

# Einige Aspekte zur Parameterschätzung in einem verallgemeinerten Ornstein-Uhlenbeck Modell

ROMY KRÄMER\*, NICO REINHOLD<sup>†</sup> und MATTHIAS RICHTER<sup>§</sup>

20. Februar 2007

Fakultät für Mathematik, Technische Universität Chemnitz,  
09107 Chemnitz

## Zusammenfassung

In der vorliegenden Arbeit werden einige Aspekte der Bestimmung gewisser reellwertiger Parameter eines Modells zur Beschreibung des Aktienpreises behandelt. Das bivariate Ornstein-Uhlenbeck Modell zeichnet sich durch die Möglichkeit aus, verschiedene Abhängigkeitsstrukturen der Returns abbilden zu können. Eine adäquate Modellidentifikation ist unumgänglich. Hier wird nun davon ausgegangen, dass eine rein zeitabhängige Volatilitätsfunktion mit geeigneten Methoden identifiziert wurde. Die restlichen (reellwertigen) Parameter werden mit geeigneten Techniken identifiziert, wobei zusätzlich zu bedenken ist, daß der Preisprozeß durch eine nichtbeobachtbare Zufallskomponente beeinflusst wird, was den Einsatz entsprechender Filtermethoden erforderlich macht. Umfangreiche numerische Fallstudien schließen die Untersuchungen ab.

**MSC2000 Klassifikation:** 60G15, 62M05, 91B70, 62M10

**Schlagworte:** verallgemeinertes bivariates Ornstein-Uhlenbeck Modell, Maximum-Likelihood Schätzung, Kalman Filter

---

\*Email: [rokra@mathematik.tu-chemnitz.de](mailto:rokra@mathematik.tu-chemnitz.de)

†Email: [nico.reinhold@s2002.tu-chemnitz.de](mailto:nico.reinhold@s2002.tu-chemnitz.de)

§Email: [richterm@mathematik.tu-chemnitz.de](mailto:richterm@mathematik.tu-chemnitz.de)

# 1 Einleitung

Um das Verhalten von Aktienpreisen zu beschreiben, wurden in den letzten Jahrzehnten viele verschiedene Modelle entwickelt. Eines davon ist das in [7] eingeführte bivariate Ornstein-Uhlenbeck Modell von Lo und Wang. Es bildet die Grundlage des hier untersuchten Modells.

In einem Zeitintervall  $[0, T]$  betrachten wir den Preisprozeß  $P_t$  einer Aktie und bezeichnen mit  $p_t$  den Logarithmus des Aktienpreisprozeß,  $p_t = \ln P_t$ . Das in [7] von Lo und Wang eingeführte Bivariate Ornstein-Uhlenbeck Modell unterstellt, daß der logarithmierte Aktienpreisprozeß  $p_t$  einen linearen deterministischen Trend  $\mu t$  besitzt. Deshalb ist es bequem den Prozeß

$$q_t := p_t - \mu t, \quad (1.1)$$

einzuführen, welcher im folgenden mit *enttrenderter logarithmierter Preisprozess* bezeichnet werden soll, da  $q_t$  keine deterministische Trendkomponente besitzt.

Das Bivariate Ornstein-Uhlenbeck Modell von Lo und Wang nimmt nun an, daß  $q_t$  der folgenden stochastischen Differentialgleichung

$$\begin{aligned} dq_t &= -(\gamma q_t - \lambda X_t) dt + \sigma dW_t^q \\ dX_t &= -\beta X_t dt + \sigma_X dW_t^X \end{aligned} \quad (1.2)$$

mit den Anfangswertbedingungen  $q_0 = c_q$ ,  $X_0 = c_X$  genügt. Dabei bezeichnen  $\gamma \geq 0$ ,  $\lambda \geq 0$ ,  $\beta \geq 0$ ,  $\sigma > 0$ ,  $\mu \in \mathbb{R}$  und  $\sigma_X > 0$  reellwertige Parameter und  $X$  einen unterliegenden stochastischen Prozeß, der den Zustand des Marktes beschreibt und somit die stochastische Drift von  $q$  beeinflusst, aber im Allgemeinen nicht beobachtbar ist. Weiterhin bezeichnen  $W^q$  und  $W^X$  korrelierte Wienerprozesse mit Korrelationskoeffizient  $\kappa$ , d. h.  $\mathbb{E}(W_t^q W_t^X) = \kappa t$ .

Um unsere Betrachtungen etwas zu vereinfachen, werden wir diese Modellklasse zunächst etwas einschränken, indem wir annehmen, daß die Wienerprozesse  $W^q$  und  $W^X$  unabhängig sind, d. h.  $\kappa = 0$ . Andererseits wollen wir das Modell etwas verallgemeinern indem wir eine zeitabhängige (aber deterministische) Volatilitätsfunktion zulassen. Diese Erweiterung ermöglicht es uns, mehr Effekte in der risikoneutralen Welt zu beschreiben.

Zusammenfassend lässt sich das in diesem Artikel untersuchte Modell wie folgt beschreiben. Im Weiteren werden wir es als verallgemeinertes Bivariates Ornstein-Uhlenbeck Modell bezeichnen.

**Modell 1.1** Wir nehmen an, daß der enttendete Log-Preisprozeß  $q_t := p_t - \mu t$  eines handelbaren Wertpapiers/Aktie der folgenden stochastischen Differentialgleichung genügt

$$\begin{aligned} dq_t &= -(\gamma q_t - \lambda X_t) dt + \sigma_t dW_t^q \\ dX_t &= -\beta X_t dt + \sigma_X dW_t^X. \end{aligned} \quad (1.3)$$

Dabei bezeichnen  $\gamma$ ,  $\lambda$ ,  $\beta \geq 0$ ,  $\sigma_X > 0$  sowie  $\mu$  reellwertige Parameter und  $\sigma : [0, T] \rightarrow \mathbb{R}$  eine zeitabhängige Volatilitätsfunktion, welche der Bedingung  $\sigma(t) > 0$ ,  $0 \leq t \leq T$  genüge. Die Anfangswerte  $q_0 = c_q$  und  $X_0 = c_X$  werden als stochastische Zufallsgrößen mit endlichen zweiten Momenten modelliert, d. h.

$$\mathbb{E} c_q^2 < \infty \quad \text{und} \quad \mathbb{E} c_X^2 < \infty.$$

Außerdem nehmen wir an, daß der Vektor  $(c_q \ c_X)^T$ , der die Anfangswerte der Prozesse  $q$  und  $X$  enthält, und die Wienerprozesse  $W^q$ ,  $W^X$  vollständig stochastisch unabhängig sind.

Es bestünde noch die Möglichkeit, den Wertpapierpreis mit einer Konstante  $D$  zu multiplizieren. Hierzu sei auf [5] verwiesen. In dieser Arbeit soll auf diese Skalierung verzichtet werden. Wir beschäftigen uns im Weiteren mit der Bestimmung der in dem Modell vorkommenden unbekannt Parameter  $\gamma, \lambda, \beta, \sigma_X$  und  $\mu$ . Beobachten können wir dazu Aktienpreise  $P_{t_k}$  bzw.  $p_{t_k}$  an diskreten Zeitpunkten

$$t_k = k\tau \quad k = 0, 1, \dots, N \quad (1.4)$$

in einem Zeitintervall  $[0, T]$  mit  $T = N\tau$ . Weiterhin setzen wir voraus, daß die Volatilitätsfunktion  $\sigma$  oder eine akzeptable Näherung  $a^\delta$  für die quadrierte Volatilitätsfunktion  $a(t) = \sigma^2(t)$  bekannt ist.

Aus diesen Daten versuchen wir mittels der Maximum-Likelihood-Methode die unbekannt Parameter zu schätzen. Wir verwenden dazu die sogenannte *Prognosefehlerzerlegung* (*prediction error decomposition form*) der Likelihood-Funktion. Diese erlaubt es uns, die Werte der Likelihood-Funktion für gegebene Parameter mit Hilfe des Kalman-Filters zu berechnen. Das bei der Schätzung der Parameter zu lösende Maximierungsproblem können wir ohne Schwierigkeiten in ein Minimierungsproblem überführen und mittels allgemeiner numerischer Optimierungsalgorithmen, z. B. in MATLAB, lösen.

In Abschnitt 4 testen wir die Verfahren anhand von numerischen Fallstudien mit simulierten Aktienpreisdaten. Dabei verdienen folgende Fragen besondere Aufmerksamkeit:

- Wie häufig und über welchen Zeitraum müssen die Aktienpreisdaten beobachtet werden, um mit der Maximum-Likelihood-Methode akzeptable Näherungen erwarten zu können?
- Gibt es Parameterkonstellationen, bei denen das Verfahren besonders gut oder schlecht funktioniert?
- Wie sensibel reagiert das Verfahren auf leichte Fehler in den Preisdaten und in der Volatilitätsfunktion  $\sigma(t)$ ?

## 2 Einführende Überlegungen

### 2.1 Uneindeutigkeitseffekte

Wir werden in diesem Abschnitt sehen, daß die Beobachtung des Aktienpreisprozesses  $P$  beziehungsweise des logarithmierter Preisprozesses  $p_t$  nicht ausreicht um alle fünf Parameter  $\gamma, \lambda, \beta, \sigma_X, \mu$  eindeutig zu schätzen. Aus diesem Grund können und werden wir die Menge der zulässigen Parameterwerte weiter einschränken. Diese Einschränkung des Parameterbereichs führt dabei nicht zu einer Einschränkung der modellierbaren Verteilungsfunktionen für  $p$  bzw.  $P$ .

Eine relativ offensichtliche Einschränkung des Parameterbereichs ergibt sich aus der Beobachtung, daß wir statt dem (nicht beobachtbaren) Prozeß  $X$  auch eine skalierte Variante  $\hat{X}_t := \frac{1}{\sigma_X} X_t$  betrachten können. (Dabei ist zu bemerken, daß in Modell 1.1  $\sigma_X > 0$  vorausgesetzt wurde.) Der Prozeß  $\hat{X}$  genügt dann der stochastischen Differentialgleichung

$$d\hat{X}_t = -\beta\hat{X}_t dt + 1 dW_t^X$$

mit der Anfangsbedingung  $\hat{X}_0 = \frac{c_X}{\sigma_X}$ .

Weiterhin ist leicht zu sehen, daß die Prozesse  $q_t$  und  $X_t$  genau dann dem System (1.3) genügen, wenn  $q_t$  und  $\hat{X}_t$  das System

$$\begin{aligned} dq_t &= -(\gamma q_t - \hat{\lambda} \hat{X}_t) dt + \sigma(t) dW_t^q \\ d\hat{X}_t &= -\beta \hat{X}_t dt + 1 dW_t^X \end{aligned} \quad (2.1)$$

mit den Anfangsbedingungen  $q_0 = c_q$ ,  $\hat{X}_0 = \frac{c_X}{\sigma_X}$  und dem Parameter  $\hat{\lambda} = \lambda \sigma_X$  erfüllen. Aus diesem Grund können wir im Weiteren stets annehmen, daß  $\sigma_X = 1$  gilt. Damit müssen lediglich vier Parameter  $\gamma, \lambda, \beta, \mu$  geschätzt werden.

Für die weiteren Betrachtungen wird es häufig sinnvoll sein, die Prozesse  $q$  und  $X$  in einem vektorwertigen Prozeß  $\underline{\alpha}$  zusammenzufassen, d. h. wir setzen  $\underline{\alpha}_t := \begin{pmatrix} q_t \\ \hat{X}_t \end{pmatrix}$ . Faßt man weiterhin die unabhängigen Wienerprozesse  $W^q$  und  $W^X$  zu einem zweidimensionalen Wienerprozeß  $W_t := \begin{pmatrix} W_t^q \\ W_t^X \end{pmatrix}$  zusammen, so kann man die Gleichungen (1.3) auch in der Form

$$d\underline{\alpha}_t = \begin{pmatrix} -\gamma & \lambda \\ 0 & -\beta \end{pmatrix} \underline{\alpha}_t dt + \begin{pmatrix} \sigma(t) & 0 \\ 0 & 1 \end{pmatrix} dW_t \quad (2.2)$$

mit der Anfangsbedingung  $\underline{\alpha}_0 = \underline{c}_\alpha := (c_q \ c_X)^T$  formulieren.

Dies ist eine (im engeren Sinne) lineare stochastische Differentialgleichung, welche die eindeutige Lösung

$$\underline{\alpha}_t = e^{tA} \underline{c}_\alpha + \int_0^t e^{(t-s)A} B(s) dW_s, \quad (2.3)$$

besitzt (siehe [1, S. 128 ff]). Hierbei wurde die Notation

$$A := \begin{pmatrix} -\gamma & \lambda \\ 0 & -\beta \end{pmatrix} \quad B(t) := \begin{pmatrix} \sigma(t) & 0 \\ 0 & 1 \end{pmatrix}.$$

eingeführt.

Zur Berechnung der Matrix  $e^{tA}$  ist eine Fallunterscheidung nötig, siehe [5]. In dem Fall  $\beta \neq \gamma$  ergibt sich

$$e^{tA} = \begin{pmatrix} e^{-\gamma t} & \frac{\lambda}{\gamma - \beta} (e^{-\beta t} - e^{-\gamma t}) \\ 0 & e^{-\beta t} \end{pmatrix}.$$

In dem Fall  $\beta = \gamma$  erhält man hingegen

$$e^{tA} = \begin{pmatrix} e^{-\gamma t} & \lambda t e^{-\gamma t} \\ 0 & e^{-\gamma t} \end{pmatrix}.$$

Damit kann der Prozeß  $q_t$  wie folgt geschrieben werden

$$\begin{aligned} q_t &= e^{-\gamma t} c_q + \frac{\lambda}{\gamma - \beta} (e^{-\beta t} - e^{-\gamma t}) c_X + \frac{\lambda}{\gamma - \beta} \int_0^t [e^{-\beta(t-s)} - e^{-\gamma(t-s)}] dW_s^X \\ &\quad + \int_0^t \sigma(s) e^{-\gamma(t-s)} dW_s^q \end{aligned} \quad (2.4)$$

für den Fall  $\gamma \neq \beta$  und als

$$q_t = e^{-\gamma t} c_q + \lambda t e^{-\gamma t} c_X + \lambda \int_0^t (t-s) e^{-\gamma(t-s)} dW_s^X + \int_0^t \sigma(s) e^{-\gamma(t-s)} dW_s^q$$

für den Fall  $\gamma = \beta$ .

In einem ersten Ansatz unserer Untersuchungen hatten wir den Fall einer konstanten Volatilitätsfunktion betrachtet und versucht, die Parameter  $\gamma$ ,  $\lambda$ ,  $\beta$  aus Momenten der logarithmischen Returns  $r_\tau(t) := \frac{\log P_t}{\log P_{t-\tau}} = p_t - p_{t-\tau}$  unter der Annahme  $\gamma \neq \beta$  zu schätzen. Dabei stellten wir fest, daß für den Fall einer konstanten Volatilitätsfunktion eine weitere Parameteruneindeutigkeit auftritt.

Um diese Überlegungen darzustellen, bemerken wir zunächst, daß für eine konstante Volatilitätsfunktion  $\sigma$  und normalverteilte Anfangswerte

$$\underline{c}_\alpha = \begin{pmatrix} c_q \\ c_X \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{2\gamma} + \frac{\lambda^2}{2\beta\gamma(\beta+\gamma)} & \frac{1}{\beta+\gamma} \frac{\lambda}{2\beta} \\ \frac{1}{\beta+\gamma} \frac{\lambda}{2\beta} & \frac{1}{2\beta} \end{pmatrix} \right) \quad (2.5)$$

der Prozeß  $\underline{q}$  (und damit auch  $q$ ) stationär ist. In diesem Fall können die Varianz und Kovarianzen der Returns  $r_\tau(t)$  in der Form

$$\mathbf{D}^2 r_\tau(t) = \frac{\lambda^2}{\beta - \gamma} \frac{\beta - \gamma + \gamma e^{-\beta\tau} - \beta e^{-\gamma\tau}}{\beta(\beta + \gamma)\gamma} + \frac{\sigma^2(1 - e^{-\gamma\tau})}{\gamma} \quad (2.6a)$$

und

$$\begin{aligned} \mathbf{Cov}(r_\tau(t), r_\tau(t + k\tau)) &= \frac{\lambda^2}{\gamma - \beta} \frac{(1 - e^{-\gamma\tau})^2 \beta e^{-\gamma(k-1)\tau} - (1 - e^{-\beta\tau})^2 \gamma e^{-\beta(k-1)\tau}}{2\beta(\beta + \gamma)\gamma} \\ &\quad - \frac{1}{2} \frac{\sigma^2(1 - e^{-\gamma\tau})^2}{\gamma} e^{-\gamma(k-1)\tau}. \end{aligned} \quad (2.6b)$$

geschrieben werden, vorausgesetzt  $\gamma \neq 0$ ,  $\beta \neq 0$ ,  $\gamma \neq \beta$  und  $k \geq 1$ . Wir bemerken, daß die Varianzen und Kovarianzen für die Grenzfälle  $\gamma = 0$ ,  $\beta = 0$  bzw.  $\gamma = \beta$  einfach durch entsprechende Grenzwertbildung in (2.6) erhalten werden können. Um jedoch die Notation zu vereinfachen und weitere Fallunterscheidungen zu vermeiden, werden wir sie im Folgenden ausschließen.

Führt man nun Hilfsvariablen

$$a := \left( \sigma^2 - \frac{\lambda^2}{\gamma^2 - \beta^2} \right) \frac{1 - e^{-\gamma\tau}}{\gamma} \quad c := \frac{\lambda^2}{\gamma^2 - \beta^2} \frac{1 - e^{-\beta\tau}}{\beta} \quad (2.7a)$$

$$g := e^{-\gamma\tau} \quad b := e^{-\beta\tau} \quad (2.7b)$$

ein, so lassen sich die Varianz und Kovarianz als

$$\mathbf{D}^2 r_\tau(t) = a + c \quad (2.8a)$$

$$\mathbf{Cov}(r_\tau(t), r_\tau(t + k\tau)) = \frac{1}{2} (a(g-1)g^{k-1} + c(b-1)b^{k-1}) \quad \text{für } k \geq 1 \quad (2.8b)$$

schreiben.

Aus der letzten Darstellung sieht man sofort, daß die Werte  $a$ ,  $c$ ,  $b$  und  $g$  nicht eindeutig durch die Momente der Returns bestimmt werden können. Genauer: Falls  $(a_1, c_1, b_1, g_1)$  eine Lösung des Systems (2.8) ist, so ist auch  $(a_2, c_2, b_2, g_2)$  mit

$$a_2 = c_1, c_2 = a_1, b_2 = g_1 \quad \text{und} \quad g_2 = b_1 \quad (2.9)$$

eine Lösung.

Weiterhin sind die Gleichungen (2.7b) für gegebene Werte  $b > 0$  und  $g > 0$  eindeutig nach  $\beta$  und  $\gamma$  auflösbar ( $\beta = -\frac{\ln b}{\tau}$ ,  $\gamma = -\frac{\ln g}{\tau}$ ) und durch Einsetzen dieser Werte in (2.7a) erhalten wir die eindeutigen Lösungen

$$\lambda = \sqrt{c(\gamma^2 - \beta^2) \frac{\beta}{1 - e^{-\beta\tau}}} \quad \text{und} \quad \sigma = \sqrt{\frac{\gamma}{1 - e^{-\gamma\tau}} a + \frac{\beta}{1 - e^{-\beta\tau}} c}$$

für  $\lambda$  und  $\sigma$ , vorausgesetzt die Terme unter den Wurzeln sind nichtnegativ.

