

**Parameterschätzung in
zeitkontinuierlichen
dynamischen Systemen**

**Dissertation
zur Erlangung des akademischen Grads
des Doktors der Sozialwissenschaften
Universität Konstanz
Fakultät für Wirtschaftswissenschaften und Statistik**

vorgelegt von

Hermann Singer

- 1. Referent: Prof. Dr. Alfred Hamerle**
- 2. Referent: Prof. Dr. Michael Kohlmann**

Inhalt

Kapitel 1: Einführung	1
1.1. Deterministische und stochastische Differentialgleichungen	1
1.2. Parameterschätzung	5
1.3. Zusammenfassende Darstellung des Inhalts und der Resultate dieser Arbeit	7
Kapitel 2 : Modellspezifikation und Kritik existierender Schätzverfahren	12
2.1. Modellspezifikation	12
2.2. Kontinuierliche Datensätze	16
2.2.1. Beispiel: ML-Schätzung des Lorenz -Attraktors	20
2.3. Die Trapez-Näherung	24
2.4. Das exakte diskrete Modell	26
2.5. Minimum Distance-Verfahren	28
2.6. Asymptotische Eigenschaften	29
2.6.1. Diskretisiertes kontinuierliches Sampling	29
2.6.2. Trapez-Methode	30
2.6.3. Exaktes diskretes Modell	31
2.6.4. Numerische Beispiele	31
2.7. Zusammenfassung	35
Kapitel 3: Ein Identifikationsproblem	37
3.1. Einführung	37
3.2. Exogene Variablen approximiert als Sprungfunktionen	39
3.3. Exogene Variablen approximiert als Polygone und allgemeinere Funktionen	39
3.4. Identifikationsresultate von Hansen u. Sargent	39
3.5. Lokale Identifizierbarkeit	44

Kapitel 4: Maximum Likelihood-Schätzung des exakten diskreten Modells mit analytischen Ableitungen	45
4.1. Modellspezifikation	45
4.2. Maximum Likelihood-Schätzung	46
4.3. Analytische Berechnung des Maximum Likelihood-Schätzers $\hat{\theta}$	48
4.4. Newton-Raphson und Scoring-Algorithmus zur numerischen Lösung der Likelihood-Gleichungen	53
4.5. Analytische Berechnung der Score-Funktion und der Fisher -Information	55
4.6. Konsistenz des Maximum-Likelihood-Schätzers	62
4.7. Asymptotische Normalität der ML-Schätzer	65
4.8. Praktische Implementierung des Scoring-Algorithmus: Das Programmpaket LSD	66
4.9. Beispiele.	69
4.9.1. Ornstein-Uhlenbeck -Prozeß	69
4.9.1.1. Exakte ML-Schätzung	72
4.9.1.2. Linearisiertes Modell	74
4.9.1.3. Eine Simulationsstudie	75
4.9.1.4. Reparametrisierung	75
4.9.2. Das Phillips -Modell	78
 Kapitel 5: ML-Schätzung des diskret/kontinuierlichen Zustandsraum-Modells mit EM, GEM und Newton-Raphson-Algorithmen	 81
5.1. Modellspezifikation	81
5.2. Parameterschätzung	81
5.3. Der EM-Algorithmus	82
5.4. Praktische Implementierung des EM-Algorithmus	84
5.5. Rekursive Berechnung der bedingten Erwartungswerte der Momentenmatrizen	87
5.6. Newton-Raphson -Algorithmus	90
5.7. IML-Module zur Schätzung des Zustandsraum-Modells	91
5.8. Beispiel: Die Wolferschen Sonnenflecken-Daten	93
5.8.1. Modellspezifikation	93
5.8.2. Identifikation	97
5.8.3. Maximum-Likelihood-Schätzung	98
 Kapitel 6: Abschließende Bemerkungen	 102

Anhänge	104
Anhang 1: Das McDonald-Swaminathan-Kalkül	105
Anhang 2: Maximum-Likelihood-Schätzung	110
Anhang 3: Analytische Ableitungen der reduzierten Parameter	112
Anhang 4: Fisher-Informations-Matrix des exakten diskreten Modells	118
Anhang 5: Asymptotische Eigenschaften der Score-Funktion und der Hesse-Matrix	122
Literatur	128

Bezeichnungen und Abkürzungen

A'	die zu A transponierte Matrix
A^+	adjungierte Matrix (transponiert und konjugiert komplex)
$\operatorname{argmax} f(x)$	Argument x , bei dem $f(x)$ maximal ist.
$\dot{x}(t)$	Ableitung von x nach t
$\partial Y / \partial X$	McDonald-Swaminathan -Ableitung
col, COL	spaltenweise Vektor-Operation
$\hat{\Theta}$	Dach (Schätzwert für Θ)
$ \Sigma $	Determinante von Σ (=Det Σ), auch Absolutbetrag.
$\delta(x)$	Diracsche Delta-Funktion
$\operatorname{Cov}(X, Y)$	$=E[(X-E(X))(Y-E(Y))']$; Kovarianzmatrix der Zufallsmatrizen X, Y
$\operatorname{Cov}(X, Y Z)$	$=E[(X-E(X Z))(Y-E(Y Z))' Z]$; bedingte Kovarianzmatrix
$E[X]$	Erwartungswert der Zufallsvariable X
$E_{\Theta}[X]$	Erwartungswert zum Wahrscheinlichkeitsmaß P_{Θ}
$E[X Y]$	bedingter Erwartungswert von X auf Y
∇	Nabla-Operator (Gradient)
$A \# B$	Hadamard -Produkt
i.i.d	unabhängig und identisch verteilt
p-lim	Konvergenz in Wahrscheinlichkeit
qm-lim	Konvergenz im Quadratmittel ($= \xrightarrow{-2}$)
\xrightarrow{D}	Konvergenz in Verteilung
$A \otimes B$	Kronecker -Produkt
$O(\varepsilon)$	Landau -Symbol: $O(\varepsilon)/\varepsilon < M$ für $\varepsilon \rightarrow 0$
log	natürlicher Logarithmus
ML	Maximum-Likelihood
$\ A\ $	Norm der Matrix A
$\ x\ _p$	p -Vektor-Norm (Hölder -Norm) ($= (\sum x_i ^p)^{1/p}$)
$\ A\ _p$	p -Matrix-Norm ($= \sup \ Ax\ _p$; $\ x\ _p = 1$)
$\ A\ _F$	Frobenius -Norm ($= (\sum A_{ij} ^2)^{1/2}$)
$N(\mu, \Sigma)$	Normal(Gauß)-Verteilung mit Erwartungswert μ und Kovarianzmatrix Σ
A^-	Pseudoinverse von A
p.d., p.s.d	positiv definit, positiv semidefinit
row, ROW	Zeilenweise Vektor-Operation
$\langle f, g \rangle$	$= f'g$; Skalarprodukt der Vektoren f und g
$\operatorname{tr}(A) = \sum A_{ii}$	Spur (Trace) der Matrix A
$\operatorname{Var}(X)$	$= \operatorname{Cov}(X, X)$: Varianz der Zufallsmatrix X
$\operatorname{Var}(X Y)$	$= \operatorname{Cov}(X, X Y)$: Bedingte Varianz
$w(t)$	Wiener -Prozeß

1. Einführung

1.1. Deterministische und stochastische Differentialgleichungen

Differentialgleichungen (sowohl "gewöhnliche" als auch "partielle") bilden eine mathematische Modellklasse, die in praktisch allen Teilgebieten der Physik eine zentrale Rolle spielt (vgl. etwa **Courant/Hilbert** 1924, 1937).

