs-Eigenschaft	11
3.4	Schätzung der Flow Response Function und ihrer Inversen	13
3.5	Sätze über das Grenzverhalten der Response Function	14
3.6	Schätzer für den effizienten Preis	15
3.7	Zusammenfassung	16
4	Implementierung des effizienten Preises	17
4.1	Aufbau der Daten	17
4.2	Die Wahl der Folge k_T	19
4.3	Die Wahl von N_T	23
4.4	Implementierung in Matlab	27

Inhaltsverzeichnis

5	Effizienter Preis und Volatilität	29
5.1	Volatilitätsschätzer	29
5.1.1	Realized Variance	30
5.1.2	Zhou Estimator	31
5.1.3	Two-Scaled Realized Volatility	31
5.1.4	Multi-Scale Realized Volatility	32
5.1.5	Realized Kernel Estimator	32
5.1.6	Vergleich der Schätzer	33
5.2	Differenz Efficient Price und Trade Price	35
5.3	Differenz Efficient Price und Best Bid Price	38
5.4	Differenz Efficient Price und Best Ask Price	40
5.5	Differenz der Werte des effizienten Preises	43
5.6	Methode der kleinsten Quadrate	44
6	Zusammenhang zwischen Delta und Volatilität	47
7	Resümee	51
8	Beweise	53
8.1	Beweis von Proposition 3.3	53
8.2	Ein hilfreiches Resultat	54
8.3	Beweis von Satz 3.4 und Satz 3.5	59
8.4	Beweis von Korollar 3.6	60
9	Programmcode	62
	Literaturverzeichnis	71

1 Einleitung

Die Preisbestimmung von komplexen Derivaten ist eine der schwierigsten Aufgaben mit denen sich Marktteilnehmer beschäftigen müssen. Der klassische Ansatz in der Finanzmathematik ist, die Preise von Standardprodukten am Markt zu beobachten und diese für die Bepreisung zu verwenden. Das Problem dabei ist, dass der Preis eines Finanzproduktes nicht eindeutig ist. Das liegt am Abstand (*Spread*) zwischen dem höchsten Preis den die Käufer zum aktuellen Zeitpunkt anbieten (*Best Bid Price*), und dem niedrigsten Preis für den der Verkäufer bereit ist seine Anteile abzugeben (*Best Ask Price*). Doch nicht nur für komplexe Derivate, sondern auch für Handelsstrategien und Algorithmen wird als Eingabe ein eindeutiger Wert benötigt.

Ein simpler Ansatz für die Definition eines möglichen fairen Preises wäre die Verwendung des Mittelwertes aus *Best Bid Price* und *Best Ask Price*. In der Praxis sind die Teilnehmer jedoch meistens nicht an einem fairen Preis interessiert, sondern an dem Wert, der höchstmöglichen Profit verspricht. Diese Größe wird oft als *effizienter Preis* bezeichnet, wobei allerdings keine exakte Definition des Begriffs existiert. Daher beziehen Marktakteure als nächsten logischen Schritt das Volumen, also die Anzahl der *Limit Orders*, in ihre Berechnung mit ein und kalkulieren ihren Preis aus einem gewichteten Mittel. Dabei lautet die goldene Regel „*the price is where the volume is not*“.

Delattre, Robert und Rosenbaum wollen in ihrem Modell keine statische Betrachtung des letzten Zeitpunktes, sondern aus dem gesamten *Order Flow* eine Größe bestimmen. In ihrem Paper *Estimating the Efficient Price from the Order Flow: a Brownian Cox Process Approach* beschreiben sie eine statistische Methode, anhand derer aus dem *Order Flow* ein Preis bestimmt wird, der den meisten Profit verspricht.¹ Die Idee dabei ist, dass die Intensität von Ordereingängen vom Abstand zwischen *effizienten Preis* und dem *Best Bid Price* abhängig ist. Bevor dieses Modell, das den Ausgangspunkt dieser Arbeit darstellt, im Kapitel 3 detailliert vorgestellt wird, werden die wichtigsten verwendeten

¹vgl. [DRR13].

1 Einleitung

Begriffe sowie mathematischen Definitionen erklärt.

Oftmals ist es schwierig ein theoretisches Modell fehlerfrei in der Praxis umzusetzen. Ziel der vorliegenden Diplomarbeit ist es zu analysieren, wie das Modell von Delattre, Robert und Rosenbaum tatsächlich angewendet werden kann und ob, beziehungsweise welche, Annahmen noch zusätzlich getroffen werden müssen. Um diese Analyse vornehmen zu können, verwende ich Daten der *Borsa Istanbul*,² die die Quelle aller folgenden Auswertungen, Abbildungen und Tabellen bildet. Die Umsetzung des Modells, der Aufbau der Daten und die Implementierung des *effizienten Preises* in Matlab werden in Kapitel 4 erklärt.

Im Kapitel 5 soll festgestellt werden, ob es eine Beziehung zwischen dem *effizienten Preis* und der Volatilität von Aktien gibt. Da die Volatilität nicht frei am Markt beobachtbar ist, muss diese geschätzt werden. Daher werden zuerst einige Ansätze zur Schätzung der Volatilität vorgestellt und anschließend versucht eine Formel zu finden, die einen Zusammenhang beschreibt. In dieser Arbeit konzentriere ich mich dabei auf die Suche nach einem einfachen, linearen Zusammenhang der mittels Berechnung von Korrelationskoeffizienten und p-Wert nachgewiesen werden soll. Unterschiedliche Berechnungsmethoden werden entwickelt und es wird versucht einen so eindeutigen Zusammenhang festzustellen, dass aus einem Wert fehlerfrei der Andere berechnet werden kann.

Vor diesem Hintergrund stellt sich die interessante Frage, ob ganz allgemein aus der Struktur des *Limit Order Book* auf die Volatilität geschlossen werden kann. Um das zu ergründen werden im Kapitel 6 die Abstände der platzierten *Limit Orders* zum *Mid Price*³ studiert und versucht, aus der sich ergebenden statistischen Verteilung Informationen zu gewinnen.

Abschließend werden die Aussagen aus Kapitel 3 bewiesen. Um die Berechnungen, die in der vorliegenden Diplomarbeit vorgenommen wurden, uneingeschränkt nachvollziehen zu können, werden im Kapitel 9 die wichtigsten Teile des verwendeten Programmcodes, der ausschließlich mit Matlab Version 7.11.0 (R2010b) erzeugt wurde, bereitgestellt.

²Einzige türkische Börse für den Handel mit Wertpapieren.

³vgl. Kapitel 2.1.

2 Begriffe und Definitionen

2.1 Limit Order Book

Mehr als die Hälfte aller Finanzmärkte verwendet ein *Limit Order Book* um den Handel abzuwickeln.¹ Durch technologischen Fortschritt haben Marktakteure weltweit in Echtzeit Zugriff auf das *Limit Order Book* und damit die Möglichkeit mittels vollständiger Information über den derzeitigen Stand des Finanzprodukts ihre Entscheidungen zu treffen. Anhand der Abbildung 2.1 will ich die für die Arbeit wichtigsten Begriffe erklären.

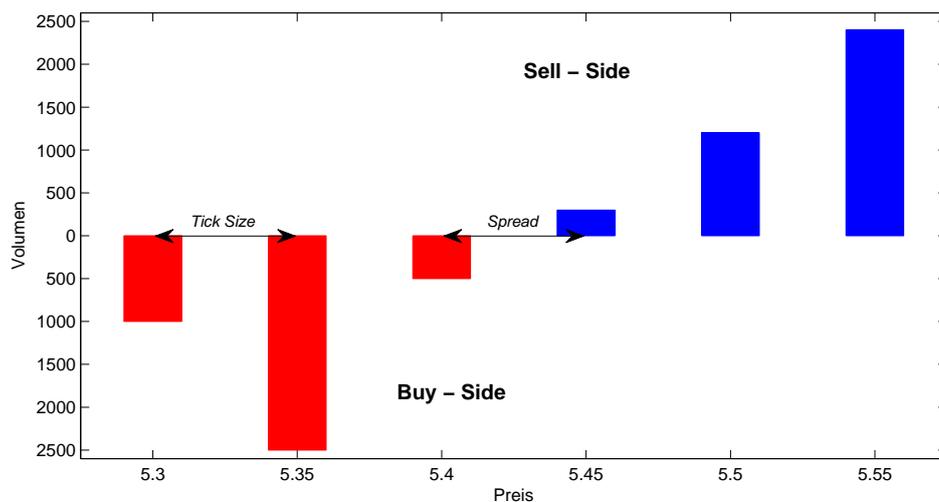


Abbildung 2.1: Schema eines *Limit Order Book*

¹vgl. [GPW⁺13], S. 1.

2 Begriffe und Definitionen

Grundsätzlich gibt es an der Börse entweder *Market Orders* oder *Limit Orders*. *Market Orders* sind Aufträge, die sofort ausgeführt werden sollen. Will ein Börsenteilnehmer Aktien mittels *Market Order* kaufen, so kennt er nicht den genauen Preis, um dem er diese erstehen wird. Je nach Organisation der Börse werden Verkaufsaufträge mit seinem Kaufauftrag ausgeglichen. Im Gegensatz dazu wird bei einer *Limit Order* nur zu diesem oder einem besseren Preis gehandelt. Wir werden Kaufaufträge mit *Buy Limit Orders* und Verkaufsaufträge mit *Sell Limit Orders* bezeichnen.

Der Preis der *Buy Limit Order* mit dem höchsten Wert wird als *Best Bid Price* bezeichnet und der Preis der *Sell Limit Order* mit dem niedrigsten Wert als *Best Ask Price*. Der Abstand zwischen diesen Größen wird *Bid-Ask Spread* oder oft auch nur *Spread* genannt. Der *Mid Price* wird als das arithmetische Mittel aus diesen Preisen definiert. Ähnlich definiere ich den *Weighted Mid Price* als gewichtetes arithmetisches Mittel dieser Preise.

Der *Trade Price* ist der Preis, um dem zuletzt gehandelt wurde. Betrachtet man die Abbildung 2.1 und nimmt an, dass eine *Buy Market Order* über ein Volumen von 100 Stück platziert wird, wäre demnach der *Trade Price* 5.45, da der Käufer um diesen Preis seine Assets erhält. Würde jemand eine *Sell Market Order* im Umfang von 100 Stück platzieren, so würde der Handel um einen Preis von 5.40 stattfinden und der *Trade Price* hätte diesen Wert. Er ist folglich vom letzten gesetzten Auftrag abhängig.

Die *Tick Size* eines *Limit Order Book* ist der kleinste zulässige Preissprung zweier verschiedener Orders. In Abbildung 2.1 beträgt dieser 0.05 und es kann damit beispielsweise kein Auftrag über ein Limit von 5.32 erteilt werden.

2.2 Mathematische Grundlagen

Folgende Begriffe werden für die Beschreibung der Schätzung des *effizienten Preises* in Kapitel 3 benötigt.

Definition 2.1. Ein *stochastischer Prozess* ist eine Familie von Zufallsobjekten.²

Die Aktienkurse an der Börse können zum Beispiel als stochastische Prozesse modelliert werden. Einer der bekanntesten stochastischen Prozesse ist die Brownsche Bewegung.

Definition 2.2. Ein reellwertiger stochastischer Prozess in stetiger Zeit heißt *Brownsche Bewegung*, wenn er

1. fast sicher in Null startet,
2. stationäre und unabhängige Inkremente hat,
3. fast sicher stetige Pfade hat,
4. und die Inkremente eine zentrierte Normalverteilung haben, deren Varianz gleich der Länge des betrachteten Intervalls ist.³

Ein weiterer stochastischer Prozess, der sehr häufig in der Finanz- und Versicherungsmathematik verwendet wird, ist der Poisson-Prozess. Ein Poisson-Prozess mit Intensitätsfunktion μ besagt, dass in einer Zeiteinheit μ Ankünfte gewisser Ereignisse erwartet werden. Eine typische Anwendung ist die Modellierung der Zeitpunkte von Großschäden einer Versicherung.

Definition 2.3. Ein stochastischer Prozess $(N_t)_{t \geq 0}$ heißt *Poisson-Prozess*, wenn er

1. fast sicher in Null startet,
2. unabhängige Inkremente hat,
3. es eine wachsende rechtsstetige Funktion $\mu : \mathbb{R}^+ \rightarrow \mathbb{R}$ mit $\mu(0) = 0$ gibt, sodass $N_t - N_s \sim \text{Poi}(\mu(t) - \mu(s))$ für $0 \leq s < t$,
4. und er càdlàg Pfade hat.⁴

²vgl. [Hub12b], S. 10.

³vgl. [Hub12b], S. 21.

⁴vgl. [Hub12a], S. 26f.

2 Begriffe und Definitionen

Definition 2.4. Eine rechtsstetige Funktion, deren linksseitigen Grenzwerte existieren, wird *càdlàg* genannt. Der Raum der càdlàg Funktionen von $[0, b]$ nach \mathbb{R} wird mit $D[0, b]$ bezeichnet.⁵

Definition 2.5. Gilt für einen Poisson-Prozess, dass $\mu(t) = t$, so heißt dieser *Standard Poisson-Prozess*.⁴

Definition 2.6. Ein Poisson-Prozess mit stochastischer Intensitätsfunktion heißt *doppelt stochastischer Poisson-Prozess* oder auch *Cox-Prozess*.⁶

Falls bei einem stochastischen Prozess die Wahrscheinlichkeit für den Übergang von einem Zustand zum Nächsten nicht von der Vergangenheit, sondern nur vom aktuellen Zustand abhängt, so wird dieser Markov-Prozess genannt. Die Brownsche Bewegung erfüllt diese Eigenschaft und ist somit auch ein Markov-Prozess.

Definition 2.7. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und \mathcal{S} eine endliche oder abzählbare Menge. Ein stochastischer Prozess $(X_t)_{t \geq 0}$ heißt *Markov-Prozess* im Zustandsraum \mathcal{S} , wenn für alle $s \leq t$ und Zustände $i, j \in \mathcal{S}$ mit $\mathbb{P}[X_s = i, X_u = x_u, 0 \leq u < s] > 0$ die *Markov-Eigenschaft* gilt:⁷

$$\mathbb{P}[X_t = j | X_s = i, X_u = x_u, 0 \leq u < s] = \mathbb{P}[X_t = j | X_s = i].$$

Definition 2.8. Ein Markov-Prozess $(X_t)_{t \geq 0}$ im Zustandsraum S heißt *Harris Recurrent*, wenn es $A \subseteq S, \lambda > 0, m > 1$ und ein Wahrscheinlichkeitsmaß ρ auf S gibt, sodass

1. $\mathbb{P}_x(T_A < \infty) = 1, x \in S$, wobei $T_A = \inf\{n \geq 0 : X_n \in A\}$,
2. $\mathbb{P}_x(X_m \in \cdot) \geq \lambda \rho(\cdot), x \in A$.⁸

Die Lokalzeit der Brownschen Bewegung stellt man sich am besten mit Hilfe eines Aktienkurses vor. Angenommen es sei $x = 26.50 \text{ €}$, dann ist $L_x(t)$ die Anzahl der Veränderungen des Aktienkurses auf genau 26.50 € bis zum Zeitpunkt t .

Definition 2.9. Die Lokalzeit der Brownschen Bewegung ist definiert als

$$L_x(t) = \lim_{\epsilon \searrow 0} \frac{1}{2\epsilon} \lambda\{0 \leq s \leq t | |W_s - x| \leq \epsilon\},$$

wobei λ für das Lebesgue-Maß steht.⁹

⁵vgl. [Hub12b], S. 23.

⁶vgl. [Lan98], S. 101.

⁷vgl. [BZ05], S. 88.

⁸vgl. [Gly13], S. 1.

⁹vgl. [Ger13], S. 55.

2 Begriffe und Definitionen

Definition 2.10. Die Skorohod Topologie J_1 auf $D[0, b]$ ist definiert durch

$$d(f, g) = \inf_{\lambda \in \Lambda} \max\{\|f \circ \lambda - g\|_\infty, \|\lambda - e\|_\infty\},$$

wobei Λ die Menge aller ordnungserhaltenden Homöomorphismen und $e \in \Lambda$ die Identität ist.¹⁰

Satz 2.11. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine reellwertige Zufallsvariable und $g : \mathbb{R} \rightarrow [0, \infty]$ messbar und nicht fallend. Dann gilt die *Markov-Ungleichung*¹¹

$$\mathbb{E}[g(X)] \geq \mathbb{E}[g(X); X \geq c] \geq g(c)\mathbb{P}[X \geq c].$$

¹⁰vgl. [Gok08], S. 469.

¹¹vgl. [Wil10], S. 59f.

3 Schätzung des effizienten Preises aus dem Order Flow

3.1 Das Modell nach Delattre, Robert, Rosenbaum

In diesem Kapitel wird das Modell wie in [DRR13] vorgestellt.

3.1.1 Hintergrund

Im Hochfrequenzhandel ist der Begriff des Preises eines Assets nicht eindeutig. Beispielsweise gibt es den *Trade Price*, den *Best Bid Price*, den *Best Ask Price*, den *Mid Price* oder den *Weighted Mid Price*. In der Praxis wissen Teilnehmer nicht, welchen Preis sie für ihre Strategie verwenden sollen. In diesem Kapitel wird der Begriff des *effizienten Preises* definiert, der den zutreffendsten Preis in der Praxis darstellen soll. Des Weiteren wird eine statistische Methodik aufgestellt, um diesen aus dem *Order Flow* zu schätzen.

Nach dem klassischen, finanzmathematischen Ansatz werden die Preise von Standardprodukten (z.B. *Future*, *Stock*) am Markt beobachtet. Diese werden dann benutzt, um komplexe Derivate zu bepreisen. Normalerweise rebalancieren Options-Händler ihr Portfolio ein- bzw. mehrmals am Tag. Die Frage nach dem „richtigen“ Preis stellt sich also einige Male täglich. Im Ultra-Frequenz-Handel ist es bereits schwierig einfache Produkte zu bepreisen. Wenn man Zugriff zum *Order Book* hat, hat man die oben erwähnte Auswahl aus verschiedensten Preisen. Algorithmen und Handelstrategien brauchen jedoch oft als Input einen eindeutigen Preis. Die Problematik für die Marktteilnehmer ist es, den richtigen Preis zu wählen. Die Wahl spielt allerdings eine entscheidende Rolle, da verschiedene Preise zu signifikant unterschiedlichen Strategien führen können. Das ist vor allem der Fall, wenn die *Tick Size* groß ist.

Des Weiteren sollte nicht unerwähnt bleiben, dass die Teilnehmer in der Praxis nicht nach einem fairen Preis suchen, sondern nach dem Preis, der den meisten Profit ver-

spricht. Dieser wird oft als der *effiziente Preis* bezeichnet. Nachfolgend wird eine statistische Vorgehensweise vorgestellt, um diesen zu schätzen.

3.1.2 Die Idee hinter dem Schätzer

Im Folgenden wird ausschließlich auf den Fall von *Large Tick* Assets eingegangen. Das sind Assets bei denen der *Bid-Ask Spread* nahe am Wert eines *Ticks* liegt. Unser Ziel ist es, auf den *effizienten Preis* solcher Assets zu schließen. Klarerweise wird dieser im *Bid-Ask Spread* liegen.

Üblicherweise wird die Differenz zwischen dem Volumen am *Best Bid* und am *Best Ask* Level betrachtet. Dabei ist die goldene Regel für die Marktteilnehmer: „*the price is where the volume is not*“. Durch die Betrachtung des *Order Flows* bekommen wir eine dynamische Version dieses Leitsatzes.

Es wird angenommen, dass die Intensität von Auftragseingängen von *Limit Orders* am *Best Bid* oder am *Best Ask* Level vom Abstand zwischen diesem Level und dem *effizienten Preis* abhängig ist. Je größer der Abstand, desto höher die Intensität. Dies kann als steigende deterministische Funktion der Distanz angenommen werden und wird als *Order Flow Response Function* bezeichnet. In diesem Ansatz ist es entscheidend, dass die Funktion in parameterfreier Weise geschätzt wird. Danach wird aus diesem Schätzer der Schätzer für den *effizienten Preis* gewonnen.