Kombinieren wir diese Überlegungen, so sehen wir, daß es zu jeder Parametersituation

$$\mathbf{Par1} : \gamma_1 > 0, \beta_1 > 0, \lambda_1 \geq 0, \sigma_1 > 0 \text{ mit } \gamma \neq \beta \text{ sowie } \lambda_1^2 - \sigma_1^2(\gamma_1^2 - \beta_1^2) \geq 0 \quad (2.10a)$$

eine zweite Parametersituation

$$\mathbf{Par2} : \gamma_2 = \beta_1, \beta_2 = \gamma_1, \lambda_2 = \sqrt{\lambda_1^2 - \sigma_1^2(\gamma_1^2 - \beta_1^2)}, \sigma_2 = \sigma_1 \quad (2.10b)$$

gibt, für die die zugehörigen Momente der Returns (2.6) gleich sind, vorausgesetzt die Anfangswerte des Prozesses  $\underline{q}$  sind durch (2.5) gegeben, wobei die Parameter  $\gamma$  etc. jeweils durch  $\gamma_1$  bzw.  $\gamma_2$  ersetzt werden müssen.

Ein ganz konkretes Beispiel für dieses Phänomen ist (Zeiteinheit Tage)

$$\mathbf{Par1}^* : \gamma_1 = 0.3748, \beta_1 = 0.0106, \lambda_1 = 0.004797, \sigma_1 = 0.0074$$

$$\mathbf{Par2}^* : \gamma_2 = 0.0106, \beta_2 = 0.3748, \lambda_2 = 0.003914709, \sigma_2 = 0.0074.$$

Die Parametersituation 1 wurde in [7, S. 107] aus den täglichen Returns eines Marktindex geschätzt. Wir werden diese Parameter später in unseren numerischen Fallstudien benutzen.

Jetzt stellt sich natürlich noch die Frage, ob die Prozesse  $q^1$  und  $q^2$ , welche durch Einsetzen der Parametersituationen  $\mathbf{Par1}^*$  bzw.  $\mathbf{Par2}^*$  in die Gleichung (2.4) und die Anfangswerte (2.5) entstehen, stochastisch identisch (in dem Sinne, dass sie die gleichen endlich-dimensionalen Verteilungen besitzen) sind. Da die Lösung (2.4) unter den getroffenen Annahmen ein Gaußprozeß ist, müssen wir zur Beantwortung dieser Frage lediglich die folgenden zwei Bedingungen überprüfen

$$(a) \mathbb{E}q_t^1 = \mathbb{E}q_t^2 \quad \text{für alle } t \geq 0$$

$$(b) \mathbb{E}q_t^1 q_s^2 = \mathbb{E}q_t^2 q_s^1 \quad \text{für alle } s, t \geq 0.$$

Da der Erwartungswert des durch (2.4) und (2.5) gegebenen Prozeß  $q_t$  stets gleich Null ist, ist die erste Bedingung offensichtlich erfüllt. Um die zweite Bedingung zu überprüfen, benötigen wir die Kovarianzfunktion von  $q$ . Da die Anfangsbedingungen  $(c_q c_X)^T$  und die

Wienerprozesse  $W^q$  und  $W^X$  unabhängig in der Gesamtheit sind, gilt für  $t \geq s$

$$\begin{aligned}
\mathbb{E} q_t q_s &:= \sigma^2 \int_0^s e^{-\gamma(t-v)} e^{-\gamma(s-v)} dv \\
&+ \frac{\lambda^2}{(\beta - \gamma)^2} \int_0^s (e^{-\gamma(t-v)} - e^{-\beta(t-v)}) (e^{-\gamma(s-v)} - e^{-\beta(s-v)}) dv \\
&+ e^{-\gamma(t+s)} \mathbf{D}^2 c_q + \frac{\lambda^2 (e^{-\gamma t} - e^{-\beta t}) (e^{-\gamma s} - e^{-\beta s})}{(\beta - \gamma)^2} \mathbf{D}^2 c_X \\
&+ \frac{\lambda (e^{-\gamma t} (e^{-\gamma s} - e^{-\beta s}) + e^{-\gamma s} (e^{-\gamma t} - e^{-\beta t}))}{\beta - \gamma} \mathbf{Cov}(c_q, c_X).
\end{aligned} \tag{2.11}$$

Einsetzen der Anfangsbedingung und Berechnung der Integrale ergibt mit der Bezeichnung  $w := t - s$  für  $w > 0$

$$\begin{aligned}
\mathbb{E} q_t q_s &:= \frac{\sigma^2 \beta e^{-\gamma w} \gamma^2 + \gamma \lambda^2 e^{-\beta w} - \sigma^2 \beta^3 e^{-\gamma w} - \lambda^2 e^{-\gamma w} \beta}{2\beta \gamma (\gamma^2 - \beta^2)} \\
&= \frac{\sigma^2 e^{-\gamma w}}{2\gamma} + \frac{\lambda^2}{\gamma^2 - \beta^2} \frac{\gamma e^{-\beta w} - \beta e^{-\gamma w}}{2\beta \gamma}
\end{aligned}$$

und für  $w = 0$ , d. h.  $t = s$

$$\mathbb{E} q_t q_t := \frac{\sigma^2}{2\gamma} + \frac{\lambda^2}{2\beta \gamma (\beta + \gamma)}.$$

Setzt man hier die Parametersituationen **Par1** und **Par2** ein, so erhält man das gleiche Ergebnis.

Schließlich sei noch bemerkt, daß die vier Gleichungen (2.8a) und (2.8b) mit  $k = 1, 2, 3$  unter der Nebenbedingung  $a \neq 0$  und  $c \neq 0$  lediglich zwei Lösungen besitzen, welche wie in (2.9) dargestellt zusammenhängen. Der Fall  $c = 0$  entspricht der Situation  $\lambda = 0$ , bei dem der unterliegende Prozeß  $X$  nicht in die Drift von  $q$  eingeht. In dieser Situation ist offensichtlich der Parameter  $\beta$  nicht schätzbar, allerdings ist er in dieser Situation auch nicht interessant.

Der Fall  $a = 0$  entspricht noch einer besonderen Situation, bei der zwar die Parameter  $\beta$  und  $\sigma$  eindeutig durch die Gleichungen (2.8a) und (2.8b) mit  $k = 1, 2, 3$  bestimmt sind. Von  $\gamma$  und  $\lambda$  ist allerdings in dieser Situation nur die Beziehung  $\frac{\lambda^2}{\gamma^2 - \beta^2} = \sigma^2$  bekannt.

Wir werden später in numerischen Fallstudien illustrieren, daß diese Parameteruneindeutigkeiten für zeitabhängige Volatilitätsfunktionen im Allgemeinen nicht auftreten. Eine kurze Motivation für diesen Effekt liefert die Situation einer stückweise konstanten Volatilitätsfunktion

$$\sigma^*(t) := \begin{cases} \sigma_1 & t \leq T/2 \\ \sigma_2 & t > T/2 \end{cases}$$

mit zwei positiven Werten  $\sigma_1 \neq \sigma_2$ . Vorausgesetzt  $T$  ist hinreichend groß, gilt nun für  $t_1 = T/2$  und  $t_2 = T$

$$\begin{aligned}
\mathbf{D}^2 r_\tau(t_1) &\approx \frac{\lambda^2}{\beta - \gamma} \frac{\beta - \gamma + \gamma e^{-\beta \tau} - \beta e^{-\gamma \tau}}{\beta(\beta + \gamma)\gamma} + \frac{\sigma_1^2(1 - e^{-\gamma \tau})}{\gamma} \\
\mathbf{D}^2 r_\tau(t_2) &\approx \frac{\lambda^2}{\beta - \gamma} \frac{\beta - \gamma + \gamma e^{-\beta \tau} - \beta e^{-\gamma \tau}}{\beta(\beta + \gamma)\gamma} + \frac{\sigma_2^2(1 - e^{-\gamma \tau})}{\gamma}
\end{aligned}$$

und damit

$$\mathbf{D}^2 r_\tau(t_1) - \mathbf{D}^2 r_\tau(t_2) \approx \frac{(\sigma_1^2 - \sigma_2^2)(1 - e^{-\gamma\tau})}{\gamma}.$$

Da die Hilfsfunktion  $h : \mathbb{R}^+ \rightarrow [0, \tau]$  mit  $h(\gamma) := \frac{1 - e^{-\gamma\tau}}{\gamma}$  streng monoton fallend in  $\gamma$  ist und somit die Umkehrfunktion  $h^{-1}$  existiert, ist also zu erwarten, daß für die Volatilitätsfunktion  $\sigma^*$  bei einem hinreichend langen Zeitintervall  $[0, T]$  auch der Parameter  $\gamma$  eindeutig identifiziert werden kann. Durch Einsetzen in (2.6) erhalten wir dann auch  $\beta$  und  $\lambda$  auf eindeutige Weise.

## 2.2 Schätzung von $\mu$

In diesem Abschnitt wollen wir die Bestimmung des Parameters  $\mu$  etwas näher betrachten. Da das verallgemeinerte bivariate Ornstein-Uhlenbeck Modell mit den Parametern  $\gamma = 0$  und  $\lambda = 0$  das Black-Scholes Modell als Spezialfall enthält, scheint es sinnvoll, zunächst die Schätzung des Parameters  $\mu$  im Black-Scholes Modell zu diskutieren.

Wir betrachten also zunächst das Modell

$$dp_t = \mu dt + \sigma dW_t$$

für den logarithmierten Aktienpreisprozeß  $p_t = \ln P_t$ . Wir nehmen an, dass wir den Preisprozess  $P$  zu diskreten Zeitpunkten  $t_k$  ( $k = 0, 1, \dots, N$ ) (siehe (1.4)) beobachten können, wobei wir in diesem Abschnitt die Schrittweite  $\tau$  als fixiert annehmen und uns für die Asymptotik  $N \rightarrow \infty$  (oder äquivalent  $T \rightarrow \infty$ ) interessieren.

Üblicherweise (vgl. zum Beispiel [3, S. 195 f.]) wird für das Black-Scholes Modell vorgeschlagen, die Drift  $\mu$  mit der Schätzfunktion

$$\mu_N = \frac{1}{N} \sum_{i=1}^N (p_{t_i} - p_{t_{i-1}}) = \frac{p_{t_N} - p_{t_0}}{N\tau} \quad (2.12)$$

zu schätzen.

Wegen  $t_N = T$  und  $t_0 = 0$  ist  $p_{t_N} - p_{t_0}$  eine  $\mathcal{N}(\mu T, T\sigma^2)$ -verteilte Zufallsgröße. Somit ist  $\mu_N$  normalverteilt mit Erwartungswert  $\mu$  und Varianz  $\frac{\sigma^2}{T}$ . Dies zeigt insbesondere, daß  $\mu_N$  eine erwartungstreue und für die Asymptotik  $T \rightarrow \infty$  konsistente Schätzfunktion ist, d. h.

$$\mathbf{E} \mu_N = \mu \quad \text{und} \quad \forall \varepsilon > 0 \text{ gilt} \quad \mathbf{P}(|\mu_N - \mu| > \varepsilon) \rightarrow 0 \quad \text{für } T \rightarrow \infty.$$

Es ist interessant zu erwähnen, daß in dieser Situation die Güte der Schätzung nur von der Länge des Beobachtungsintervalls, d. h. nur von  $T$  abhängt. Ließe man  $T$  konstant und vergrößerte  $N$  (durch Verkleinerung von  $\tau$ ) würde sich die Güte der Schätzung nicht verbessern.

Wir wenden uns nun wieder dem verallgemeinerten Bivariaten Ornstein-Uhlenbeck Modell (1.3) zu. Im nächsten Abschnitt werden wir die in [8, S. 415 ff.] dargestellte Theorie zu dem Kalman Filter und der Maximum Likelihood Schätzung anwenden. Da dort der beobachtbare Prozeß mit  $y$  bezeichnet wird, wollen wir im folgenden den logarithmierten Aktienpreisprozess  $p$  mit  $y$  bezeichnen, d. h. wir setzen

$$y_t := p_t = q_t + \mu t. \quad (2.13)$$

Wir versuchen also jetzt den Parameter  $\mu$  aus den zu den Zeitpunkten  $t_0, t_1, \dots, t_N$  beobachteten Kursdaten

$$\underline{y} := (y_{t_0}, y_{t_1}, \dots, y_{t_N})^T \quad (2.14)$$

zu schätzen. Zur Vereinfachung beschränken wir uns dabei (d. h. für den Rest dieses Abschnitts) auf Parameterkonstellationen, welche der folgenden Annahme genügen.

**Annahme 2.1** Unser Modell erfülle die folgenden Bedingungen:

1. Es seien  $\gamma > 0$  und  $\beta > 0$  reelle Parameter mit  $\beta \neq \gamma$ .
2. Die Volatilität  $\sigma(t)$  sei für alle Zeitpunkte konstant, d. h. es gilt  $\sigma(t) = \sigma > 0$  für alle  $t \in [0, T]$ .
3. Für die Anfangsbedingungen gelte  $c_q = c_X = 0$ .

Wir wollen im weiteren zwei Schätzer vergleichen. Zum einen den in (2.12) eingeführten Schätzer  $\mu_N$  und weiterhin den durch

$$\hat{\mu}(\underline{y}) := \frac{2}{T(N+1)} \sum_{k=1}^N y_{t_k} \quad (2.15)$$

definierten Schätzer  $\hat{\mu}$ . Als Motivation für den Schätzer  $\hat{\mu}$  bemerken wir, daß unter den Voraussetzungen aus Annahme 2.1

$$\mathbf{E} p_{t_k} = \mu t_k \quad \text{und somit} \quad \mathbf{E} \left( \sum_{k=1}^N p_{t_k} \right) = \mu \sum_{k=1}^N k\tau = \mu \frac{T(N+1)}{2}$$

gilt. Hieraus folgt, daß  $\hat{\mu}$  eine erwartungstreue Schätzfunktion für den Parameter  $\mu$  ist. Offensichtlich berücksichtigt der Schätzer  $\hat{\mu}$  im Gegensatz zu dem oben betrachteten Schätzer  $\hat{\mu}_N$  alle Beobachtungswerte  $p_{t_k}$  ( $k = 1, \dots, N$ ), nicht nur den ersten und letzten. Wir werden im Folgenden sehen, daß dies unter bestimmten Voraussetzungen (insbesondere Annahme 2.1) zu einer besseren Asymptotik für die Varianz des Schätzers führt.

Bevor wir Aussagen zur Konsistenz und den Varianzen von  $\hat{\mu}$  und  $\mu_N$  angeben, berechnen wir die Varianz von  $q_t$ . Durch Einsetzen der Anfangsbedingungen in (2.11) folgt

$$\begin{aligned} \mathbf{D}^2 q_{t_k} &= \frac{\lambda^2}{2\beta\gamma(\beta+\gamma)} - \frac{\lambda^2}{(\gamma-\beta)^2} \left( \frac{1}{2\beta} e^{-2\beta t_k} - \frac{2}{\beta+\gamma} e^{-(\beta+\gamma)t_k} + \frac{1}{2\gamma} e^{-2\gamma t_k} \right) \\ &+ \frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma t_k}) \quad (k = 1, \dots, N). \end{aligned} \quad (2.16)$$

Damit ergibt sich die Varianz der Schätzfunktion  $\mu_N$  als

$$\begin{aligned} \mathbf{D}^2 \mu_N &= \frac{1}{t_N^2} \cdot \frac{\lambda^2}{2\beta\gamma(\beta+\gamma)} - \frac{1}{t_N^2} \cdot \frac{\lambda^2}{(\gamma-\beta)^2} \left( \frac{1}{2\beta} e^{-2\beta t_N} - \frac{2}{\beta+\gamma} e^{-(\beta+\gamma)t_N} + \frac{1}{2\gamma} e^{-2\gamma t_N} \right) \\ &+ \frac{1}{t_N^2} \cdot \frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma t_N}). \end{aligned} \quad (2.17)$$

Da der Prozeß  $p_{t_N}$  ein Gauß-Prozeß ist, sind  $p_{t_N}$  und  $\mu_N = \frac{p_{t_N}}{t_N}$  normalverteilte Zufallsgrößen. Da weiterhin die Varianz des Schätzers für immer größer werdende Beobachtungszeiträume gegen Null konvergiert, d. h.  $\lim_{N \rightarrow \infty} \mathbf{D}^2 \mu_N = 0$  gilt, folgt die Konsistenz des Schätzers.

Nun wenden wir uns dem Schätzer  $\hat{\mu}$  zu. Wir zeigen zunächst die Konsistenz.

**Satz 2.2**

Die Bedingungen aus Annahme 2.1 seien erfüllt. Dann ist die Schätzfunktion

$$\hat{\mu}(\underline{y}) = \frac{2}{T(N+1)} \sum_{k=1}^N y_{t_k}$$

konsistent, d. h. für beliebiges  $\varepsilon > 0$  gilt

$$\lim_{N \rightarrow \infty} \mathbf{P} \left( \left| \hat{\mu}(\underline{y}) - \mu \right| \geq \varepsilon \right) = 0.$$

**Beweis:** Wegen  $c_q = c_X = 0$  gilt  $\mathbf{E} q_{t_k} = 0$  für alle  $k = 1, \dots, N$ . Für die Varianz von  $\hat{\mu}$  ergibt sich dann mit  $\frac{\mathbf{Cov}(q_{t_i}, q_{t_j})}{\sqrt{\mathbf{D}^2 q_{t_i}} \sqrt{\mathbf{D}^2 q_{t_j}}} \leq 1$  für  $i, j = 1, \dots, N$  die Abschätzung

$$\begin{aligned} \mathbf{D}^2 \hat{\mu}(\underline{y}) &= \frac{4}{T^2(N+1)^2} \mathbf{D}^2 \left( \sum_{k=1}^N y_{t_k} \right) = \frac{4}{N^2(N+1)^2 \tau^2} \mathbf{D}^2 \left( \sum_{k=1}^N (q_{t_k} + \mu t_k) \right) \\ &= \frac{4}{N^2(N+1)^2 \tau^2} \mathbf{D}^2 \left( \sum_{k=1}^N q_{t_k} \right) = \frac{4}{N^2(N+1)^2 \tau^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{Cov}(q_{t_i}, q_{t_j}) \\ &\leq \frac{4}{N^2(N+1)^2 \tau^2} \underbrace{\left( \sqrt{\mathbf{D}^2 q_{t_1}} + \dots + \sqrt{\mathbf{D}^2 q_{t_N}} \right)^2}_{\leq N^2 C_N} \\ &\leq \frac{4C_N}{(N+1)^2 \tau^2}, \end{aligned} \tag{2.18}$$

wobei  $C_N := \max_{k=1, \dots, N} \mathbf{D}^2 q_{t_k}$ . Aus der Gleichung (2.16) erhalten wir wegen

$$\left( \frac{1}{2\beta} e^{-2\beta t_k} - \frac{2}{\beta + \gamma} e^{-(\beta + \gamma)t_k} + \frac{1}{2\gamma} e^{-2\gamma t_k} \right) = \frac{(\gamma e^{-\beta t_k} - \beta e^{-\gamma t_k})^2 + \beta\gamma(e^{-\beta t_k} - e^{-\gamma t_k})^2}{2\beta\gamma(\beta + \gamma)} \geq 0$$

für  $k = 1, \dots, N$  die folgende Abschätzung für  $C_N$

$$C_N = \max_{k=1, \dots, N} \mathbf{D}^2 q_{t_k} \leq \frac{\lambda^2}{2\beta\gamma(\beta + \gamma)} + \frac{\sigma^2}{2\gamma} =: C. \tag{2.19}$$

Sei nun  $\varepsilon > 0$  beliebig. Aufgrund der Erwartungstreue ergibt sich dann mit der Ungleichung von Tschebyscheff

$$0 \leq \mathbf{P} \left( \left| \hat{\mu}(\underline{y}) - \mu \right| \geq \varepsilon \right) \leq \frac{4C}{(N+1)^2 \tau^2 \varepsilon^2}.$$

Mit Grenzwertbildung für  $N \rightarrow \infty$  folgt daraus schließlich die Behauptung. ■

Die Varianz des Schätzers  $\hat{\mu}$  für die hier betrachtete Situation explizit auszurechnen ist sehr aufwendig. Dies wird durch die Tatsache verursacht, daß die Größen  $q_{t_k}$  nicht stochastisch unabhängig sind. Wir beschränken uns deshalb neben den bereits vorgenommenen Einschränkungen noch auf den Fall  $\lambda = 0$  (vgl. auch [4]). In diesem Fall ist  $q$  ein univariater Ornstein-Uhlenbeck-Prozeß. Auch hier sind die Werte  $q_{t_k}$  nach wie vor stochastisch abhängig. Allerdings finden wir für diesen Fall neben der expliziten Formel relativ leicht eine schärfere obere Schranke für die Varianz der Schätzfunktion  $\hat{\mu}$ .