Ausgehend von der klassischen Mechanik (**Newton-**, **Lagrange-** und **Hamilton-Formalismus**), der Elektrodynamik (**Maxwell-Gleichungen**) bis hin zur Quantenmechanik (**Schrödinger-Gleichung**) und Relativitätstheorie (**Einsteinsche Feldgleichungen**), überall sind die dynamischen Grundgesetze in differentieller Gestalt formuliert worden.

Bemerkenswert hierbei ist, daß man jahrhundertlang mit deterministischen Methoden auskam; erst die Behandlung von Makrosystemen (bestehend aus sehr vielen, typischerweise 10^{23} Komponenten) erforderte die Benutzung statistischer Konzepte (durch **Boltzmann**, **Gibbs** und **Einstein**). Insbesondere wurde die Behandlung der **Brownschen Bewegung** durch **Langevin** (vgl. **Nelson** 1967, **van Kampen** 1981) ein Kristallisationspunkt für die spätere Theorie der stochastischen Differentialgleichungen, bei denen bestimmte Teile (z.B. Anfangs- und Randbedingungen, Inhomogenitäten) zufälligen Einflüssen unterliegen .

Betrachten wir etwa die **Newtonsche Bewegungsgleichung** für den Ort $x(t)$ eines Teilchens der Masse m (siehe z.B. V.I. **Arnold** 1978)

$$(1.1) \quad m \ddot{x} = K (x, \dot{x}, t)$$

Hierbei ist die Beschleunigung $\ddot{x} = d^2/dt^2 x(t)$ proportional zur auf das Teilchen wirkenden Kraft K ("Kraft = Masse mal Beschleunigung"), wobei die Kraft wiederum von Ort x und Geschwindigkeit \dot{x} (bei Reibungskräften) bzw. explizit von der Zeit abhängt. Man bekommt so eine selbstkonsistentes System von Gleichungen für den Ort und dessen Ableitungen (Differentialquotienten) : eine Differentialgleichung.

Ein Pendel kann etwa (für kleine Auslenkungen x) durch die Gleichung

$$(1.2) \quad m \ddot{x} + \gamma \dot{x} + Dx = 0$$

beschrieben werden (siehe etwa V. **Arnold** 1973, Kap. 1), deren Lösung (siehe Kap. 4.8) auf gedämpfte Schwingungen führt. Unkontrollierbare Einflüsse (etwa die Ablenkung des Pendels durch Windböen) können nun durch die Hin-

zunahme zufälliger Kräfte beschrieben werden; etwa war die Idee von **Langevin**, die zufällige Kraft $F(t)$ als Inhomogenität mit verschwindendem Erwartungswert und sehr kurzer Autokorrelationszeit τ zu modellieren, d.h. $E[F(t)] = 0$ und $E[F(t)F(s)] = 0$ für $|t-s| > \tau$. Für ein freies **Brownsches** Teilchen erhält man so die **Langevin-Gleichung** ($v = \dot{x}$)

$$(1.3) \quad m \dot{v} + \gamma v = F(t)$$

die sozusagen den Prototyp einer stochastischen Differentialgleichung darstellt (insbesondere können Differentialgleichungen höherer Ordnung als mehrdimensionale Version von Glg. (1.3) dargestellt werden).

Stochastische Kräfte, bei denen die Autokorrelation $C(t)$ nach sehr kurzer Zeit verschwindet (ideal $C(t) = E[F(t)F(s)] \sim \delta(t-s)$) werden als weißes Rauschen bezeichnet, da die Spektraldichte (die als **Fourier-Transformierte** der Autokorrelation die verallgemeinerte Funktion 1 ist) auf der gesamten reellen Achse konstant ist (und somit wie weißes Licht alle Frequenzen enthält).

Gleichung (1.3) bietet jedoch durch Erweiterung des Zustandsvektors auch die Möglichkeit, zufällige Kräfte mit allgemeinerer Autokorrelation zu behandeln.

Aufgrund der Erfolge der mathematischen Physik in der Behandlung von Naturvorgängen ist es nicht weiter verwunderlich, daß auch in anderen Gebieten, wie z.B. Ökonometrie (für eine historische Übersicht siehe **Bergstrom** 1988), Soziologie (**Tuma u. Hannan** 1984, **Doreian u. Hummon** 1976) und Psychologie (**Möbus u. Nagl** 1983, **Deppe** 1977, **Singer u. Hautzinger** 1988) in neuerer Zeit versucht wird, Differentialgleichungsmodelle für zeitlich und/oder räumlich variierende Phänomene heranzuziehen. Vorteilhaft erscheint hierbei, daß man nicht vorgegebene Zeitverläufe (etwa eine periodische Funktion) an die Daten anpaßt, sondern eine dynamische Wechselwirkung der Variablen und ihrer zeitlichen Veränderungen (Ableitungen) in Form einer Differentialgleichung postuliert wird, welche die beobachtete Klasse von Lösungen (etwa Oszillationen) erzeugt. Beispielsweise erhält man als Lösung Exponentialfunktionen, wenn die Veränderung Δx in einem Zeitintervall Δt dieser Variable proportional ist, also $\Delta x / \Delta t \sim x$ ($\Delta t \rightarrow 0$). Dann gilt

$$x(t) = e^{\lambda t} x(0)$$

mit einer **Rate** λ , die als prozentuale Veränderung von x pro Zeiteinheit interpretiert werden kann. Kommt also ein konstanter Anteil $\lambda x \Delta t$ in der Zeiteinheit hinzu, so erhält man das bekannte explosionsartige Anwachsen (etwa einer Tierart, CO_2 -Konzentration etc.).

Ist dagegen die Veränderung der Veränderung (d.h. die Beschleunigung) dem

Systemzustand x proportional, so erhält man oszillatorische Lösungen (in statistischer Terminologie: autoregressive Prozesse 2. Ordnung).

Die in den Bewegungsgleichungen enthaltenen zufälligen Einflüsse trüben die ideale Regularität der deterministischen Lösungen durch überlagerte Fluktuationen, die jedoch in den realen Daten praktisch immer enthalten sind, insbesondere in Fachgebieten mit komplexem Gegenstandsbereich.

An dieser Stelle sind zwei Dinge zu bemerken:

1. auch rein deterministische nichtlineare Gleichungen können (schon bei 3 Variablen) zu extrem irregulären (chaotischen) Lösungen führen (für eine Übersicht siehe **Ruelle 1980, Schuster 1984** ; vgl. auch Kap. 2).

2. Der statistische Erwartungswert der Lösung ($\mu(t) = E[x(t)]$) erfüllt nicht unbedingt den deterministischen Teil der stochastischen Differentialgleichung. Die Interaktion des Systemzustands mit den stochastischen Fluktuationen führt zu essentiell stochastischen Effekten (etwa rausch-induzierte Drift), die auch im statistischen Mittel übrigbleiben (vgl. **Van Kampen 1981** und Kap. 2). Daher sind immer dann Schwierigkeiten zu erwarten, wenn zu einer phänomenologisch gewonnenen deterministischen Differentialgleichung stochastische Fluktuationen **ad hoc** hinzugefügt werden. I.a. sind daher die stochastische Differentialgleichung und daraus abgeleitete deterministische Gleichungen für die Momente (etwa $\mu(t)=E[x(t)]$ und $\sigma(t)=\text{var}(x(t))$) aus einem detaillierten Modell des betrachteten Vorgangs zu gewinnen.