3.2 Beschreibung des Modells

Es wird angenommen, dass die *Tick Size* gleich eins ist. Das bedeutet ein Asset kann nur Werte in \mathbb{N}_0 annehmen.

Der *effiziente Preis* ist gegeben durch

$$P_t = P_0 + \sigma W_t, \quad t \in [0, T],$$

wobei $(W_t)_{t \geq 0}$ eine Brownsche Bewegung auf einen filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ und P_0 eine \mathcal{F} -messbare Zufallsvariable, unabhängig von $(W_t)_{t \geq 0}$, ist. P_0 ist weiters stetig gleichverteilt auf $[p_0, p_0 + 1)$, mit $p_0 \in \mathbb{N}$.

Sei Y_t der gebrochene Anteil von P_t , geschrieben als $\{P_t\}$, also

$$Y_t = \{P_t\} = P_t - \lfloor P_t \rfloor.$$

3 Schätzung des effizienten Preises aus dem Order Flow

Ohne Beschränkung der Allgemeinheit wird ab jetzt der *Limit Order Flow* am *Best Bid Level* betrachtet. Es wird angenommen, dass, wenn eine *Limit Order* am *Best Bid Level* zum Zeitpunkt t platziert wird, der Preis gerade $B_t = \lfloor P_t \rfloor$ beträgt. Somit ist der *effiziente Preis* zum Zeitpunkt t

$$P_t = B_t + Y_t. \quad (3.1)$$

N_t bezeichnet die Gesamtanzahl an *Limit Orders*, die im Zeitraum $[0, t]$ platziert wurden. Um der Regel „*the price is where the volume is not*“ zu folgen wird davon ausgegangen, dass $(N_t)_{t \geq 0}$ ein Cox-Prozess mit der Intensität

$$\mu h(Y_t)$$

zum Zeitpunkt t ist. Dabei ist μ eine positive Konstante und die *Response Function* eine Funktion $h: [0, 1) \rightarrow \mathbb{R}_+$. Die nachfolgenden Annahmen werden von h erfüllt.

Annahme 3.1. *Response Function.* Für die Funktion h gilt $h \in C^1$ auf $[0, 1)$. Für die Ableitung h' existieren positive Konstanten c_1 und c_2 , sodass für alle x in $[0, 1)$ gilt

$$c_1 \leq h'(x) \leq c_2.$$

Weiters gilt die Identifizierungs-Bedingung:

$$\int_0^1 h(x) dx = 1.$$

Annahme 3.1 impliziert die Beschränktheit von h auf $[0, 1)$. Weiters ist N_t damit sicher steigend bezüglich Y_t . Die *Response Function* drückt den Einfluss des *effizienten Preises* auf den *Order Flow* aus. Falls h eine konstante Funktion ist, so ergibt sich ein Standard Poisson Prozess. Oft ist es technisch praktisch diesen Fall anzunehmen, obwohl er nicht sehr realistisch ist. Derzeit ist Y_t der einzige Einfluss auf die Fluktuation der Intensität. Das Modell soll allerdings auch saisonale Gegebenheiten und den möglichen Einfluss von exogenen Variablen darstellen.

Wir studieren den Punktprozess (N_t) auf $[0, T]$. Um asymptotisches Verhalten zu beobachten, wird $T \rightarrow \infty$ betrachtet. Es ist außerdem notwendig, dass μ von T abhängt, $\mu = \mu_T$. Das führt zur zweiten Annahme:

Annahme 3.2. *Asymptotisches Verhalten.* Für ein $\varepsilon > 0$ gilt

$$\frac{T^{\frac{5}{2} + \varepsilon}}{\mu_T} \xrightarrow{T \rightarrow \infty} 0.$$

Anstatt $T \rightarrow \infty$, kann T fixiert werden und die *Tick Size* läuft gegen 0.

3.3 Eigenschaften von Y_t

Die Intensität des *Order Flows* ist $\mu_T h(Y_t)$. Dieser Prozess $(Y_t)_{t \geq 0}$ hat einige hilfreiche Eigenschaften.

3.3.1 Markov-Eigenschaft

Zuerst sollte man sich in Erinnerung rufen, dass, wenn eine Zufallsvariable U stetig gleichverteilt auf $[0, 1]$ und X eine reelle, von U unabhängige, Zufallsvariable ist, auch $\{U + X\}$ stetig gleichverteilt auf $[X, X + 1]$ ist. Da P_0 stetig gleichverteilt auf $[p_0, p_0 + 1)$ ist, wird klar, dass (Y_t) ein stationärer Markov Prozess ist. Aus Eigenschaften der Brownschen Bewegung folgt, dass (Y_t) *Harris Recurrent* ist und somit für eine beschränkte positive messbare Funktion f fast sicher

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(Y_s) ds = \int_0^1 f(s) ds$$

gilt.¹

3.3.2 Erneuerungs-Eigenschaft

Der Prozess (Y_t) erfüllt auch die Erneuerungs-Eigenschaft. Definiere eine Folge von Stoppzeiten $(\nu_n)_{n \in \mathbb{N}}$ wie folgt: $\nu_0 = 0$, $\nu_1 = \inf\{t > 0 : P_t \in \mathbb{N}\}$ und für $n \geq 2$:

$$\nu_n = \inf\{t > \nu_{n-1} : P_t = P_{\nu_{n-1}} \pm 1\} = \inf\{t > \nu_{n-1} : W_t = W_{\nu_{n-1}} \pm \frac{1}{\sigma}\}. \quad (3.2)$$

Die Zyklen $(Y_{t+\nu_n})_{0 \leq t < \nu_{n+1} - \nu_n}$ sind unabhängig und identisch verteilt für $n \geq 1$. Demnach ist (Y_t) ein Erneuerungsprozess. Man beachte, dass $\nu_2 - \nu_1$ dem selben Gesetz folgt, wie

$$\tau_1 = \inf\{t \geq 0 : |W_t| = \frac{1}{\sigma}\}.$$

Die Laplace-Transformierte von τ_1 ist gegeben durch

$$\mathbb{E}[e^{-\gamma \tau_1}] = \frac{1}{\cosh(\frac{\sqrt{2\gamma}}{\sigma})}, \quad \gamma \geq 0.$$

Dadurch ergeben sich die Momente $\mathbb{E}[\tau_1] = \frac{1}{\sigma^2}$ und $\mathbb{E}[\tau_1^2] = \frac{5}{3\sigma^4}$. Ist f eine beschränkte positive messbare Funktion so erhalten wir

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(Y_s) ds = \frac{1}{\mathbb{E}[\nu_2 - \nu_1]} \mathbb{E}\left[\int_{\nu_1}^{\nu_2} f(Y_t) dt\right] = \sigma^2 \mathbb{E}\left[\int_0^{\tau_1} f(\{\sigma W_t\}) dt\right]$$

¹vgl. [RY91], S. 397f.

3 Schätzung des effizienten Preises aus dem Order Flow

fast sicher.² Insbesondere folgt damit

$$\sigma^2 \mathbb{E} \left[\int_0^{\tau_1} f(\{\sigma W_t\}) dt \right] = \int_0^1 f(s) ds.$$

Weiters ergibt sich die folgende Konvergenz in Verteilung³

$$\sqrt{T} \left(\frac{1}{T} \int_0^T f(Y_t) dt - \sigma^2 \mathbb{E} \left[\int_0^{\tau_1} f(\{\sigma W_t\}) dt \right] \right) \xrightarrow[T \rightarrow \infty]{d} \mathbb{N}(0, \sigma^2 \text{Var}[Z^f]), \quad (3.3)$$

wobei Z^f eine zentrierte Zufallsvariable ist, die definiert wird durch

$$Z^f = \int_0^{\tau_1} f(\{\sigma W_t\}) dt - \tau_1 \sigma^2 \mathbb{E} \left[\int_0^{\tau_1} f(\{\sigma W_t\}) dt \right].$$

Nun betrachten wir zwei sehr nützliche Spezialfälle für die Funktion f . Zuerst wählen wir $f = h$ und mit Annahme 3.1 folgt fast sicher, dass

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T h(Y_t) dt = \sigma^2 \mathbb{E} \left[\int_0^{\tau_1} h(\{\sigma W_t\}) dt \right] = 1.$$

Somit

$$\begin{aligned} Z^h &= \int_0^{\tau_1} \left(\mathbb{I}_{\{W_t < 0\}} (h(1 + \sigma W_t) - 1) + \mathbb{I}_{\{W_t > 0\}} (h(\sigma W_t) - 1) \right) dt \\ &= \int_{-\frac{1}{\sigma}}^{\frac{1}{\sigma}} \left(\mathbb{I}_{\{u < 0\}} (h(1 + \sigma u) - 1) + \mathbb{I}_{\{u > 0\}} (h(\sigma u) - 1) \right) L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(u) du, \end{aligned}$$

wobei $L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}$ die Lokalzeit von $(W_t)_{t \geq 0}$ bezüglich τ_1 ist.

Wählt man die Funktion $f(x) = \mathbb{I}_{\{h(x) \leq t\}}$, $t \in [0, h(1^-))$, so ergibt sich

$$\int_0^1 \mathbb{I}_{\{h(s) \leq t\}} ds = h^{-1}(t).$$

Wir definieren $Z_e(t) = Z^f$ durch

$$\begin{aligned} Z_e(t) &= \int_0^{\tau_1} \left(\mathbb{I}_{\{W_s < 0, h(1 + \sigma W_s) \leq t\}} + \mathbb{I}_{\{W_s > 0, h(\sigma W_s) \leq t\}} - h^{-1}(t) \right) ds \\ &= \int_{-\frac{1}{\sigma}}^{\frac{1}{\sigma}} \left(\mathbb{I}_{\{u < 0, h(1 + \sigma u) \leq t\}} + \mathbb{I}_{\{u > 0, h(\sigma u) \leq t\}} - h^{-1}(t) \right) L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(u) du. \end{aligned}$$

Es wird hilfreich sein, $(Z_e(t))_{t \in [0, h(1^-))}$ als Prozess mit Index t zu betrachten. Wir führen die Kovarianzfunktion ein als

$$\rho(t_1, t_2) = \text{Cov}[Z_e(t_1), Z_e(t_2)] = \mathbb{E}[Z_e(t_1)Z_e(t_2)]. \quad (3.4)$$

²vgl. [Asm03], S. 178.

³vgl. [Asm03], S. 178f.

3 Schätzung des effizienten Preises aus dem Order Flow

Diese Funktion und auch $\text{Var}[Z^h]$ können als mehrdimensionales Integral auf $[-\frac{1}{\sigma}, \frac{1}{\sigma}]^2$ berechnet werden. Tatsächlich gibt es für $(u, v) \in [-\frac{1}{\sigma}, \frac{1}{\sigma}]^2$ sogar einen expliziten Ausdruck für die bivariate Laplace-Transformierte von⁴

$$(L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(u), L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(v)).$$

Mittels Differentiation kann $\mathbb{E}[L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(u), L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(v)]$ hergeleitet werden und damit auch $\rho(t_1, t_2)$ und $\text{Var}[Z^h]$.

3.4 Schätzung der Flow Response Function und ihrer Inversen

Im Nachfolgenden soll die *Flow Response Function* geschätzt werden. Bevor diese Schätzung vorgenommen werden kann, wird ein Schätzer für μ_T benötigt. Da

$$\mathbb{E}\left[\frac{N_T}{\mu_T T}\right] = \mathbb{E}\left[\frac{1}{T} \int_0^T h(Y_t) dt\right] = \frac{1}{T} \int_0^T \mathbb{E}[h(Y_t)] dt = 1$$

gilt, arbeiten wir mit dem Schätzer

$$\hat{\mu}_T = \frac{N_T}{T}. \quad (3.5)$$

Es gilt folgende Proposition.

Proposition 3.3. Gilt $\mu_T \rightarrow \infty$ wenn $T \rightarrow \infty$, so folgt

$$\sqrt{T} \left(\frac{\hat{\mu}_T}{\mu_T} - 1 \right) \xrightarrow{d} N(0, \sigma^2 \text{Var}[Z^h]).$$

Unten wird darauf eingegangen, wie die *Flow Response Function* geschätzt wird. Sei k_T eine bekannte deterministische Folge von positiven natürlichen Zahlen. Dann definieren wir für $j = 1, \dots, k_T$

$$\hat{\theta}_j = k_T \frac{N_{j \frac{T}{k_T}} - N_{(j-1) \frac{T}{k_T}}}{\hat{\mu}_T T} = \frac{k_T}{N_T} (N_{j \frac{T}{k_T}} - N_{(j-1) \frac{T}{k_T}}). \quad (3.6)$$

Näherungsweise ist $\hat{\theta}_j$ gleich

$$\frac{1}{\mu_T \frac{T}{k_T}} \sum_{i=1}^{\lfloor \mu_T \frac{T}{k_T} \rfloor} (N_{(j-1) \frac{T}{k_T} + \frac{i}{\mu_T}} - N_{(j-1) \frac{T}{k_T} + \frac{(i-1)}{\mu_T}}).$$

⁴vgl. [BS96], S. 182.

3 Schätzung des effizienten Preises aus dem Order Flow

Bedingt man auf den Pfad von (Y_t) , so sind die Variablen in der Summe unabhängig und wenn $\frac{T}{k_T}$ klein genug ist, so folgen sie näherungsweise einer Poisson-Verteilung mit Parameter $\mu_T \frac{h(Y_{(j-1)T/k_T})}{k_T}$. Weiters kann man erwarten, dass, wenn $\mu_T \frac{T}{k_T}$ hinreichend groß ist, $\hat{\theta}_j$ nahe bei $h(Y_{(j-1)T/k_T})$ liegt. In dieser Arbeit wird angenommen, dass k_T so gewählt wird, damit für ein $p > 0$ gilt

$$\frac{T^{p+\frac{1}{2}}}{k_T^{\frac{p}{2}}} \xrightarrow{T \rightarrow \infty} 0, \quad \frac{k_T T^{\frac{1}{2}}}{\mu_T} \xrightarrow{T \rightarrow \infty} 0.$$

Wegen $\frac{T^{\frac{5}{2}+\varepsilon}}{\mu_T} \rightarrow 0$ existiert so eine Folge k_T . Die eingeführten $\hat{\theta}_j$ sind k_T Schätzer von der Form $h(u_j)$, $u_j \in [0, 1]$. Die Werte von u_j sind unbekannt, jedoch wissen wir, dass sie stetig gleichverteilt auf $[0, 1]$ sind. Wir können die $\hat{\theta}_j$ ordnen: $\hat{\theta}_{(1)} \leq \hat{\theta}_{(2)} \leq \dots \leq \hat{\theta}_{(k_T)}$. Für $u \in [0, 1)$ definieren wir den Schätzer von $h(u)$ auf folgende Weise:

$$\hat{h}(u) = \hat{\theta}_{(\lfloor uk_T \rfloor + 1)}.$$

Die Identifizierungs-Bedingung gilt dabei, da

$$\int_0^1 \hat{h}(u) \, du = \frac{1}{k_T} \sum_{j=1}^{k_T} \hat{\theta}_j = 1.$$

Der Schätzer der Inversen der *Flow Response Function* ist in natürlicher Weise gegeben durch

$$\hat{h}^{-1}(t) = \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\hat{\theta}_j \leq t\}}. \quad (3.7)$$

In der Tat ist \hat{h} die rechtsstetige verallgemeinerte Umkehrfunktion von \hat{h}^{-1} .

3.5 Sätze über das Grenzverhalten der Response Function

In diesem Unterkapitel wird auf zwei Sätze eingegangen, die in Zusammenhang mit \hat{h}^{-1} und \hat{h} stehen. Für ein $b > 0$ bezeichnet $D[0, b)$ (bzw. $D[0, b]$) den Raum der càdlàg Funktionen von $[0, b)$ (bzw. $[0, b]$) nach \mathbb{R} . Die Konvergenz (in Verteilung) in $D[0, b)$ wird definiert als die Konvergenz (in Verteilung) der Einschränkung eines stochastischen Prozesses auf ein kompaktes Intervall $[0, a]$, $0 < a < b$ bezüglich der Skorohod Topologie J_1 . Damit gilt für \hat{h}^{-1} der folgende Satz.

Satz 3.4. Unter den Annahmen 3.1 und 3.2 gilt

$$\sqrt{T}(\hat{h}^{-1}(\cdot) - h^{-1}(\cdot)) \xrightarrow[T \rightarrow \infty]{d} \sigma G(\cdot) - \frac{(\cdot)}{h'(h^{-1}(\cdot))} \int_0^{h(1^-)} \sigma G(v) dv$$

in $D[0, h(1^-))$, wobei $G(\cdot)$ ein zentrierter stetiger Gauß-Prozess mit Kovarianzfunktion ρ , wie in (3.4) definiert, ist.

Obwohl sich k_T Terme in der Summe der Definition von h^{-1} , vgl. Formel (3.7), befinden, muss beachtet werden, dass die Konvergenzrate in Satz 3.4 gleich \sqrt{T} und damit langsamer als $\sqrt{k_T}$ ist. Dies folgt aus der starken Abhängigkeit der $\hat{\theta}_j$ in jedem Zyklus und der Tatsache, dass die Anzahl der Zyklen von der Ordnung T ist. Ein ähnliches Resultat gilt auch für \hat{h} .

Satz 3.5. Unter den Annahmen 3.1 und 3.2 gilt

$$\sqrt{T}(\hat{h}(\cdot) - h(\cdot)) \xrightarrow[T \rightarrow \infty]{d} -h'(\cdot)\sigma G(h(\cdot)) + h(\cdot) \int_0^{h(1^-)} \sigma G(v) dv$$

in $D[0, 1)$.

Nicht vergessen werden darf, dass sowohl in Satz 3.4 als auch in Satz 3.5, die Kovarianzfunktion des Grenzprozesses explizit berechnet werden kann, vgl. Abschnitt 3.3.

3.6 Schätzer für den effizienten Preis

Sei $t \in (0, T]$ die Ankunftszeit einer Order. Eine hilfreiche Folgerung aus Satz 3.4 ist die Möglichkeit der Schätzung von Y_t , welcher der gebrochene Anteil des *effizienten Preises* zur Zeit t ist. Durch die Kenntnis des *Best Bid Price*, lässt sich der *effiziente Preis* nun einfach berechnen.⁵ Da die Schätzer von \hat{h}^{-1} und h^{-1} bekannt sind und sich auch $h(Y_t)$ durch

$$\widehat{h}(Y_t) = k_T \frac{N_t - N_{t - \frac{T}{k_T}}}{\hat{\mu}_T T} \quad (3.8)$$

schätzen lässt, wählen wir für den Schätzer \hat{Y}_t von Y_t

$$\hat{Y}_t = \hat{h}^{-1}(\widehat{h}(Y_t)). \quad (3.9)$$

Das folgende Korollar zeigt, dass dies ein wirklich guter Schätzer für den *effizienten Preis* Y_t ist.

⁵vgl. Formel (3.1).

Korollar 3.6. Für ein $t > 0$, gilt unter den Annahmen 3.1 und 3.2, dass

$$\sqrt{T}(\hat{Y}_t - Y_t) \xrightarrow[T \rightarrow \infty]{d} \sigma G(h(Y_t)),$$

wobei G unabhängig von Y_t ist.

Dieser Ansatz ermöglicht es also, den *effizienten Preis* mit der Konvergenzrate \sqrt{T} aus dem *Order Flow* zu erhalten. Die Verteilung von $\sigma G(h(Y_T))$ kann explizit berechnet werden. Bedingt man auf Y_t , so erhält man sogar eine zentrierte normalverteilte Zufallsvariable mit Varianz $\sigma^2 \rho(h(Y_t), h(Y_t))$. Hier ist es hilfreich zu wissen, dass Y_t unabhängig von G und stetig gleichverteilt ist.

3.7 Zusammenfassung

Im Unterschied zum Vorgehen in den klassischen Ansätzen wird in diesem Modell die dynamische Information des *Order Flows* verwendet um den *effizienten Preis* zu schätzen. Das Grundprinzip dabei lautet „*the price is where the volume is not*“. Es wird angenommen, dass die Intensität des *Order Flows* als deterministische Funktion der Distanz zwischen dem betrachteten Level (*Best Bid* oder *Best Ask*) und dem *Efficient Price* beschrieben werden kann. Danach wird diese Funktion in parameterfreier Weise, das heißt ohne Annahmen an die Wahrscheinlichkeitsverteilung, geschätzt und daraus der Schätzer des *effizienten Preises* abgeleitet.