**Satz 2.3**

Neben der Annahme 2.1 sei noch  $\lambda = 0$ . Dann gilt für die Varianz des Schätzers  $\mu_N$

$$\mathbf{D}^2 \mu_N = \frac{1}{N^2 \tau^2} \cdot \frac{\sigma^2}{2\gamma} \cdot (1 - e^{-2\gamma N \tau}) \quad (2.20)$$

und für die Varianz des Schätzers  $\hat{\mu}$

$$\mathbf{D}^2 \hat{\mu} < \frac{4}{N^3 \tau^2} \cdot \frac{\sigma^2}{2\gamma} \cdot \frac{1 + e^{-\gamma \tau}}{1 - e^{-\gamma \tau}}. \quad (2.21)$$

**Beweis:** Die Beziehung (2.20) erhalten wir unmittelbar aus der Formel (2.17) durch Einsetzen von  $\lambda = 0$ .

Zum Beweis der Ungleichung (2.21) setzen wir zunächst  $g := e^{-\gamma \tau}$ . Wegen  $\gamma > 0$  gilt dann offensichtlich  $0 < g < 1$ . Ferner definieren wir uns Zufallsgrößen  $\xi_k$  mit

$$\xi_k := \sigma \int_{t_{k-1}}^{t_k} e^{-\gamma(t_k-s)} dW_s^q \quad (k = 1, \dots, N).$$

Diese sind für alle  $k = 1, \dots, N$  vollständig stochastisch unabhängig und normalverteilt mit Erwartungswert Null und den Varianzen

$$\mathbf{D}^2 \xi_k = \sigma^2 \int_{t_{k-1}}^{t_k} e^{-2\gamma(t_k-s)} ds = \frac{\sigma^2}{2\gamma} (1 - e^{-2\gamma \tau}) = \frac{\sigma^2}{2\gamma} (1 - g^2) \quad (k = 1, \dots, N).$$

Nach Gleichung (2.4) können wir dann  $q_{t_k}$  wie folgt zerlegen

$$\begin{aligned} q_{t_k} &= g q_{t_{k-1}} + \xi_k \\ &= g(g q_{t_{k-2}} + \xi_{k-1}) + \xi_k = g^2 q_{t_{k-2}} + g \xi_{k-1} + \xi_k \\ &= \dots = \underbrace{g^k c_q}_{=0} + \sum_{j=1}^k g^{k-j} \xi_j \quad (k = 1, \dots, N). \end{aligned}$$

Wir summieren nun die Werte  $q_{t_k}$  für  $k = 1, \dots, N$  auf und erhalten

$$\begin{aligned} \sum_{k=1}^N q_{t_k} &= \xi_1 + (g \xi_1 + \xi_2) + \dots + (g^{N-1} \xi_1 + \dots + g \xi_{N-1} + \xi_N) \\ &= \xi_1 (1 + g + \dots + g^{N-1}) + \dots + \xi_{N-1} (1 + g) + \xi_N. \end{aligned}$$

Mit den vollständig unabhängigen Zufallsgrößen

$$\zeta_k := \xi_k \cdot \sum_{j=0}^{N-k} g^j \quad (k = 1, \dots, N)$$

können wir jetzt die Summe abhängiger Zufallsgrößen  $q_{t_k}$  in eine Summe unabhängiger Zufallsgrößen  $\zeta_k$  überführen. Es gilt dann gerade

$$\sum_{k=1}^N q_{t_k} = \sum_{k=1}^N \zeta_k.$$

Die  $\zeta_k$  sind ebenfalls normalverteilt mit Erwartungswert Null. Wegen  $0 < g < 1$  finden wir mit dem Grenzwert der unendlichen geometrischen Reihe für die Varianzen die folgende Abschätzung

$$\mathbf{D}^2 \zeta_k = \mathbf{D}^2 \xi_k \cdot \left( \sum_{j=0}^{N-k} g^j \right)^2 < \mathbf{D}^2 \xi_k \cdot \frac{1}{(1-g)^2} = \frac{\sigma^2}{2\gamma} \cdot \frac{1+g}{1-g} \quad (k = 1, \dots, N). \quad (2.22)$$

Unter Ausnutzung von  $\frac{1}{(N+1)^2} < \frac{1}{N^2}$  für  $N \geq 1$  erhalten wir aus der Unabhängigkeit der  $\zeta_k$  und der Abschätzung (2.22) die Behauptung.  $\blacksquare$

**Bemerkung 2.4** Aus der Zerlegung von  $q_{t_k}$  mittels der Zufallsgrößen  $\xi_k$  erhält man nach einigen Umformungen auch den folgenden etwas unhandlichen Ausdruck für die Varianz des Schätzers  $\hat{\mu}$  (bei Gültigkeit der Annahme 2.1 und  $\lambda = 0$ )

$$\mathbf{D}^2 \hat{\mu} = \frac{4}{N^2(N+1)^2\tau^2} \cdot \frac{\sigma^2}{2\gamma} \cdot \frac{1+e^{-\gamma\tau}}{1-e^{-\gamma\tau}} \left( (N+1) - 2 \cdot \frac{1-e^{-\gamma(N+1)\tau}}{1-e^{-\gamma\tau}} + \frac{1-e^{-2\gamma(N+1)\tau}}{1-e^{-2\gamma\tau}} \right).$$

**Bemerkung 2.5** Aus dem Satz 2.3 können wir schlussfolgern, daß, bei Gültigkeit der Annahme 2.1 und  $\lambda = 0$  für  $\gamma > 0$  die Varianz des Schätzers  $\hat{\mu}$  eine bessere Asymptotik besitzt als die Varianz des Schätzers  $\mu_N$ . Andererseits kann die Konstante  $\frac{1+e^{-\gamma\tau}}{1-e^{-\gamma\tau}}$  für kleine Werte  $\gamma$  und  $\tau$  sehr groß werden. Für einen festen Beobachtungszeitraum  $[0, T]$  könnte es also für kleine Werte  $\gamma$  durchaus passieren, daß der Schätzer  $\mu_N$  dem Schätzer  $\hat{\mu}$  überlegen ist.

**Beispiel 2.6** Wir betrachten den Fall  $\lambda = 0$  und  $\gamma = 0.5$  mit der konstanten Volatilität  $\sigma = 0.01$ . Ferner nehmen wir an, daß uns für die Lags  $\tau = 1$  bzw.  $\tau = 10$  jeweils  $N = 100$  Beobachtungswerte zur Verfügung stehen.

Nach Satz 2.3 können wir erwarten, daß für  $\tau = 1$  die Varianz von  $\mu_N$  mindestens um den Faktor 6 größer ist als die der Schätzfunktion  $\hat{\mu}$ . Für  $\tau = 10$  sollte der Faktor sogar fast 25 betragen. Wir haben für diese beiden Fälle jeweils 1000 Trajektorien simuliert und die entsprechenden empirischen Varianzen der Schätzer verglichen. Mit 6.38 und 25.85 lagen die Ergebnisse für den relativen Größenunterschied im Rahmen unserer Erwartungen (vgl. Tabelle 2.1). Es sei aber auch erwähnt, daß die absoluten Werte der empirischen Varianzen mit Größenordnungen von  $10^{-8}$  und darunter für dieses Beispiel ohnehin sehr klein ausfallen.

| $\tau$ | $\mathbf{D}^2 \mu_N / \mathbf{D}^2 \hat{\mu}$ |                  |
|--------|---|------------------|
|        | Untere Schranke                               | Empirischer Wert |
| 1      | 6.12  | 6.38             |
| 10     | 24.67   | 25.85            |

Tabelle 2.1: Vergleich der Varianzen der Schätzfunktionen  $\mu_N$  und  $\hat{\mu}$

Wir wollen noch untersuchen, welche Varianz der Schätzer  $\hat{\mu}$  im Black-Scholes Modell (d. h. für  $\gamma = \lambda = 0$ ) besitzt. Das nächste Lemma gibt eine Antwort. Für große  $N$  gilt  $\mathbf{D}^2 \hat{\mu} \approx \frac{4}{3} \mathbf{D}^2 \mu_N$ . Dies zeigt, dass der Schätzer  $\mu_N$  im Black-Scholes Modell dem Schätzer  $\hat{\mu}$  überlegen ist.

**Lemma 2.7**

Seien  $\gamma = \lambda = 0$  in (2.1) sowie  $c_q = c_X = 0$ . Sei weiterhin die Volatilitätsfunktion konstant, d. h.  $\sigma(t) = \sigma > 0$  für alle  $t \in [0, T]$ . Dann gilt für die Varianzen der Schätzer  $\hat{\mu}$  und  $\mu_N$

$$\mathbf{D}^2 \hat{\mu} = \frac{4}{3} \frac{\sigma^2}{N\tau} \frac{N + \frac{1}{2}}{N + 1} \quad \text{und} \quad \mathbf{D}^2 \mu_N = \frac{\sigma^2}{N\tau}. \quad (2.23)$$

**Beweis:** Die Varianz für den Schätzer  $\mu_N$  wurde bereits am Anfang dieses Abschnitts untersucht. Es bleibt die Formel für  $\mathbf{D}^2 \hat{\mu}$  zu zeigen. Unter den getroffenen Annahmen gilt  $q_{t_k} = \int_0^{t_k} \sigma dW_s^q$  und damit

$$\mathbf{Cov}(q_{t_i}, q_{t_j}) = \sigma^2 \min(t_i, t_j).$$

Weiterhin gilt

$$\begin{aligned} \mathbf{D}^2 \hat{\mu} &= \frac{4}{N^2(N+1)^2\tau^2} \mathbf{D}^2 \left( \sum_{k=1}^N (q_{t_k} + \mu^* t_k) \right) \\ &= \frac{4}{N^2(N+1)^2\tau^2} \sum_{i=1}^N \sum_{j=1}^N \mathbf{Cov}(q_{t_i}, q_{t_j}). \end{aligned}$$

Um die Beziehung (2.23) zu beweisen, müssen wir also lediglich

$$\sum_{i=1}^N \sum_{j=1}^N \min(i, j) = \frac{1}{3} N(N+1) \left( N + \frac{1}{2} \right)$$

zeigen. Mit den bekannten Summenformeln

$$\sum_{i=1}^N i = \frac{N(N+1)}{2} \quad \text{und} \quad \sum_{i=1}^N i^2 = \frac{N(N+1)(2N+1)}{6}$$

geschieht dies wie folgt,

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^N \min(i, j) &= \sum_{i=1}^N \left( \underbrace{\sum_{j=1}^i \min(i, j)}_{=j} + \sum_{j=i+1}^N \underbrace{\min(i, j)}_{=i} \right) \\ &= \sum_{i=1}^N \left( \frac{i(i+1)}{2} + (N-i)i \right) = \left( N + \frac{1}{2} \right) \sum_{i=1}^N i - \frac{1}{2} \sum_{i=1}^N i^2 \\ &= \frac{1}{3} N(N+1) \left( N + \frac{1}{2} \right). \end{aligned}$$

■

Für unser Modell mit verschwindenden Anfangsbedingungen und konstanter Volatilität haben wir mit Hilfe numerischer Fallstudien noch andere Parameterkonstellationen untersucht, auch solche mit  $\lambda > 0$ . In einer Vielzahl dieser Versuche hat sich gezeigt, daß der Schätzer  $\hat{\mu}$  eine kleinere empirische Varianz besitzt als der Schätzer  $\mu_N$ . Wenn dieser Effekt auch für viele Parameterkonstellationen bei weitem nicht so ausgeprägt ist wie in Beispiel 2.6, so ist er doch meist vorhanden.

Selbstverständlich muß eingeräumt werden, daß es auch Situationen gibt, in denen der Schätzer  $\hat{\mu}$  etwas schlechter abschneidet als  $\mu_N$ , so z. B. für den Fall, daß der Parameter  $\gamma$  sehr klein ist ( $\gamma \ll 1$ ). Dies ist in Anbetracht der obigen theoretischen Untersuchungen auch zu erwarten, vergleiche Bemerkung 2.5.

Zusammenfassend läßt sich sagen, daß der Schätzer  $\hat{\mu}$  für zahlreiche Parameterkonstellationen deutlich bessere Ergebnisse (im Sinne der empirischen Varianz) liefert, während er bei keiner Parameterkonstellation gravierend schlechter abschneidet als  $\mu_N$ . Wir konnten in keinem der von uns untersuchten Fälle eine empirische Varianz von  $\hat{\mu}$  beobachten, die mehr als doppelt so groß war wie diejenige von  $\mu_N$ .

### 3 Maximum-Likelihood-Methode

In Abschnitt 2 haben wir gesehen, dass der Parameter  $\sigma_X$  ohne Einschränkung der Allgemeinheit auf Eins gesetzt werden kann. Weiterhin haben wir eine Schätzfunktion für den Parameter  $\mu$  hergeleitet. In diesem Abschnitt werden wir nun sehen, wie wir auf dieser Basis die restlichen Parameter mit Hilfe der Maximum-Likelihood-Methode schätzen können.

Dazu stellen wir im ersten Abschnitt zunächst die Grundidee der Maximum-Likelihood Schätzung kurz dar und führen einige Bezeichnungen ein, die wir im Weiteren benötigen werden. Anschließend überführen wir das verallgemeinerte Bivariate Ornstein-Uhlenbeck Modell (1.3) in ein Zustandsraummodell, auf das wir den Kalman Filter anwenden können. Dieser liefert uns Hilfsgrößen (bedingte Erwartungswerte und bedingte Varianzen), welche wir zur Herleitung der Likelihoodfunktion für die beobachtbaren logarithmierten Aktienpreise in Abschnitt 3.3 benötigen.

#### 3.1 Allgemeine Einführung

Wir betrachten einen Zufallsvektor  $\underline{u} = (u_1, \dots, u_n)^T$ ,  $n \geq 1$ . Die einzelnen Komponenten können stochastisch unabhängig sein, müssen dies aber nicht zwingend. Wir nehmen an, daß die Verteilung von  $\underline{u}$  durch einen Parametervektor  $\underline{\theta} = (\theta_1, \dots, \theta_m)^T \in \Theta$ ,  $m \geq 1$  bestimmt wird.  $\Theta \subseteq \mathbb{R}^m$  ist der sogenannte Parameterraum.

##### Definition 3.1 (Likelihood-Funktion)

Sei  $\underline{u} = (u_1, \dots, u_n)^T$ ,  $n \geq 1$  ein Zufallsvektor wie oben. Für eine konkrete Realisierung  $\underline{u}^* = (u_1^*, \dots, u_n^*)^T$  von  $\underline{u}$  heißt die durch

$$L(\underline{u}^* | \underline{\theta}) := g_{\underline{u}}(\underline{u}^*, \underline{\theta}) \tag{3.1}$$

definierte Funktion Likelihood-Funktion von  $\underline{u}^*$ . Dabei ist  $g_{\underline{u}}(\cdot, \underline{\theta})$  die Dichtefunktion von  $\underline{u}$  unter dem Parametervektor  $\underline{\theta} = (\theta_1, \dots, \theta_m)^T \in \Theta$ .

**Bemerkung 3.2** Die Funktion  $L(\underline{u}^* | \underline{\theta})$  ist eine Funktion in  $\underline{\theta}$ .

Zur Bestimmung eines Schätzwertes  $\underline{\theta}^* = (\theta_1^*, \dots, \theta_m^*)^T \in \Theta$  für den unbekanntem Parametervektor  $\underline{\theta}$  betrachten wir das Optimierungsproblem

$$L(\underline{u}^* | \underline{\theta}) \longrightarrow \max \quad \text{für } \underline{\theta} \in \Theta. \tag{3.2}$$

Wir wählen den Schätzwert  $\underline{\theta}^*$  als Lösung von (3.2). Dabei kann es vorkommen, daß dieser nicht eindeutig bestimmt ist.

Da die Funktion  $\ln(\cdot) : (0, \infty) \rightarrow \mathbb{R}$  streng monoton wachsend ist, können wir anstelle von (3.2) auch die dazu äquivalente Aufgabe

$$\ln L(\underline{u}^* | \underline{\theta}) \longrightarrow \max \quad \text{für } \underline{\theta} \in \Theta \quad (3.3)$$

lösen. Die Funktion  $\ln L(\underline{u}^* | \underline{\theta})$  wird als *Loglikelihood-Funktion* bezeichnet.

**Bemerkung 3.3** Die Loglikelihood-Funktion läßt sich in vielen Fällen sowohl explizit als auch bei numerischen Rechnungen leichter handhaben.

Wir kehren nun wieder zu unserem Modell zurück und führen die folgende Bezeichnung ein. Um die Notation ähnlich zu der in [8] zu halten, haben wir in (2.13) den logarithmierten Aktienpreisprozess  $p$  mit  $y$  und eine konkrete Realisierung dieses Prozesses mit  $y^*$  bezeichnet.

Weiterhin fassen wir die drei zu identifizierenden Parameter  $\gamma$ ,  $\lambda$  und  $\beta$  in einem Vektor  $\underline{p} = (\gamma, \lambda, \beta)^T$  zusammen. Den Definitionsbereich für  $\underline{p}$  beschränken wir dabei auf die Menge

$$D = \{\underline{p} \in \mathbb{R}^3 : \gamma \geq 0, \lambda \geq 0, \beta \geq 0\} \quad (3.4)$$

ein.