Zusammenfassend können mehrere Gründe für die Notwendigkeit einer stochastischen Behandlung von Differentialgleichungen genannt werden:

1. Zufällige Anfangsbedingungen

Die Anfangsbedingungen sind nicht genau bekannt oder es ist nicht möglich, einen genau definierten Anfangszustand zu präparieren. Z.B. kann in einem Vielteilchensystem möglicherweise nur die Temperatur kontrolliert werden; dies führt zu einer Verteilungsfunktion für die Orts- und Impulskoordinaten, die als **kanonische Verteilung** bezeichnet wird. Ein anderes Beispiel ist eine soziologische oder psychologische Panelstudie; auch hier können nur wenige Variablen kontrolliert werden (etwa Alter oder Geschlecht der Probanden), die verbleibenden extrem vielfältigen Gesichtspunkte führen zu einer statistischen Verteilung der Anfangsbedingungen.

2. Zufällige externe Einflüsse

Selbst wenn die Modelldynamik für ein abgeschlossenes System bekannt ist, führt der Kontakt zur Außenwelt zu einer Interaktion mit einer Vielzahl von Einflußgrößen, von denen nur wenige gemessen und explizit berücksichtigt werden können. Die verbleibenden Faktoren können dann als stochastische

Einflußgrößen betrachtet werden, deren exakter Verlauf durch statistische Charakterisierungen wie Erwartungswert und Autokorrelation approximiert wird.

3. Komplexe Systemdynamik

Im Rahmen der klassischen Mechanik sind alle Systemtrajektorien (im Prinzip) durch die **Hamiltonschen** Bewegungsgleichungen und die Anfangswerte bestimmt (deterministisches System). Die häufig geäußerte Auffassung, die Methode der statistischen Mechanik sei durch die Unkenntnis der Anfangsbedingungen des Makrosystems begründet, wird von einigen Autoren in Frage gestellt (vgl. **Van Kampen** 1987). Auch bei genauer Präparation des Anfangszustands wäre man gezwungen, die äußerst komplizierten, nichtlinearen Hamiltongleichungen durch vereinfachte Gleichungen für wenige Variablen zu ersetzen, die dann jedoch irreguläre Anteile enthalten, die durch stochastische Prozesse approximiert werden können. Anstelle der Lösung riesiger deterministischer Gleichungssysteme (was praktisch unmöglich ist) rekuriert man stattdessen auf statistische und informationstheoretische Prinzipien (Maximierung der Entropie oder des 4-dimensionalen Kaliber-Funktional, vgl. **Jaynes** 1978, **Giuliasu u. Shenitzer** 1985), um die Gesetze der reversiblen und irreversiblen Thermodynamik zu erhalten.

In den "weichen" Wissenschaftszweigen sind jedoch die Grundprinzipien unbekannt, sodaß man von vorneherein nicht die Möglichkeit hat, von diesen auf approximative Systemgleichungen zu schließen.

4. Approximative Systemgleichungen

Die gewählte Differentialgleichung (Modellspezifikation) muß i.a. als mehr oder weniger gute Näherung für die wirklichen Verhältnisse angesehen werden (etwa sind bei linearen Gleichungen die nichtlinearen Anteile vernachlässigt, die jedoch bei starker Anregung des Systems wichtig werden können ; Einteilchenmodelle vernachlässigen die kollektiven Interaktionen der Systemteile). Hier bietet sich an, weggelassene Anteile durch stochastische Parameter zu modellieren.

Andererseits liegt der Erfolg einer deterministischen Theorie in einer geschickten Modellwahl begründet, bei der sich eine Vielzahl zufälliger (bzw. irregulärer) Größen zu nahezu deterministischen Variablen ergänzen (formalisiert durch die Grenzwertsätze der Wahrscheinlichkeitstheorie). Kollektive Interaktionen können dann als phänomenologische Konstanten ("mittlere Felder") modelliert werden. Umgekehrt kann aber auch die Hinzunahme einer stochastischen Fluktuation die Weiterbenutzung einer approximativen Systemgleichung ermöglichen (siehe z.B. **Lewis** 1986, S. 85 ff).

5. Meßfehler

I.a. kann der betrachtete Systemzustand nur unvollständig (etwa nur bestimmte Komponenten) und mit einem Meßfehler gemessen werden. Daher sind Anfangsbedingungen und spätere Systemzustände nur unvollständig bekannt (man weiß z.B. nur, daß $y(0)$ aus dem Intervall $[y-\varepsilon, y+\varepsilon]$ stammt). Entsprechend kann dann auch für den späteren Zustand $y(t)$ nur ein Intervall angegeben werden. Insbesondere im Falle instabiler Lösungen (Chaos) laufen anfangs benachbarte Trajektorien exponentiell auseinander, sodaß auch hier statistische Konzepte (etwa Autokorrelation, stationäre Verteilung (invariantes Maß)) wichtig werden (vgl. hierzu **Schuster** 1984). Wenn also der Systemzustand $y(t)$ der deterministischen Gleichung

$$\dot{y} = f(y)$$

genügt, jedoch nur mit Meßfehler gemessen werden kann:

$$z(t) = y(t) + \varepsilon(t)$$

erfüllt die **Meßgröße** $z(t)$ folgende Gleichung

$$\begin{aligned}\dot{z} &= f(z-\varepsilon) + \dot{\varepsilon} \\ &= f(z) - f'(z) \varepsilon + O(\varepsilon^2) + \dot{\varepsilon}\end{aligned}$$

Systeme, bei denen der gleiche Anfangszustand $z(t)$ **gemessen** wurde, werden sich also auseinanderentwickeln, dies umso mehr, je nichtlinearer das Vektorfeld $f(y)$ beschaffen ist.

1.2. Parameterschätzung

Bisher wurde davon ausgegangen, daß die Parameter des Systems (Koeffizienten der stochastischen Differentialgleichung) bekannt sind (oder unabhängig von der Dynamik gemessen werden können, etwa die Masse des Pendels). Die traditionelle Vorgehensweise der Physik besteht darin, daß die Lösungen der Systemgleichungen als Funktion der Parameter ausgedrückt und entsprechende Meßgrößen (z.B. Zeitkorrelationsfunktionen) bereitgestellt werden. Experimentelle Daten werden dann an diese Modelle angepaßt (im Sinne eines nicht-linearen kleinsten-Quadrate-Fits). Dies entspricht einer Kombination aus kleinster Quadrate- und Momentenmethode der mathematischen Statistik.