Um den Lesefluß nicht zu stören, werden die Aussagen in Kapitel 8 bewiesen.

4 Implementierung des effizienten Preises

Im nachfolgenden Kapitel wollen wir das vorgestellte theoretische Modell auch praktisch nutzen. Dafür werden reale Börsendaten benötigt. In dieser Arbeit werden ausschließlich Daten der *Borsa Istanbul*, kurz *BIST* oder *ISE* für *Istanbul Stock Exchange*, verwendet.

Die *BIST* ist ein zentralisierter und rein auftragsgesteuerter Markt und verfügt über eine einzige elektronische Handelsplattform. Da die *BIST* eine vollkommen automatisierte Börse ist und alle Aufträge elektronisch ohne *Market Maker* übermittelt werden, enthalten die vorhandenen Daten den gesamten *Order Flow* der Aktien des *ISE-30* Index für die Monate Juni und Juli 2008. Dieser Index enthält 75% des gesamten Handelsvolumens in dieser Periode. Jeder Marktteilnehmer hat Zugriff zu genau diesen Daten. Damit hat man die selben Informationen wie die Marktakteure und somit eine perfekte Nachbildung der Realität. Gehandelt wird zwischen 9:30 und 12:00 und nach einer Mittagspause von 14:00 bis 17:00 Uhr. Die meistgehandelte Aktie und damit diejenige mit den meisten Datensätzen ist die der *T.Garanti Bankasi A.Ş.*¹

4.1 Aufbau der Daten

Um die Implementierung vollständig zu verstehen, wird zunächst der Aufbau der Datenbank erklärt.

Die Daten stehen in Matlab *.mat* - Files und beinhalten jeweils eine Struktur, deren erste Zeile wieder Strukturen enthält. Die restlichen Zellen sind leer. Jeder diese Strukturen enthält Informationen über eine Aktie.

¹vgl. [VZFR13], S. 2ff.

4 Implementierung des effizienten Preises

	1	2	3	...	9	10
1	1x1 struct	1x1 struct	1x1 struct	...	1x1 struct	1x1 struct
2						
⋮						

Tabelle 4.1: Aufbau der Matlab-Dateien

Sehen wir uns eine dieser Strukturen im Detail an. Diese enthält:

- eine Struktur *order*,
- eine Struktur *trade*,
- eine Struktur *tick*,
- und den Wert *ticker*, der den Namen der Aktie beinhaltet.

Die Struktur *tick* besitzt genauso viele Zeilen wie die Struktur *order*, allerdings nur eine Spalte, in welcher die aktuelle *Tick Size* steht.

Die Tabelle *order* hat insgesamt 77 Spalten. Allerdings werden für den in dieser Arbeit vorgestellten Code nur wenige davon verwendet.

	...	10	11	...	15	16	...	26	27	...
1	...	34200	93000	...	NaN	NaN	...	1	1	...
2	...	34200	93000	...	NaN	NaN	...	1	1	...
⋮	...	⋮	⋮	...	⋮	⋮	...	⋮	⋮	...
27	...	34201	93001	...	5.3500	5.4500	...	1	1	...
28	...	34201	93001	...	5.3500	5.4500	...	1	1	...
⋮	...	⋮	⋮	...	⋮	⋮	...	⋮	⋮	...
538	...	36090	100130	...	5.3500	5.4000	...	0	0	...
539	...	36092	100132	...	5.3500	5.4000	...	0	1	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Tabelle 4.2: Aufbau der *order* Tabelle

4 Implementierung des effizienten Preises

Jede Zeile beschreibt einen eingetroffenen Auftrag. In der Spalte 26 ist die Information enthalten, um welche Order-Art es sich handelt. Der Wert 1 steht für eine *Limit Order* und Null für eine *Market Order*. Wenn in Spalte 27 die Zahl 1 steht, handelt es sich um eine *Buy Order*, anderenfalls um eine *Sell Order*.

Spalte 10 beinhaltet den Handelszeitpunkt in Sekunden. Zum Handelsstart um 9:30 Uhr sind $9,5 * 60 * 60 = 34200$ Sekunden seit Mitternacht vergangen. Die elfte Spalte enthält die Handelszeit im Format HHMMSS.

In den Spalten 15 und 16 stehen der *Best Bid Price* bzw. der *Best Ask Price*. Bei Tagesbeginn stehen diese jedoch noch nicht fest. Daher beinhalten die Zellen den Wert NaN (*Not a Number*).

4.2 Die Wahl der Folge k_T

Für die Berechnung des *effizienten Preises* muss eine Folge k_T gewählt werden, die folgenden Voraussetzungen genügt:

$$\frac{T^{p+\frac{1}{2}}}{k_T^{\frac{p}{2}}} \xrightarrow{T \rightarrow \infty} 0, \quad \frac{k_T T^{\frac{1}{2}}}{\mu_T} \xrightarrow{T \rightarrow \infty} 0.$$

Damit diese Bedingungen erfüllt sind, muss k_T mindestens die Konvergenzrate cT^2 haben, wobei c eine positive Konstante ist. Allerdings ergeben sich bei dieser Wahl zwei Probleme:

- Das Matlab-Programm zur Berechnung des *effizienten Preises* ist sehr ressourcenintensiv. Das Laufzeitverhalten ist also entsprechend schlecht.
- Der *Efficient Price* nähert sich immer dem *Best Ask Level* an.

4 Implementierung des effizienten Preises

Wird nur ein Wert des *Efficient Price* benötigt, ist das Laufzeitverhalten noch relativ überschaubar. Wird der *effiziente Preis* für verschiedene Zeitpunkte benötigt, kann die Berechnung mehrere Stunden in Anspruch nehmen. Das größere Übel ist jedoch, dass selbst bei der Wahl einer kleinen Konstanten c der *effiziente Preis* sehr schnell konvergiert. Dies wird in der folgenden Abbildung gut ersichtlich.

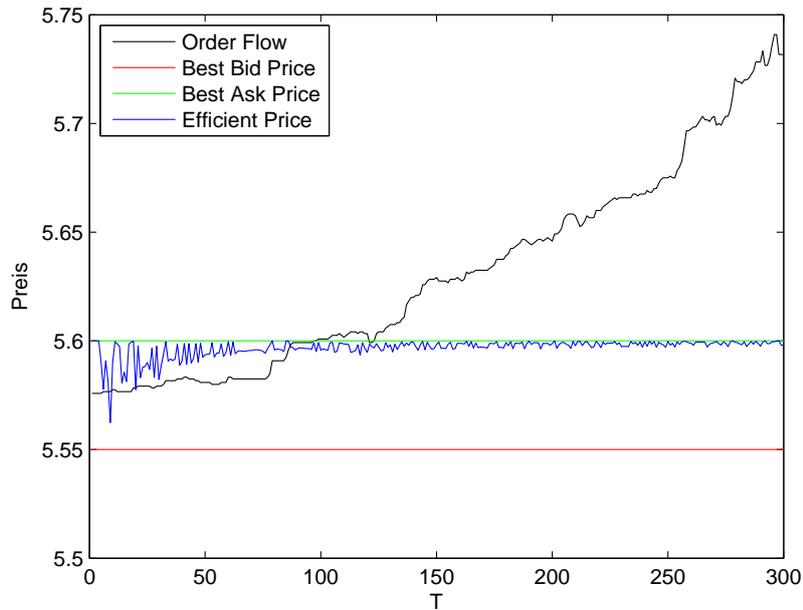


Abbildung 4.1: Entwicklung des effizienten Preises mit $k_T = 0.05 * T^2$

Der *Order Flow* in den Abbildungen 4.1, 4.2 und 4.3 beschreibt, ob mehr Verkaufs- oder Kaufaufträge im betrachteten Zeitraum platziert wurden, vgl. Programm 9.5. In den Grafiken wurde dieser stark skaliert. Ein Blick auf die Abbildung zeigt, dass mehr *Sell Limit Orders* in Auftrag gegeben wurden und daher der *Efficient Price* eigentlich näher am *Best Bid Price* sein sollte.

4 Implementierung des effizienten Preises

Wird die Folge k_T in Abhängigkeit vom Poisson-Prozess N_T gewählt, so ergibt sich eine schwächere Tendenz zur Annäherung an das *Best Ask* Level. Allerdings folgt die Berechnung des *Efficient Price* trotzdem noch nicht dem Grundsatz „*the price is where the volume is not*“.

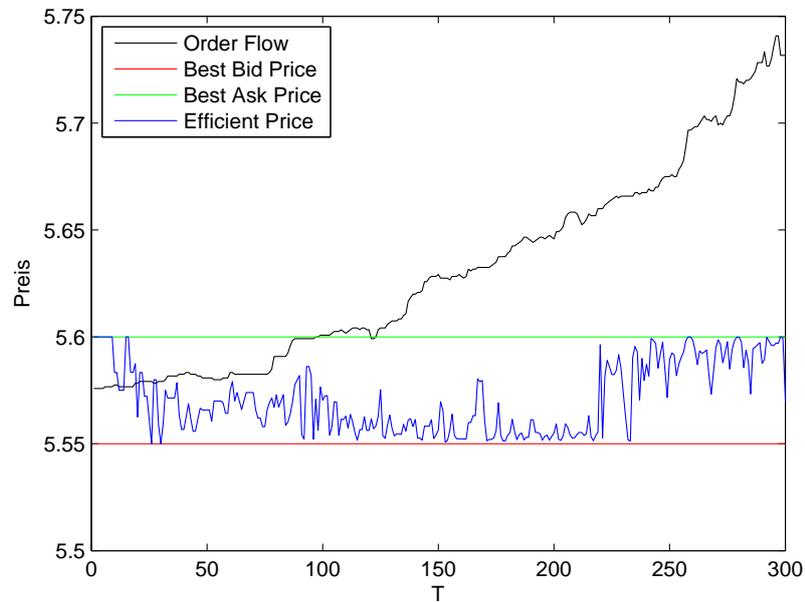


Abbildung 4.2: Entwicklung des effizienten Preises mit $k_T = \frac{N_T}{2}$

Bis zu einer *Order Flow* Länge von 220 verhält sich der *effiziente Preis* so, wie man es erwarten würde. Danach nähert er sich allerdings wieder dem *Best Ask* Level an. Im Vergleich zu $k_T = 0.05 * T^2$ liefert diese Wahl aber zweifellos sinnvollere Werte.

4 Implementierung des effizienten Preises

Setzt man die Folge als $k_T = T$, so variiert der *effiziente Preis* sehr stark. Es gibt keine Tendenz zum *Best Ask*, der goldenen Regel folgt der Wert jedoch auch nicht.

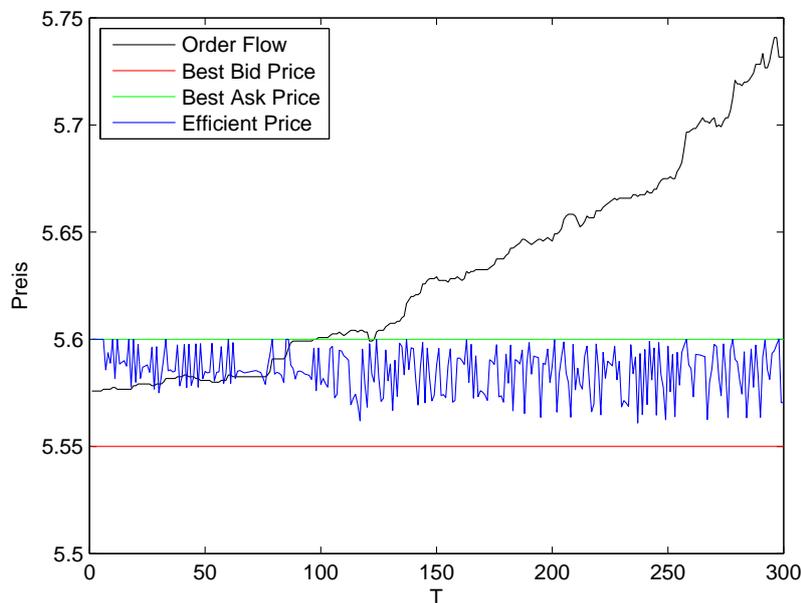


Abbildung 4.3: Entwicklung des effizienten Preises mit $k_T = T$

Nun stellt sich die Frage, wie die Tendenz zum *Best Ask Price* erklärt werden kann. Im Modell von [DRR13] wird der *effiziente Preis* aus der Anzahl der platzierten *Limit Orders* geschätzt, vgl. Formeln (3.8), (3.9) und (3.1). Allerdings wird dabei nur die Intensität der Ankünfte der *Limit Orders*, und nicht das Verhältnis von Ankünften am *Best Ask* und Ankünften am *Best Bid* berücksichtigt. Die Autoren erklären, dass es eine Verbindung zwischen den Auftragseingängen am *Best Bid* bzw. *Best Ask* durch den *Efficient Price* gibt.² Diese konnte mit den Daten der Istanbuler Börse allerdings nicht nachgewiesen werden.

Nachfolgende Abbildungen bzw. die dazugehörigen Berechnungen wurden mit $k_t = \frac{N_T}{2}$ durchgeführt. Bei der Erstellung der Arbeit wurden auch weitere Möglichkeiten für die Wahl der Folge k_T berücksichtigt, allerdings liefert diese Kalkulation die besten Ergebnisse.

²vgl. [DRR13], S. 2611.

4.3 Die Wahl von N_T

Delattre, Robert und Rosenbaum definieren in ihrem Modell N_T als die Gesamtanzahl an *Limit Orders*, die im Zeitraum $[0, T]$ platziert wurden. Davor nehmen sie allerdings ohne Beschränkung der Allgemeinheit an, dass nur mehr der *Limit Order Flow* am *Best Bid* Level betrachtet wird. Durch diese missverständliche sprachliche Formulierung wird nicht klar, wie N_T gewählt werden soll. Die Ergebnisse, die sich durch die Berechnung des *effizienten Preises* aus einer der folgenden Varianten

- N_T als Anzahl aller platzierten *Limit Orders*
- N_T als Anzahl aller platzierten *Buy Limit Orders*
- N_T als Anzahl aller *Limit Orders*, platziert am *Best Bid* Level

ergeben, werden in den Abbildungen 4.5, 4.6 und 4.7 dargestellt.

Für eine Aktie, die alle 100 Sekunden betrachtet wird, ergeben sich signifikante Unterschiede in der berechneten Anzahl der *Limit Orders*.

T	100	200	300	400	500	600	700	800	900	1000
<i>LO am Best Bid</i>	7	9	9	9	9	10	10	10	12	18
<i>Buy Limit Orders</i>	133	168	177	180	183	185	185	185	188	209
<i>Limit Orders</i>	368	494	542	569	577	585	585	585	591	665

Tabelle 4.3: Werte für N_T

4 Implementierung des effizienten Preises

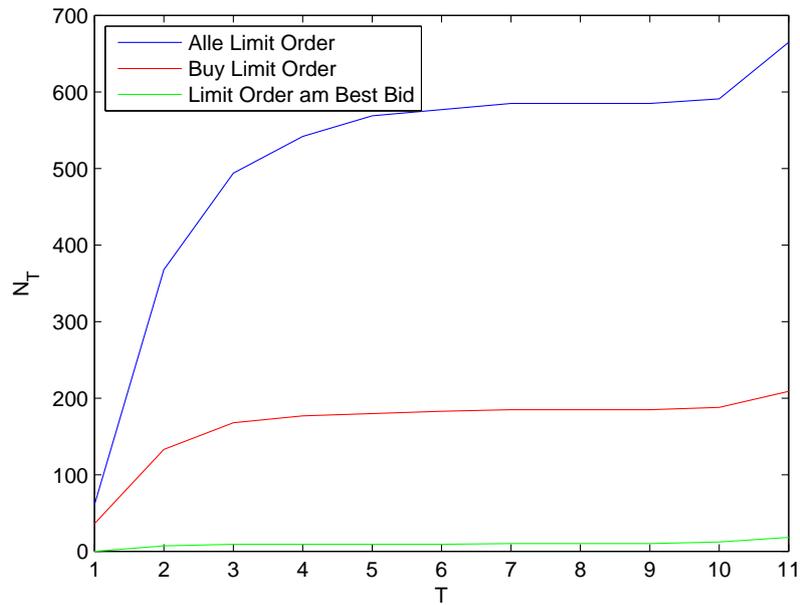


Abbildung 4.4: Entwicklung von N_T

Wird N_T als Anzahl aller platzierten *Limit Orders* gewählt, so sieht man die Entwicklung des *effizienten Preises* in Abbildung 4.5, wobei N_T in den folgenden Grafiken zur besseren Lesbarkeit stark skaliert wurde. In Abbildung 4.6 wird die Entwicklung für die Wahl von N_T als Anzahl aller platzierten *Buy Limit Orders* dargestellt.

4 Implementierung des effizienten Preises

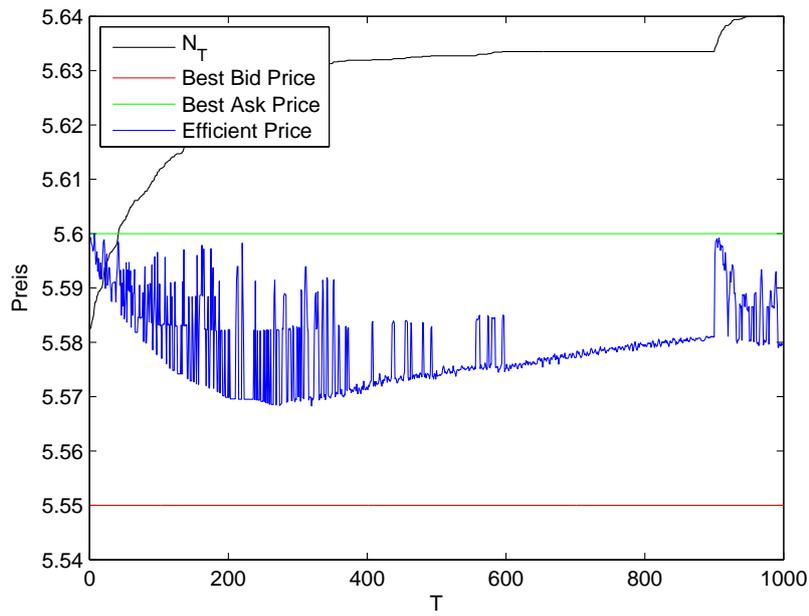


Abbildung 4.5: Entwicklung des effizienten Preises bei Betrachtung aller *Limit Orders*

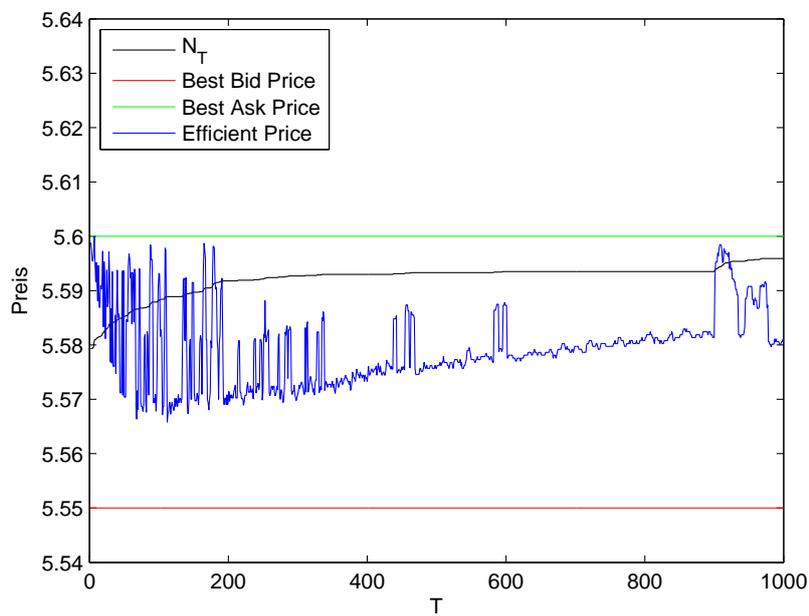


Abbildung 4.6: Entwicklung des effizienten Preises bei Betrachtung aller *Buy Limit Orders*

4 Implementierung des effizienten Preises

Werden für die Berechnung nur die *Limit Orders* am *Best Bid Level* verwendet, so sieht der Verlauf des *Efficient Price* wie folgt aus:

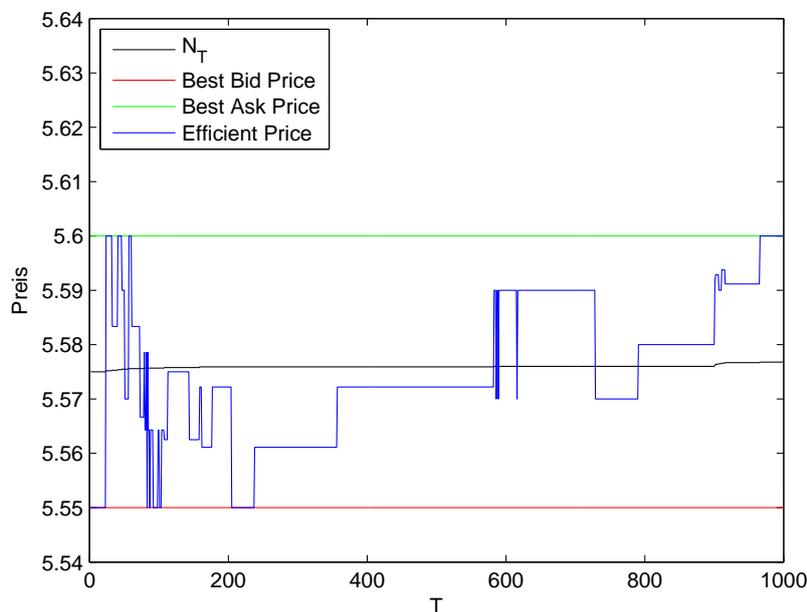


Abbildung 4.7: Entwicklung des effizienten Preises bei Betrachtung aller *Limit Orders* am *Best Bid Level*

Da keine allzu großen Unterschiede zwischen den ersten zwei betrachteten Wahlen für N_T liegen, halte ich mich nachfolgend an das Modell von Delattre, Robert und Rosenbaum, und wähle N_T als Anzahl aller platzierten *Limit Orders*. Bei der Ausarbeitung wurden jedoch die vielen verschiedenen Möglichkeiten der Wahl von k_T und N_T berücksichtigt. Der *Efficient Price* wurde für verschiedene Zeitpunkte und mit unterschiedlichen *Order Flow* Längen berechnet. Zur besseren Lesbarkeit der Diplomarbeit werden die Ergebnisse und die Implementierung nur für die sinnvollste Wahl gezeigt.