Jedem Parametervektor  $\underline{p} \in D$  wird ein stochastischer Prozeß  $y$  zugeordnet. Die Eigenschaften dieses Prozesses hängen natürlich von dem jeweiligen Parametervektor ab. Für eine konkrete Zufallssituation  $\omega \in \Omega$  erhalten wir dann eine Trajektorie des Prozesses.

In den nächsten Abschnitten gehen wir davon aus, dass der Parameter  $\mu$  bekannt ist und werden mittels Maximum-Likelihood-Methode den exakten Parametervektor  $\underline{p}$  aus den zu den Zeitpunkten  $t_0, t_1, \dots, t_N$  beobachteten Kursdaten (2.14) schätzen. Die Dichtefunktion von  $\underline{y}$  unter dem Parametervektor  $\underline{p}$  bezeichnen wir mit  $g_{\underline{y}}(\cdot, \underline{p})$ . Zur Herleitung dieser Dichtefunktion formulieren wir das Modell 1.1 zunächst als Zustandsraummodell, für das wir den Kalman Filter anwenden können. Dieser liefert uns wesentliche Größen, die wir zur Berechnung der Likelihoodfunktion benötigen.

## 3.2 Zustandsraummodell und Kalman-Filter

Wie in Abschnitt 2 fassen wir die Werte  $q_{t_k}$  und  $X_{t_k}$  in einem Zufallsvektor  $\underline{\alpha}$  zusammen, d. h. wir setzen  $\underline{\alpha}_{t_k} := \begin{pmatrix} q_{t_k} \\ X_{t_k} \end{pmatrix}$ . Aus der Darstellung (2.3) erhält man sofort die rekursive Darstellung

$$\underline{\alpha}_{t_k} = e^{\tau A} \underline{\alpha}_{t_{k-1}} + \int_{t_{k-1}}^{t_k} e^{(t_k-s)A} B(s) dW_s. \quad (3.5)$$

Um Fallunterscheidungen zu vermeiden, werden wir die weiteren Berechnungen nur für den Fall  $\gamma \neq \beta$  darstellen. Wir bemerken jedoch, daß der Fall  $\gamma = \beta$  vollkommen analog behandelt werden kann. Mit der Notation  $T := e^{\tau A}$  sowie  $R := \begin{pmatrix} 1 & \frac{\lambda}{\gamma-\beta} \\ 0 & 1 \end{pmatrix}$  und

$$\underline{\eta}_{t_k} := \begin{pmatrix} \int_{t_{k-1}}^{t_k} \sigma(s) e^{-\gamma(t_k-s)} dW_s^q - \frac{\lambda}{\gamma-\beta} \int_{t_{k-1}}^{t_k} e^{-\gamma(t_k-s)} dW_s^X \\ \int_{t_{k-1}}^{t_k} e^{-\beta(t_k-s)} dW_s^X \end{pmatrix} \quad (k = 1, \dots, N)$$

können wir (3.5) auch als

$$\underline{\alpha}_{t_k} = T \underline{\alpha}_{t_{k-1}} + R \underline{\eta}_{t_k} \quad (k = 1, \dots, N)$$

schreiben.

Diese Darstellung zeigt, daß  $\underline{\alpha}$  ein Markov-Prozeß erster Ordnung in diskreter Zeit ist, d. h. ein Prozeß, bei dem die Abhängigkeit des Zustands zum Zeitpunkt  $t_k$  von der Vergangenheit vollständig durch die Abhängigkeit zum Zeitpunkt  $t_{k-1}$  beschrieben wird. Die Vektoren  $\underline{\eta}_{t_k}$  sind unabhängige normalverteilte Zufallsvektoren. Es gilt

$$\underline{\eta}_{t_k} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, Q_{t_k} \right),$$

wobei  $Q_{t_k} := \mathbf{E} \underline{\eta}_{t_k} \underline{\eta}_{t_k}^T$  die Kovarianzmatrix bezeichnet. Mit der Notation

$$H(t_1, t_2) := \int_{t_1}^{t_2} \sigma^2(s) e^{-2\gamma(t_2-s)} ds$$

und der Hilfsfunktion  $h_\tau(\cdot) : [0, \infty) \rightarrow \mathbb{R}$ , definiert durch

$$h_\tau(c) := \int_0^\tau e^{-c(\tau-s)} ds = \begin{cases} \frac{1}{c} (1 - e^{-c\tau}) & c > 0 \\ \tau & c = 0 \end{cases}, \quad (3.6)$$

können wir die Matrizen  $Q_{t_k}$  wie folgt berechnen

$$Q_{t_k} = \begin{pmatrix} H(t_{k-1}, t_k) + \left(\frac{\lambda}{\gamma-\beta}\right)^2 h_\tau(2\gamma) & -\frac{\lambda}{\gamma-\beta} h_\tau(\beta + \gamma) \\ -\frac{\lambda}{\gamma-\beta} h_\tau(\beta + \gamma) & h_\tau(2\beta) \end{pmatrix} \quad (k = 1, \dots, N). \quad (3.7)$$

Fassen wir die bisherigen Überlegungen zusammen und führen die Notation  $\underline{z} := (1, 0)^T$  und  $d_{t_k} := \mu t_k$  ein, so können wir  $y_{t_k} = q_{t_k} + \mu t_k$  wie folgt beschreiben.

$$y_{t_k} = \underline{z}^T \underline{\alpha}_{t_k} + d_{t_k} \quad (k = 1, \dots, N), \quad (3.8a)$$

wobei  $\underline{\alpha}_{t_k}$  der Gleichung

$$\underline{\alpha}_{t_k} = T \underline{\alpha}_{t_{k-1}} + R \underline{\eta}_{t_k} \quad (k = 1, \dots, N) \quad (3.8b)$$

genügt.

Diese Darstellung bezeichnet man auch als *Zustandsraummodell (state space model)*. Da der Prozeß  $y$  beobachtet (gemessen) werden kann, wird die Gleichung (3.8a) *Beobachtungsgleichung (measurement equation)* genannt. Die Werte  $y_{t_k}$  hängen von den Zuständen  $\underline{\alpha}_{t_k}$  ab, die nicht direkt beobachtet werden können. Der Vektor  $\underline{\alpha}_{t_k}$  wird *Zustandsvektor (state vector)* und die Gleichung (3.8b) *Zustandsgleichung (transition equation)* genannt (siehe z. B. [2]).

Zur Vervollständigung des Zustandsraummodells für den logarithmierten Preisprozeß  $p_t$  benötigen wir abschließend noch eine Annahme über den Anfangszustand  $\underline{c}_\alpha$ . Wir werden im Folgenden annehmen, daß der Anfangszustand  $\underline{\alpha}_{t_0} = \underline{c}_\alpha$  normalverteilt ist mit Erwartungswertvektor  $\tilde{\underline{a}}$  und Kovarianzmatrix  $\tilde{P}$ , d. h. es soll gelten

$$\underline{c}_\alpha \sim \mathcal{N}(\tilde{\underline{a}}, \tilde{P}). \quad (3.9)$$

Weiterhin erinnern wir daran, dass der Anfangszustand und die Vektoren  $\underline{\eta}_{t_k}$  für alle  $k = 1, \dots, N$  unkorreliert sind.

**Definition 3.4**

Es seien  $y_{t_0}, y_{t_1}, \dots, y_{t_k}$  die zufälligen Werte des logarithmierten Aktienpreisprozesses  $y = p$  auf dem in (1.4) eingeführten Zeitgitter bis einschließlich zum Zeitpunkt  $t_k$ ,  $k = 0, 1, \dots, N$ . Die durch diese Werte erzeugte kleinste  $\sigma$ -Algebra bezeichnen wir mit  $\mathfrak{F}_{t_k}$ , d. h. es gilt

$$\mathfrak{F}_{t_k} := \sigma(y_{t_0}, y_{t_1}, \dots, y_{t_k}) \quad (k = 0, 1, \dots, N). \quad (3.10)$$

Wir bemerken, daß die Familie  $(\mathfrak{F}_{t_k})_{k=0, \dots, N}$  von  $\sigma$ -Algebren eine Filtration bildet.

Man definiert für  $k, l = 0, 1, \dots, N$  und  $l \leq k$

$$\underline{a}_{t_k | t_l} := \mathbf{E}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_l}) \quad \text{und} \quad (3.11a)$$

$$\begin{aligned} P_{t_k | t_l} &:= \mathbf{Cov}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_l}) \\ &= \mathbf{E} \left[ (\underline{\alpha}_{t_k} - \mathbf{E}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_l})) (\underline{\alpha}_{t_k} - \mathbf{E}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_l}))^T | \mathfrak{F}_{t_l} \right]. \end{aligned} \quad (3.11b)$$

$\underline{a}_{t_k | t_l}$  ist der bedingte Erwartungswert und  $P_{t_k | t_l}$  die bedingte Kovarianzmatrix von  $\underline{\alpha}_{t_k}$  unter der Bedingung  $\mathfrak{F}_{t_l}$ . Für  $k = l$  schreiben wir kurz  $\underline{a}_{t_k}$  und  $P_{t_k}$ .

**Bemerkung 3.5** Die so definierten bedingten Erwartungswerte bezüglich einer  $\sigma$ -Algebra sind im allgemeinen Zufallsgrößen. Erst bei Fixierung eines  $\omega_0 \in \Omega$  können wir die bedingten Erwartungswerte als konkrete Zahlen (Realisierungen dieser Zufallsgrößen) auffassen. Für unseren konkreten Gauß-Fall, werden wir jedoch im Weiteren sehen, dass die Matrizen  $P_{t_k | t_{k-1}}$  bzw.  $P_{t_k}$  nicht zufällig sind. Insbesondere werden wir Formeln herleiten, mit denen diese Größen für jedes  $k$  berechnet werden können, vorausgesetzt die Parameter  $\gamma, \lambda, \beta$ , die Volatilitätsfunktion  $\sigma$  und  $\tilde{a}$  sowie  $\tilde{P}$  sind bekannt. Die bedingten Erwartungswerte  $\underline{a}_{t_k | t_l}$  sind jedoch auch in dem hier betrachteten Fall nichtentartete Zufallsgrößen, d. h. sie hängen von der konkreten Trajektorie  $\omega$  ab.

Wir betrachten nun den bedingten Erwartungswert und die bedingte Kovarianzmatrix der Zustände  $\underline{\alpha}_{t_k}$  unter der Bedingung  $\mathfrak{F}_{t_{k-1}}$ , kurz

$$\underline{a}_{t_k | t_{k-1}} = \mathbf{E}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_{k-1}}) \quad \text{und} \quad (3.12a)$$

$$P_{t_k | t_{k-1}} = \mathbf{Cov}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_{k-1}}). \quad (3.12b)$$

Wie wir wissen, sind die Vektoren  $\underline{\eta}_{t_k}$  und der Vektor  $\underline{\alpha}_{t_0}$  für alle  $k = 1, \dots, N$  vollständig stochastisch unabhängig (vgl. oben). Mit (3.8b) erhalten wir dann für  $k = 1, \dots, N$

$$\begin{aligned} \underline{a}_{t_k | t_{k-1}} &= \mathbf{E}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_{k-1}}) = \mathbf{E} \left( T \underline{\alpha}_{t_{k-1}} + R \underline{\eta}_{t_k} | \mathfrak{F}_{t_{k-1}} \right) \\ &= T \underbrace{\mathbf{E}(\underline{\alpha}_{t_{k-1}} | \mathfrak{F}_{t_{k-1}})}_{=\underline{a}_{t_{k-1}}} + R \underbrace{\mathbf{E}(\underline{\eta}_{t_k} | \mathfrak{F}_{t_{k-1}})}_{=\mathbf{E}\underline{\eta}_{t_k}=0} \\ &= T \underline{a}_{t_{k-1}} \quad \text{und} \end{aligned} \quad (3.13a)$$

$$\begin{aligned}
P_{t_k | t_{k-1}} &= \mathbf{Cov}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_{k-1}}) = \mathbf{E} \left[ \left( \underline{\alpha}_{t_k} - \underline{a}_{t_k | t_{k-1}} \right) \left( \underline{\alpha}_{t_k} - \underline{a}_{t_k | t_{k-1}} \right)^T \middle| \mathfrak{F}_{t_{k-1}} \right] \\
&= \mathbf{E} \left[ \left( T \underline{\alpha}_{t_{k-1}} + R \underline{\eta}_{t_k} - T \underline{a}_{t_{k-1}} \right) \left( T \underline{\alpha}_{t_{k-1}} + R \underline{\eta}_{t_k} - T \underline{a}_{t_{k-1}} \right)^T \middle| \mathfrak{F}_{t_{k-1}} \right] \\
&= \mathbf{E} \left[ \left( T \left( \underline{\alpha}_{t_{k-1}} - \underline{a}_{t_{k-1}} \right) + R \underline{\eta}_{t_k} \right) \left( T \left( \underline{\alpha}_{t_{k-1}} - \underline{a}_{t_{k-1}} \right) + R \underline{\eta}_{t_k} \right)^T \middle| \mathfrak{F}_{t_{k-1}} \right] \\
&= T \underbrace{\mathbf{E} \left[ \left( \underline{\alpha}_{t_{k-1}} - \underline{a}_{t_{k-1}} \right) \left( \underline{\alpha}_{t_{k-1}} - \underline{a}_{t_{k-1}} \right)^T \middle| \mathfrak{F}_{t_{k-1}} \right]}_{= \mathbf{Cov}(\underline{\alpha}_{t_{k-1}} | \mathfrak{F}_{t_{k-1}}) = P_{t_{k-1}}} T^T \\
&\quad + T \cdot 0 \cdot R^T + R \cdot 0 \cdot T^T + R \underbrace{\mathbf{E} \left( \underline{\eta}_{t_k} \underline{\eta}_{t_k}^T \middle| \mathfrak{F}_{t_{k-1}} \right)}_{= \mathbf{E} \underline{\eta}_{t_k} \underline{\eta}_{t_k}^T = Q_{t_k}} R^T \\
&= T P_{t_{k-1}} T^T + R Q_{t_k} R^T.
\end{aligned} \tag{3.13b}$$

Bei den verschwindenden Termen in den obigen Rechnungen haben wir die Unabhängigkeit von  $\underline{\alpha}_{t_{k-1}}$  und  $\underline{\eta}_{t_k}$  ausgenutzt. Diese lässt sich wie folgt zeigen. Nach der Gleichung (3.8b) können wir  $\underline{\alpha}_{t_{k-1}}$  schreiben als

$$\underline{\alpha}_{t_{k-1}} = T^{k-1} \underline{\alpha}_{t_0} + T^{k-2} R \underline{\eta}_{t_1} + \dots + T R \underline{\eta}_{t_{k-2}} + R \underline{\eta}_{t_{k-1}}. \tag{3.14}$$

Da die Vektoren  $\underline{\eta}_{t_1}, \dots, \underline{\eta}_{t_k}$  und  $\underline{\alpha}_{t_0}$  vollständig stochastisch unabhängig sind, ergibt sich aus der Darstellung (3.14) auch die stochastische Unabhängigkeit von  $\underline{\alpha}_{t_{k-1}}$  und  $\underline{\eta}_{t_k}$ .

Die Gleichungen (3.13a) und (3.13b) erlauben es, anhand der Daten  $y_{t_0}, y_{t_1}, \dots, y_{t_{k-1}}$  eine Schätzung für den Zustand  $\underline{\alpha}_{t_k}$  zu berechnen. Man nennt sie deshalb *Prognosegleichungen* (*prediction equations*).

Mit  $\underline{a}_{t_k | t_{k-1}}$  kann man auch eine Schätzung (Prognose) für  $y_{t_k}$  angeben. Man definiert analog für  $k = 1, \dots, N$

$$\begin{aligned}
y_{t_k | t_{k-1}} &:= \mathbf{E}(y_{t_k} | \mathfrak{F}_{t_{k-1}}) \\
&= \mathbf{E}(\underline{z}^T \underline{\alpha}_{t_k} + d_{t_k} | \mathfrak{F}_{t_{k-1}}) \\
&= \underline{z}^T \mathbf{E}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_{k-1}}) + d_{t_k} \\
&= \underline{z}^T \underline{a}_{t_k | t_{k-1}} + d_{t_k}.
\end{aligned} \tag{3.15}$$

Der Fehler der Schätzung (3.15) (Prognosefehler) wird definiert durch

$$\begin{aligned}
v_{t_k} &:= y_{t_k} - y_{t_k | t_{k-1}} \\
&= \underline{z}^T \left( \underline{\alpha}_{t_k} - \underline{a}_{t_k | t_{k-1}} \right) \quad (k = 1, \dots, N).
\end{aligned} \tag{3.16a}$$

Für den bedingten Erwartungswert und die bedingte Varianz von  $v_{t_k}$  unter der Bedingung

$\mathfrak{F}_{t_{k-1}}$  erhalten wir für  $k = 1, \dots, N$

$$\begin{aligned}
\mathbf{E}(v_{t_k} | \mathfrak{F}_{t_{k-1}}) &= \mathbf{E}(y_{t_k} - y_{t_k | t_{k-1}} | \mathfrak{F}_{t_{k-1}}) \\
&= \mathbf{E}(y_{t_k} | \mathfrak{F}_{t_{k-1}}) - y_{t_k | t_{k-1}} = 0 \quad \text{und} \\
f_{t_k} &:= \mathbf{E}(v_{t_k}^2 | \mathfrak{F}_{t_{k-1}}) = \mathbf{E}(v_{t_k} v_{t_k}^T | \mathfrak{F}_{t_{k-1}}) \\
&= \underline{z}^T \underbrace{\mathbf{E} \left[ \left( \underline{\alpha}_{t_k} - \underline{a}_{t_k | t_{k-1}} \right) \left( \underline{\alpha}_{t_k} - \underline{a}_{t_k | t_{k-1}} \right)^T \middle| \mathfrak{F}_{t_{k-1}} \right]}_{= \mathbf{Cov}(\underline{\alpha}_{t_k} | \mathfrak{F}_{t_{k-1}}) = P_{t_k | t_{k-1}}} \underline{z} \\
&= \underline{z}^T P_{t_k | t_{k-1}} \underline{z}.
\end{aligned} \tag{3.16b}$$

**Bemerkung 3.6** Die Größen  $v_{t_k}$  und  $f_{t_k}$  sind gerade die Resultate des Kalman-Filters, welche wir später zur Berechnung der Likelihood-Funktion benötigen. Fixieren wir  $\omega_0 \in \Omega$ , d. h. wir betrachten die durch  $\omega_0$  erzeugte Trajektorie, so sind die Werte  $v_{t_k}^* := v_{t_k}(\omega_0)$  und  $f_{t_k}^* := f_{t_k}(\omega_0)$  konkrete Zahlen.