In dieser Arbeit wird die sogenannte **Maximum-Likelihood-Methode** (siehe Anhang 2) benutzt, um effiziente und asymptotisch normalverteilte Schätzungen

für die Systemmatrizen zu erhalten. Vorteilhaft ist auch die Möglichkeit, simultane Konfidenzintervalle zu konstruieren, um die Genauigkeit der Schätzungen zu beurteilen. Hierzu muß die Wahrscheinlichkeitsdichte $f(Z|\Theta)$ der Messungen Z , die durch den Parametervektor Θ parametrisiert ist, berechnet werden. Als **Maximum-Likelihood-Schätzer** $\hat{\Theta}$ bezeichnet man das Maximum von $f(Z|\Theta)$ für gegebene Daten, d.h.

$$(1.4) \quad \hat{\Theta}(Z) = \underset{\Theta}{\operatorname{argmax}} f(Z|\Theta)$$

Als Funktion von Θ wird $L(\Theta|Z)=f(Z|\Theta)$ auch als **Likelihood-Funktion** bezeichnet. Im Falle kontinuierlicher Messungen $z(t)$, $t \in [t_0, t_T]$ hat man die Wahrscheinlichkeitsverteilung eines **stochastischen Prozesses** zu spezifizieren, die durch die Familie der sog. **endlich-dimensionalen Verteilungen** gegeben ist (vgl. L. Arnold 1973, S. 36 ff). Für den überabzählbaren Datensatz $z(t)$ kann man dann ein **Likelihood-Funktional (Radon-Nikodym-Ableitung)** angeben, das von der gesamten Trajektorie abhängt (siehe hierzu Liptser u. Shirayev 1978, Kap. 17).

In der Praxis ist es allerdings so, daß Daten nur zu bestimmten Zeitpunkten gemessen werden können, die z.B. in der Sozialwissenschaft sehr weit auseinanderliegen können, etwa 1 Jahr. Hier läßt sich der endliche Datensatz $\{Z(t_0), \dots, Z(t_T)\}$ durch die Verteilung

$$F_{t_0, \dots, t_T}(z_0, \dots, z_T) = P(Z_0 \leq z_0, \dots, Z_T \leq z_T)$$

bzw. die Dichte $f(z_0, \dots, z_T|\Theta)$ charakterisieren.

Durch die zeitdiskrete Messung (sampling) geht jedoch Information über die Trajektorie verloren; dies drückt sich in Mehrdeutigkeiten aus, d.h. im Allgemeinen passen mehrere verschiedene Systemmodelle (die durch Parameter $\Theta_1, \Theta_2, \Theta_3, \dots$ charakterisiert sind) zum gleichen Datensatz (**Aliasing-Problem**). Man wird so auf das Problem der **Parameteridentifikation** geführt; häufig koexistieren **beobachtungsäquivalente Strukturen** mit qualitativ widersprüchlichen Kausalinterpretationen (vgl. die weiteren Kapitel, Phillips 1973, Tigelaar 1982, Hansen u. Sargent 1983). Durch Kenntnisse über die Struktur des Systems (Restriktionen) oder genügend feines Zeitraster können diese Probleme jedoch überwunden werden.

An dieser Stelle ist allerdings anzumerken, daß die diskrete Datensituation in den Wirtschafts- und Sozialwissenschaften dazu geführt hat, daß (wenn überhaupt), vorwiegend Differenzgleichungsmodelle in diskreter Zeit benutzt werden (Zeitreihen). Aufgrund der verschiedenen Zeitabstände Δt in verschiedenen Studien können diese jedoch nicht verglichen werden (Kreuzvalidierung), da auf der diskreten Ebene keine Möglichkeit existiert, die Parameter ineinander umzurechnen. In einer zeitstetigen Methodik, bei der auf die differenziellen Para-

meter zugegriffen wird, können jedoch verschiedene Studien verglichen und die Parameter beliebiger diskreter Modelle als Funktionen der Zeit und der differentiellen System-Matrizen explizit angegeben werden. Man erhält somit eine Theorie-Ebene, die reichhaltigere empirische Fragestellungen zuläßt (etwa unregelmäßige Zeitabstände, Verlauf der Systemtrajektorie zwischen den Messungen (Interpolation), momentane Änderungsrate einer Variable (Geschwindigkeit), Prognose für beliebige Zeitpunkte etc.; vgl. die Diskussion in **Möbus u. Nagl** 1983). Weiterhin erscheint es als natürlich, die in stetiger Zeit ablaufenden Phänomene auch als solche zu formulieren und die System-Modellierung vom Meßvorgang gedanklich abzutrennen (Quanteneffekte sollen in diesem Kontext nicht diskutiert werden). Auch aggregierte ("flows") oder über endliche Zeiträume geglättete Daten können so im Rahmen eines Modells des Meßvorgangs (Meßmodell) behandelt werden.

1.3. Zusammenfassende Darstellung des Inhalts und der Resultate dieser Arbeit

Kapitel 2 befaßt sich mit der allgemeinen Modellspezifikation (nichtlineare stochastische Differentialgleichungen mit kontinuierlichen oder diskreten Messungen) und der Diskussion existierender Schätzverfahren.

Die Probleme, die bei der Behandlung des "weißen Rauschens" als Prozeßfehler auftreten, werden kurz diskutiert (**Itô**-Integral, **Stratonovich**-Integral). Es wird der **Vorschlag** gemacht, in einer empirischen Vorgehensweise von einem Prozeßfehler mit endlicher Korrelationszeit auszugehen, dann auf die korrespondierende **Itô**-Gleichung überzugehen (Korrekturterm in der Drift) und anschließend die Konstruktion der Likelihood mit dieser Darstellung vorzunehmen. Dies ist durch die Resultate der mathematischen Literatur begründet, die auf der Martingaleigenschaft des Stochastischen Integrals im Sinne von **Itô** beruhen (**kontinuierliches Sampling**, **Liptser u. Shirayev**, 1977, 1978). Weiterhin wird gezeigt, daß auch die Likelihood für die diskreten Meßzeitpunkte in Grenzwert auf stochastische **Itô**-Integrale führt (**diskretisiertes kontinuierliches Sampling**, **Le Breton**, 1976).

Das **Lorenz**-Modell, das in der Chaos-Theorie eine wichtige Rolle spielt, dient als Beispiel für die Methodik. Der Prozeßfehler, der als Approximation weglassener Terme der **Navier-Stokes**-Gleichungen angesehen werden kann, führt zu einem früheren Einsetzen des irregulären (chaotischen) Teils der Trajektorie. Weiterhin zeigt sich, daß die Systemparameter mit der Maximum-Likelihood (ML)-Methode in effektiver Weise geschätzt werden können.

Die **Trapez**-Methode (**Bergstrom** 1966), ein in ökonometrischen Anwendungen

häufig benutztes Näherungsverfahren, geht von der Integralform der linearen Differentialgleichung aus und approximiert die Integrale durch eine Trapezformel. Der entstehende Spezifikationsfehler ist eine Funktion des Zeitabstands Δt der Messungen. Der approximative ML-Schätzer wird anschließend explizit abgeleitet. Dabei zeigt sich, daß wiederum Integralsummen im Sinne von $It\hat{0}$ erscheinen, die Zeitintegrale jedoch vom diskretisierten kontinuierlichen Sampling abweichen. Dies hat zur Konsequenz, daß die asymptotische Verzerrung nur von Ordnung Δt^3 für kleine Δt ist, während im Falle des diskretisierten kontinuierlichen Samplings (bzw. Rechtecknäherung statt Trapez) die Verzerrung linear in Δt herauskommt.