4.4 Implementierung in Matlab

Die Implementierung erfolgt nach dem im Kapitel 3 beschriebenen Modell und wurde angepasst an den Aufbau der vorhandenen Daten. Im Folgenden werden nur Ausschnitte aus der Berechnung des *effizienten Preis* betrachtet. Zur vollständigen Nachvollziehbarkeit kann der Programmcode aller verwendeten Funktionen in Kapitel 9 eingesehen werden. Die Zeilennummerierung bezieht sich auf den vollständigen Code des Programms 9.1. Die ganze Analyse des *Efficient Price* in meiner Diplomarbeit beruht auf dieser Implementierung.

```
23 %Bestimmt Mue.
24 mu_T = N_T/tau;
```

In Zeile 24 des Programmcodes wird Formel (3.5) für $\hat{\mu}_T$ verwendet.

```
33 %Genaue Berechnung von Theta.
34 [jfirstRow jlastRow] = findRows(file, index, day, month, year, T-tau, 0);
35 N_jT = numberLOAll(file, index, jfirstRow, jlastRow);
36 for j=1:K_T
37     N_iT = N_jT;
38     [jfirstRow jlastRow] = findRows(file, index, day, month, year, ...
        T-tau+j*tau/K_T, j*tau/K_T);
39     N_jT = numberLOAll(file, index, jfirstRow, jlastRow);
40     helptheta = K_T/N_T * (N_jT - N_iT);
41     theta(j) = helptheta;
42 end;
```

In den Zeilen 34 bis 42 Zeile werden die $\hat{\theta}_j$ analog zu Formel (3.6) bestimmt.

```
44 %Berechnung der Order Flow Response Function.
45 [hfirstRow hlastRow] = findRows(file, index, day, month, year, T-tau + ...
        tau - tau/K_T, tau - tau/K_T);
46 help = numberLOAll(file, index, hfirstRow, hlastRow);
47 hYt = K_T * (N_T - help)/(mu_T*tau);
```

Der Schätzer für die *Order Flow Response Function* (vgl. Formel (3.8)) wird in den Programmzeilen 45 bis 47 implementiert.

4 Implementierung des effizienten Preises

```
49 %Gebrochenen Anteil zum Zeitpunkt T berechnen.
50 sort(theta);
51 fractionalPart = numel(find(theta ≤ hYt))/K.T;
52
53 %Skalierungsfaktor auslesen.
54 factor = file(1,index).tick(lastRow,1);
55
56 %Im Modell wird von einer Tick Size von eins ausgegangen.
57 %Daher muss der gebrochene Anteil noch angepasst werden.
58 fractionalPart = fractionalPart * factor;
```

In den Zeilen 50 bis 58 wird schließlich der gebrochene Anteil Y_t mit Hilfe der Formeln (3.7) beziehungsweise (3.9) berechnet. Dabei muss berücksichtigt werden, dass im Modell eine *Tick Size* von eins angenommen wird und an der Börse weit niedrigere Werte üblich sind. Um den richtigen Wert zu erhalten, wird mit der aktuellen *Tick Size* skaliert. Dies ist zulässig, da im Modell der *Efficient Price* parameterfrei geschätzt wird und damit Ergebnisse invariant gegenüber einer Transformation mit einer streng monotonen Funktion sind.

```
60 %Best Bid zum Zeitpunkt T.
61 bestBid = file(1,index).order(lastRow,15);
62
63 %Best Ask zum Zeitpunkt T.
64 bestAsk = file(1,index).order(lastRow,16);
65
66 %Efficient Price setzt sich aus Best Bid und dem gebrochenen Anteil
67 %zusammen.
68 EP = bestBid + fractionalPart;
```

Schlussendlich wird in Zeile 61 der *Best Bid Price* ausgelesen und in Zeile 68 der gebrochenen Anteil addiert. Dadurch erhält man den *effizienten Preis* wie in Formel (3.1) definiert.

5 Effizienter Preis und Volatilität

Die Frage, ob es einen einfachen linearen Zusammenhang zwischen *effizientem Preis* und der Volatilität gibt, versuche ich in diesem Kapitel zu beantworten. Auf den folgenden Seiten werden einige Volatilitätsschätzer vorgestellt und versucht auf Grundlage der vorhandenen Daten der *Borsa Istanbul* Strukturen bzw. Abhängigkeiten zwischen den Ergebnissen zu erkennen.

5.1 Volatilitätsschätzer

Wird ein stetiger stochastischer Prozess $(X_t)_{t \in \mathbb{R}^+}$ über ein festgelegtes Zeitintervall beobachtet und in $n \in \mathbb{N}$ kleine angrenzende Intervalle unterteilt, so liefert die quadratische Variation¹

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n (X_i - X_{i-1})^2$$

einen Schätzer für die Volatilität. Dabei gilt der Limes in Wahrscheinlichkeit. Erfüllen die Preise die stochastische Differentialgleichung

$$dX_t = \mu_t dt + \sigma_t dW_t, \quad (5.1)$$

wobei W_t eine Brownsche Bewegung, $\mu_t \in \mathbb{R}$, $\sigma_t > 0$ und $t > 0$ sind, so stimmt die quadratische Variation mit der integrierten Varianz

$$\int_0^T \sigma_t^2 dt,$$

für ein $T > 0$, überein.²

Im Hochfrequenzhandel kann in Gleichung (5.1) $\mu_t = 0$ gesetzt werden, da der *Drift* im Sekundenbereich zu vernachlässigen ist. Es tritt jedoch ein *Market Microstructure*

¹vgl. [Øk00], S. 19.

²vgl. [ZMAS05], S. 1394.

5 Effizienter Preis und Volatilität

Noise Process auf, wodurch die beobachteten Preise Y_t folgenden Aufbau nachkommen:

$$Y_t = X_t + \epsilon_t.$$

Dabei wird angenommen, dass die nicht beobachtbaren logarithmischen Preise X_t die stochastische Differentialgleichung (5.1) erfüllen und der *Noise Process* ϵ_t unabhängig von X_t ist.

Dieser *Noise* Term ϵ_t wird durch eine Reihe von Marktunstimmigkeiten verursacht. Beispielsweise Sprünge zwischen *Bid-* und *Ask Price*, ungewöhnliche Preisänderungen und Rundungen, Unterschiede in Handelsvolumen, Effekte des Portfoliomanagement, sowie Mess- und Aufzeichnungsfehler bzw. Fehler bei der Platzierung einer Order. Das heißt, vom eigentlichen Interesse ist die Volatilität des verunreinigten Preisprozesses X_t .

Die nachfolgenden Volatilitätsschätzer sind alle Schätzer für die integrierte Varianz der X_t . Die geschätzte Volatilität kann beispielsweise zum *Hedgen* von Portfolios oder zur Berechnung der integrierten Varianz des nächsten Tages verwendet werden.³

5.1.1 Realized Variance

Ein einfacher Schätzer ist der *Realized Variance Estimator (RV)*.⁴

$$RV_{\text{all}} = \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2.$$

Dieser ist ohne *Microstructure Noise* ein konsistenter und effizienter Schätzer der integrierten Varianz. Mit *Market Microstructure Noise* divergiert er allerdings, wenn die Anzahl der Beobachtungen n erhöht wird.⁵

Es ist möglich die beste Samplingfrequenz zu berechnen und dadurch bessere Ergebnisse zu erhalten. Dabei kann es aber sein, dass sehr viele Daten nicht berücksichtigt werden. Aus statistischen Gründen sollte jedoch immer versucht werden, möglichst alle Daten zu verwenden. Daher liefert der *RV* Schätzer auch nicht die besten Ergebnisse und seine Werte weichen von den Werten der anderen Schätzer ab.⁶ Bei der empirischen Studie wird der Schätzer mit optimaler Frequenz kalkuliert und als RV_{sparse} bezeichnet.

³vgl. [ASMZ11], S. 160f.

⁴vgl. [HL06], S. 129.

⁵vgl. [GO10], S. 260.

⁶vgl. Tabelle 5.1 und Abbildung 5.1.

5.1.2 Zhou Estimator

Eine Erweiterung des RV Schätzers ist der *Zhou Estimator*, der durch Einbeziehung der Autokovarianzen erster Ordnung eine einfache Biaskorrektur durchführt.

$$RV_{\text{Zhou}} = \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})^2 + \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})(Y_{t_{i-1}} - Y_{t_{i-2}}) + \sum_{i=1}^n (Y_{t_i} - Y_{t_{i-1}})(Y_{t_{i+1}} - Y_{t_i}).$$

Nachteile des *Zhou Estimator* sind, dass er ein inkonsistenter Schätzer ist und Daten außerhalb des betrachteten Intervalls benötigt werden.⁷ In Abbildung 5.1 lässt sich allerdings leicht erkennen, dass er sich ähnlich der nachfolgenden Volatilitätsschätzer verhält.

5.1.3 Two-Scaled Realized Volatility

Ein konsistenter, asymptotisch unverzerrter Schätzer ist der *Two Scaled Realized Volatility Estimator (TSRV)*. Die Daten werden mit zwei unterschiedlichen Frequenzen abgetastet und daraus wird jeweils die integrierte Varianz geschätzt. Der *TSRV* wird als Linearkombination dieser Werte definiert.

$$TSRV = \underbrace{RV_{\text{avg}}^{(K)}}_{\text{slow time scale}} - \frac{\bar{n}}{n} \underbrace{RV_{\text{all}}}_{\text{fast time scale}},$$

wobei

$$RV_{\text{avg}}^{(K)} = \frac{1}{K} \sum_{k=1}^K RV_{\text{sparse}}^{(k)} \quad \text{und} \quad \bar{n} = \frac{n - K + 1}{K}.$$

Der Beobachtungszeitraum wird in K nicht überlappende Teilnetze aufgeteilt. Für jedes Teilnetz k wird der $RV_{\text{sparse}}^{(k)}$ Schätzer berechnet und danach über alle gemittelt. Auch hier ist es möglich sich die optimale Anzahl der Teilnetze K zu bestimmen.⁸

⁷vgl. [Zho96], S. 47.

⁸vgl. [ZMAS05], S. 1395ff.

5.1.4 Multi-Scale Realized Volatility

Anstatt wie beim *TSRV* nur zwei Skalen zu betrachten, werden beim *Multi-Scale Realized Volatility (MSRV)* mehrere Skalen in die Berechnung miteinbezogen, wobei diese auch unterschiedlich gewichtet werden. Der *MSRV* ist ein gegenüber dem *Microstructure Noise Process* robuster Schätzer und weist eine bessere Konvergenzrate auf.⁹ Allerdings ist er durch seine Komplexität aufwändiger zu implementieren und auch der Rechenaufwand erhöht sich wesentlich.

5.1.5 Realized Kernel Estimator

Des Weiteren gibt es den *Realized Kernel Estimator*. Sei δ die Länge der betrachteten Zeitintervalle im gesamten Intervall $[0, T]$, dann ist der *Realized Kernel Estimator* definiert durch

$$K(Y_\delta) = \gamma_0(Y_\delta) + \sum_{h=1}^H k\left(\frac{h-1}{H}\right) [\gamma_h(Y_\delta) + \gamma_{-h}(Y_\delta)],$$

wobei der Kern $k(x)$ für $x \in [0, 1]$ eine deterministische Gewichtsfunktion ist und die h -te *Realized Autocovariance* bestimmt ist als

$$\gamma_h(Y_\delta) = \sum_{j=1}^n (Y_{\delta j} - Y_{\delta(j-1)})(Y_{\delta(j-h)} - Y_{\delta(j-h-1)}), \quad (5.2)$$

mit $h = -n, \dots, -1, 0, 1, \dots, n$ und $n = \lfloor \frac{T}{\delta} \rfloor$. Diese Klasse von Schätzern hat durch die freie Wahl des Kerns eine gewisse Anpassungsfreiheit. Deswegen kann durch optimale Wahl der Parameter ein konsistenter Schätzer mit einer guten Konvergenzrate erreicht werden.¹⁰

⁹vgl. [Zha06], S. 1019ff.

¹⁰vgl. [BNHLS08], S. 1482, 1520f.

5.1.6 Vergleich der Schätzer

Ohne näher auf die Implementierung der Volatilitätsschätzer in Matlab einzugehen, möchten ich hier die Resultate aus denselben kurz vergleichen. In der nachfolgenden Abbildung ist gut erkennbar, dass sich der Schätzer RV_{sparse} ein wenig anders verhält als RV_{Zhou} , $TSRV$ und $MSRV$, die sich nur geringfügig voneinander unterscheiden.

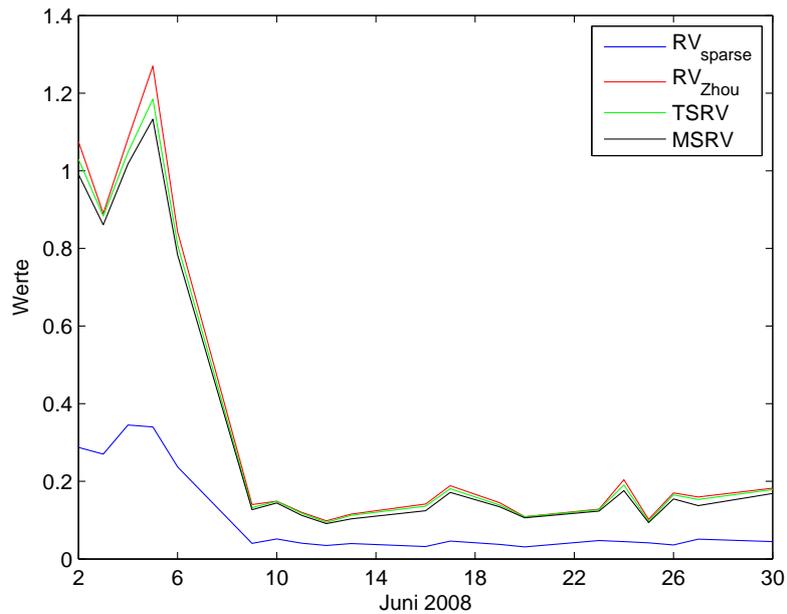


Abbildung 5.1: Vergleich der Volatilitätsschätzer

5 Effizienter Preis und Volatilität

In Tabelle 5.1 stehen die dazugehörigen Werte der Volatilitätsschätzer, die später zum Vergleich mit dem *Efficient Price* herangezogen werden.

Datum	RV_{sparse}	RV_{Zhou}	$TSRV$	$MSRV$
2.06.08	0.2875	1.0750	1.0295	0.9902
3.06.08	0.2700	0.8900	0.8830	0.8607
4.06.08	0.3450	1.0825	1.0475	1.0176
5.06.08	0.3400	1.2700	1.1850	1.1329
6.06.08	0.2375	0.8425	0.8095	0.7838
9.06.08	0.0400	0.1400	0.1325	0.1270
10.06.08	0.0516	0.1488	0.1483	0.1442
11.06.08	0.0408	0.1200	0.1180	0.1125
12.06.08	0.0348	0.0984	0.0945	0.0912
13.06.08	0.0396	0.1156	0.1123	0.1032
16.06.08	0.0320	0.1416	0.1362	0.1245
17.06.08	0.0460	0.1888	0.1810	0.1715
19.06.08	0.0376	0.1448	0.1389	0.1342
20.06.08	0.0312	0.1088	0.1089	0.1061
23.06.08	0.0476	0.1284	0.1270	0.1233
24.06.08	0.0448	0.2044	0.1904	0.1760
25.06.08	0.0416	0.1020	0.0970	0.0937
26.06.08	0.0360	0.1700	0.1654	0.1549
27.06.08	0.0512	0.1596	0.1532	0.1375
30.06.08	0.0448	0.1824	0.1789	0.1686

Tabelle 5.1: Vergleich der Volatilitätsschätzer

5.2 Differenz Efficient Price und Trade Price

In den nächsten Absätzen wird die Summe der Absolutbeträge der Differenz des *Efficient Price* und des *Trade Price* dargestellt. Dieser resultierende Wert wird als V_{TP} definiert und tageweise berechnet. Der *effiziente Preis* zum Zeitpunkt t mit betrachteter *Order Flow* Länge von $\tau \in \mathbb{N}$ in Sekunden wird bezeichnet mit $EP_{t,\tau}$. Der *Trade Price* zum Zeitpunkt t mit TP_t , wobei $t \in \mathbb{N}$ jeweils die Sekunden seit Beginn des Tages beschreibt. Sei $S \in \mathbb{N}$ die Gesamtanzahl der Sekunden an einem Handelstag. Damit ergibt sich

$$V_{\text{TP}} = \sum_{i=1}^{\lfloor \frac{S}{\tau} \rfloor} |EP_{i\tau,\tau} - TP_{i\tau}|.$$

Die folgende Abbildung veranschaulicht die Berechnung.