Sobald uns ein neuer Beobachtungswert  $y_{t_k}$  zur Verfügung steht, können wir die Schätzung für  $\underline{\alpha}_{t_k}$  aktualisieren, d. h. wir berücksichtigen die zusätzliche Information, die uns der Wert  $y_{t_k}$  verschafft. Dies geschieht mit Hilfe der sogenannten *Korrekturgleichungen* (*updating equations*) (vgl. [2], [6])

$$\begin{aligned}
\underline{a}_{t_k} &= \underline{a}_{t_k | t_{k-1}} + P_{t_k | t_{k-1}} \underline{z} f_{t_k}^{-1} \left( y_{t_k} - \underline{z}^T \underline{a}_{t_k | t_{k-1}} - d_{t_k} \right) \quad (k = 1, \dots, N) \\
&= \underline{a}_{t_k | t_{k-1}} + P_{t_k | t_{k-1}} \underline{z} f_{t_k}^{-1} v_{t_k} \quad \text{und}
\end{aligned} \tag{3.17a}$$

$$P_{t_k} = P_{t_k | t_{k-1}} - P_{t_k | t_{k-1}} \underline{z} f_{t_k}^{-1} \underline{z}^T P_{t_k | t_{k-1}} \quad (k = 1, \dots, N). \tag{3.17b}$$

Die Herleitung der Korrekturgleichungen (3.17a) und (3.17b) ist ziemlich umfangreich, weshalb wir diese hier nicht vornehmen werden. Sie kann in [2] oder [8] nachgelesen werden.

Die Initialisierung des Kalman Filters geschieht wie folgt. Als Input wird der Erwartungswert  $\underline{\tilde{a}}$  und die Kovarianzmatrix  $\tilde{P}$  des Anfangszustands  $\underline{c}_\alpha$  benötigt, siehe auch (3.9). Nach der ersten Beobachtung (zum Zeitpunkt  $t_0$ ) berechnen wir  $\underline{a}_{t_0}$  und  $P_{t_0}$  durch

$$\begin{aligned}
\underline{a}_{t_0} &:= \mathbf{E}(\underline{\alpha}_{t_0} | \mathfrak{F}_{t_0}) = \underline{\tilde{a}} + \tilde{P} \underline{z} \tilde{f}^{-1} [y_{t_0} - \underline{z}^T \underline{\tilde{a}}] \\
P_{t_0} &:= \mathbf{Cov}(\underline{\alpha}_{t_0} | \mathfrak{F}_{t_0}) = \tilde{P} - \tilde{P} \underline{z} \tilde{f}^{-1} \underline{z}^T \tilde{P}.
\end{aligned}$$

mit  $\tilde{f} := \underline{z}^T \tilde{P} \underline{z}$  (siehe auch [6, S. 73]).

Anschließend werden die Prognoseschritte und der Korrekturschritte abwechselnd für  $k = 1, \dots, N$  ausgeführt, indem die entsprechenden Werte rekursiv berechnet werden. Die in den Rekursionen verwendeten Größen sind in der Tabelle 3.1 noch einmal übersichtlich zusammengefaßt.

**Bemerkung 3.7** Bei der Berechnung von  $P_{t_k | t_{k-1}}$  tritt das Produkt  $RQ_{t_k}R^T$  auf. Für die numerische Auswertung sollte man dabei berücksichtigen, dass der Term  $\frac{\lambda}{\gamma - \beta}$  in  $R$  für  $\gamma \approx \beta$  sehr groß wird. Da dies zu numerischen Problemen führen kann, ist es empfehlenswert, bei der numerischen Umsetzung die Matrizen  $R$ ,  $Q_{t_k}$  nicht getrennt aufzustellen und dann zu

| Größe                           | Definition  | Berechnung   |
|---------------------------------|---|--|
| $\tilde{\underline{a}}$         | $\mathbf{E} \underline{c}_\alpha = \mathbf{E} \underline{\alpha}_{t_0}$     | —  |
| $\tilde{P}$                     | $\mathbf{Cov} \underline{c}_\alpha = \mathbf{Cov} \underline{\alpha}_{t_0}$ | —  |
| $\tilde{f}$                     | $\mathbf{Cov} y_{t_0}$  | $\underline{z}^T \tilde{P} \underline{z}$  |
| $\underline{a}_{t_0}$           | $\mathbf{E} (\underline{\alpha}_{t_0}   \mathfrak{F}_{t_0})$                | $\tilde{\underline{a}} + \tilde{P} \underline{z} \tilde{f}^{-1} [y_{t_0} - \underline{z}^T \tilde{\underline{a}}]$ |
| $P_{t_0}$                       | $\mathbf{Cov} (\underline{\alpha}_{t_0}   \mathfrak{F}_{t_0})$              | $\tilde{P} - \tilde{P} \underline{z} \tilde{f}^{-1} \underline{z}^T \tilde{P}$                                     |
| $\underline{a}_{t_k   t_{k-1}}$ | $\mathbf{E} (\underline{\alpha}_{t_k}   \mathfrak{F}_{t_{k-1}})$            | $T \underline{a}_{t_{k-1}}$  |
| $P_{t_k   t_{k-1}}$             | $\mathbf{Cov} (\underline{\alpha}_{t_k}   \mathfrak{F}_{t_{k-1}})$          | $T P_{t_{k-1}} T^T + R Q_{t_k} R^T$  |
| $\underline{a}_{t_k}$           | $\mathbf{E} (\underline{\alpha}_{t_k}   \mathfrak{F}_{t_k})$                | $\underline{a}_{t_k   t_{k-1}} + P_{t_k   t_{k-1}} \underline{z} f_{t_k}^{-1} v_{t_k}$                             |
| $P_{t_k}$                       | $\mathbf{Cov} (\underline{\alpha}_{t_k}   \mathfrak{F}_{t_k})$              | $P_{t_k   t_{k-1}} - P_{t_k   t_{k-1}} \underline{z} f_{t_k}^{-1} \underline{z}^T P_{t_k   t_{k-1}}$               |
| $y_{t_k   t_{k-1}}$             | $\mathbf{E} (y_{t_k}   \mathfrak{F}_{t_{k-1}})$                             | $\underline{z}^T \underline{a}_{t_k   t_{k-1}} + d_{t_k}$  |
| $y_{t_k}$                       | —   | Werte können beobachtet werden   |
| $v_{t_k}$                       | $y_{t_k} - y_{t_k   t_{k-1}}$   | $\underline{z}^T (\underline{\alpha}_{t_k} - \underline{a}_{t_k   t_{k-1}})$ bzw. wie in der Definition            |
| $f_{t_k}$                       | $\mathbf{E} (v_{t_k}^2   \mathfrak{F}_{t_{k-1}})$                           | $\underline{z}^T P_{t_k   t_{k-1}} \underline{z}$  |

Tabelle 3.1: Größen in den Kalman-Filter-Rekursionen

multiplizieren, sondern sofort die Matrix  $\Sigma_{t_k} := RQ_{t_k}R^T$  aufzustellen. Die Einträge von  $\Sigma_{t_k}$  sehen wie folgt aus

$$\begin{aligned}\Sigma_{t_k}(1, 1) &= H(t_{k-1}, t_k) + \left(\frac{\lambda}{\gamma - \beta}\right)^2 \int_0^\tau (e^{-\gamma(\tau-s)} - e^{-\beta(\tau-s)})^2 ds \\ \Sigma_{t_k}(2, 1) = \Sigma_{t_k}(1, 2) &= \frac{\lambda}{\gamma - \beta} \int_0^\tau e^{-\beta(\tau-s)} (e^{-\beta(\tau-s)} - e^{-\gamma(\tau-s)}) ds \\ \Sigma_{t_k}(2, 2) &= \int_0^\tau e^{-2\beta(\tau-s)} ds.\end{aligned}$$

Wegen  $\lim_{\gamma \rightarrow \beta} \frac{e^{-\gamma(\tau-s)} - e^{-\beta(\tau-s)}}{\gamma - \beta} = (s - \tau)e^{-\beta(\tau-s)}$  sind diese Ausdrücke beschränkt.

### 3.3 Likelihood-Funktion

Bisher sind wir davon ausgegangen, daß der Anfangszustand  $\underline{a}_{t_0} = \underline{c}_\alpha$  normalverteilt ist mit Erwartungswertvektor  $\tilde{\underline{a}}$  und Kovarianzmatrix  $\tilde{P}$ . Um die weiteren Betrachtungen etwas zu vereinfachen, nehmen wir nun an, daß die Anfangsbedingung  $\underline{c}_\alpha$  identisch Null ist, d. h.  $\tilde{\underline{a}} = \mathbf{0}$  und  $\tilde{P} = \mathbf{0}$ . Der Wert  $y_{t_0}$  kann in diesem Fall als eine Zufallsgröße, die fast sicher den Wert Null annimmt, aufgefaßt werden. Ohne diese Einschränkung würden eventuell in  $\tilde{\underline{a}}$  bzw.  $\tilde{P}$  enthaltene zusätzliche Parameter, welche die Verteilung von  $q_0$  und  $X_0$  charakterisieren, mit in die Likelihood-Funktion eingehen.

#### 3.3.1 Gemeinsame Dichtefunktion der Beobachtungen

In Abschnitt 3.1 hatten wir den Vektor  $\underline{y} = (y_{t_0}, \dots, y_{t_N})$  von Beobachtungsdaten einer beliebigen Trajektorie des Prozesses  $y$  definiert. Aufgrund der in diesem Abschnitt hinzugefügten Anfangsbedingung  $y_{t_0} = 0$  ist die Beobachtung zum Zeitpunkt  $t_0$  nicht nötig. Die Initialisierung des Kalman Filters erfolgt in diesem Fall einfach durch  $\underline{a}_{t_0} = \mathbf{0}$  und  $P_{t_0} = \mathbf{0}$ .

Wir versuchen jetzt, eine Darstellung für die Dichtefunktion  $g_{y_{t_1}, \dots, y_{t_N}}(x_1, \dots, x_N, \underline{p})$  dieses zufälligen Vektors unter den Parametern  $\underline{p}$  zu finden. Dafür benötigen wir den folgenden Satz über die normale Korrelation an. Für den Beweis verweisen wir auf [6, S. 61ff.] oder auf [9].

#### Satz 3.8

Sei  $\underline{u} = (u_1, \dots, u_p)^T \sim \mathcal{N}_p(\underline{\mu}, \Sigma)$  ein  $p$ -dimensional normalverteilter Zufallsvektor mit Erwartungswertvektor  $\underline{\mu} := \mathbf{E} \underline{u}$  und Kovarianzmatrix

$$\Sigma := \mathbf{Cov}(\underline{u}) = \mathbf{E} \left[ (\underline{u} - \underline{\mu})(\underline{u} - \underline{\mu})^T \right].$$

Ferner seien

$$\underline{u}_1 = (u_1, \dots, u_r)^T \sim \mathcal{N}_r(\underline{\mu}_1, \Sigma_{11}) \quad \text{und} \quad \underline{u}_2 = (u_{r+1}, \dots, u_p)^T \sim \mathcal{N}_{p-r}(\underline{\mu}_2, \Sigma_{22})$$

mit  $1 \leq r < p$  zwei Teilvektoren von  $\underline{u}$ , d. h. es gilt

$$\underline{u} = \begin{pmatrix} \underline{u}_1 \\ \underline{u}_2 \end{pmatrix}, \quad \underline{\mu} = \begin{pmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{E} \underline{u}_1 \\ \mathbf{E} \underline{u}_2 \end{pmatrix} \quad \text{und}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{Cov}(\underline{u}_1, \underline{u}_1) & \mathbf{Cov}(\underline{u}_1, \underline{u}_2) \\ \mathbf{Cov}(\underline{u}_2, \underline{u}_1) & \mathbf{Cov}(\underline{u}_2, \underline{u}_2) \end{pmatrix}.$$

Dann ist die bedingte Verteilung von  $\underline{u}_2$  bei gegebenem  $\underline{u}_1$  eine  $(p-r)$ -dimensionale Normalverteilung mit Erwartungswertvektor

$$\tilde{\underline{\mu}} := \underline{\mu}_2 + \Sigma_{21}\Sigma_{11}^- (\underline{u}_1 - \underline{\mu}_1) \quad (3.18a)$$

und Kovarianzmatrix

$$\tilde{\Sigma} := \Sigma_{22} - \Sigma_{21}\Sigma_{11}^- \Sigma_{12}, \quad (3.18b)$$

wobei  $\Sigma_{11}^-$  eine verallgemeinerte Inverse von  $\Sigma_{11}$  ist.

**Bemerkung 3.9** Für eine verallgemeinerte Inverse  $\Sigma_{11}^-$  von  $\Sigma_{11}$  gilt  $\Sigma_{11}\Sigma_{11}^-\Sigma_{11} = \Sigma_{11}$ . Falls die Inverse  $\Sigma_{11}^{-1}$  existiert, so gilt  $\Sigma_{11}^- = \Sigma_{11}^{-1}$ .

Wir nehmen an, daß die Inverse  $\tilde{\Sigma}^{-1}$  existiert. Mit den Vektoren  $\underline{x}_1 = (x_1, \dots, x_r)^T$  und  $\underline{x}_2 = (x_{r+1}, \dots, x_p)^T$  als Teilvektoren von  $\underline{x} = (x_1, \dots, x_p)^T$  erhalten wir damit für die bedingte Dichte von  $\underline{u}_2$  unter der Bedingung  $\underline{u}_1 = \underline{x}_1$

$$g_{\underline{u}_2}(\underline{x}_2 | \underline{u}_1 = \underline{x}_1) = \frac{1}{(2\pi)^{\frac{r-p}{2}} (\det \tilde{\Sigma})^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2} (\underline{x}_2 - \tilde{\underline{\mu}})^T \tilde{\Sigma}^{-1} (\underline{x}_2 - \tilde{\underline{\mu}})\right). \quad (3.19)$$

Aus den Herleitungen in [9] wird auch ersichtlich, daß gilt

$$g_{\underline{u}_2}(\underline{x}_2 | \underline{u}_1 = \underline{x}_1) = \frac{g_{\underline{u}}(\underline{x})}{g_{\underline{u}_1}(\underline{x}_1)} \quad \text{für } g_{\underline{u}_1}(\underline{x}_1) > 0. \quad (3.20)$$

Wir kehren nun wieder zu unserem Modell zurück. Man kann zeigen, daß die gemeinsame Verteilung des zufälligen Vektors  $(y_{t_1}, \dots, y_{t_k})^T$  eine Normalverteilung ist. Der Satz 3.8 liefert uns, daß dann die bedingte Verteilung von  $y_{t_k}$ ,  $k \geq 2$  bei gegebenen Werten  $y_{t_1}, \dots, y_{t_{k-1}}$  ebenfalls eine Normalverteilung ist.

Mit

$$g_{y_{t_1}, \dots, y_{t_k}}(\cdot, \underline{p}) \quad (k = 1, \dots, N)$$

bezeichnen wir die gemeinsame Dichtefunktion des zufälligen Vektors  $(y_{t_1}, \dots, y_{t_k})^T$  unter dem Parametervektor  $\underline{p}$  und mit

$$g_{y_{t_k}}(\cdot, \underline{p} | y_{t_1} = x_1, \dots, y_{t_{k-1}} = x_{k-1}) \quad (k = 2, \dots, N)$$

die bedingte Dichtefunktion von  $y_{t_k}$  bei gegebenen Werten  $y_{t_1} = x_1, \dots, y_{t_{k-1}} = x_{k-1}$  unter dem Parametervektor  $\underline{p}$ .

Zur Vereinfachung und um die weiteren Rechnungen übersichtlicher zu halten, vereinbaren wir die folgenden Schreibweisen:

$$g_{y_{t_k}}(x_k, \underline{p} | x_1, \dots, x_{k-1}) := \begin{cases} g_{y_{t_k}}(x_k, \underline{p} | y_{t_1} = x_1, \dots, y_{t_{k-1}} = x_{k-1}) & k = 2, \dots, N \\ g_{y_{t_1}}(x_1, \underline{p}) & k = 1 \end{cases}$$

$$m(x_1, \dots, x_{k-1}) := \begin{cases} \mathbf{E}(y_{t_k} | y_{t_1} = x_1, \dots, y_{t_{k-1}} = x_{k-1}) & k = 2, \dots, N \\ \mathbf{E} y_{t_1} & k = 1 \end{cases}$$

$$s^2(x_1, \dots, x_{k-1}) := \begin{cases} \mathbf{D}^2(y_{t_k} | y_{t_1} = x_1, \dots, y_{t_{k-1}} = x_{k-1}) & k = 2, \dots, N \\ \mathbf{D}^2 y_{t_1} & k = 1 \end{cases}$$

Für  $k = 1, \dots, N$  können wir die Dichten  $g_{y_{t_k}}(x_k, \underline{p} | x_1, \dots, x_{k-1})$  nun wie folgt explizit angeben

$$g_{y_{t_k}}(x_k, \underline{p} | x_1, \dots, x_{k-1}) = \frac{1}{\sqrt{2\pi} s(x_1, \dots, x_{k-1})} \cdot \exp\left(-\frac{(x_k - m(x_1, \dots, x_{k-1}))^2}{2s^2(x_1, \dots, x_{k-1})}\right). \quad (3.21)$$

Außerdem folgt aus (3.20) mit  $g_{y_{t_1}, \dots, y_{t_{k-1}}}(x_1, \dots, x_{k-1}, \underline{p}) > 0$  für  $k = 2, \dots, N$  die Beziehung

$$g_{y_{t_k}}(x_k, \underline{p} | x_1, \dots, x_{k-1}) = \frac{g_{y_{t_1}, \dots, y_{t_k}}(x_1, \dots, x_k, \underline{p})}{g_{y_{t_1}, \dots, y_{t_{k-1}}}(x_1, \dots, x_{k-1}, \underline{p})}. \quad (3.22)$$

Durch wiederholte Anwendung der Beziehung (3.22) ergibt sich schließlich für die gesuchte Dichte  $g_{y_{t_1}, \dots, y_{t_N}}(x_1, \dots, x_N, \underline{p})$

$$\begin{aligned} g_{y_{t_1}, \dots, y_{t_N}}(x_1, \dots, x_N, \underline{p}) &= g_{y_{t_N}}(x_N, \underline{p} | x_1, \dots, x_{N-1}) \cdot g_{y_{t_1}, \dots, y_{t_{N-1}}}(x_1, \dots, x_{N-1}, \underline{p}) \\ &= \dots \\ &= g_{y_{t_N}}(x_N, \underline{p} | x_1, \dots, x_{N-1}) \cdot \dots \cdot g_{y_{t_2}}(x_2, \underline{p} | x_1) \cdot g_{y_{t_1}}(x_1, \underline{p}) \\ &= \prod_{k=1}^N g_{y_{t_k}}(x_k, \underline{p} | x_1, \dots, x_{k-1}) \\ &= \prod_{k=1}^N \frac{1}{\sqrt{2\pi} s(x_1, \dots, x_{k-1})} \cdot \exp\left(-\frac{(x_k - m(x_1, \dots, x_{k-1}))^2}{2s^2(x_1, \dots, x_{k-1})}\right). \end{aligned} \quad (3.23)$$