Das **exakte diskrete Modell** ist ein aus der linearen Differentialgleichung abgeleitetes autoregressives Schema, dessen Parameter jedoch in nichtlinearer Weise von den Systemmatrizen des kontinuierlichen Modells abhängen. Die von verschiedenen Autoren (**Doreian u. Hummon 1976, Tuma u. Hannan 1984, Arminger 1986**) vorgeschlagene Schätzung des exakten diskreten Modells und anschließende Rückrechnung kann aufgrund eines Identifikationsproblems, das durch die Nichtlinearität zustandekommt, nicht empfohlen werden (ausführlich dazu Kap. 3 und 4). Im Grenzfall $\Delta t \rightarrow 0$ erhält man wieder die Formeln des kontinuierlichen Samplings ($It\hat{0}$ -Integrale), obwohl im linearen Modell (zustandsunabhängiger Diffusionskoeffizient) $It\hat{0}$ - und **Stratonovich**-Lösung für die Trajektorie übereinstimmen.

Das **Minimum-Distance-Verfahren** (**Phillips 1972, Hauptmann u. Schmid 1980**), das als nichtlinearer verallgemeinerter kleinster Quadrate-Schätzer bezeichnet werden kann, erlaubt es nicht, Restriktionen in der Diffusionsmatrix zu implementieren, die jedoch aus theoretischen Gründen gegeben sein können und zusätzlich zur Identifikation beitragen.

Anschließend werden die asymptotischen Eigenschaften anhand eines numerischen Beispiels illustriert.

Aufgrund der Beschränkungen der referierten Schätzverfahren:

- Gültigkeit nur für kleines Meßintervall Δt (diskretisiertes kontinuierliches Sampling, Trapezmethode)
- Nichtbeachtung des Identifikationsproblems (Rückrechnung, "LISREL -Methode")
- Fehlspezifikation des Diffusionsterms (Minimum Distance-Methode)

wird ein **exaktes Maximum-Likelihood-Verfahren** vorgeschlagen, das direkt auf die Parameter des zeitstetigen Systems zugreift. Zuvor muß jedoch der Identifikationsstatus des Modell geklärt werden.

In **Kap. 3** wird zuerst das Identifikationsproblem erläutert, das durch die Mehrdeutigkeit des Matrix-Logarithmus zustandekommt. Der Identifikationssta-

tus des Modells ist jedoch abhängig von den Annahmen, die über die exogenen Einflüsse getroffen wurden. Approximiert man nämlich die exogenen Variablen durch Treppenfunktionen (konstant zwischen den Messungen), so existieren beobachtungsäquivalente Strukturen ("Aliasing"), während bei der Approximation durch Polygone die Identifizierbarkeit des Modells gezeigt werden kann. Trotzdem ist der Fall stückweise konstanter oder verschwindender exogener Einflüsse (vollständige endogene Modellierung) ein wichtiger Fall, der in Kap. 4 behandelt wird. Die Identifikationsresultate von **Hansen u. Sargent** (1983) zeigen, daß auch in diesem Fall aufgrund der positiven Definitheit der Diffusionsmatrix der Satz beobachtungsäquivalenter Strukturen im wesentlichen auf endlich viele eingeschränkt werden kann (die genaue Anzahl hängt jedoch vom wahren Parameter ab, der in statistischen Anwendungen unbekannt ist). Daher muß weiterhin auf Restriktionen zurückgegriffen werden, die alle Aliasing-Matrizen ausschalten. In einem praktischen Kontext kann auf das Theorem von **Rothenberg** (1971) zurückgegriffen werden, das im wesentlichen die Äquivalenz von lokaler Identifizierbarkeit und Nichtsingularität der Informationsmatrix aufzeigt.

Kapitel 4 stellt den zentralen Punkt der Arbeit dar. Ausgehend von der Spezifikation des linearen zeitstetigen System-Modells mit diskreten Messungen wird die Likelihood-Funktion abgeleitet, die bei konstanter oder **Gauß**-verteilter Anfangsbedingung aus einer multivariaten Normalverteilungs-Dichte hervorgeht, jedoch in komplizierter Weise von den Matrizen des zeitstetigen Systems abhängt.

Approximiert man die exogenen Variablen als Sprungfunktionen, so führen die Maximum-Likelihood-Gleichungen auf Ausdrücke, welche nur die Parameter des exakten diskreten Modells enthalten (dies ist ein Ausdruck der Nichtidentifizierbarkeit des Modells). Man erhält dann Lösungen, die mit dem gewöhnlichen kleinsten Quadrate-Schätzer übereinstimmen (analog zur Zeitreihenanalyse), jedoch sind aufgrund der Abhängigkeit des Fehlerterms vom autoregressiven Parameter weitere Lösungen möglich (dies kann durch die numerische Analyse bestätigt werden). Nach Einführung von Restriktionen, die eine Identifikation ermöglichen, erhält man komplizierte nichtlineare Gleichungen, die numerisch gelöst werden müssen (das gleiche gilt im Falle der Approximation der exogenen Größen durch Sprungfunktionen).

Die numerische Lösung der Likelihoodgleichungen wird mit einem Scoring-Algorithmus durchgeführt, bei dem aus Effizienzgründen exakte analytische Ausdrücke für die Ableitungen der Log-Likelihood (Score) und die **Fisher**-Information benutzt werden.

Da die exogenen Variablen für verschiedene Trajektorien i.a. unterschiedlich sind, muß die Konvergenz der **Fisher**-Informations-Matrix und die Konsistenz

des Maximum-Likelihood-Schätzers bei nicht identisch verteilten Beobachtungen untersucht werden. Unter etwas verschiedenen Forderungen folgt die Quadratmittelkonvergenz und die starke Konvergenz der geschätzten Informations-Matrix bzw. die starke Konsistenz des ML-Schätzers.

Anschließend wird die praktische Implementierung des Scoring-Verfahrens im Programmpaket LSD (Lineare Stochastische Differentialgleichungen) beschrieben. Ziel ist es, Simulation und Schätzung von Datensätzen im Rahmen der zeitstetigen Methode einem weiteren Anwenderkreis zugänglich zu machen. Da die Module in der Matrixsprache SAS/IML implementiert wurden, ist es möglich, die Anwenderprogramme mit wenigen Aufrufen in einer der Mathematik nahen Notation zu erstellen (eine ausführliche Programmbeschreibung ist in Vorbereitung). Zusätzlich kann auf existierende SAS-Datensätze zugegriffen und andere SAS-Software leicht eingebunden werden.

Numerisch studiert wird die Methode mit Hilfe simulierter Datensätze, die einem autoregressiven Prozeß 2. Ordnung entsprechen (stochastischer Oszillator). Dabei zeigt sich in überzeugender Weise, daß die exakte ML-Methode auch bei großen Zeitabständen der Messungen zu guten Schätzungen führt, während das linearisierte Modell hier völlig unbrauchbar ist. Weiterhin lassen sich gut die Anomalien untersuchen, die bei der Rückrechnung aus dem exakten diskreten Modell auftreten. Für die Simulation und Schätzung eines Datensatzes mit 50 Replikationen und 6 Zeitpunkten (exaktes **und** linearisiertes Modell) wurde eine durchschnittliche Rechenzeit von 6.7 Sekunden benötigt (BASF 7/73). Damit kann der Algorithmus als effizient bezeichnet werden.

Eine weitere Simulationsstudie hat das **Phillips-Modell** (**Phillips** 1972) zum Thema. Auch hier zeigt sich wiederum die Überlegenheit der exakten ML-Methode in der Genauigkeit gegenüber Linearisierung und Trapez-Methode.