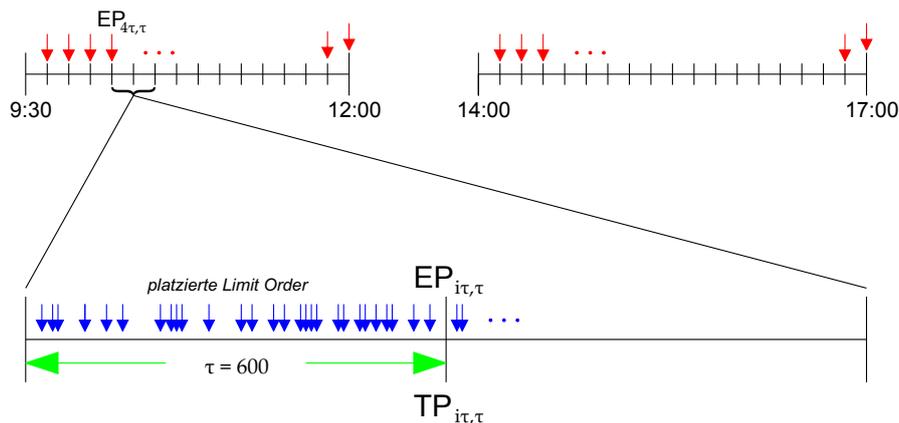


Abbildung 5.2: Berechnung von V_{TP}

Es wird jeweils ein *Order Flow* von beispielsweise $\tau = 600$ Sekunden, also 10 Minuten, betrachtet. Aus den Daten der letzten 10 Minuten wird der *effiziente Preis* berechnet und der aktuelle *Trade Price* ausgelesen. Die Summe über die Absolutbeträge der Differenz dieser Werte ergibt schließlich V_{TP} . Dass zwischen 12:00 und 14:00 Uhr kein Handel stattfindet, wird bei der Kalkulation berücksichtigt. Es werden dabei alle Daten verwendet.

5 Effizienter Preis und Volatilität

In der Tabelle 5.2 stehen die Werte für V_{TP} für die Aktie der *T.Garanti Bankasi A.Ş.* am *ISE* im Juni 2008 bei unterschiedlicher Wahl von τ .

Datum	$\tau = 120$	$\tau = 300$	$\tau = 600$	$TSRV$
2. Juni 2008	4.1133	1.6554	0.6521	1.0295
3. Juni 2008	4.1713	1.4596	0.6826	0.8830
4. Juni 2008	3.8293	1.4687	0.7374	1.0475
5. Juni 2008	3.6351	1.4895	0.7485	1.1850
6. Juni 2008	3.9781	1.4774	0.7258	0.8095
9. Juni 2008	1.5366	0.6017	0.3028	0.1325
10. Juni 2008	1.4982	0.6552	0.3311	0.1483
11. Juni 2008	1.5940	0.6852	0.2955	0.1180
12. Juni 2008	1.6595	0.6600	0.3159	0.0945
13. Juni 2008	1.5875	0.6004	0.2776	0.1123
16. Juni 2008	1.4977	0.5525	0.2908	0.1362
17. Juni 2008	1.5879	0.6476	0.3040	0.1810
19. Juni 2008	1.5291	0.5847	0.3036	0.1389
20. Juni 2008	1.7464	0.8143	0.3592	0.1089
23. Juni 2008	1.7449	0.6574	0.3184	0.1270
24. Juni 2008	1.6332	0.6390	0.3010	0.1904
25. Juni 2008	1.5895	0.7305	0.2697	0.0970
26. Juni 2008	1.5331	0.6962	0.3678	0.1654
27. Juni 2008	1.6124	0.6881	0.3567	0.1532
30. Juni 2008	1.7088	0.6333	0.3077	0.1789

Tabelle 5.2: Werte für V_{TP} und für den Volatilitätsschätzer $TSRV$

Die Korrelationskoeffizienten ρ_τ und die dazugehörigen p-Werte¹¹ können der folgenden Tabelle entnommen werden:

¹¹vgl. [Gur11], S. 10, 89.

5 Effizienter Preis und Volatilität

$TSRV$	$\tau = 120$	$\tau = 300$	$\tau = 600$
ρ_τ	0.9631	0.9687	0.9745
p-Wert	0	0	0

RV_{sparse}	$\tau = 120$	$\tau = 300$	$\tau = 600$
ρ_τ	0.9628	0.9657	0.9756
p-Wert	0	0	0

Tabelle 5.3: Korrelationskoeffizienten ρ_τ und p-Werte

Durch den p-Wert von Null ist die Signifikanz des Ergebnisses gegeben und es kann von einer bestehenden linearen Abhängigkeit von V_{TP} und $TSRV$ bzw. RV_{sparse} ausgegangen werden. Die Berechnung für den Volatilitätsschätzer RV_{Zhou} liefert ähnliche Größen. Sieht man sich jedoch den Korrelationsplot der Daten für $\tau = 600$ genauer an, ist keine eindeutige lineare Abhängigkeit in der Form $f(x) = kx + d$ zu erkennen.

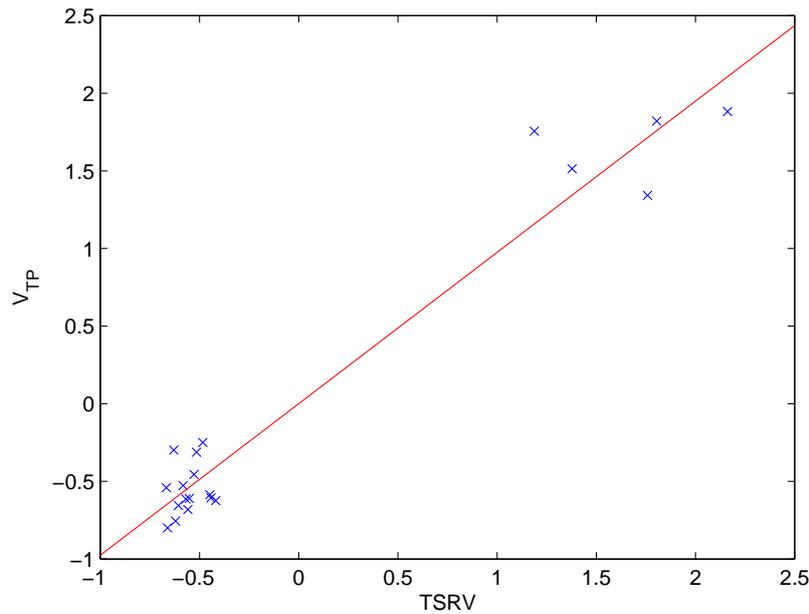


Abbildung 5.3: Korrelation zwischen V_{TP} , mit $\tau = 600$, und dem Volatilitätsschätzer $TSRV$

Da dies schon sehr hohe Werte sind, gibt es eine starke lineare Abhängigkeit. Idealerweise gäbe es jedoch einen eindeutiger Zusammenhang, der mittels einer Formel und ohne Fehlerabweichung von der einen zur anderen Größe führt.

5.3 Differenz Efficient Price und Best Bid Price

Im zweiten Versuch habe ich, den *Efficient Price* mit dem *Best Bid Price* verglichen. Es gelten die gleichen Bezeichnungen wie auf den vorgehenden Seiten. Zusätzlich wird der Preis am *Best Bid Level* als BB_t und das Ergebnis als V_{BB} definiert.

$$V_{BB} = \sum_{i=1}^{\lfloor \frac{S}{\tau} \rfloor} |EP_{i\tau, \tau} - BB_{i\tau}|.$$

Wird τ unterschiedlich gewählt, ergeben sich für die Aktie der *T.Garanti Bankasi A.Ş.* am *ISE* im Juni 2008 folgende Werte für V_{BB} :

Datum	$\tau = 120$	$\tau = 300$	$\tau = 600$	$MSRV$
2. Juni 2008	4.9876	1.8739	0.9698	0.9902
3. Juni 2008	4.9808	1.7660	0.8909	0.8607
4. Juni 2008	4.7594	2.0600	0.9213	1.0176
5. Juni 2008	4.6316	1.8752	0.8192	1.1329
6. Juni 2008	5.1114	1.8876	0.8514	0.7838
9. Juni 2008	2.0616	0.7504	0.3756	0.1270
10. Juni 2008	1.9007	0.7396	0.3360	0.1442
11. Juni 2008	1.9746	0.7490	0.3654	0.1125
12. Juni 2008	1.8820	0.6932	0.3364	0.0912
13. Juni 2008	1.9930	0.7153	0.3755	0.1032
16. Juni 2008	2.0626	0.7351	0.3594	0.1245
17. Juni 2008	1.9164	0.7020	0.3508	0.1715
19. Juni 2008	1.9262	0.7474	0.4179	0.1342
20. Juni 2008	1.9432	0.7499	0.3451	0.1061
23. Juni 2008	1.9408	0.6936	0.3851	0.1233
24. Juni 2008	1.8017	0.7426	0.3686	0.1760
25. Juni 2008	1.9618	0.7144	0.3622	0.0937
26. Juni 2008	2.0253	0.7570	0.3786	0.1549
27. Juni 2008	1.9180	0.7457	0.3629	0.1375
30. Juni 2008	2.0682	0.7825	0.3977	0.1686

Tabelle 5.4: Werte für V_{BB} und für den Volatilitätsschätzer $MSRV$

5 Effizienter Preis und Volatilität

Der Korrelationskoeffizient ρ_{120} ergibt 0.9694 bei Wahl von $\tau = 120$. Für die weiteren Intervalle ergeben sich die Werte $\rho_{300} = 0.9849$ und $\rho_{600} = 0.9744$. In allen Fällen ergibt der p-wert Null, wodurch die Signifikanz der Ergebnisse gegeben ist. Es besteht wieder eine lineare Abhängigkeit zwischen V_{BB} und der Volatilität. Die Werte gelten für den $MSRV$, da bei diesem Schätzer der größte Zusammenhang festgestellt werden konnte.

Die folgende Abbildung zeigt allerdings, dass es leider keine eindeutige Funktion $f(x) = kx + d$ gibt, mit der aus dem *effizienten Preis* auf die Volatilität geschlossen werden kann.

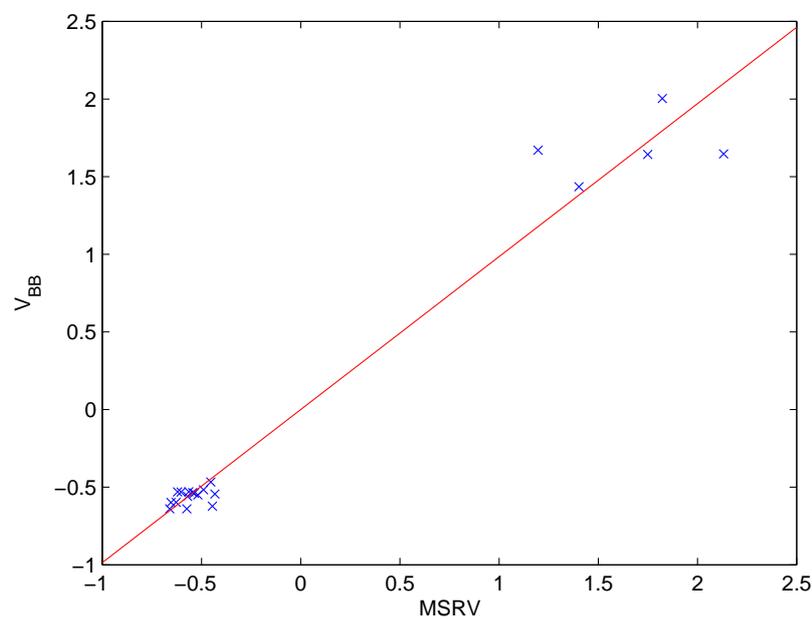
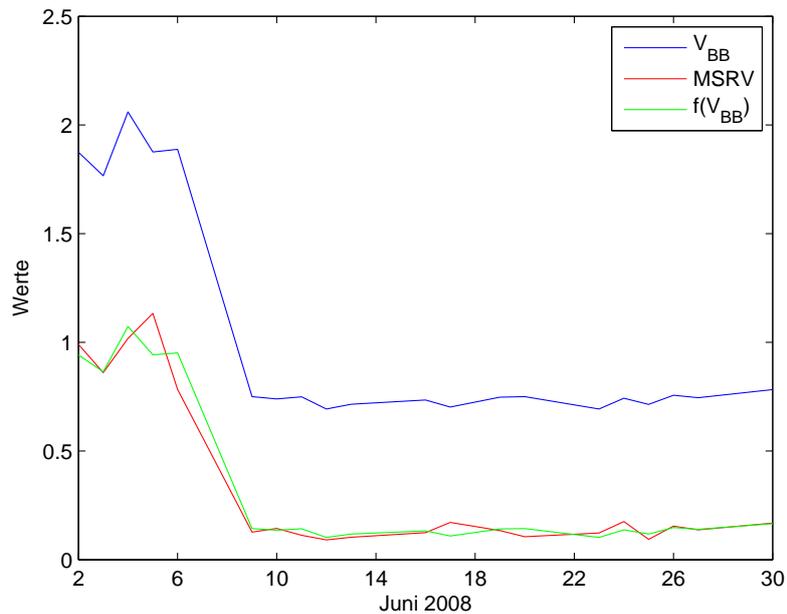


Abbildung 5.4: Korrelation zwischen V_{BB} , mit $\tau = 300$, und dem Volatilitätsschätzer $MSRV$

Abbildung 5.5: Schätzung des $MSRV$ aus den Werten von V_{BB}

Auch in diesem Versuch reicht die Abhängigkeit nicht aus, um auf fehlerfreier Weise mit den Werten des *Efficient Price* die Volatilität zu bestimmen. In Abbildung 5.5 sieht man, dass dieser Zusammenhang jedoch bereits zu einem guten Schätzer für die Volatilität führt. Die Koeffizienten der Funktion $\widehat{MSRV} = f(V_{BB}) = 0.7102 + 0.3896 V_{BB}$ wurden dabei mittels linearer Regression ermittelt.

5.4 Differenz Efficient Price und Best Ask Price

Anstatt das *Best Bid Level* BB_t zu betrachten, wird hier mit dem höheren *Best Ask Price* verglichen, den wir als BA_t definieren wollen. Die Kalkulation und die weiteren Bezeichnungen folgen dem obigen Schema.

$$V_{BA} = \sum_{i=1}^{\lfloor \frac{S}{\tau} \rfloor} |EP_{i\tau, \tau} - BA_{i\tau}|.$$

Wir betrachten wieder die Aktienwerte der *T.Garanti Bankasi A.Ş.* an der *Borsa Istanbul*.

5 Effizienter Preis und Volatilität

Datum	$\tau = 120$	$\tau = 300$	$\tau = 600$	RV_{sparse}
2. Juni 2008	3.1624	1.3261	0.5802	0.2875
3. Juni 2008	3.3692	1.4840	0.6591	0.2700
4. Juni 2008	3.3906	1.1400	0.6287	0.3450
5. Juni 2008	3.5184	1.3248	0.7308	0.3400
6. Juni 2008	3.0386	1.3124	0.6986	0.2375
9. Juni 2008	1.1984	0.5296	0.2444	0.0400
10. Juni 2008	1.3793	0.5604	0.3040	0.0516
11. Juni 2008	1.2854	0.5310	0.2546	0.0408
12. Juni 2008	1.4580	0.6068	0.2836	0.0348
13. Juni 2008	1.2670	0.5647	0.2445	0.0396
16. Juni 2008	1.2574	0.5649	0.2606	0.0320
17. Juni 2008	1.3436	0.5780	0.2692	0.0460
19. Juni 2008	1.3338	0.5326	0.2021	0.0376
20. Juni 2008	1.3168	0.5301	0.2749	0.0312
23. Juni 2008	1.3192	0.5864	0.2349	0.0476
24. Juni 2008	1.4583	0.5374	0.2514	0.0448
25. Juni 2008	1.2982	0.5656	0.2578	0.0416
26. Juni 2008	1.2347	0.5230	0.2414	0.0360
27. Juni 2008	1.3420	0.5343	0.2571	0.0512
30. Juni 2008	1.1918	0.4975	0.2223	0.0448

Tabelle 5.5: Werte für V_{BA} und für den Volatilitätsschätzer RV_{sparse}

Dabei ergibt sich die größte Abhängigkeit für den Schätzer RV_{sparse} .

RV_{sparse}	$\tau = 120$	$\tau = 300$	$\tau = 600$
ρ_τ	0.9895	0.9500	0.9651
p-Wert	0	0	0

Tabelle 5.6: Korrelationskoeffizienten und p-Werte

Auch in dieser Berechnung ergibt sich nicht der gewünschte einfache, strukturelle Zusammenhang (vgl. Abbildung 5.6). Allerdings ist ein Wert von $\rho_{120} = 0.9895$ der bislang beste und kommt dem gewünschten Ergebnis schon sehr nahe.

5 Effizienter Preis und Volatilität

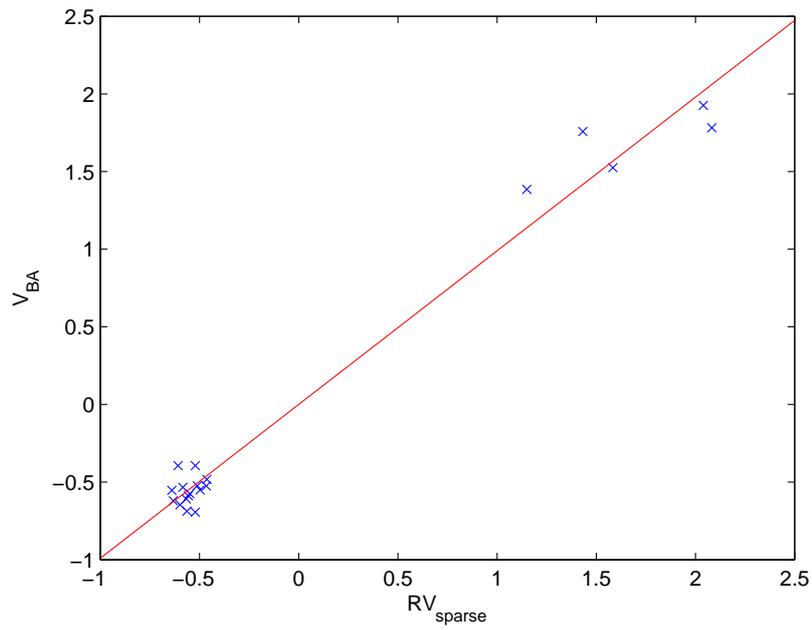


Abbildung 5.6: Korrelation zwischen V_{BA} , mit $\tau = 120$, und dem Volatilitätsschätzer RV_{sparse}

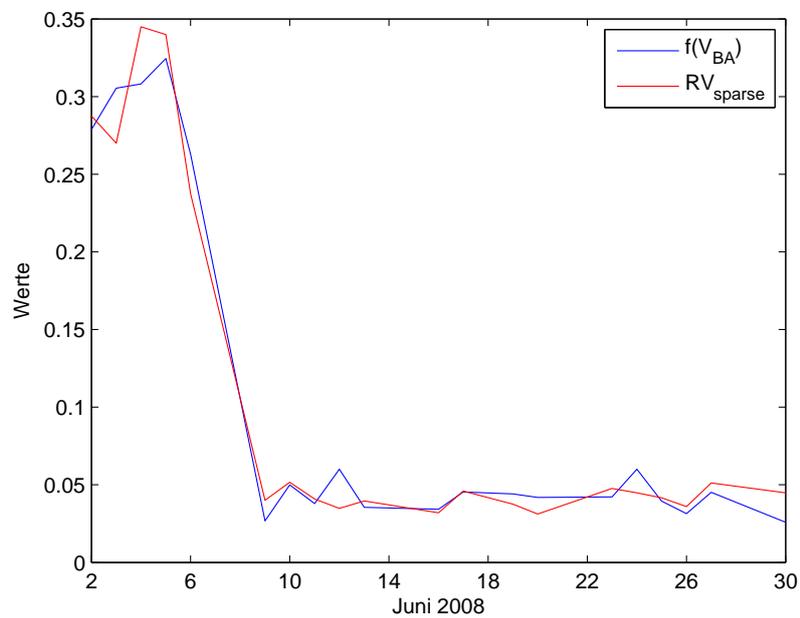


Abbildung 5.7: Schätzung des RV_{sparse} aus den Werten von V_{BA}

5 Effizienter Preis und Volatilität

Mit Hilfe von linearer Regression erhält man den Zusammenhang $\widehat{RV}_{\text{sparse}} = f(V_{BA}) = -0.1271 + 0.1284 V_{BA}$. In Abbildung 5.7 sieht man den RV_{sparse} und die Schätzung aus den Werten von V_{BA} . Es ist deutlich zu erkennen, dass diese Berechnung eine sehr gute Approximation der Volatilität zur Folge hat.