**Bemerkung 3.10** Falls man statt der Anfangsbedingung  $\alpha_{t_0} = 0$  weiterhin lediglich (3.9) mit einer regulären Kovarianzmatrix  $\tilde{P}$  fordern möchte, ist eine Beobachtung zum Zeitpunkt  $t_0$  sinnvoll. In diesem Fall bestimmt man die gemeinsame Dichtefunktion

$$g_{y_{t_0}, \dots, y_{t_k}}(\cdot, \underline{p}) \quad (k = 0, \dots, N)$$

des zufälligen Vektors  $(y_{t_0}, \dots, y_{t_k})^T$  unter dem Parametervektor  $\underline{p}$ . Dazu definiert man die bedingte Dichtefunktion von  $y_{t_k}$  bei gegebenen Werten  $y_{t_0} = x_0, \dots, y_{t_{k-1}} = x_{k-1}$  unter dem Parametervektor  $\underline{p}$  durch

$$g_{y_{t_k}}(\cdot, \underline{p} | y_{t_0} = x_0, \dots, y_{t_{k-1}} = x_{k-1}) \quad (k = 1, \dots, N).$$

Mit der Schreibweise

$$g_{y_{t_k}}(x_k, \underline{p} | x_0, \dots, x_{k-1}) := \begin{cases} g_{y_{t_k}}(x_k, \underline{p} | y_{t_0} = x_0, \dots, y_{t_{k-1}} = x_{k-1}) & k = 1, \dots, N \\ g_{y_{t_0}}(x_0, \underline{p}) & k = 0 \end{cases}$$

sowie

$$m(x_0, \dots, x_{k-1}) := \mathbf{E}(y_{t_k} | y_{t_0} = x_0, \dots, y_{t_{k-1}} = x_{k-1})$$

$$s^2(x_0, \dots, x_{k-1}) := \mathbf{D}^2(y_{t_k} | y_{t_0} = x_0, \dots, y_{t_{k-1}} = x_{k-1})$$

für  $k = 1, \dots, N$  können die Dichten  $g_{y_{t_k}}(x_k, \underline{p} | x_0, \dots, x_{k-1})$  für  $k = 1, \dots, N$  wie folgt explizit angegeben werden

$$g_{y_{t_k}}(x_k, \underline{p} | x_0, \dots, x_{k-1}) = \frac{1}{\sqrt{2\pi} s(x_0, \dots, x_{k-1})} \cdot \exp\left(-\frac{(x_k - m(x_0, \dots, x_{k-1}))^2}{2s^2(x_0, \dots, x_{k-1})}\right). \quad (3.24)$$

Weiterhin gilt mit  $\tilde{f} := \mathbf{D}^2 y_0$

$$g_{y_{t_0}}(x_0, \underline{p}) = \frac{1}{\sqrt{2\pi \tilde{f}}} \cdot \exp\left(-\frac{(x_0 - \tilde{a}(1))^2}{2\tilde{f}}\right).$$

Wie in (3.23) erhält man durch wiederholte Anwendung der Beziehung

$$g_{y_{t_k}}(x_k, \underline{p} | x_0, \dots, x_{k-1}) = \frac{g_{y_{t_0}, \dots, y_{t_k}}(x_0, \dots, x_k, \underline{p})}{g_{y_{t_0}, \dots, y_{t_{k-1}}}(x_0, \dots, x_{k-1}, \underline{p})} \quad (k = 1, \dots, N) \quad (3.25)$$

die gesuchte Dichte  $g_{y_{t_0}, \dots, y_{t_N}}(x_0, \dots, x_N, \underline{p})$

$$\begin{aligned} & g_{y_{t_0}, \dots, y_{t_N}}(x_0, \dots, x_N, \underline{p}) \\ &= g_{y_{t_0}}(x_0, \underline{p}) \prod_{k=1}^N \frac{1}{\sqrt{2\pi} s(x_0, \dots, x_{k-1})} \cdot \exp\left(-\frac{(x_k - m(x_0, \dots, x_{k-1}))^2}{2s^2(x_0, \dots, x_{k-1})}\right). \end{aligned}$$

### 3.3.2 Berechnung und Maximierung der Likelihood-Funktion

Wir betrachten jetzt die beobachteten Werte  $y_{t_1}^*, \dots, y_{t_N}^*$  der konkreten Trajektorie  $y^*$  und fixieren dasjenige  $\omega_0 \in \Omega$ , welches gerade unsere beobachtete Trajektorie  $y^*$  erzeugt hat, d. h. es gilt  $y^* = y(\omega_0)$ . Damit können wir  $m(y_{t_1}^*, \dots, y_{t_{k-1}}^*)$  und  $s^2(y_{t_1}^*, \dots, y_{t_{k-1}}^*)$  für  $k = 1, \dots, N$  folgendermaßen explizit berechnen (vgl. Abschnitt 3.2)

$$\begin{aligned} m(y_{t_1}^*, \dots, y_{t_{k-1}}^*) &= \mathbf{E}(y_{t_k} | \mathfrak{F}_{t_{k-1}})(\omega_0) = y_{t_k | t_{k-1}}(\omega_0) \\ &=: y_{t_k | t_{k-1}}^* \end{aligned} \quad (3.26a)$$

und

$$\begin{aligned} s^2(y_{t_1}^*, \dots, y_{t_{k-1}}^*) &= \mathbf{D}^2(y_{t_k} | \mathfrak{F}_{t_{k-1}})(\omega_0) = \mathbf{E}\left[(y_{t_k} - y_{t_k | t_{k-1}})^2 | \mathfrak{F}_{t_{k-1}}\right](\omega_0) \\ &= \mathbf{E}(v_{t_k}^2 | \mathfrak{F}_{t_{k-1}})(\omega_0) = f_{t_k}(\omega_0) \\ &=: f_{t_k}^*. \end{aligned} \quad (3.26b)$$

Zusammen mit der Notation  $v_{t_k}^* = y_{t_k}^* - y_{t_k | t_{k-1}}^*$  erhalten wir die folgende Darstellung für die gesuchte Likelihood-Funktion.

### Korollar 3.11

Für die Likelihood-Funktion der beobachteten Werte  $\underline{y}^* = (y_{t_1}^*, \dots, y_{t_N}^*)^T$  der konkreten Trajektorie  $y^*$  unter dem Parametervektor  $\underline{p}$  gilt

$$L(\underline{y}^* | \underline{p}) = \prod_{k=1}^N \frac{1}{\sqrt{2\pi f_{t_k}^*}} \cdot \exp\left(-\frac{v_{t_k}^{*2}}{2f_{t_k}^*}\right). \quad (3.27)$$

Die Werte  $v_{t_k}^*$  und  $f_{t_k}^*$  erhält man direkt aus den Kalman-Filter-Rekursionen. Da  $v_{t_k}^*$  gerade der Prognosefehler für  $y_{t_k}^*$  und  $f_{t_k}^*$  die Varianz dieses Fehlers ist (vgl. oben), bezeichnet man diese Darstellung der Likelihood-Funktion auch als *Prognosefehlerzerlegung* (*prediction error decomposition form*).

Wir formen (3.27) weiter um und erhalten

$$\begin{aligned} L(\underline{y}^* | \underline{p}) &= (2\pi)^{-\frac{N}{2}} \prod_{k=1}^N f_{t_k}^{*-\frac{1}{2}} \cdot \exp\left(-\frac{v_{t_k}^{*2}}{2f_{t_k}^*}\right) \\ &= (2\pi)^{-\frac{N}{2}} \cdot \left(\prod_{k=1}^N f_{t_k}^{*-\frac{1}{2}}\right) \cdot \exp\left(-\frac{1}{2} \sum_{k=1}^N f_{t_k}^{*-1} v_{t_k}^{*2}\right). \end{aligned}$$

Der Übergang zur Loglikelihood-Funktion  $\ln L(\underline{y}^* | \underline{p})$  liefert uns schließlich das folgende Maximierungsproblem zur Bestimmung des gesuchten Parametervektors  $\underline{p}$  aus den uns zur Verfügung stehenden Daten  $\underline{y}^*$

$$\ln L(\underline{y}^* | \underline{p}) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{k=1}^N \ln f_{t_k}^* - \frac{1}{2} \sum_{k=1}^N f_{t_k}^{*-1} v_{t_k}^{*2} \longrightarrow \max \quad \text{für } \underline{p} \in D. \quad (3.28)$$

Indem wir in (3.28) das Absolutglied weglassen und anschließend mit  $(-2)$  durchmultiplizieren, erhalten wir die dazu äquivalente Minimierungsaufgabe

$$\sum_{k=1}^N \ln f_{t_k}^* + \sum_{k=1}^N f_{t_k}^{*-1} v_{t_k}^{*2} \longrightarrow \min \quad \text{für } \underline{p} \in D. \quad (3.29)$$

Dieses Minimierungsproblem können wir nun mittels einer geeigneten Optimierungsroutine lösen.

## 4 Numerische Fallstudien

In diesem Abschnitt werden wir die Ergebnisse aus den vorangegangenen Abschnitten anhand numerischer Fallstudien überprüfen. Wir geben uns dazu Parametervektoren  $\underline{p}^*$  (die exakten Werte) vor und versuchen dann wiederum genau diese Parameter mittels der erarbeiteten Schätzverfahren aus simulierten Trajektorien des logarithmierten Aktienpreisprozesses  $p_t$  zurückzugewinnen. Bevor wir uns jedoch der Parameterschätzung zuwenden, erörtern wir im nächsten Abschnitt eine Möglichkeit, die für unsere Versuche benötigten Trajektorien zu simulieren. Wir gehen dabei auch auf die Möglichkeit der Skalierung des Modells ein.

## 4.1 Simulation der Trajektorien

Es gibt verschiedene Methoden, Trajektorien des hier untersuchten bivariaten Ornstein-Uhlenbeck Modells zu simulieren. So sind einerseits Simulationsalgorithmen denkbar, die auf der expliziten Lösung des Ausgangssystems (1.3) basieren und andererseits solche, die eine Diskretisierung dieses Systems stochastischer Differentialgleichungen zur Grundlage haben. In [10] wird für beide Varianten jeweils ein Algorithmus vorgestellt. Wir verwenden ein Verfahren, das auf dem in Abschnitt 3.2 eingeführten Zustandsraummodell des logarithmierten Preisprozesses  $p_t$  und damit auf der expliziten Lösung beruht. Der größte Vorteil liegt hierbei darin, daß die Rechengenauigkeit unabhängig von der Anzahl der Stützstellen ist. Bei approximativen Verfahren, wie z. B. der Diskretisierung, ist dies nicht der Fall (vgl. [10]).

Wir beschränken uns wieder auf den Fall verschwindender Anfangsbedingungen, d. h. es soll gelten  $q_0 = X_0 = 0$ . Damit ergibt sich unmittelbar  $y_{t_0} = 0$ . Beginnend mit diesem Startwert simulieren wir die Werte  $y_{t_k}$  für die Zeitpunkte  $t_1, \dots, t_N$  rekursiv anhand der Beobachtungsgleichung (3.8a) und der Zustandsgleichung (3.8b). Die Matrizen  $T$ ,  $R$  und  $Q_{t_k}$  sowie der Wert  $d_{t_k}$  sind für alle Zeitpunkte  $t_k$  eindeutig durch die vorgegebenen Parameter bestimmt (vgl. Abschnitt 3.2). Wir benötigen lediglich noch ein Verfahren zur Simulation der normalverteilten Störungsvektoren  $\underline{\eta}_{t_k}$ . Dabei hilft uns der folgende wohlbekanntes Satz.

### Satz 4.1

Sei  $\underline{u} = (u_1, u_2)^T$  ein zweidimensionaler normalverteilter Zufallsvektor mit Erwartungswertvektor  $\mathbf{E} \underline{u} = 0$  und Kovarianzmatrix  $\mathbf{E} \underline{u} \underline{u}^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$ . Sei weiterhin  $\underline{a} \in \mathbb{R}^2$  beliebig und  $B \in \mathbb{R}^{2 \times 2}$  eine symmetrische positiv definite Matrix mit der Cholesky-Zerlegung  $B = LL^T$ . Die Matrix  $L \in \mathbb{R}^{2 \times 2}$  ist dabei eine linke untere Dreiecksmatrix. Dann ist der Zufallsvektor  $\underline{v} = \underline{a} + L\underline{u}$  zweidimensional normalverteilt mit Erwartungswertvektor  $\underline{a}$  und Kovarianzmatrix  $B$ , d. h. es gilt

$$\underline{v} = \underline{a} + L\underline{u} \sim \mathcal{N}_2(\underline{a}, B).$$

Bei der Simulation des Prozesses  $\underline{q}$  werden wir die Cholesky-Zerlegung der Matrizen  $Q_{t_k}$  für  $k = 1, \dots, N$  benötigen. Die Symmetrie dieser Matrizen ist offensichtlich. Die positive Definitheit zeigen wir in dem folgenden Lemma.

### Lemma 4.2

Die durch (3.7) gegebenen Matrizen  $Q_{t_k}$  sind unter den in Modell 1.1 getroffenen Voraussetzungen für alle  $k = 1, \dots, N$  positiv definit.

**Beweis:** Zunächst bemerken wir, dass für beliebiges  $k = 1, \dots, N$

$$H(t_{k-1}, t_k) := \int_{t_{k-1}}^{t_k} \sigma^2(s) e^{-2\gamma(t_k-s)} ds > 0$$

gilt, da der Integrand für alle  $s \in [t_{k-1}, t_k]$  größer als 0 ist. Weiterhin hatten wir die Funktion  $h_\tau : [0, \infty) \rightarrow \mathbb{R}$  in (3.6) gerade so definiert, dass

$$h_\tau(c) = \int_0^\tau e^{-c(\tau-s)} ds$$

gilt. Insbesondere folgt daraus  $h_\tau(c) > 0$  für alle  $c \geq 0$ . Weiterhin erhalten wir mit der Cauchy-Schwartzschen Ungleichung

$$\left( \int_0^\tau e^{-(\beta+\gamma)(\tau-s)} ds \right)^2 \leq \int_0^\tau e^{-2\beta(\tau-s)} ds \int_0^\tau e^{-2\gamma(\tau-s)} ds$$

und damit  $h_\tau(\beta + \gamma)^2 \leq h_\tau(2\beta)h_\tau(2\gamma)$ .

Mit Hilfe dieser Überlegungen folgt nun sofort

$$Q_{t_k}(1, 1) = H(t_{k-1}, t_k) + \left(\frac{\lambda}{\gamma - \beta}\right)^2 h_\tau(2\gamma) > 0$$

sowie

$$\det Q_{t_k} = H(t_{k-1}, t_k)h_\tau(2\beta) + \left(\frac{\lambda}{\gamma - \beta}\right)^2 h_\tau(2\gamma)h_\tau(2\beta) - \left(\frac{\lambda}{\gamma - \beta}\right)^2 h_\tau(\beta + \gamma)^2 > 0$$

und damit die positive Definitheit von  $Q_{t_k}$ . ■

Seien nun  $u_{k,j}$  ( $k = 1, \dots, N$ ,  $j = 1, 2$ ) vollständig unabhängige standardnormalverteilte Zufallszahlen. Dann können wir mit der Folge

$$\{\underline{u}_k = (u_{k,1}, u_{k,2})^T\}_{k=1, \dots, N}$$

vollständig unabhängiger auf  $\mathbb{R}^2$  standardnormalverteilter Zufallszahlenvektoren und der Cholesky-Zerlegung  $L_{t_k}L_{t_k}^T$  der Kovarianzmatrix  $Q_{t_k}$  die Vektoren  $\underline{\eta}_{t_k}$  folgendermaßen simulieren

$$\underline{\eta}_{t_k} = L_{t_k}\underline{u}_k \quad (k = 1, \dots, N). \quad (4.1)$$

### Skalierung des Modells

Aus numerischen Gründen verwenden wir nicht die Werte  $y_t = q_t + \mu t$ , sondern die mit einem reellen Faktor  $c > 1$  skalierten Größen

$$\tilde{y}_t := cy_t = cq_t + c\mu t \quad (k = 1, \dots, N). \quad (4.2)$$

In Versuchsrechnungen haben wir bemerkt, daß sich in vielen Fällen mit  $c = 100$  bessere Ergebnisse einstellen, als dies ohne die Skalierung der Fall wäre. Auf die Ursachen dieses Verhaltens werden wir an dieser Stelle jedoch nicht näher eingehen.

Wir erreichen die Skalierung, indem wir den Parameter  $\mu$  durch  $\tilde{\mu} := c\mu$  und den Prozeß  $q_t$  durch den skalierten Prozeß  $\tilde{q}_t := cq_t$  ersetzen. Zu untersuchen bleibt, wie wir die anderen Parameter und die Volatilitätsfunktion  $\sigma(t)$  verändern müssen, um eine Multiplikation des Prozesses  $q$  mit dem Faktor  $c$  zu erreichen. Dazu betrachten wir wieder das Ausgangssystem (1.3).

Es gilt

$$\begin{aligned} dq_t &= -(\gamma q_t - \lambda X_t) dt + \sigma(t) dW_t^q \\ \iff q_t - q_0 &= -\int_0^t (\gamma q_s - \lambda X_s) ds + \int_0^t \sigma(s) dW_s^q \\ \iff \tilde{q}_t - \tilde{q}_0 &= -\int_0^t (\gamma \tilde{q}_s - c\lambda X_s) ds + \int_0^t c\sigma(s) dW_s^q \\ \iff d\tilde{q}_t &= -(\gamma \tilde{q}_t - c\lambda X_t) dt + c\sigma(t) dW_t^q \end{aligned}$$

mit der Anfangsbedingung  $\tilde{q}_0 = cq_0$ . Folglich müssen wir zur Skalierung des Prozesses  $q$  mit dem Faktor  $c$  noch den Parameter  $\lambda$  durch  $\tilde{\lambda} := c\lambda$  und die Volatilitätsfunktion  $\sigma(t)$  durch die Funktion  $\tilde{\sigma}(t) := c\sigma(t)$  ersetzen. Die Parameter  $\gamma$  und  $\beta$  lassen wir unverändert.