Da häufig nicht alle Komponenten der Systemvariablen meßbar und durch Meßfehler überlagert sind, wird in **Kapitel 5** ein Meßmodell eingeführt. Die Parameterschätzung erfolgt dann mit Hilfe eines EM-Algorithmus, bei dem die Erwartungswertbildung (E) mit Hilfe eines rekursiven Glätters (**Rauch, Tung u. Striebel** 1965) und die Maximierung (M) mit dem in Kap. 4 diskutierten Scoring-Algorithmus durchgeführt wird. Die bekannten schlechten Konvergenzeigenschaften der EM-Methode in der Nähe des Maximums der Likelihood-Funktion werden dort durch Nachiteration mit einem **Newton-Raphson**-Algorithmus verbessert. Der so resultierende kombinierte EM/**Newton-Raphson**-Algorithmus (implementiert in IML-Modulen) wird dann auf die **Wolferschen** Sonnenflecken-daten angewandt. Hierbei zeigt sich, daß die Resultate zu den klassischen Untersuchungen von **Yule** und **Bartlett** vergleichbar sind. Insbesondere ist die Rolle des Meßfehlers herauszustellen, der mit der vorliegenden Methode expliziert modelliert wurde. Die Modellidentifikation kann dann mit Hilfe der "kom-

pensierten Likelihood" durchgeführt werden.

Abschließend wird im **6. Kapitel** ein Resümee gezogen und Erweiterungen der diskutierten Methoden auf verschiedene Themen (unregelmäßige Zeitabstände, beliebige Approximation der exogenen Variablen, stochastische Parameter (Personeneffekte), raum-zeitliche Panelmodelle etc.) diskutiert..

2. Modellspezifikation und Kritik existierender Schätzverfahren

2.1. Modellspezifikation

Wir nehmen an, daß der p -dimensionale Systemzustand $y(t)$ durch Eigendynamik, exogene Einflüsse und zufällige Störungen ξ einer Zeitentwicklung unterworfen ist, die durch folgende zufällige Differentialgleichung beschrieben werden kann:

$$(2.1.1) \quad d/dt y(t) = f(t, y(t)) + g(t, y(t)) \xi(t) ;$$

$$y(0) = y_0 ; t \in [t_0, t_T]$$

wobei y nur teilweise und mit Meßfehler beobachtet werden kann (Meß-Modell)

$$(2.1.2) \quad z_i = H(y(t_i), t_i) + \varepsilon_i ; \quad i=0, \dots, T$$

Insbesondere wird im folgenden fast ausschließlich der lineare Fall mit konstanten Koeffizienten behandelt

$$(2.2) \quad d/dt y(t) = Ay(t) + Bx(t) + G\xi(t) ; y(0) = y_0 \sim N(\mu, \Sigma)$$

$$z_i = Hy_i + Dx_i + \varepsilon_i ; i=0, \dots, T ;$$

Hierbei sind A ($p \times p$), B ($p \times q$), G ($p \times r$), H ($k \times p$) und D ($k \times q$) konstante, i.A. unbekannte Matrizen, welche Eigendynamik, Input (Kontrollvariable), Prozeßfehler und Output charakterisieren. Die Meßfehler ε_i werden als identisch verteilte, unabhängige **Gauß**- Variablen mit Kovarianzmatrix R : $k \times k$ und Erwartungswert 0 modelliert.

Solange es sich bei der Störgröße (Prozeßfehler $\xi(t)$) um einen trajektorienweise stetigen stochastischen Prozeß handelt, können obige Gleichungen trajektorienweise als gewöhnliche Differentialgleichungen (genaue Bedingungen werden bei **Ruymgaart u. Soong** 1971 angegeben) interpretiert oder mit dem Quadratmittelkalkül behandelt werden (vgl. **Soong** 1973). Will man jedoch Prozesse mit sehr kurzer Korrelationszeit betrachten (im Grenzfall weißes Rauschen), so ergeben sich charakteristische, in der Literatur wohlbekannte Schwierigkeiten, die eine vorsichtige Interpretation obiger Gleichungen erfordern.

Selbverständlich wird in der Natur ideales weißes Rauschen mit der Autokorrelation

$$E[\zeta(t)\zeta(s)] = \delta(t-s)$$

nicht vorkommen, ähnlich wie die δ -Funktion nur eine mathematische Idealisierung für sehr scharfe Impulse (Spikes) darstellt (dieser Standpunkt wird z.B. von **Van Kampen** 1987 vertreten). In diesem Sinne kann man also die Größe $\zeta(t)$ in obigen Gleichungen als Folge $\zeta_i(t)$ von Störungen mit sehr kurzer, aber endlicher Korrelationszeit auffassen, d.h.

$$E[\zeta_i(t)\zeta_i(s)] = \delta_i(t-s)$$

wobei $\delta_i(t-s)$ eine die δ -Funktion darstellende Funktionenfolge ist (vgl. hierzu **Lighthill** 1966, Kap. 2, **Urbanik** 1958, **Gelfand u. Wilenkin** 1964, S. 242).

Um die Schwierigkeiten zu umgehen, die mit der Nichtexistenz des weißen Rauschens als gewöhnlichem stochastischem Prozeß (bzw. der δ -Funktion als gewöhnlicher Funktion) zusammenhängen, wurde von **Ito** (1951) ein Kalkül von Integralgleichungen bzw. stochastischen Integralen entwickelt, das obigen Gleichungen einen präzisen Sinn verleiht. Setzt man formal $\zeta(t) = dw(t)/dt$, wobei $w(t)$ der **Wiener-Prozeß** (auch **Brownsche Bewegung**), ein stetiger stochastischer Prozeß mit unabhängigen Zuwächsen ist, so kann (2.1.1) in folgende Gleichung umgeschrieben werden

$$(2.3) \quad dy(t) = f(t,y(t))dt + g(t,y(t)) dw(t)$$

was nur eine Kurzschreibweise für die Integralgleichung

$$(2.4) \quad y(t) - y(t_0) = \int_{t_0}^t f(s,y(s))ds + \int_{t_0}^t g(s,y(s))dw(s)$$

darstellt.

Da $w(t)$ eine Funktion von nicht beschränkter Schwankung ist, muß dem formalen **Riemann-Stieltjes-Integral** eine Bedeutung verliehen werden, da die Integralsumme

$$\sum_{i=0}^{T-1} g(\tau_i, y(\tau_i)) [w(t_{i+1}) - w(t_i)] \quad ; \quad \tau_i \in [t_i, t_{i+1}]$$

je nach Wahl der Zwischenpunkte τ_i verschiedene Werte annehmen kann. Etwa

gilt für das stochastische Integral

$$(2.5) \quad \int_{t_0}^t w(s)dw(s) = \text{qm-lim}_{\Delta t \rightarrow 0} \sum_{i=0}^{T-1} w(\tau_i) [w(t_{i+1}) - w(t_i)] ; \Delta t = \max_i (t_{i+1} - t_i)$$

$$= (w^2(t) - w^2(t_0)) / 2 + (a-1/2)(t-t_0)$$

wobei für die Zwischenpunkte $\tau_i = t_i + a \Delta t$ gewählt wurde (siehe L. **Arnold** 1973, Kap. 4). Die Wahl $a=0$ (asymmetrisch bzgl. der Zukunft) ergibt die **Ito**sche Integraldefinition

$$(2.6) \quad (I) \int_{t_0}^t g(s, y(s)) dw(s) = \text{qm-lim}_{\Delta t \rightarrow 0} \sum_{i=0}^{T-1} g(t_i, y(t_i)) [w(t_{i+1}) - w(t_i)]$$