5.5 Differenz der Werte des effizienten Preises

In folgendem Ansatz wird versucht, aus der Summe der Absolutbeträge der Differenz des *Efficient Price* auf die Volatilität zu schließen. Ich benenne diesen Wert mit V_{EP} , da dieser ausschließlich vom *effizienten Preis* abhängig ist.

$$V_{EP} = \sum_{i=1}^{\lfloor \frac{S}{\tau} \rfloor} |EP_{i\tau, \tau} - EP_{(i-1)\tau, \tau}|.$$

Auf die Wertetabelle wird in dieser Analyse verzichtet, der Vergleich wird lediglich in einer Grafik dargestellt.

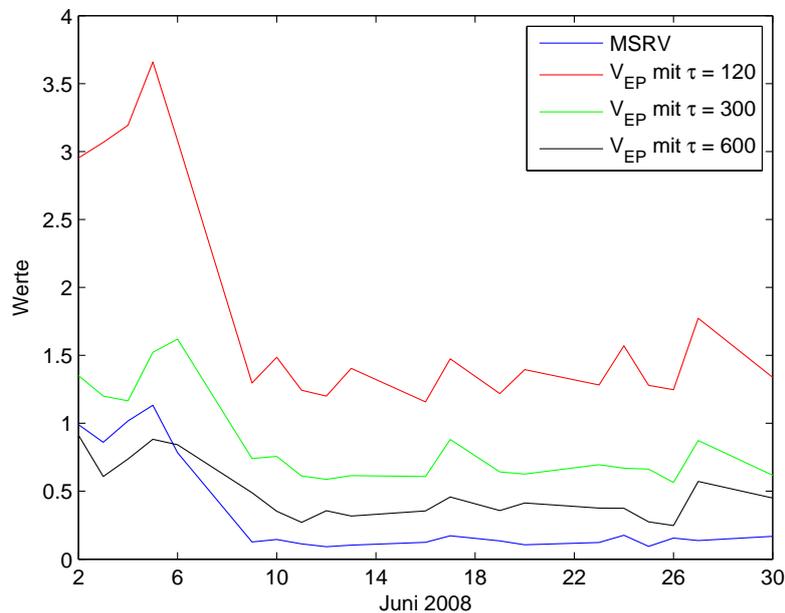


Abbildung 5.8: Vergleich der Ergebnisse

5 Effizienter Preis und Volatilität

Man kann erkennen, dass die V_{EP} annähernd den Verlauf des $MSRV$ haben. Dieser wurde als Referenzkurs gewählt, da bei dieser Berechnung die Volatilitätsschätzer $TSRV$ und $MSRV$ die besten Ergebnisse liefern.

$TSRV$	$\tau = 120$	$\tau = 300$	$\tau = 600$
ρ_τ	0.9806	0.9144	0.9035
p-Wert	0	0	0

$MSRV$	$\tau = 120$	$\tau = 300$	$\tau = 600$
ρ_τ	0.9804	0.9137	0.9017
p-Wert	0	0	0

Tabelle 5.7: Korrelationskoeffizienten und p-Werte

Durch die Tabelle entsteht der Eindruck, dass, je höher die Frequenz gewählt wird, desto stärker wird die lineare Abhängigkeit zwischen dem Volatilitätsschätzer und V_{EP} . Infolgedessen habe ich auch den Korrelationskoeffizienten für $\tau = 60$ berechnet und mir einen noch besseren Wert erhofft. Das Resultat von $\rho_{60} = -0.7137$ ist jedoch der bislang schlechteste Wert für den linearen Zusammenhang.

5.6 Methode der kleinsten Quadrate

Die Vorgehensweise ist der im letzten Ansatz ähnlich. Diesmal werden allerdings große Abweichungen stärker bestraft. Wir werden sehen ob durch diese Anpassung eine bessere Verbindung zu einem Volatilitätsschätzer hergestellt werden kann. Dabei verwende ich wieder die Bezeichnungen wie oben und definiere V_{KQ} als

$$V_{KQ} = \sqrt{\sum_{i=1}^{\lfloor \frac{S}{\tau} \rfloor} (EP_{i\tau,\tau} - EP_{(i-1)\tau,\tau})^2}.$$

5 Effizienter Preis und Volatilität

Es werden die Aktienwerte der *T.Garanti Bankasi A.Ş.* an der *Borsa Istanbul* betrachtet und dabei ergeben sich folgende Größen:

Datum	$\tau = 120$	$\tau = 30$	$\tau = 600$	RV_{Zhou}
2. Juni 2008	0.3155	0.2386	0.2180	1.0750
3. Juni 2008	0.3174	0.1945	0.1550	0.8900
4. Juni 2008	0.3151	0.1914	0.1695	1.0825
5. Juni 2008	0.3596	0.2600	0.2206	1.2700
6. Juni 2008	0.3339	0.2857	0.2014	0.8425
9. Juni 2008	0.1485	0.1550	0.1279	0.1400
10. Juni 2008	0.1526	0.1251	0.0863	0.1488
11. Juni 2008	0.1219	0.0989	0.0701	0.1200
12. Juni 2008	0.1232	0.1042	0.0907	0.0984
13. Juni 2008	0.1453	0.0981	0.0796	0.1156
16. Juni 2008	0.1213	0.1014	0.0831	0.1416
17. Juni 2008	0.1493	0.1442	0.1168	0.1888
19. Juni 2008	0.1262	0.1012	0.0845	0.1448
20. Juni 2008	0.1493	0.1068	0.0958	0.1088
23. Juni 2008	0.1292	0.1160	0.0934	0.1284
24. Juni 2008	0.1662	0.1185	0.0970	0.2044
25. Juni 2008	0.1308	0.1016	0.0656	0.1020
26. Juni 2008	0.1277	0.0896	0.0643	0.1700
27. Juni 2008	0.1762	0.1371	0.1323	0.1596
30. Juni 2008	0.1377	0.1027	0.1007	0.1824

Tabelle 5.8: Werte für V_{KQ} und für den Volatilitätsschätzer RV_{Zhou}

Die größte Abhängigkeit ergibt sich bei der Wahl von $\tau = 120$.

	RV_{sparse}	RV_{Zhou}	$TSRV$	$MSRV$
ρ_{120}	0.9712	0.9747	0.9758	0.9758
p-Wert	0	0	0	0

Tabelle 5.9: Korrelationskoeffizienten und p-Werte

5 Effizienter Preis und Volatilität

Der Korrelationsplot zeigt auch hier, dass es keinen einfachen Zusammenhang der Form $f(x) = kx + d$ gibt.

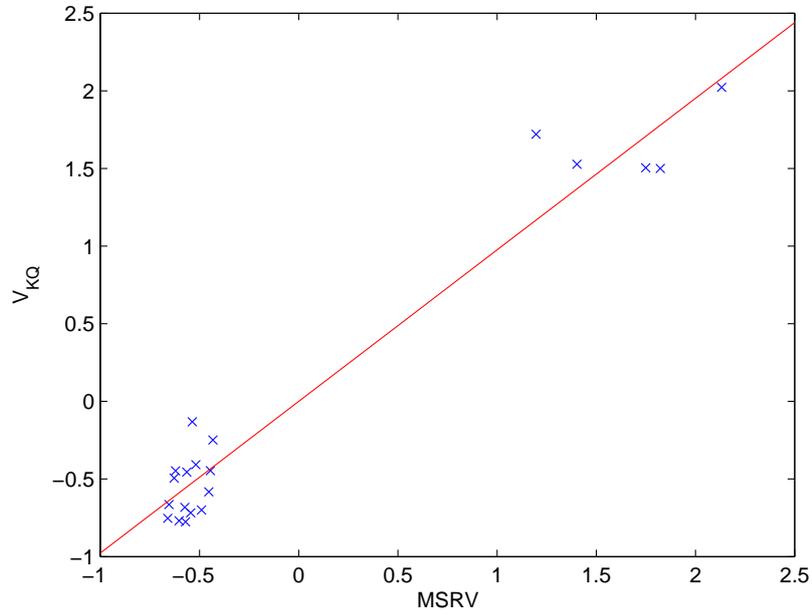


Abbildung 5.9: Korrelation zwischen V_{KQ} , mit $\tau = 120$, und dem Volatilitätsschätzer $MSRV$

Eine hundertprozentige Abhängigkeit konnte ich nicht nachweisen, allerdings bestätigen alle Ergebnisse einen sehr starken linearen Zusammenhang zwischen den Volatilitätsschätzern und dem *effizienten Preis*.

6 Zusammenhang zwischen Delta und Volatilität

Ziel dieses Kapitels ist es zu untersuchen, ob es einen Zusammenhang zwischen den nachfolgend definierten Delta und der Volatilität gibt.¹ Die Delta beschreiben den Preisabstand zwischen dem Wert der platzierten *Limit Order* P_{limit} und dem aktuellen *Mid Price* P_{mid} , wobei mit der derzeitigen *Tick Size* skaliert wird.²

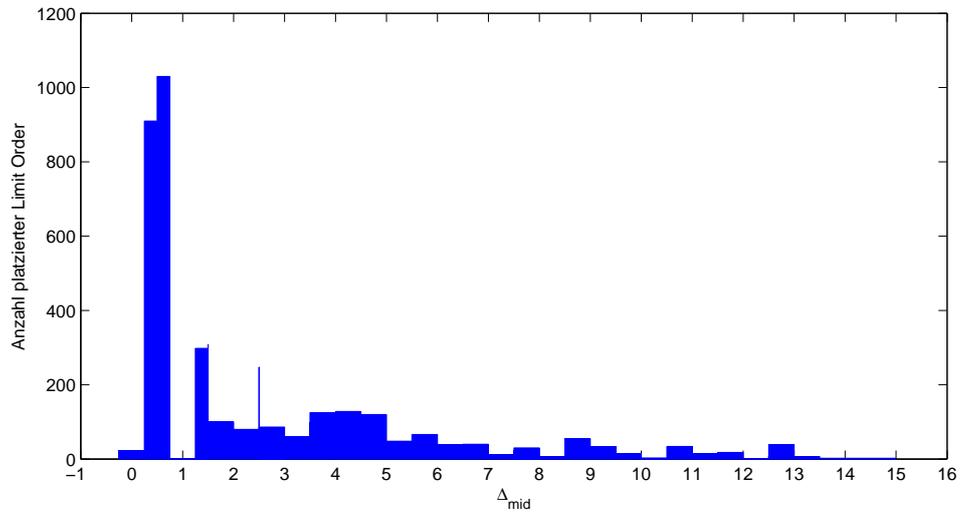
$$\Delta_{\text{mid}} = \frac{|P_{\text{limit}} - P_{\text{mid}}|}{\text{Tick}}.$$

Ein $\Delta_{\text{mid}} = 3$ bedeutet, dass die platzierte *Limit Order* um drei *Ticks* vom *Mid Price* abweicht. Es war naheliegend davon auszugehen, dass die Volatilität umso höher sein wird, je weiter die *Limit Orders* vom *Mid Price* entfernt platziert werden. Dazu haben wir uns die Delta tages- sowie monatsweise berechnet und haben von dieser statistischen Verteilung der *Limit Orders* (vgl. Abbildung 6.1) den Mittelwert, Median und den am häufigsten vorkommenden Wert betrachtet.

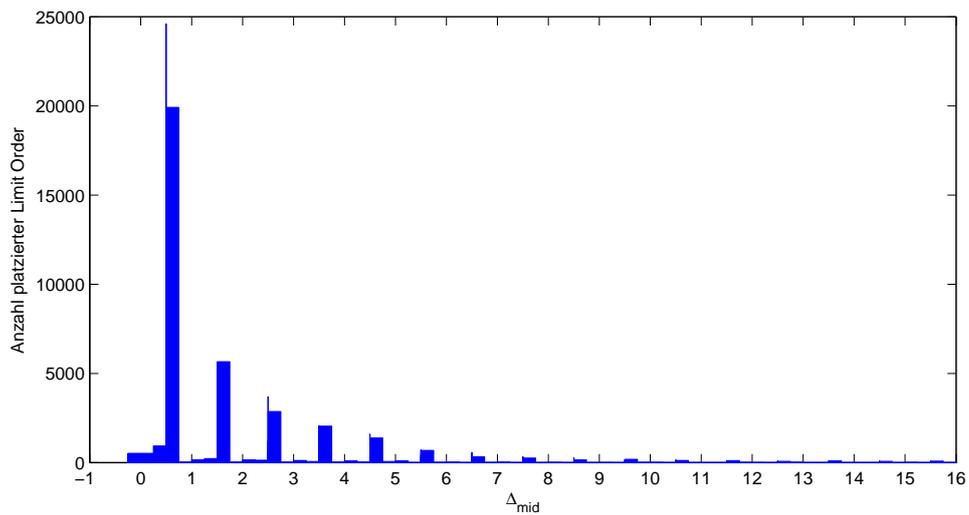
¹Die Implementierung der Delta in Matlab und die Analyse der Ergebnisse wurde gemeinsam mit meinem Studienkollegen Rainer Hirk durchgeführt.

²vgl. [VZFR13], S. 6.

6 Zusammenhang zwischen Delta und Volatilität



(a) Verteilung am Tag



(b) Verteilung im Monat

Abbildung 6.1: Verteilung der Δ_{mid}

In obigen Graphiken wurden die Daten des *T.Garanti Bankasi A.Ş.* an der *Borsa Istanbul* ausgewertet. Es ist erkennbar, dass die *Limit Orders* nicht sehr weit entfernt vom *Mid Price* platziert werden.

6 Zusammenhang zwischen Delta und Volatilität

Bei der tageweisen Betrachtung der Δ_{mid} lässt sich kein linearer Zusammenhang zu den Volatilitätsschätzern feststellen. Das ist bereits sehr gut an folgender Abbildung auszumachen und lässt sich zusätzlich durch die Korrelationskoeffizienten und p-Werte bestätigen.

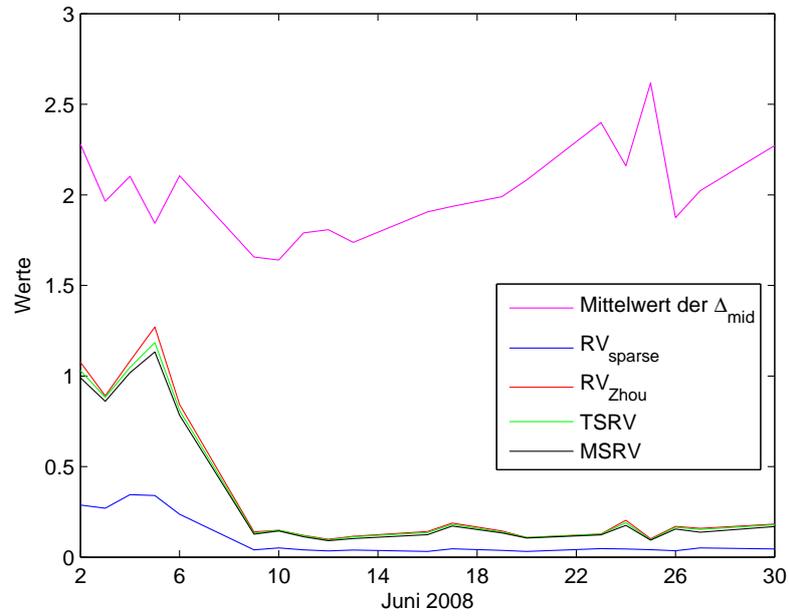


Abbildung 6.2: Vergleich der Delta mit den Volatilitätsschätzern

Auch bei der Analyse eines größeren Zeitraums von einem Monat ergibt sich kein Zusammenhang. In der Tabelle 6.1 sieht man den Vergleich von 10 verschiedenen Aktien der *Borsa Istanbul*.

6 Zusammenhang zwischen Delta und Volatilität

<i>Aktie</i>	<i>Mittelwert der Δ_{mid}</i>	<i>RV_{Zhou}</i>
AKBNK	2.5228	4.1410
AKGRT	3.1191	1.6207
ARCLK	4.0773	0.9882
ASYAB	3.0192	27.0774
DOHOL	3.1052	0.2445
DYHOL	3.0416	0.3132
EREGL	3.0177	13.3900
GARAN	2.0093	12.7298
GSDHO	2.2239	0.2528
HALKB	2.7768	4.6675

Tabelle 6.1: Mittelwert der Δ_{mid} und RV_{Zhou} für die Aktien der *Borsa Istanbul* im Juni 2008

Weitere Versuche mit Bid- oder Ask-Preis führten leider auch zu keinem positiven Ergebnis. Daraus schließen wir, dass es für die Daten der *Borsa Istanbul* keinen simplen linearen Zusammenhang zwischen der Gestalt des *Limit Order Book* und der Volatilität gibt.

Möglicherweise enthalten die Delta zu wenig Information über das *Limit Order Book*. Eine umfangreichere Untersuchung, zum Beispiels mittels *Principal Component Analysis*, würde eventuell zu besseren Ergebnissen führen, jedoch den Rahmen dieser Arbeit übertreffen.

7 Resümee

Da in der Finanzwirtschaft der Preis eines Assets nicht eindeutig definiert ist, habe ich mich in der vorliegenden Diplomarbeit intensiv mit einem neuen Ansatz beschäftigt, der den Preis aus dem zugrundeliegenden *Order Flow* desselbigen berechnet. Um die theoretische Ausarbeitung des Modells zur Schätzung des *Efficient Price* verständlich darlegen zu können, wurden in den ersten zwei Kapiteln wichtige ökonomische und mathematische Begriffe vorgestellt.

Anschliessend wurde darauf eingegangen, wie mit Hilfe von Informationen aus dem *Order Flow* der *Efficient Price* geschätzt werden kann. Das erläuterte Modell nimmt an, dass die Intensität der Auftragseingänge vom Abstand zwischen *Efficient Price* und *Best Bid Price* abhängt. Diese Abhängigkeit wurde mit Hilfe der *Order Flow Response Function* beschrieben und schlussendlich wurde daraus der Schätzer konstruiert.

Ziel von Kapitel 4 war es, das Modell in der Praxis anzuwenden. Mit Hilfe von Daten der *Borsa Istanbul* wurde die Implementierung getestet und analysiert. Dabei erwies sich vor allem die Wahl der Folge k_T als Problem. Wurde die Folge wie im Modell gefordert gewählt, so tendierte der *Efficient Price* immer sehr schnell zum *Best Ask Price*. Dieses Verhalten ist zwar nicht sinnvoll, kann aber dadurch erklärt werden, dass im Modell in erster Linie die Intensität der Ankünfte von neuen *Limit Orders* betrachtet wird, dem Verhältnis von Kauf- zu Verkaufsaufträgen jedoch zu wenig Beachtung geschenkt wurde.

Mit einer an meine Daten angepassten Wahl von k_T habe ich die Analyse fortgesetzt und konnte mittels Korrelationskoeffizient und p-Wert einen sehr starken linearen Zusammenhang zwischen dem *Efficient Price* und der Volatilität der Aktie feststellen. Da die Volatilität geschätzt werden muss, wurden zu Beginn des Kapitels einige Volatilitätsschätzer vorgestellt und danach verschiedenste Berechnungen unternommen, um eine möglichst hohe Abhängigkeit zu erreichen.

7 Resümee

Ob nun auch aus den Delta des *Limit Order Books* direkt auf die Volatilität geschlossen werden kann, konnte allerdings nicht positiv beantwortet werden. Es konnte kein linearer Zusammenhang festgestellt werden, da die leicht zu berechnenden Delta für diese einfache Verbindung vermutlich zu wenig Information enthalten.

Im Finanzwesen wird wohl immer nach dem Preis gesucht werden, der den maximalen Profit verspricht. Insofern bleibt zu hoffen, dass weiter an diesem aktuellen und spannenden Ansatz geforscht wird.

8 Beweise

Es werden die Beweise aus Kapitel 3 wie in [DRR13] gezeigt. In allen Beweisen gelten die Annahmen 3.1 und 3.2. Des Weiteren beschreibt c eine Konstante die sich in jeder, sogar in derselben, Zeile unterscheiden kann.