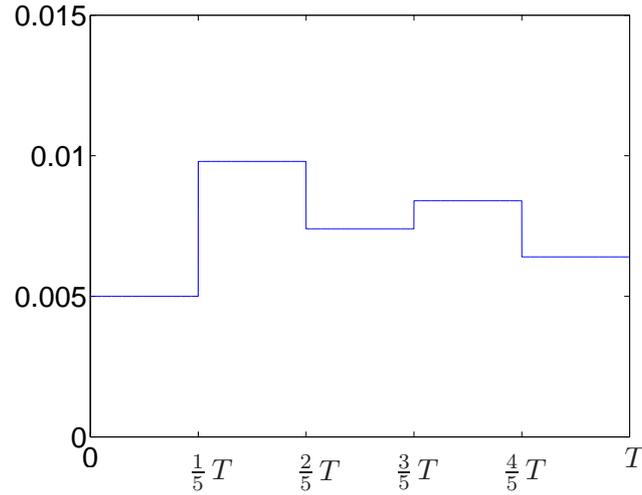


Abbildung 4.1: verwendete Volatilitätsfunktion  $\sigma_2(t)$

|                               |   |
|-------------------------------|---|
| $\gamma, \lambda, \beta, \mu$ | zu schätzende Modellparameter               |
| $\sigma_X$                    | Volatilität des Driftprozesses $X(t)$       |
| $\sigma_1$ bzw. $\sigma_2$    | Volatilität des Log-Preisprozesses $q(t)$   |
| $q_0, X_0$                    | Anfangsbedingungen                          |
| $T$                           | Länge des Beobachtungszeitraumes in Tagen   |
| $N$                           | Anzahl der Beobachtungen des Preisprozesses |
| $\tau$                        | Lag   |
| $c$                           | Skalierungsfaktor                           |

Tabelle 4.1: Bezeichnungen

## 4.2 Parameterschätzung

Die in dieser Arbeit entwickelten Schätzverfahren haben wir in verschiedenen Fallstudien insbesondere auf die am Ende von Abschnitt 1 formulierten Fragestellungen hin überprüft. Zur Vereinfachung sind wir dabei stets von verschwindenden Anfangsbedingungen  $q_0$  und  $X_0$  sowie einer (stückweise) konstanten Volatilitätsfunktion ausgegangen. Unsere ersten Rechnungen haben wir mit einer konstanten Volatilitätsfunktion  $\sigma_1 \equiv 0.0074$  durchgeführt. Für die in diesem Paper präsentierten numerischen Ergebnisse haben wir die stückweise konstante Volatilitätsfunktion

$$\sigma_2(t) = \begin{cases} 0.0050 & t \in [0, 0.2T) \\ 0.0098 & t \in [0.2T, 0.4T) \\ 0.0074 & t \in [0.4T, 0.6T) \\ 0.0084 & t \in [0.6T, 0.8T) \\ 0.0064 & t \in [0.8T, T] \end{cases} \quad (4.3)$$

verwendet (siehe Bild 4.1). Ferner wurde, wie in Abschnitt 2.1 erläutert,  $\sigma_X = 1$  gesetzt. Die in den numerischen Fallstudien verwendeten Bezeichnungen sind noch einmal in Tabelle 4.1 zusammengefaßt.

| $\gamma$ | $\lambda$ | $\beta$ | $\mu$      | $\sigma_X$ | $q_0$ | $X_0$ | $\tau$ | $c$ |
|----------|-----------|---------|------------|------------|-------|-------|--------|-----|
| 0.3748   | 0.004797  | 0.0106  | 1.6483e-04 | 1          | 0     | 0     | 1      | 100 |

Tabelle 4.2: Parameterwerte und weitere Einstellungen

Da der Parameter  $\mu$  in unserem Modell eine gewisse Sonderstellung einnimmt, wollen wir diesen auch gesondert schätzen, d. h. bei der Bestimmung der Parameter aus den simulierten Trajektorien sind wir in der nachstehenden Reihenfolge vorgegangen:

1. Schätzung des Parameters  $\mu$  mit der Schätzfunktion  $\hat{\mu}$  und
2. Schätzung der verbliebenen Parameter  $\gamma$ ,  $\lambda$  und  $\beta$  mittels der Maximum-Likelihood-Methode. Dazu werden die in Abschnitt 3.2 vorgestellten Kalman-Filter-Rekursionen angewendet und anschließend das Minimierungsproblem (3.29) numerisch in MATLAB gelöst. Als Wert für den Parameter  $\mu$  wird dem Algorithmus der zuvor im ersten Schritt ermittelte Schätzwert übergeben.

Der für unsere Fallstudien verwendete Parametervektor  $\underline{p}^*$  sowie die weiteren Einstellungen sind in Tabelle 4.2 angegeben. Die Werte haben wir dabei größtenteils in Anlehnung an [7] gewählt (siehe auch Abschnitt 2.1). Zur Untersuchung der in der Einleitung formulierten Fragestellungen werden wir gegebenenfalls einzelne Größen ceteris paribus variieren. Diese veränderten Werte geben wir dann an den entsprechenden Stellen noch einmal gesondert an.

Als Startwerte für die Maximierung der Likelihood-Funktion wählen wir, soweit nichts anderes angegeben ist, die Werte

$$(\gamma_0, \lambda_0, \beta_0)^T = (0.5, 0.002, 0.03)^T.$$

#### 4.2.1 Einfluß von Beobachtungsdauer und Beobachtungshäufigkeit

In dieser Fallstudie haben wir mit den Daten aus Tabelle 4.2 jeweils 100 Trajektorien für verschiedene Beobachtungszeiträume  $T$  simuliert. Aus diesen Daten wurden anschließend die Parameter  $\gamma$ ,  $\lambda$ ,  $\beta$  und  $\mu$  geschätzt. In den Tabellen 4.3 und 4.4 sind die empirischen Mittelwerte und die empirischen Varianzen der Ergebnisse angegeben.

Neben der Auswertung mittels Tabellen, welche Zahlenwerte wie empirische Momente oder die Anteile der Schätzungen mit einem gewissen relativen Fehler wiedergeben, verwenden wir auch Diagramme wie z.B. Abbildung 4.2. Für jeden der vier Parameter  $\gamma$ ,  $\lambda$ ,  $\beta$  und  $\mu$  ist darin der exakte Wert mit einem roten Strich gekennzeichnet und jeder einzelne Schätzwert wurde durch eine blaue Linie dargestellt. Weiterhin bezeichnen die rot gestrichelten Linien die 5% und die 95%-Quantile.

Wir erkennen, daß längere Beobachtungszeiträume bessere Ergebnisse liefern als kürzere. Angesichts der größeren Anzahl an verfügbaren Preisdaten sollte dies auch nicht überraschen. Die Resultate fallen aber für die einzelnen Parameter unterschiedlich aus. So werden  $\gamma$  und  $\lambda$  auch für kurze Beobachtungsdauern im Mittel bereits sehr gut geschätzt, während sich für  $\beta$  und  $\mu$  noch Probleme ergeben. Bei kurzen Beobachtungszeiträumen reichte vermutlich die in den Aktienpreisen enthaltene Information über den Parameter  $\beta$  nicht aus, um eine gute Näherung für diesen Parameter zu erhalten. In dem Diagramm 4.2, welches die Ergebnisse für unser Experiment mit  $T = 1000$  wiedergibt, sieht man deutlich, daß der

| Beobachtungs-<br>zeitraum $T$ | Empirischer Mittelwert der Schätzungen |               |             |           |
|-------------------------------|--|---------------|-------------|-----------|
|                               | für $\gamma$                           | für $\lambda$ | für $\beta$ | für $\mu$ |
| 1000                          | 0.37969                                | 0.0048555     | 0.016964    | 0.016028  |
| 5000                          | 0.37205                                | 0.00479       | 0.011494    | 0.016625  |
| 10000                         | 0.37729                                | 0.00482       | 0.010877    | 0.016502  |
| 20000                         | 0.37312                                | 0.00477       | 0.010749    | 0.016488  |

Tabelle 4.3: Empirische Mittelwerte der Schätzungen für verschiedene Beobachtungszeiträume  $T$

| Beobachtungs-<br>zeitraum $T$ | Empirische Varianz der Schätzungen |               |             |           |
|-------------------------------|------------------------------------|---------------|-------------|-----------|
|                               | für $\gamma$                       | für $\lambda$ | für $\beta$ | für $\mu$ |
| 1000                          | 0.011421                           | 9.6736e-7     | 1.0232e-5   | 4.6774e-5 |
| 5000                          | 0.0018243                          | 1.5602e-7     | 5.0364e-6   | 4.8782e-7 |
| 10000                         | 0.00086512                         | 7.1921e-8     | 2.1500e-6   | 5.4673e-8 |
| 20000                         | 0.00042841                         | 3.7073e-8     | 1.4192e-6   | 8.1588e-9 |

Tabelle 4.4: Empirische Varianzen der Schätzungen für verschiedene Beobachtungszeiträume  $T$

geschätzte Parameter  $\hat{\beta}$  häufig zu nah an dem Startwert der Iteration  $\beta_0 = 0.03$  liegt. Für einen Zeitraum von  $T = 5000$  Tagen traten diese Probleme nicht mehr auf.

Da uns in der Praxis nur eine Trajektorie zur Verfügung steht, interessiert uns vor allem die Streuung der Ergebnisse. Tabelle 4.5 zeigt uns, daß, abgesehen von  $\mu$ , bei einer Beobachtungsdauer von 5000 Tagen noch ein nicht unwesentlicher Teil aller Schätzungen relative Fehler von 20 Prozent und mehr aufweist. Am schwierigsten erscheint die Schätzung des Rückzugsparmeters  $\beta$ , bei der selbst für einen Beobachtungszeitraum von 20000 Tagen noch ca. 40% der Schätzwerte einen relativen Fehler von mehr als 20 Prozent besitzen. Die geringste Schwankungsbreite weisen zweifellos die Ergebnisse für  $\mu$  auf. Für  $T \geq 5000$  konnten wir bei nahezu keiner Schätzung einen Fehler von mehr als 20 Prozent nachweisen.

Bei der Bewertung dieser Ergebnisse ist generell zu bemerken, daß ein Beobachtungszeitraum von 5000 Tagen (ca. 14 Jahre) für praktische Untersuchungen unrealistisch lang ist. Andererseits hat uns das in Abschnitt 2 dargestellte Beispiel für die Schätzung des Driftparameters  $\mu$  im Black-Scholes Modell gelehrt, dass es bei der Schätzung von Driftparametern nicht ausreicht, viele Beobachtungen auf einem kurzen Zeitintervall durchzuführen. Konsistenz des Schätzers erreicht man im Allgemeinen nur für die Asymptotik  $T \rightarrow \infty$ . Weiterhin ist zu bemerken, dass ein relativer Fehler von 20 Prozent schon ein recht gutes Ergebnis darstellt. Insbesondere sind natürlich die Parameter  $\gamma$ ,  $\lambda$ ,  $\beta$  recht klein, so daß ein relativer Fehler von 20 Prozent einem sehr kleinen absoluten Fehler entspricht. In der Praxis nimmt man vielleicht auch größere Fehler in Kauf.

Die Verringerung der Schwankungsbreite für wachsenden Beobachtungszeitraum kann auch sehr schön anhand der Diagramme 4.2 und 4.3, welche jeweils die Schätzungen mit den exakten Parametern vergleichen, abgelesen werden.

Für die konstante Volatilität  $\sigma_1$  haben wir auch mit größeren Beobachtungsabständen ( $\tau = 2, \dots, 5$ ) experimentiert. Dabei haben wir für hinreichend lange Beobachtungszeiträume im Mittel akzeptable Schätzungen erhalten. Die Streuung nahm aber im Vergleich zu täglichen Beobachtungen zu, da schlicht weniger Preisdaten zur Schätzung der Parameter verfügbar

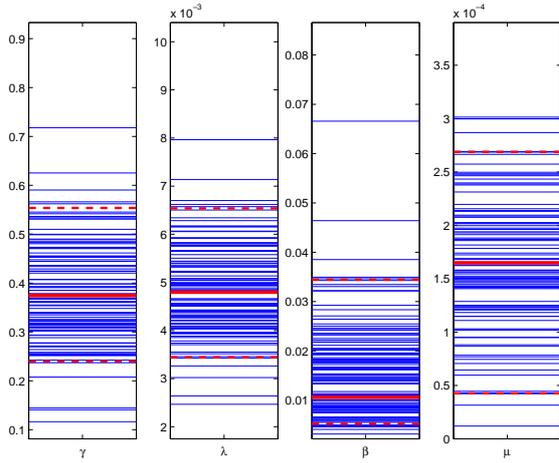


Abbildung 4.2: Ergebnisse für die Situation  $T = 1000$

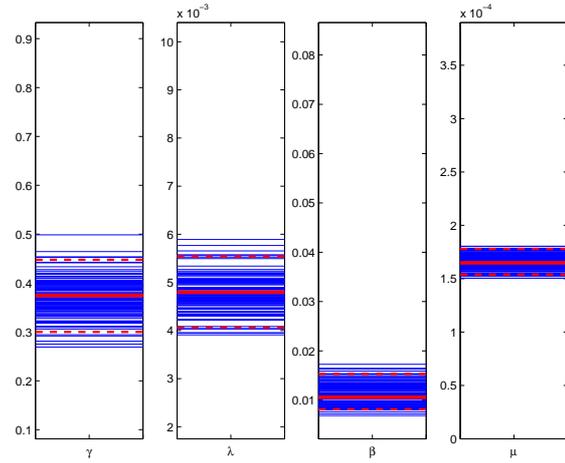


Abbildung 4.3: Ergebnisse für die Situation  $T = 5000$

| Beobachtungs-<br>zeitraum $T$ | Anteil der Schätzungen mit einem relativen Fehler $\leq 20\%$ |               |             |           |
|-------------------------------|---|---------------|-------------|-----------|
|                               | für $\gamma$  | für $\lambda$ | für $\beta$ | für $\mu$ |
| 1000                          | 0.26  | 0.35          | 0.09        | 0.20      |
| 5000                          | 0.65  | 0.83          | 0.33        | 1.00      |
| 10000                         | 0.80  | 0.92          | 0.57        | 1.00      |
| 20000                         | 0.89  | 0.98          | 0.59        | 1.00      |

Tabelle 4.5: Anteile der Schätzungen mit relativen Fehlern von maximal 20 Prozent für verschiedene Beobachtungszeiträume  $T$

sind. Bei sehr großen Lags ( $\tau > 10$ ) verschlechterten sich die Schätzungen allerdings deutlich und wurden zum Teil ganz unbrauchbar.

Wir wollen noch kurz auf den für die Schätzungen notwendigen Rechenaufwand eingehen. Es ist plausibel, anzunehmen, daß dieser mit wachsender Anzahl  $N$  der Beobachtungen des Preisprozesses ebenfalls steigt. Die Frage ist, wie diese Zunahme erfolgt. Bei unseren Simulationen mit Tagesdaten ( $\tau = 1$ ) und einer konstanten Volatilität  $\sigma_1 \equiv 0.0074$  haben wir festgestellt, daß der Rechenaufwand im wesentlichen linear mit der Beobachtungsdauer zunimmt. Die durchschnittlichen Rechenzeiten pro Trajektorie beliefen sich für Beobachtungszeiträume von 1000, 5000 bzw. 10000 Tagen auf gerundet 16, 85 bzw. 163 Sekunden<sup>1</sup>.

Weiterhin bemerken wir, dass wir für eine konstante Volatilitätsfunktion  $\sigma_1 \equiv 0.0074$  auch die Parameter  $\gamma$ ,  $\beta$  und  $\lambda$  variiert und die Güte der Schätzungen verglichen haben. Die Ergebnisse sollen hier nicht alle mit Tabellen dargestellt werden, sondern nur kurz qualitativ ausgewertet und interpretiert werden. Aufgrund unserer theoretischen Überlegungen in Abschnitt 2 hatten wir folgende Effekte erwartet.

- Für kleine Werte  $\lambda$  wird der Parameter  $\beta$  schlecht geschätzt, da in diesen Situationen sein Einfluß auf den Preisprozess gering ist.
- Die deterministische Drift  $\mu$  wird immer dann besonders gut geschätzt, wenn die Rückzugsparameter  $\gamma$  und  $\beta$  groß und  $\lambda$  klein ist.

<sup>1</sup>Verwendeter Rechner: Microstar 1860 Professional, Pentium 4 1,8 GHz, 100 (400 QDR) MHz FSB, 256 KB L2-Cache, 1 GB RAM, Windows XP SP2 (Home Edition).

Neben diesen erwarteten Effekten haben wir festgestellt, daß es für kleine Parameter  $\gamma$  bei einigen Trajektorien zu sehr langen Rechenzeiten kam, d. h. bei der Optimierung wurden sehr viele Iterationen durchgeführt. Die genaue Ursache dieser Effekte konnte noch nicht geklärt werden. Wir vermuten jedoch, daß die Zielfunktion für diese Trajektorien vielleicht Sattelpunkte besitzen könnte. Wie auch die folgenden Untersuchungen andeuten, scheint die Zielfunktion (negative Loglikelihoodfunktion) in manchen Fällen mehrere lokale Minima zu besitzen, wodurch die Optimierung erschwert wird.

#### 4.2.2 Parameteruneindeutigkeiten

Aufgrund unserer Betrachtungen in Abschnitt 2 haben wir vermutet, dass die dort beschriebenen Parameteruneindeutigkeiten für zeitabhängige Volatilitätsfunktionen im Allgemeinen nicht auftreten. Wir betrachten dazu nochmals die in Abschnitt 2 eingeführten Parametersituationen **Par1\*** und **Par2\***.

Für unser Experiment haben wir Trajektorien mit der Parametersituation **Par1\*** und der konstanten Volatilitätsfunktion  $\sigma_1 \equiv 0.0074$  bzw. der in Abbildung 4.1 dargestellten stückweise konstanten Volatilitätsfunktion  $\sigma_2$  simuliert. Dabei sollte bemerkt werden, dass die Funktion  $\sigma_2$  um den Wert 0.0074 schwankt. Für jede Trajektorie haben wir die Optimierung der negativen Likelihoodfunktion mehrmals durchgeführt, wobei wir unterschiedliche Startwerte  $\underline{p}_0^k$  ( $k = 1, \dots, 4$ ) für die Iteration verwendet haben, siehe Tabelle 4.6. Die Startwerte  $\underline{p}_0^1$  und  $\underline{p}_0^2$  entsprechen dabei genau den Parametersituationen **Par1\*** und **Par2\***. Die Startwerte  $\underline{p}_0^3$  bzw.  $\underline{p}_0^4$  liegen in der Nähe von  $\underline{p}_0^1$  bzw.  $\underline{p}_0^2$ . Die am Ende der Optimierung mit Startwert  $\underline{p}_0^k$  erhaltenen Ergebnisse bezeichnen wir mit  $\hat{p}^k$ .

Für die konstante Volatilitätsfunktion würden wir aufgrund unserer theoretischen Untersuchungen erwarten, dass die Ergebnisse  $\hat{p}^1$  und  $\hat{p}^3$  nahezu identisch sind und in der Nähe der Parametersituation **Par1\*** liegen, d. h. es gilt  $\hat{p}^1 \approx \hat{p}^3 \approx \underline{p}_0^1$ . Analog würden wir  $\hat{p}^2 \approx \hat{p}^4 \approx \underline{p}_0^2$  erwarten. Diese Erwartung hat sich bestätigt und die Werte der Zielfunktion

$$Z(\underline{p}) := \sum_{k=1}^N \ln f_{t_k}^* + \sum_{k=1}^N f_{t_k}^{*-1} v_{t_k}^{*2}$$

in den Punkten  $\hat{p}^k$  waren nahezu identisch. Die linke Grafik in Abbildung 4.4 stellt für zehn simulierte Trajektorien die relativen Abweichungen der Zielfunktionswerte

$$\frac{Z(\hat{p}^1) - Z(\hat{p}^2)}{Z(\hat{p}^1)} \quad (4.4)$$

dar. Wie man deutlich sieht, sind die Abweichungen generell sehr klein (in der Größenordnung  $10^{-5}$  und können sowohl positive als auch negative Werte annehmen).