Vorteilhaft ist, daß das **Ito**-Integral als Funktion der oberen Grenze ein Martingal ist, die Formeln für die ersten beiden Momente einfach sind und die Berechnung der Likelihood-Funktion (**Radon-Nikodym**-Ableitung) auf **Ito**-Integrale führt (siehe Kap. 2.2). Auf der anderen Seite erhält man andere Rechenregeln als im klassischen Kalkül, wie Formel 2.5 zeigt. Dieser Nachteil kann durch die symmetrische Wahl $a=1/2$ beseitigt werden, zusammen mit den aufgeführten Vorteilen (ausführlich L. **Arnold** 1973, Kap. 4 und 10). Man erhält so das stochastische Integral im Sinne von **Stratonovich**

$$(2.7) \quad (S) \int_{t_0}^t g(s, y(s)) dw(s) = \text{qm-lim}_{\Delta t \rightarrow 0} \sum_{i=0}^{T-1} g(t_i, (y_i + y_{i+1})/2) [w(t_{i+1}) - w(t_i)]$$

Die Integrale können mit Hilfe einer **Taylor**-Entwicklung ineinander umgerechnet werden (L. **Arnold** 1973, Kap. 10)

$$(2.8) \quad (S) \int H(s, y(s)) dy(s) = (I) \int H(s, y(s)) dy(s) + 1/2 \sum_{jk} \int \partial H_{ij} / \partial x_k |_{y(s)} b_{jk}(s, y(s)) ds$$

wobei b_{jk} die Diffusionsmatrix des Diffusionsprozesses $y(t)$ darstellt (weitere Bedingungen an H und B siehe L. **Arnold** 1973, Satz 10.2.5).

Wichtig für die **empirische Modellbildung** und **Parameterschätzung** ist nun die richtige Interpretation der zufälligen Differentialgleichung (2.1), da der in (2.8) auftretende, sog. **Korrekturterm** auf unterschiedliche Lösungen führt.

Ich schlage folgende **empirische Prozedur** vor:

1. Die "reale" Differentialgleichung

$$\dot{y}_i = m(t, y_i) + g(t, y_i) \zeta_i$$

enthält einen Prozeßfehler ζ_i mit kurzer, aber endlicher Autokorrelationszeit τ_i .

2. Im Grenzfall $\tau_i \rightarrow 0$, $i \rightarrow \infty$ erhält man eine Folge von Lösungen $y_i(t)$. Diese Lösungen konvergieren (wie von **Wong u. Zakai** (1965) gezeigt), gegen die Lösung der **Stratonovich-Gleichung**

$$(S) \quad dy(t) = m(t, y(t))dt + g(t, y(t))dw(t)$$

die zur Lösung der **Ito-Gleichung**

$$(I) \quad dy(t) = \left[m(t, y(t)) + 1/2 \sum_{kj} \frac{\partial g_{ij}}{\partial x_k} \Big|_{y(t)} g_{kj} \right] dt + g(t, y(t))dw(t) \\ = f(t, y(t))dt + g(t, y(t))dw(t)$$

äquivalent ist (dies folgt aus Formel (2.8), wenn man dort für y den Vektor $[y', w']$ und $H=[0, g]$ setzt). Der Driftkoeffizient wird also durch die Interaktion des Systemzustandes y mit den stochastischen Fluktuationen "renormalisiert" (man spricht auch von rausch-induzierter Drift).

3. Zur Berechnung des Likelihood-Quotienten wird die **Ito-Gleichung** (I) verwendet.

Dieser ist (für kontinuierliche Datensätze) als

$$L(y) = \exp \left\{ \int \langle f, \Omega^{-1} dy \rangle - 1/2 \int \langle f, \Omega^{-1} f \rangle ds \right\}$$

gegeben (**Liptser u. Shirayev** 1978, Kap. 17, **Basawa u. Prakasa Rao** 1980, Kap. 9). Hierbei ist \langle , \rangle ein Skalarprodukt von p -Vektoren und $\int \langle f, \Omega^{-1} dy \rangle$ ein stochastisches **Ito-Integral**.

Insbesondere zeigt sich hier, daß auch im Falle eines zustandsunabhängigen Diffusionskoeffizienten $g(t, y) = g(t)$, bei dem **Stratonovich-** und **Ito-Lösung** übereinstimmen (vgl. (2.8)), die Likelihood-Funktion trotzdem auf Ito-Integrale führt (ausführlich Kap. 2.2).

Beispielsweise erhält man im Falle eines linearen, skalaren Systems

$$dy(t) = Ay(t)dt + dw(t)$$

die Likelihood $L = \exp\{A \int y dy - 1/2 A^2 \int y^2 dt\}$.

Das Ito-Integral $\int y dy$ ist hier durch $1/2 [y^2(t) - y^2(t_0) - (t - t_0)]$ gegeben, während die Nichtbeachtung der Rechenregeln zum Ausdruck $1/2 [y^2(t) - y^2(t_0)]$ führen würde. Der entsprechende ML-Schätzer $\hat{A} = (s) \int y dy / \int y^2 dt$ ist inkonsistent (vgl. **Singer** 1988).

Das eben vorgestellte Schema zeigt, daß sowohl **Ito**- als auch **Stratonovich**-Definition des stochastischen Integrals im Kontext von Modellbildung und Parameterschätzung ihre Bedeutung haben.

2.2. Kontinuierliche Datensätze

Wir nehmen im folgenden an, daß der Systemzustand $y(t)$ im Intervall $[t_0, t_T]$ kontinuierlich beobachtet werden kann. In diesem Fall kann die (konstante) Diffusionsmatrix $\Omega = GG'$ mit Hilfe der Statistik

$$(2.10) \quad \hat{\Omega} = \lim_{T \rightarrow \infty} 1/\tau \sum_{i=1}^{2^T} \Delta y_i \Delta y_i' = \Omega \quad (\text{fast sicher})$$

berechnet werden, wobei $\tau = t_T - t_0$ und $\Delta y_i = y(i\tau/2^T) - y((i-1)\tau/2^T)$ die Zuwächse des Systemzustands sind (siehe **Le Breton** 1976, **Basawa u. Prakasa Rao** 1980 Kap.9).

Mit Hilfe maßtheoretischer Methoden kann man zeigen, daß der Likelihood-Quotient (**Radon-Nikodym**-Ableitung) für das Modell $dy = f(\alpha, t, y)dt + g(t, y)dw$, $y(0) = y_0$ (deterministisch) mit der durch den Vektor α parametrisierten Drift f durch folgenden Ausdruck gegeben ist

$$(2.11) \quad L = (d\mu_\alpha / d\mu_\beta)(y) = \exp \left\{ \int \langle f(\alpha, y, t) - f(\beta, y, t), \Omega^{-1} dy(t) \rangle - 1/2 \int \langle f(\alpha, t, y) - f(\beta, t, y), \Omega^{-1} [f(\alpha, t, y) + f(\beta, y, t)] \rangle dt \right\}$$

Das stochastische Integral von $f(y(t))$ bzgl. dy kann dabei in Analogie zum Integral bzgl. dem **Wiener**-Prozeß definiert werden (**Liptser u. Shirayev** 1977, Kap.4).