8.1 Beweis von Proposition 3.3

Die charakteristische Funktion von $\sqrt{T}(\frac{\hat{\mu}_T}{\mu_T} - 1)$ erfüllt

$$\mathbb{E}\left[\exp\left(it\sqrt{T}\left(\frac{\hat{\mu}_T}{\mu_T} - 1\right)\right)\right] = \exp(-it\sqrt{T}) \mathbb{E}\left[\mathbb{E}_Y\left[\exp\left(it\frac{N_T}{\mu_T\sqrt{T}}\right)\right]\right],$$

wobei \mathbb{E}_Y die bedingte Erwartung bezüglich dem Pfad (Y_t) ist. Damit gilt, dass N_T eine Poisson-verteilte Zufallsvariable mit Parameter

$$\mu_T \int_0^T h(Y_t) dt$$

ist. Da bekannt ist, wie die charakteristische Funktion einer Poisson-verteilten Zufallsvariable aussieht, folgt insgesamt

$$\exp(-it\sqrt{T}) \mathbb{E}\left[\exp\left(\left(e^{it(\mu_T\sqrt{T})^{-1}} - 1\right)\mu_T \int_0^T h(Y_t) dt\right)\right].$$

Für die Exponentialfunktion gilt weiters

$$|e^{it(\mu_T\sqrt{T})^{-1}} - 1 - it(\mu_T\sqrt{T})^{-1}| \leq t^2(\mu_T\sqrt{T})^{-2}.$$

Durch die Beschränktheit von h und der Tatsache, dass $\mu_T \rightarrow \infty$, erhalten wir für die charakteristische Funktion

$$\lim_{T \rightarrow \infty} \mathbb{E}\left[\exp\left(\left(it\sqrt{T}\right)\left(\frac{1}{T} \int_0^T h(Y_t) dt - 1\right)\right)\right].$$

Wegen Gleichung (3.3) stimmt das mit

$$\exp\left(-\frac{\sigma^2 t^2}{2} \text{Var}[Z^h]\right)$$

überein. □

8.2 Ein hilfreiches Resultat

Aus folgendem Resultat kann man Satz 3.4 und Satz 3.5 einfach herleiten. Zuerst definieren wir für $j = 1, \dots, k_T$

$$\theta_j = \frac{k_T}{\mu_T T} (N_{\frac{jT}{k_T}} - N_{\frac{(j-1)T}{k_T}}) = \frac{\hat{\mu}_T}{\mu_T} \hat{\theta}_j$$

und die Funktion \hat{h}_e für $u \in [0, 1)$ durch

$$\hat{h}_e(u) = \theta_{(\lfloor uk_T \rfloor + 1)} = \hat{h}(u) \frac{\hat{\mu}_T}{\mu_T}.$$

Zu beachten ist, dass

$$\frac{\hat{\mu}_T}{\mu_T} = \frac{N_T}{T\mu_T} = \frac{1}{k_T} \sum_{j=1}^{k_T} \theta_j = \int_0^1 \hat{h}_e(u) du$$

gilt. Weiters wird

$$\hat{h}_e^{-1}(t) = \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\theta_j \leq t\}}$$

definiert. Es gilt

$$\hat{h}^{-1}(t) = \hat{h}_e^{-1}\left(t \frac{\hat{\mu}_T}{\mu_T}\right).$$

Proposition 8.1. In $D[0, h(1^-))$ gilt

$$\sqrt{T}(\hat{h}_e^{-1}(\cdot) - h^{-1}(\cdot)) \xrightarrow[T \rightarrow \infty]{d} \sigma G(\cdot).$$

Es genügt zu zeigen, dass die endlich-dimensionalen Verteilungen konvergieren und ein Dichtheits-Kriterium gilt.¹ Der Beweis ist in zwei Lemmas unterteilt.

Lemma 8.2. Sei $n \in \mathbb{N}$ und $(t_1, \dots, t_n) \in [0, h(1^-))^n$ beliebig. Dann gilt

$$\sqrt{T}(\hat{h}_e^{-1}(t_i) - h^{-1}(t_i))_{i=1, \dots, n} \xrightarrow[T \rightarrow \infty]{d} N(0, \sigma^2 \nu),$$

wobei $\nu_{i,j} = \rho(t_i, t_j)$ und ρ die Kovarianzfunktion, wie in Gleichung (3.4) definiert, ist.

¹vgl. [Bil68], S. 124.

8 Beweise

Beweis. Es wird nur der Fall $n = 1$ betrachtet. Den Beweis für größere n erhält man aus Linearkombination der Komponenten von $\sqrt{T}(\hat{h}_e^{-1}(t_i) - h^{-1}(t_i))_{i=1,\dots,n}$. Als erstes schreibt man $\hat{h}_e^{-1}(t) - h^{-1}(t) = T_1 + T_2 + T_3$, mit

$$\begin{aligned} T_1 &= \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\theta_j \leq t\}} - \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du \leq t\}}, \\ T_2 &= \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{I}_{\{\frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du \leq t\}} - \frac{1}{T} \sum_{j=1}^{k_T} \int_{(j-1)T/k_T}^{jT/k_T} \mathbb{I}_{\{h(Y_u) \leq t\}} du, \\ T_3 &= \frac{1}{T} \int_0^T \mathbb{I}_{\{h(Y_u) \leq t\}} du - h^{-1}(t). \end{aligned}$$

Es wird jede Komponente getrennt betrachtet. Es gilt

$$T_1 = \frac{1}{k_T} \sum_{j=1}^{k_T} v_{j,k_T,T}$$

mit

$$v_{j,k_T,T} = \begin{cases} 1 & , \text{wenn } \theta_j \leq t < \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt, \\ -1 & , \text{wenn } \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt \leq t < \theta_j, \\ 0 & , \text{sonst.} \end{cases}$$

Sei $\varepsilon_T > 0$, es gilt $\mathbb{E}[|v_{j,k_T,T}|] \leq A_1 + A_2$, mit

$$\begin{aligned} A_1 &= \mathbb{P}\left[\left|\frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du - t\right| \leq \varepsilon_T\right], \\ A_2 &= \mathbb{E}\left[\mathbb{I}_{\{v_{j,k_T,T} = \pm 1\}} \mathbb{I}_{\left\{ \left| t - \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt \right| > \varepsilon_T \right\}}\right]. \end{aligned}$$

Es gilt

$$\left| \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt - h(Y_{(j-1)T/k_T}) \right| \leq c \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} |Y_t - Y_{(j-1)T/k_T}| dt.$$

Aus der Markow-Ungleichung und da Y_s stetig gleichverteilt ist folgt für beliebige $p > 0$, $\delta > 0$, $\lambda > 0$:

$$\mathbb{P}\left[\sup_{t \in [s, s+\delta]} |Y_t - Y_s| \geq \lambda\right] \leq c_p (\delta^{\frac{p}{2}} \lambda^{-p} + \delta^{\frac{1}{2}}).$$

Somit gilt für $p > 0$

$$\mathbb{P}\left[\left|\frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt - h(Y_{(j-1)T/k_T})\right| \geq \varepsilon_T\right] \leq c \frac{(T/k_T)^{\frac{p}{2}}}{\varepsilon_T^p} + c(T/k_T)^{\frac{1}{2}}.$$

8 Beweise

Folglich ist

$$A_1 \leq \mathbb{P}[|h(Y_{(j-1)T/k_T}) - t| \leq 2\varepsilon_T] + c \frac{(T/k_T)^{\frac{p}{2}}}{\varepsilon_T^p} + c(T/k_T)^{\frac{1}{2}}.$$

Unter Verwendung der stationären Verteilung von $Y_{(j-1)T/k_T}$ erhalten wir für ε_T klein genug

$$A_1 \leq h^{-1}(t + 2\varepsilon_T) - h^{-1}(t - 2\varepsilon_T) + c \frac{(T/k_T)^{\frac{p}{2}}}{\varepsilon_T^p} + c(T/k_T)^{\frac{1}{2}} \leq c\varepsilon_T + c \frac{(T/k_T)^{\frac{p}{2}}}{\varepsilon_T^p} + c(T/k_T)^{\frac{1}{2}}.$$

Jetzt wird A_2 betrachtet. Aus $|v_{j,k_T,T}| = 1$ folgt

$$|\theta_j - \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt| \geq |t - \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt|.$$

Man erhält

$$A_2 \leq \mathbb{P}[|\theta_j - \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt| > \varepsilon_T].$$

Wenn bezüglich dem Pfad von (Y_t) bedingt wird, gilt $\theta_j \stackrel{d}{=} Z \frac{k_T}{\mu_T T}$, wobei Z eine Poissonverteilte Zufallsvariable mit Parameter

$$\mu_T \int_{(j-1)T/k_T}^{jT/k_T} h(Y_t) dt$$

ist. Insgesamt ergibt sich die Abschätzung

$$A_2 \leq c \frac{k_T}{\mu_T T \varepsilon_T^2}.$$

Unter Verwendung von

$$\mathbb{E}[|T_1|] \leq \frac{1}{k_T} \sum_{j=1}^{k_T} \mathbb{E}[|v_{j,k_T,T}|]$$

erhält man schließlich

$$\mathbb{E}[|\sqrt{T}T_1|] \leq c\varepsilon_T \sqrt{T} + c \frac{T^{\frac{p+1}{2}}}{\varepsilon_T^p k_T^{\frac{p}{2}}} + c \frac{k_T}{\mu_T \sqrt{T} \varepsilon_T^2} + c \frac{T}{\sqrt{k_T}}.$$

Wird $\varepsilon_T = \frac{\zeta_T}{\sqrt{T}}$ gewählt, wobei $\zeta_T \rightarrow 0$, so folgt

$$\mathbb{E}[|\sqrt{T}T_1|] \leq c\zeta_T + c \frac{T^{p+\frac{1}{2}}}{\zeta_T^p k_T^{\frac{p}{2}}} + c \frac{k_T \sqrt{T}}{\mu_T \zeta_T^2} + c \frac{T}{\sqrt{k_T}}.$$

8 Beweise

Werden die Annahmen an k_T berücksichtigt, findet sich eine Folge ζ_T , die nach Null konvergiert, sodass $\mathbb{E}[|\sqrt{T}T_1|]$ ebenfalls gegen Null strebt. Wenden wir uns nun T_2 zu. Es gilt

$$\begin{aligned} T_2 &= \frac{1}{k_T} \sum_{j=1}^{k_T} \int_{(j-1)T/k_T}^{jT/k_T} (\mathbb{I}_{\{\frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du \leq t\}} - \mathbb{I}_{\{h(Y_s) \leq t\}}) ds \\ &= \frac{1}{T} \sum_{j=1}^{k_T} \int_{(j-1)T/k_T}^{jT/k_T} w_{j,k_T,s,T} ds \end{aligned}$$

mit

$$w_{j,k_T,s,T} = \begin{cases} 1 & , \text{ wenn } \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du \leq t < h(Y_s), \\ -1 & , \text{ wenn } h(Y_s) \leq t < \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du, \\ 0 & , \text{ sonst.} \end{cases}$$

Nun impliziert $|w_{j,k_T,s,T}| = 1$, dass

$$|h(Y_s) - \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du| \geq |t - \frac{k_T}{T} \int_{(j-1)T/k_T}^{jT/k_T} h(Y_u) du|.$$

Durch Verwendung der gleichen Berechnungsschritte wie für T_1 mit ε'_T , positiv und klein genug, und $p > 0$ ergibt sich

$$\mathbb{P}[|w_{j,k_T,s,T}| = 1] \leq c\varepsilon'_T + c \frac{(T/k_T)^{\frac{p}{2}}}{(\varepsilon'_T)^p} + c(T/k_T)^{\frac{1}{2}}.$$

Genau wie vorhin erhält man, dass $\mathbb{E}[|\sqrt{T}T_2|]$ gegen Null strebt. Schlussendlich, vgl. Gleichung (3.3), konvergiert $\sqrt{T}T_3$ in Verteilung gegen eine zentrierte Normalverteilte Zufallsvariable mit Varianz $\sigma^2 \text{Var}[Z_e(t)]$. \square

Um Proposition 8.1 zu beweisen, wird das folgende Lemma benötigt.

Lemma 8.3. Sei

$$\alpha_T(t) = \sqrt{T} \left(\frac{1}{T} \int_0^T \mathbb{I}_{\{h(Y_s) \leq t\}} ds - h^{-1}(t) \right).$$

Die Folge $(\alpha_T(t))_{0 \leq t < h(1^-)}$ ist dicht in $D[0, h(1^-))$.

Beweis. Zuerst sollte man sich das folgende klassische C-Dichtheits-Kriterium in Erinnerung rufen:² Wenn für $p > 0$, $p_1 > 0$ und für alle $0 \leq t_1, t_2 < h(1^-)$

$$\mathbb{E}[|\alpha_T(t_1) - \alpha_T(t_2)|^p] \leq c |t_1 - t_2|^{p_1}$$

²vgl. [Bil68], S. 128.

8 Beweise

gilt, dann ist die Folge $(\alpha_T(t))_{0 \leq t < h(1^-)}$ dicht in $D[0, h(1^-)]$. Nun wird die Folge von Stoppzeiten $(\nu_n)_{n \in \mathbb{N}}$, definiert in (3.2), verwendet. Sei

$$n_T = \inf\{i : \nu_i \geq T\}.$$

Wir wissen, dass

$$\mathbb{E}[\nu_{n_T}] \leq cT \tag{8.1}$$

und

$$\mathbb{E}\left[\left(\frac{n_T}{\sigma^2} - T\right)^2\right] \leq cT \tag{8.2}$$

gilt.³ Definiere

$$Y_i(t_1, t_2) = \frac{1}{\sqrt{T}} \left(\int_{\nu_{i-1}}^{\nu_i} \mathbb{I}_{\{t_1 < h(Y_t) \leq t_2\}} dt - \frac{1}{\sigma^2} (h^{-1}(t_2) - h^{-1}(t_1)) \right)$$

und

$$\tilde{Y}_{n_T+1}(t_1, t_2) = \frac{1}{\sqrt{T}} \int_T^{\nu_{n_T}} \mathbb{I}_{\{t_1 < h(Y_t) \leq t_2\}} dt.$$

Für $i \geq 2$ sind die $Y_i(t_1, t_2)$ zentriert und unabhängig identisch verteilt. Mit Hilfe der *Occupation Formula*⁴ und der Taylor-Entwicklung erhält man

$$\mathbb{E}[(Y_2(t_1, t_2))^2] \leq T^{-1} \mathbb{E}\left[\left(\int_{\nu_{i-1}}^{\nu_i} \mathbb{I}_{\{t_1 < h(Y_t) \leq t_2\}} dt\right)^2\right] \leq cT^{-1} |t_2 - t_1|^2 \mathbb{E}[(L^*)^2],$$

wobei

$$L^* = \sup_{u \in [-\frac{1}{\sigma}, \frac{1}{\sigma}]} (L_{-\frac{1}{\sigma}, \frac{1}{\sigma}}(u)).$$

Wir wissen aus der Ray-Knight Version der Burkholder-Davis-Gundy Ungleichung⁵, dass die *Polynomial Moments* von L^* endlich sind und damit

$$\mathbb{E}[(Y_2(t_1, t_2))^2] \leq cT^{-1} |t_2 - t_1|^2$$

gilt. Auf die selbe Weise zeigt man, dass

$$\mathbb{E}[(Y_1(t_1, t_2))^2] + \mathbb{E}[(\tilde{Y}_{n_T+1}(t_1, t_2))^2] \leq cT^{-1} |t_2 - t_1|^2.$$

³vgl. [Gut88], S. 54ff.

⁴vgl. [Ber03], S. 2.

⁵vgl. [RY91], S. 420f.

8 Beweise

Mit diesen Abschätzungen bestätigt man, dass das Dichtheits-Kriterium erfüllt ist. Es gilt

$$\alpha_T(t_1) - \alpha_T(t_2) = B_1 + B_2 - B_3 + B_4,$$

wobei

$$B_1 = Y_1(t_1, t_2), \quad B_2 = \sum_{i=2}^{n_T} Y_i(t_1, t_2), \quad B_3 = \tilde{Y}_{n_T+1}(t_1, t_2),$$

$$B_4 = \sqrt{T} \left(\frac{n_T - 1}{\sigma^2 T} - 1 \right) (h^{-1}(t_2) - h^{-1}(t_1)).$$

Daraus ergibt sich

$$\mathbb{E}[|\alpha_T(t_1) - \alpha_T(t_2)|^2] \leq c \sum_{i=1}^4 \mathbb{E}[|B_i|^2].$$

Es folgt⁶

$$\mathbb{E}\left[\left|\sum_{i=2}^{n_T} Y_i(t_1, t_2)\right|^2\right] \leq c \mathbb{E}[n_T] \mathbb{E}[(Y_2(t_1, t_2))^2].$$

Weiters, unter Verwendung von (8.1), gilt

$$\mathbb{E}[n_T] \mathbb{E}[|Y_2(t_1, t_2)|^2] \leq c |t_1 - t_2|^2.$$

Schließlich folgt aus Ungleichung (8.2)

$$\mathbb{E}[(B_4)^2] \leq c |t_1 - t_2|^2,$$

womit das Dichtheits-Kriterium erfüllt ist. □

8.3 Beweis von Satz 3.4 und Satz 3.5

Damit diese Sätze gezeigt werden können, wird ein Korollar von Proposition 8.1 benötigt.

Korollar 8.4. Wenn $T \rightarrow \infty$, dann konvergiert

$$\sqrt{T} \left(\hat{h}_e^{-1}(\cdot) - h^{-1}(\cdot), \hat{h}_e(\cdot) - h(\cdot), \int_0^1 \hat{h}_e(u) du - 1 \right)$$

in Verteilung gegen

$$\left(\sigma G(\cdot), -h'(\cdot) \sigma G(h(\cdot)), - \int_0^{h(1^-)} \sigma G(v) dv \right)$$

in $D[0, h(1^-)] \times D[0, 1] \times \mathbb{R}$ (bezüglich der Produkttopologie).

Beweis. Das folgt direkt aus Proposition 8.1 zusammen mit Theorem 13.7.2 aus [Whi02]⁷

⁶vgl. [Gut88], S. 20f.

⁷vgl. [Whi02], S. 450.

8 Beweise

und dem *Continuous Mapping Theorem*⁸. □

Nun kommen wir zum Beweis von Satz 3.4. Wir schreiben

$$\sqrt{T}(\hat{h}^{-1}(\cdot) - h^{-1}(\cdot)) = V_1 + V_2,$$

wobei

$$V_1 = \sqrt{T}(\hat{h}_e^{-1}(\cdot \frac{\hat{\mu}_T}{\mu_t}) - h^{-1}(\cdot \frac{\hat{\mu}_T}{\mu_T})),$$

$$V_2 = \sqrt{T}(\hat{h}^{-1}(\cdot \frac{\hat{\mu}_T}{\mu_t}) - h^{-1}(\cdot)).$$

Aus Proposition 8.1 und der Stetigkeit der *Composition Map*⁹ folgt

$$\sqrt{T}V_1 \xrightarrow{d} \sigma G(\cdot)$$

in $D[0, h(1^-))$. Wenn man jetzt beachtet, dass

$$\int_0^1 \hat{h}_e(u) du = \frac{\hat{\mu}_T}{\mu_T},$$

folgt mit Hilfe von Korollar 8.4 und Theorem 13.3.3 aus [Whi02]¹⁰

$$\sqrt{T}V_2 \xrightarrow{d} -\frac{(\cdot)}{h'(h^{-1}(\cdot))} \int_0^{h(1^-)} \sigma G(v) dv$$

in $D[0, h(1^-))$. Die Konvergenz von $\sqrt{T}V_1$ zusammen mit der von $\sqrt{T}V_2$ beendet den Beweis. □

Aus Satz 3.4 und Theorem 13.7.2 aus [Whi02] folgt Satz 3.5. □

8.4 Beweis von Korollar 3.6

Sei $\widehat{h}_e(Y_t)$ der Schätzer von $h(Y_t)$, der genauso definiert ist wie $\widehat{h}(Y_t)$, nur statt $\hat{\mu}_T$ wird μ_T verwendet. Nun benützt man, dass

$$\hat{h}^{-1}(\cdot \frac{\mu_T}{\hat{\mu}_T}) = \hat{h}_e^{-1}(\cdot)$$

⁸vgl. [Whi02], S. 85f.