Für eine zeitabhängige Volatilitätsfunktion würden wir bei einer exakten Minimierung erwarten, dass die Ergebnisse  $\hat{p}^k$  für alle  $k = 1, \dots, 4$  in der Nähe der Parametersituation **Par1\*** liegen. Dies zeigte sich jedoch nicht in den numerischen Experimenten. Genau wie bei der konstanten Volatilitätsfunktion galt für jede Trajektorie  $\hat{p}^1 \approx \hat{p}^3 \approx \underline{p}_0^1$  sowie  $\hat{p}^2 \approx \hat{p}^4 \approx \underline{p}_0^2$ . Allerdings war hier ein deutlicher Unterschied in den Werten der Zielfunktion in den Punkten  $\hat{p}^k$  festzustellen. Die rechte Abbildung 4.4 stellt die relativen Abweichungen der Zielfunktionswerte dar (vergleiche (4.4)). Man sieht deutlich, dass die relativen Abweichungen in diesem Fall alle negativ sind, d. h. es gilt für alle Trajektorien  $Z(\hat{p}^1) < Z(\hat{p}^2)$

| $\underline{p}_0^k$ | unterschiedliche Startwerte für die Iterationen |               |             |
|---------------------|---|---------------|-------------|
|                     | $\gamma_0^k$                                    | $\lambda_0^k$ | $\beta_0^k$ |
| $\underline{p}_0^1$ | 0.3748  | 0.4797        | 0.0106      |
| $\underline{p}_0^2$ | 0.0106  | 0.3915        | 0.3748      |
| $\underline{p}_0^3$ | 0.3000  | 0.1000        | 0.0300      |
| $\underline{p}_0^4$ | 0.0300  | 0.4200        | 0.3000      |

Tabelle 4.6: zu dem Experiment zu den Parameteruneindeutigkeiten: unterschiedliche Startwerte  $\underline{p}_0^k$  für die Iterationen

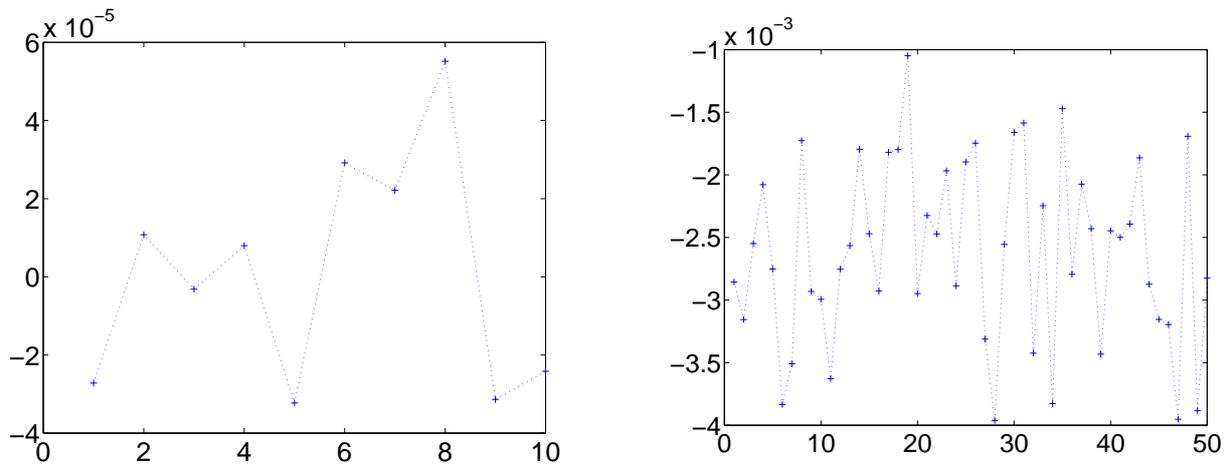


Abbildung 4.4: relative Abweichungen der Zielfunktionswerte siehe (4.4): links konstante Volatilitätsfunktion  $\sigma_1$ , rechts: zeitabhängige Volatilitätsfunktion  $\sigma_2$

und die Abweichung ist mit der Größenordnung  $10^{-3}$  deutlich höher als die Abweichungen für die konstante Volatilitätsfunktion.

Aus diesen Beobachtungen können wir die Vermutung ableiten, dass in der Nähe der Parametersituationen **Par1\*** und **Par2\*** jeweils lokale Minima der Zielfunktion liegen. Für eine zeitabhängige Volatilitätsfunktion unterscheidet sich die Höhe der Werte der Zielfunktion in den lokalen Minimum, jedoch schaffte es der benutzte Optimierungsalgorithmus `fminsearchbnd` aus MATLAB nicht, aus dem höher gelegenen lokalen Minimum „herauszukommen“ und das (vermutlich) globale Minimum zu finden. Um dieses Problem zu beseitigen, wäre eine intensivere Untersuchung des Minimierungsalgorithmus durchzuführen. Ein erster Ansatz könnte darin bestehen, nach der ersten Minimierung mit Hilfe der Formeln (2.10) eine Vermutung für das zweite lokale Minimum zu bestimmen und dort eine weitere Minimierung zu starten. Eine solche Untersuchung geht jedoch über das Ziel dieses Papers hinaus.

### 4.2.3 Ergebnisse im leicht gestörten Fall

Im Rahmen von zwei abschließenden Fallstudien wollen wir das Verhalten des Verfahrens bei leichten Fehlern in den Preisdaten bzw. in der Volatilität  $\sigma$  untersuchen. Dazu simulieren wir jeweils 30 Trajektorien für die Parameter aus Tabelle 4.2. Die Beobachtungsdauer soll 10000 Tage bei täglichen Beobachtungen umfassen, d. h. es gilt  $T = N = 10000$ .

|                  | Anteil der Schätzungen mit einem relativen Fehler $\leq 10\%$ |               |             |           |
|------------------|---|---------------|-------------|-----------|
|                  | für $\gamma$  | für $\lambda$ | für $\beta$ | für $\mu$ |
| exakte Preise    | 0.83  | 0.90          | 0.50        | 1.00      |
| $\delta = 0.001$ | 0.83  | 0.83          | 0.47        | 1.00      |
| $\delta = 0.005$ | 0.57  | 0.77          | 0.47        | 1.00      |

Tabelle 4.7: Anteile der Schätzungen mit relativen Fehlern von maximal 10 Prozent für exakte und mit  $\delta = 0.001$  gestörte Preise, Beobachtungszeitraum  $T = 10000$

### (a) Auswirkung von Störungen in den Preisdaten

Durch den Bid-Ask-Spread, also die Spanne zwischen Geld- und Briefkurs, aber auch durch mehrere Notierungen ein und derselben Aktie an verschiedenen (internationalen) Börsenplätzen können Aktienkurse unter Umständen etwas verfälscht sein. Wir gehen deshalb davon aus, daß wir nicht die exakten Preise  $P(t_i)$  beobachten können, sondern nur die mit kleinen Fehlern behafteten Werte  $P^\delta(t_i)$ .

Mit den relativen Fehlern

$$e_i := \frac{P^\delta(t_i) - P(t_i)}{P(t_i)} \quad (i = 1, \dots, N) \quad (4.5)$$

erhalten wir für die gestörten logarithmierten Preise  $p^\delta(t_i) = \ln P^\delta(t_i)$

$$p^\delta(t_i) = p(t_i) + \ln(1 + e_i) \quad (i = 1, \dots, N). \quad (4.6)$$

Wir nehmen an, daß die zufälligen Fehler  $e_i$  stochastisch unabhängig und normalverteilt sind mit Erwartungswert 0 und Varianz  $\delta > 0$ . Die Annahme ist durchaus realistisch, da wir auf diese Weise unabhängige und unterschiedlich starke Störungen zulassen. Dabei wird die Varianz so klein gewählt, daß der Fall  $e_i < -1$  vernachlässigt werden kann.

Um unser Schätzverfahren zu testen haben wir die Parameter aus Preisdaten geschätzt, welche wir zuvor mittels entsprechender Fehler gestört haben, und anschließend die Schätzwerte mit denen vom ungestörten Fall verglichen. Die Ergebnisse sind in Tabelle 4.7 zu sehen. Der Unterschied zwischen der Schätzung mit exakten Preisdaten und leicht gestörten Preisdaten ( $\delta = 0.001$ ) ist gering. Für größere Störungen in den Preisdaten ( $\delta = 0.005$ ) tritt zwar eine Verschlechterung, insbesondere bei der Schätzung von  $\gamma$ , ein. Allerdings zeigt die folgende Betrachtung, daß  $\delta = 0.005$  schon einem relativ großen Fehler in den Preisdaten entspricht.

Die Wahrscheinlichkeit einer Abweichung in den Aktienpreisen, die betragsmäßig größer ist als der Wert  $x$ , beträgt für unsere  $\mathcal{N}(0, \delta)$ -verteilten Fehler  $e_i$ ,  $i = 1, \dots, N$  gerade

$$\mathbf{P}(|e_i| > x) = 2 \left( 1 - \Phi \left( \frac{x}{\sqrt{\delta}} \right) \right) \quad (x \geq 0), \quad (4.7)$$

wobei  $\Phi(\cdot)$  die Verteilungsfunktion der Standardnormalverteilung bezeichnet.

Für zunehmende Werte von  $\delta$  nehmen auch die Wahrscheinlichkeiten für das Auftreten vergleichsweise hoher prozentualer Störungen in den Aktienpreisen zu. Kommen unrealistisch hohe Abweichungen von 10 Prozent und mehr für  $\delta = 0.001$  gerade noch mit einer Wahrscheinlichkeit von 0.0016 vor, beträgt diese Wahrscheinlichkeit bei  $\delta = 0.002$  sogar

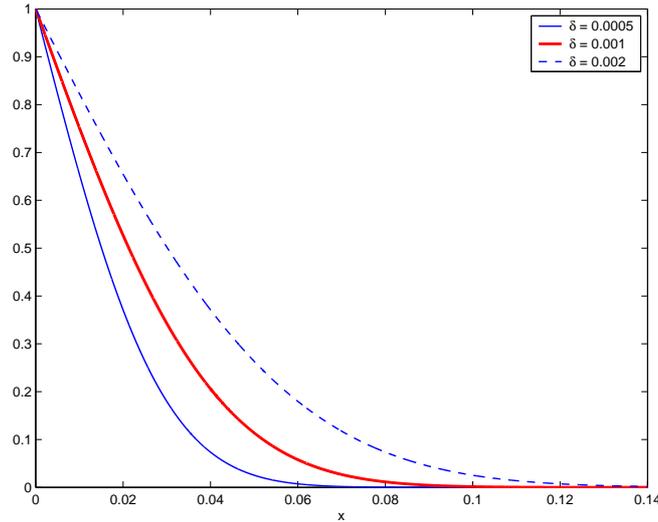


Abbildung 4.5: Wahrscheinlichkeit  $\mathbf{P}(|e_i| > x)$  zu verschiedenen Datenfehlervarianzen  $\delta$

deutlich mehr als das Zehnfache. Demzufolge ist die Betrachtung derartig großer Fehlervarianzen nicht mehr sinnvoll. Ganz im Gegenteil können bei liquiden Titeln noch kleinere Werte (z. B.  $\delta = 0.0005$ ) vollkommen ausreichen. Eine graphische Darstellung des Zusammenhangs zwischen Fehlervarianz und Fehlerwahrscheinlichkeiten findet sich in Abbildung 4.5.

### **(b) Auswirkung von Störungen in der Volatilität $\sigma$**

Für die Volatilität  $\sigma$  wird uns im Allgemeinen auch kein exakter, sondern lediglich ein zuvor selbst aus empirischen Daten ermittelter Schätzwert vorliegen. Dieser wird in aller Regel einen mehr oder minder großen Fehler aufweisen. Um derartige Situationen in unseren Versuchen nachzustellen, stören wir den Wert  $a = \sigma^2$  bei der Parameterschätzung (nicht jedoch bei der Simulation) wie folgt

$$a^\delta(t) := a(t)(1 + \delta). \quad (4.8)$$

Analog dazu definieren wir die gestörte Volatilität

$$\sigma^\delta(t) := \sigma(t)\sqrt{1 + \delta} = \sqrt{a^\delta(t)}. \quad (4.9)$$

$\delta$  kann also als relativer Fehler in der quadrierten Volatilität  $a$  aufgefaßt werden. Zu bemerken ist, daß wir die Fälle  $\delta \geq 0$  und  $\delta \leq 0$  getrennt untersucht haben. Man könnte noch den Fall unabhängiger Fehler auf den einzelnen Intervallen untersuchen. Da sich dort die Fehler ausgleichen können, kann man in diesem Fall sicherlich mit einem kleineren Fehler rechnen.

Für 30 simulierte Trajektorien haben wir jeweils die Schätzungen für verschiedene  $a^\delta$  mit  $\delta \leq 0$  durchgeführt. Die Bilder 4.6-4.9 stellen die Ergebnisse für verschiedene Störungen dar. Für sehr kleine Abweichungen in der Volatilität stellen wir nur geringe Auswirkungen auf die Genauigkeit der Schätzungen im Vergleich zum ungestörten Fall fest. Für mittlere Fehler treten allerdings zunehmend Verschlechterungen bei  $\gamma$  und  $\lambda$  auf. Die Bilder 4.8 und 4.9 zeigen, daß diese Parameter für negative  $\delta$ , d. h. in dem Fall  $a^\delta < a$ , tendenziell zu groß geschätzt werden. Wir bemerken, daß für das Experiment mit positiven  $\delta$ -Werten ( $a^\delta \geq a$ ) die Parameter  $\gamma$  und  $\beta$  tendenziell zu gering geschätzt wurden. In beiden Fällen stiegen die Abweichung für wachsendes  $|\delta|$ .

Auf den ersten Blick scheint die Schätzung des Parameters  $\beta$  recht robust zu sein. Dies kann

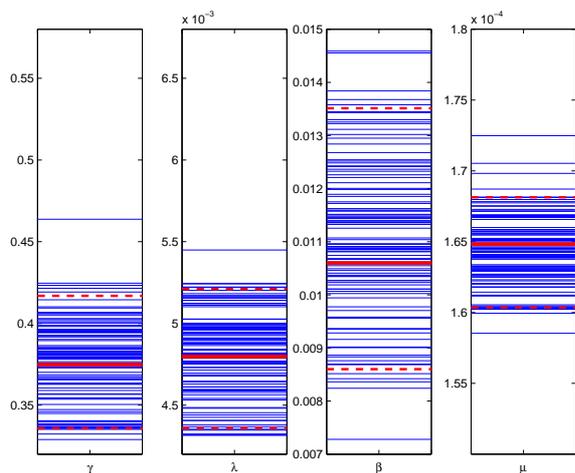


Abbildung 4.6: Ergebnisse für die Situation  $\delta = 0$

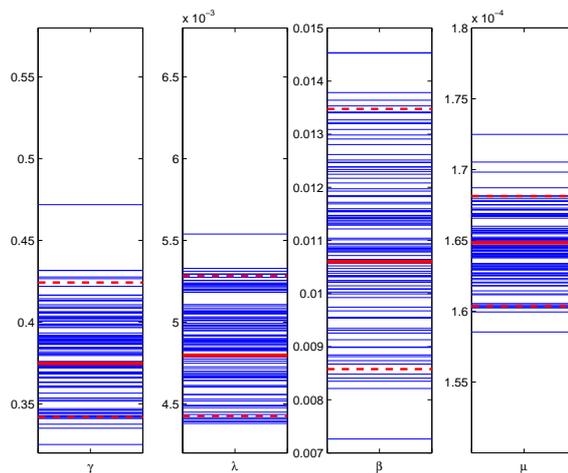


Abbildung 4.7: Ergebnisse für die Situation  $\delta = -0.01$

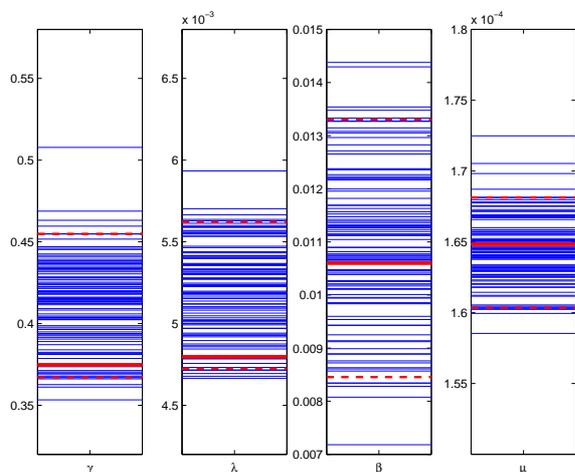


Abbildung 4.8: Ergebnisse für die Situation  $\delta = -0.05$

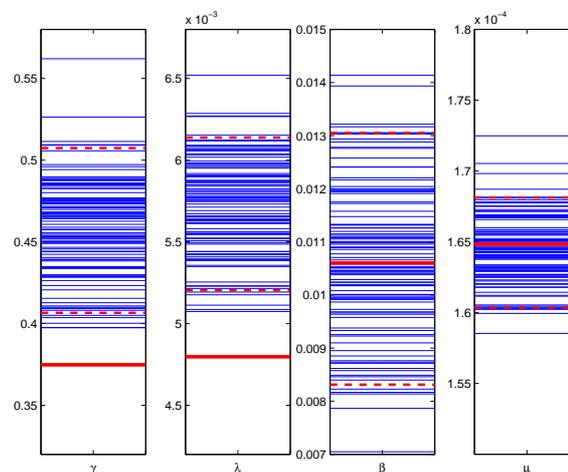


Abbildung 4.9: Ergebnisse für die Situation  $\delta = -0.1$

aber dadurch erklärt werden, dass durch die generell höhere Schwankungsbreite der leichte Trend zu zu kleinen Schätzwerten nicht ins Auge fällt.

Der Parameter  $\mu$  nimmt hier wiederum eine Sonderstellung ein. Da wir diesen nicht mit der Maximum-Likelihood-Methode ermitteln, ist für seine Bestimmung auch keine Kenntnis von  $\sigma$  notwendig. Folglich können Ungenauigkeiten in der Volatilität auch keinen Einfluß auf die Schätzwerte haben.

## Literatur

- [1] L. Arnold. *Stochastic Differential Equations: Theory and Applications*. Wiley, New York, 1992.
- [2] A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

- [3] W. Hausmann, K. Diener, J. Käsler. *Derivate, Arbitrage und Portfolio-Selection. Stochastische Finanzmarktmodelle und ihre Anwendungen*. Vieweg, Braunschweig, Wiesbaden, 2002.
- [4] R. Krämer, M. Richter, B. Hofmann. Parameter estimation in a generalized bivariate Ornstein-Uhlenbeck model. In J. vom Scheidt, Hrsg., *Tagungsband zum Workshop „Stochastische Analysis“ 27.09.2004 – 29.09.2004*, Chemnitz, 2005.
- [5] R. Krämer, M. Richter. A Generalized Bivariate Ornstein-Uhlenbeck Model for Financial Assets. *Preprint 2007-1*, Technische Universität Chemnitz, 2007.
- [6] R. S. Liptser, A. N. Shiryaev. *Statistics of Random Processes, Band 2*. Springer, Berlin 2001.
- [7] A. W. Lo, J. Wang. Implementing Option Pricing Models When Asset Returns Are Predictable. *The Journal of Finance*, **50**, 87-129, 1995.
- [8] H. Lütkepohl. *Introduction to Multiple Time Series Analysis*. Springer, Berlin, 1991.
- [9] C. R. Rao. *Lineare statistische Methoden und ihre Anwendungen*. Akademie-Verlag, Berlin, 1973.
- [10] N. Reinhold. *Monte-Carlo-Simulation verallgemeinerter bivariater Ornstein-Uhlenbeck-Prozesse*. Semesterarbeit, TU Chemnitz, Fakultät für Mathematik, 2005.