⁹vgl. [Whi02], S. 430.

¹⁰vgl. [Whi02], S. 434f.

8 Beweise

und dass die Zyklen von $\widehat{h_e(Y_t)}$ bei der Schätzung von $h^{-1}(\cdot)$ und μ_T vernachlässigbar sind. Dann folgt aus Proposition 8.1

$$\left(\sqrt{T} \left(\widehat{h}^{-1} \left(\cdot \frac{\mu_T}{\widehat{\mu}_T} \right) - h^{-1}(\cdot) \right), \widehat{h_e(Y_t)} \right) \xrightarrow{d} (\sigma G(\cdot), h(Y_t))$$

in $D[0, h(1^-)) \times \mathbb{R}^+$, wobei G unabhängig von Y_t ist. Mit Hilfe von Theorem 13.2.1 aus [Whi02]¹¹ und

$$\widehat{h(Y_t)} = \frac{\mu_T}{\widehat{\mu}_T} \widehat{h_e(Y_t)}$$

erhält man

$$\sqrt{T} \left(\widehat{h}^{-1}(\widehat{h(Y_t)}) - h^{-1}(\widehat{h_e(Y_t)}) \right) \xrightarrow{d} \sigma G(h(Y_t)).$$

Nun erinnere man sich an

$$|h^{-1}(\widehat{h_e(Y_t)}) - h^{-1}(h_e(Y_t))| \leq c |\widehat{h_e(Y_t)} - h_e(Y_t)|$$

und dass bedingt auf den Pfad von (Y_t) gilt: $\widehat{h_e(Y_t)} = Z \frac{k_T}{\mu_T T}$, wobei Z ein Poisson-verteilte Zufallsvariable mit Parameter

$$\mu_T \int_{t-T/k_T}^t h(Y_u) du$$

ist. Somit erhält man unter Benutzung der Taylor-Entwicklung und wegen

$$\mathbb{P} \left[\sup_{u \in [t - \frac{T}{k_T}, t]} |Y_u - Y_t| \geq 1 \right] \leq c \left(\frac{T}{k_T} \right)^{\frac{1}{2}},$$

dass

$$\mathbb{E} \left[|\widehat{h_e(Y_t)} - h_e(Y_t)| \right]$$

kleiner ist als

$$c \mathbb{E} \left[\sup_{u \in [t - \frac{T}{k_T}, t]} |Y_u - Y_t| \right] + c \left(\frac{k_T}{\mu_T T} \right)^{\frac{1}{2}} \leq c \left(\frac{T}{k_T} \right)^{\frac{1}{2}} + c \left(\frac{k_T}{\mu_T T} \right)^{\frac{1}{2}}.$$

Letztendlich, da $\frac{T}{\sqrt{k_T}}$ und $\frac{k_T \sqrt{T}}{\mu_T}$ gegen Null streben, gilt

$$\sqrt{T} \mathbb{E} \left[|\widehat{h_e(Y_t)} - h_e(Y_t)| \right] \leq c \frac{T}{\sqrt{k_T}} + c \left(\frac{k_T}{\mu_T} \right)^{\frac{1}{2}} \rightarrow 0.$$

Womit der Beweis abgeschlossen ist. □

¹¹vgl. [Whi02], S. 429.

9 Programmcode

Vorrangiges Ziel dieses Kapitels ist, die Berechnungen der Diplomarbeit nachvollziehen zu können. Daher wird nachfolgend nur ein kleiner Teil der programmierten Funktionen abgebildet. Speziell durch die unterschiedlichen Wahlen der Folgen k_T und N_T in der Analyse ergab sich sehr viel Programmcode, der sich teilweise jedoch nur um ein paar Zeilen unterschieden hat. Des Weiteren werden die Programme zur Erstellung der Abbildungen nicht vorgestellt.

Implementierung des *Efficient Price*:

```
1 function [bestBid, bestAsk, value, N.T, EP] = EPAll(file, stock, T, ...
    tau, day, month, year)
2 %Schaetzung des Efficient Price.
3 %file ... Name der Struktur
4 %stock ... Name der Aktie ... zB.: AKBNK
5 %T ... [0,T] ... T in Sekunden
6 %tau ... der zu betrachtende Order Flow in Sekunden
7 %day ... Tag fuer den der Efficient Price berechnet werden soll.
8 %month ... Monat der Berechnung
9 %year ... Jahr der Berechnung
10
11 %Gibt die Spaltennummer der Aktie in der Struktur zurueck.
12 index = indexStock(file,stock);
13
14 %Die erste bzw. letzte relevante Zeile in der Struktur.
15 [firstRow lastRow] = findRows(file, index, day, month, year, T, tau);
16
17 %Berechnet den Order Flow.
18 value = getValue(file, index, firstRow, lastRow);
19 %Bestimmt die gesamte Anzahl an Limit Orders.
20 N.T = numberLOAll(file, index, firstRow, lastRow);
21 %Bestimmt Mue.
22 mu.T = N.T/tau;
```

9 Programmcode

```
23
24 %Bestimme Folge K.T.
25 K_T = tau;
26 if K_T < 1
27     K_T = 1;
28 end;
29
30 theta(1) = 0;
31 %Genaue Berechnung von Theta.
32 [jfirstRow jlastRow] = findRows(file, index, day, month, year, T-tau, 0);
33 N_jT = numberLOAll(file, index, jfirstRow, jlastRow);
34 for j=1:K_T
35     N_iT = N_jT;
36     [jfirstRow jlastRow] = findRows(file, index, day, month, year, ...
37         T-tau+j*tau/K_T, j*tau/K_T);
38     N_jT = numberLOAll(file, index, jfirstRow, jlastRow);
39     helptheta = K_T/N_T * (N_jT - N_iT);
40     theta(j) = helptheta;
41 end;
42
43 %Berechnung der Order Flow Response Function.
44 [hfirstRow hlastRow] = findRows(file, index, day, month, year, T-tau + ...
45     tau - tau/K_T, tau - tau/K_T);
46 help = numberLOAll(file, index, hfirstRow, hlastRow);
47 hYt = K_T * (N_T - help)/(mu_T*tau);
48
49 %Gebrochenen Anteil zum Zeitpunkt T berechnen.
50 sort(theta);
51 fractionalPart = numel(find(theta <= hYt))/K_T;
52
53 %Skalierungsfaktor auslesen.
54 factor = file(1,index).tick(lastRow,1);
55
56 %Im Modell wird von einer Tick Size von eins ausgegangen.
57 %Daher muss der gebrochene Anteil noch angepasst werden.
58 fractionalPart = fractionalPart * factor;
59
60 %Best Bid zum Zeitpunkt T.
61 bestBid = file(1,index).order(lastRow,15);
62
63 %Best Ask zum Zeitpunkt T.
64 bestAsk = file(1,index).order(lastRow,16);
65
```

9 Programmcode

```
64 %Efficient Price setzt sich aus Best Bid und dem gebrochenen Anteil
65 %zusammen.
66 EP = bestBid + fractionalPart;
67 end
```

Programm 9.1: *EPAll.m*

Die Berechnung von N_t wird folgendermaßen in Matlab implementiert:

```
1 function [N_t] = numberLOAll(file, index, firstRow, lastRow)
2 %Berechnet die Gesamtanzahl der Limit Orders zwischen den uebergebenen
3 %Zeilen.
4 %file ... Name der Struktur
5 %index ... Spaltennummer der Aktie
6 %firstRow ... erste relevante Datenzeile
7 %lastRow ... letzte relevante Datenzeile
8
9 %Gesamtanzahl der Limit Orders auf Null setzen.
10 N_t = 0;
11 for j=firstRow:lastRow
12     %Wenn die Order eine Limit Order ist steht in Spalte 26 eine eins.
13     %Bei einer Market Order steht eine Null.
14     N_t = N_t + file(1,index).order(j,26);
15 end;
16 end
```

Programm 9.2: *numberLOAll.m*

9 Programmcode

Die Hilfsfunktion *findRows* gibt die erste und die letzte Zeile im relevanten Datensatz zurück und sieht folgendermaßen aus:

```
1 function [firstRow lastRow] = findRows(file, index, day, month, year, ...
    T, tau)
2 %Berechnet den Order Flow.
3 %file ... Name der Struktur
4 %index ... Spaltennummer der Aktie
5 %day ... Tag fuer den der Efficient Price berechnet werden soll
6 %month ... Monat der Berechnung
7 %year ... Jahr der Berechnung
8 %T ... [0,T] ... T in Sekunden
9 %tau ... der zu betrachtende Order Flow in Sekunden
10
11 %In welcher Zeile steht T.
12 firstTime = file(1,index).order(1,10);
13 lastRow = find(file(1,index).order(:,10) == (T+firstTime) & ...
    file(1,index).order(:,1) == day & file(1,index).order(:,2) == ...
    month & file(1,index).order(:,3) == year,1,'last');
14
15 %Falls es keinen Eintrag mit Wert = T gibt wird mit der Zeile mit der ...
    letzten Order vor T weitergerechnet. Es wird trotzdem ein gleich ...
    langer Order Flow betrachtet. Tau bleibt unveraendert.
16 if (isempty(lastRow) == 1)
17     lastRow = find(file(1,index).order(:,10) < (T+firstTime) & ...
        file(1,index).order(:,1) == day & file(1,index).order(:,2) == ...
        month & file(1,index).order(:,3) == year,1,'last');
18 end;
19
20 %Startzeit
21 startTime = file(1,index).order(lastRow,10) - tau;
22
23 %Zeile finden, ab der der Order Flow betrachtet wird.
24 firstRow = find(file(1,index).order(:,10) == startTime & ...
    file(1,index).order(:,1) == day & file(1,index).order(:,2) == ...
    month & file(1,index).order(:,3) == year,1,'first');
25
26 %Falls es keinen Eintrag mit Wert = startwert gibt, wird die Zeile mit der
27 %naechsten Order gesucht.
28 if (isempty(firstRow) == 1)
29     firstRow = find(file(1,index).order(:,10) > startTime & ...
        file(1,index).order(:,1) == day & file(1,index).order(:,2) == ...
```

9 Programmcode

```
        month & file(1,index).order(:,3) == year,1, 'first');  
30 end;  
31 end
```

Programm 9.3: *findRows.m*

Die Hilfsfunktion *indexStock* gibt die Spaltennummer der Aktie in der Struktur zurück:

```
1 function index = indexStock(file,stock)  
2 %Gibt die Spaltennummer der Aktie in der Struktur zurueck.  
3 %file ... Name der Struktur  
4 %stock ... Name der Aktie  
5  
6 %Wird die Aktie nicht gefunden, wird die Erste genommen.  
7 index = 1;  
8 for i=1:length(file)  
9     if stock == file(1,i).ticker(1,:)   
10         index = i;  
11     end;  
12 end;  
13 end
```

Programm 9.4: *indexStock.m*

9 Programmcode

Die Hilfsfunktion *getValue* berechnet den *Order Flow* im Sinne von:

- Der Rückgabewert steigt, wenn die Order eine *Sell Limit Order* ist.
- Der Rückgabewert fällt, wenn die Order eine *Buy Limit Order* ist.

Dieser Wert ist für die Berechnung des *effizienten Preises* nicht relevant. Er dient rein der Veranschaulichung und zeigt an, ob der *Order Flow* in Richtung *Best Bid* oder *Best Ask* Level tendiert.

```
1 function [value] = getValue(file, index, firstRow, lastRow)
2 %Berechnet den Order Flow
3 %file ... Name der Struktur
4 %index ... Spaltennummer der Aktie
5 %firstRow ... erste relevante Datenzeile
6 %lastRow ... letzte relevante Datenzeile
7
8 value = 0;
9 for j=firstRow:lastRow
10     %Wenn es eine Sell Limit Order ist, soll value um 1 steigen.
11     %Wenn es eine Buy Limit Order ist, soll value um 1 fallen.
12     value = value + file(1,index).order(j,26) * ...
13         (-1)^file(1,index).order(j,27);
14 end;
```

Programm 9.5: *getValue.m*

9 Programmcode

Nachfolgend Funktionen, die für die Kalkulationen im Kapitel 6 verwendet wurden. Das dort verwendete Programm *deltaDay.m* berechnet die Δ_{mid} für einen Tag:

```
1 function [delta] = deltaDay(file, stock, day)
2 %Berechnet die Delta fuer einen Tag.
3 %file ... Name der Struktur
4 %stock ... Name der Aktie
5 %day ... Tag fuer den die Delta berechnet werden sollen
6
7 index = indexStock(file, stock);
8 counter = 1;
9 delta(counter) = 0;
10 month = 6;
11 year = 2008;
12
13 [firstRow lastRow] = findRows(file, index, day, month, year, 26999, 26999);
14
15 for j=firstRow:lastRow
16     if file(1, index).order(j, 26) == 1 && ...
17         isnan(file(1, index).order(j, 25)) == 0
18         if isnan(file(1, index).order(j, 15)) == 0 && ...
19             isnan(file(1, index).order(j, 16)) == 0
20             delta(counter) = abs(file(1, index).order(j, 8) - ...
21                 (file(1, index).order(j, 16)+file(1, index).order(j, 15))/2);
22             delta(counter) = delta(counter) / file(1, index).tick(j, 1);
23             counter = counter + 1;
24         end
25     end;
26 end;
27 end
```

Programm 9.6: *deltaDay.m*

9 Programmcode

Das Programm *deltaDayHist.m* berechnet die Δ_{mid} für einen Tag und zeichnet die Verteilung:

```
1 function delta = deltaDayHist(file, stock, day)
2 %Berechnet die Delta fuer einen Tag und zeichnet die Verteilung.
3 %file ... Name der Struktur
4 %stock ... Name der Aktie
5 %day ... Tag fuer den die Delta berechnet werden sollen
6 delta = deltaDay(file,stock,day);
7 hist(delta,unique(delta));
8 h = findobj(gca,'Type','patch');
9 set(h,'FaceColor','b','EdgeColor','w')
10 end
```

Programm 9.7: *deltaDayHist.m*

Das Programm *deltaMonthHist.m* berechnet die Δ_{mid} für einen Monat und zeichnet die Verteilung:

```
1 function delta = deltaMonthHist(file, stock)
2 %Berechnet die Delta fuer einen Monat und zeichnet die Verteilung.
3 %file ... Name der Struktur
4 %stock ... Name der Aktie
5 index = indexStock(file,stock);
6 delta = [];
7 for j=1:31
8     test = find(file(1,index).order(:,1) == j);
9     if (isempty(test) == 0)
10         delta = [delta deltaDay(file,stock,j)];
11     end;
12 end;
13 figure;
14 hist(delta,unique(delta));
15 h = findobj(gca,'Type','patch');
16 set(h,'FaceColor','b','EdgeColor','w')
17 t = get(gca,'ytick');
18 set(gca,'yticklabel',num2str(t));
19 end
```

Programm 9.8: *deltaDayHist.m*

9 Programmcode

Das Programm *deltaMean.m* berechnet den Mittelwert der Δ_{mid} eines Tages für ein Monat:

```
1 function [day, delta_mean_day] = deltaMean(file, stock)
2 %Berechnet den Mittelwert ueber die Delta eines Tages fuer ein Monat.
3 %file ... Name der Struktur
4 %stock ... Name der Aktie
5
6 index = indexStock(file, stock);
7
8 counter = 1;
9 delta_mean_day(counter) = 0;
10 day(counter) = 0;
11
12 for j=1:31
13     test = find(file(1,index).order(:,1) == j);
14     if (isempty(test) == 0)
15         day(counter) = j;
16         delta_mean_day(counter) = mean(deltaDay(file, stock, j));
17         counter = counter + 1;
18     end;
19 end;
20 end
```

Programm 9.9: *deltaMean.m*

Literaturverzeichnis

- [Asm03] ASMUSSEN, S.: *Applied Probability and Queues*. Springer, New York, 2. Auflage, 2003.
- [ASMZ11] AÏT-SAHALIA, Y., P. A. MYKLAND und L. ZHANG: *Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise*. *Journal of Econometrics*, 160:160–175, 2011.
- [Ber03] BERESTYCKI, N.: *Introduction to the Theory of Brownian Local Times*, 2003.
- [Bil68] BILLINGSLEY, P.: *Convergence of Probability Measures*. Wiley, New York, 1968.
- [BNHLS08] BARNDORFF-NIELSEN, O. E, P. R. HANSEN, A. LUNDE und N. SHEPHARD: *Designing Realized Kernels to Measure the Ex Post Variation of Equity Prices in the Presence of Noise*. *Econometrica*, 76:1481–1536, 2008.
- [BS96] BORODIN, A. N. und P. SALMINEN: *Handbook of Brownian Motion - Facts and Formulae*. Birkhäuser, 1996.
- [BZ05] BRZEŃNIAK, Z. und T. ZASTAWNIAK: *Basic Stochastic Processes*. Springer, London, 7. Auflage, 2005.
- [DRR13] DELATTRE, S., C. Y. ROBERT und M. ROSENBAUM: *Estimating the Efficient Price from the Order Flow: a Brownian Cox Process Approach*. *Stochastic Processes and their Applications*, 123:2603–2619, 2013.
- [Ger13] GERHOLD, S.: *Finanzmathematik 2: zeitstetige Modelle*. Vorlesungsmitschrift, 2013.
- [Gly13] GLYNN, P. W.: *Stochastic Systems - Section 6: Harris Recurrence*. Vorlesungsskript, Stanford, June 2013.
- [GO10] GATHERAL, J. und R. C. A. OOMEN: *Zero-intelligence Realized Variance Estimation*. *Finance and Stochastics*, 14:249–283, 2010.

Literaturverzeichnis

- [Gok08] GOKHMAN, D.: *Topologies for Hybrid Solutions*. Nonlinear Analysis: Hybrid Systems, Seiten 468–473, 2008.
- [GPW⁺13] GOULD, M. D., M. A. PORTER, S. WILLIAMS, M. McDONALD, D. J. FENN und S. D. HOWISON: *Limit Order Books*, 2013.
- [Gur11] GURKER, W.: *Angewandte Statistik*. Vorlesungsskript, 2011.
- [Gut88] GUT, A.: *Stopped Random Walks. Limit Theorems and Applications*. Springer, New York, 1988.
- [HL06] HANSEN, P. R. und A. LUNDE: *Realized Variance and Market Microstructure Noise*. Journal of Business and Economic Statistics, 24:127–161, 2006.
- [Hub12a] HUBALEK, F.: *Risiko- und Ruintheorie*. Vorlesungsmitschrift, 2012.
- [Hub12b] HUBALEK, F.: *Stochastische Analysis für FVM 1*. Vorlesungsmitschrift, 2012.
- [Lan98] LANDO, D.: *On Cox Processes and Credit Risky Securities*. Review of Derivatives Research, 2:99–120, 1998.
- [Øk00] ØKSENDAL, B.: *Stochastic Differential Equations*. Springer, Heidelberg, 5. Auflage, 2000.
- [RY91] REVUZ, D. und M. YOR: *Continuous Martingales and Brownian Motion*. Springer, Berlin, 1991.
- [VZFR13] VALENZUELA, M., I. ZER, P. FRYZLEWICZ und T. RHEINLÄNDER: *Relative Liquidity and Future Volatility*, 2013.
- [Whi02] WHITT, W.: *Stochastic-Process Limits*. Springer, New York, 2002.
- [Wil10] WILLIAMS, D.: *Probability with Martingales*. Cambridge University Press, Cambridge, 12. Auflage, 2010.
- [Zha06] ZHANG, L.: *Efficient Estimation of Stochastic Volatility using Noisy Observations: a Multi-Scale Approach*. Bernoulli, 12:1019–1043, 2006.
- [Zho96] ZHOU, B.: *High-Frequency Data and Volatility in Foreign-Exchange Rates*. Journal of Business and Economic Statistics, 14:45–52, 1996.

Literaturverzeichnis

- [ZMAS05] ZHANG, L., P. A. MYKLAND und Y. AÏT-SAHALIA: *A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data*. Journal of the American Statistical Association, 100:1394–1411, 2005.