Formel (2.11) kann in heuristischer Weise erhalten werden, wenn man die endlich-dimensionale Verteilung für die Zeitpunkte $\{t_0, t_1, \dots, t_T\}$ aufstellt und im Likelihood-Quotienten

$$L = p_\alpha(y_0, y_1, \dots, y_T) / p_\beta(y_0, y_1, \dots, y_T)$$

den Zeitabstand $\Delta t = \max(t_{i+1} - t_i)$ gegen 0 gehen läßt (Basawa u. Prakasa Rao 1980, Kap. 8, Wong u. Hajek 1985, Kap. 6).

Betrachten wir hierzu die Lösung der Ito-Gleichung mit dem fixen Startwert y bei t :

$$y(s) = y + \int_t^s f(s, y(s)) ds + \int_t^s g(s, y(s)) dw(s)$$

Für eine genügend feine Zerlegung kann man also schreiben:

$$(2.12) \quad y_{i+1} \approx y_i + f(t_i, y_i) \Delta t_i + g(t_i, y_i) \Delta w_i, \quad i=0, \dots, T-1$$

Da die Lösung $y(t)$ ein Markoff-Prozeß ist, kann die gemeinsame Wahrscheinlichkeit (bedingt auf y_0) als Produkt der Übergangswahrscheinlichkeiten dargestellt werden:

$$p(y_0, \dots, y_T) = \prod_{i=0}^{T-1} p(y_{i+1} | y_i)$$

Für genügend kleines Δt sind die Übergangswahrscheinlichkeiten approximative Normalverteilungen mit Erwartungswert $\mu_i = E[y_{i+1} | y_i]$ und Varianz $\sigma_i = V[y_{i+1} | y_i]$. Da in (2.12) y_{i+1} aus dem Startwert y_i entstanden ist, können die bedingten Momente aus den unbedingten Momenten von y_{i+1} berechnet werden, d.h.

$$\mu_i = y_i + f(t_i, y_i) \Delta t$$

$$\text{und} \quad \sigma_i = g(t_i, y_i) g'(t_i, y_i) \Delta t = \Omega_i \Delta t$$

Die gemeinsame Dichte der Variablen y_0, \dots, y_T lautet also näherungsweise ($f_i = f(t_i, y_i)$, $\Omega_i = \Omega(t_i, y_i)$, $\Delta y_i = y_{i+1} - y_i$ gesetzt)

$$\begin{aligned} p(y_0, \dots, y_T) &= \prod_i |2\pi\sigma_i|^{-1/2} \exp \left\{ -1/2 \sum_i (y_{i+1} - \mu_i)' \sigma_i^{-1} (y_{i+1} - \mu_i) \right\} \\ &= \prod_i |2\pi\Omega_i \Delta t|^{-1/2} \exp \left\{ -1/2 \sum_i \langle \Delta y_i, (\Omega_i \Delta t)^{-1} \Delta y_i \rangle \right. \\ &\quad \left. - 2 \langle f_i, \Omega_i^{-1} \Delta y_i \rangle + \langle f_i, \Omega_i^{-1} f_i \rangle \Delta t \right\} \end{aligned}$$

Teilt man durch die Dichte der Lösung mit Drift $f(\beta, t, y)$, um den divergenten ersten Term im Exponenten wegzuschaffen, so ergibt sich für den Likelihood-Quotienten ($f_\alpha = f(\alpha, t_i, y_i)$)

$$\begin{aligned} L &= p_\alpha(y_0, y_1, \dots, y_T) / p_\beta(y_0, y_1, \dots, y_T) \\ &= \exp \left\{ \sum \langle f_\alpha - f_\beta, \Omega^{-1} \Delta y \rangle - 1/2 \sum \langle f_\alpha, \Omega^{-1} f_\alpha \rangle \Delta t - \langle f_\beta, \Omega^{-1} f_\beta \rangle \Delta t \right\} \end{aligned}$$

und somit im Grenzfall $\Delta t \rightarrow 0$

$$L = \exp \left\{ \int \langle f_\alpha - f_\beta, \Omega^{-1} dy \rangle - 1/2 \int \langle f_\alpha, \Omega^{-1} f_\alpha \rangle dt - \langle f_\beta, \Omega^{-1} f_\beta \rangle dt \right\}$$

In Anlehnung an **Basawa u. Prakasa Rao** 1980, Kap. 9 sei im folgenden die Drift f mit Hilfe einer **Fourier**-Entwicklung parametrisiert:

$$f(\alpha, t, y) = \Phi_0(t, y) + \sum_{l=1}^k \alpha_l \Phi_l(t, y)$$

Wählt man als Referenzpunkt im Parameterraum $\alpha=0$ (o.B.d.A.), so gilt (die Diffusionsmatrix $\Omega(t, y)$ wird als bekannt vorausgesetzt)

$$L = \exp \left\{ \int \sum \alpha_l \langle \Phi_l, \Omega^{-1}(dy - \Phi_0 ds) \rangle - 1/2 \int \sum \alpha_l \alpha_k \langle \Phi_l, \Omega^{-1} \Phi_k \rangle ds \right\}$$

Für die Score-Funktion $s = \partial l / \partial \alpha$ ($l = \log L$) erhält man dann

$$s = \partial l / \partial \alpha_k = \int \langle \Phi_k, \Omega^{-1}(dy - \Phi_0 ds) \rangle - \int \sum \alpha_l \langle \Phi_k, \Omega^{-1} \Phi_l \rangle ds$$

An der Nullstelle von s (notwendig für das Maximum von l) gilt für den Maximum-Likelihood-Schätzer $\hat{\alpha}$:

$$(2.13) \quad \hat{\alpha} = B^{-1} \eta$$

wobei $\eta_k = 1/\tau \int \langle \Phi_k, \Omega^{-1}(dy - \Phi_0 ds) \rangle$ und $B_{kl} = 1/\tau \int \langle \Phi_k, \Omega^{-1} \Phi_l \rangle ds$ abgekürzt wurde.

Unter der Voraussetzung, daß B für $\tau = t_T - t_0 \rightarrow \infty$ in Wahrscheinlichkeit gegen eine nichtsinguläre Matrix J konvergiert, zeigen **Basawa u. Prakasa Rao**, daß $\hat{\alpha}$ ein konsistenter Schätzer mit asymptotischer Normalverteilung $N(\alpha, 1/\tau J^{-1})$ ist (Theorem 5.1, Kap. 9).

Spezialisiert man obige Resultate auf ein lineares Modell mit Drift $f = Ay + Bx$ und konstanter Diffusionsmatrix $\Omega = GG'$ so gilt

$$f = [A, B] \begin{bmatrix} y \\ x \end{bmatrix} = (I \otimes [y', x']) \text{ row } [A, B] := (I \otimes z') \alpha \\ = [\Phi_1, \dots, \Phi_k] \alpha \\ = \Phi \alpha$$

Dann findet man

$$\eta = 1/\tau \int \Phi' \Omega^{-1} dy = 1/\tau \int (I \otimes z) \Omega^{-1} dy = 1/\tau \int (\Omega^{-1} \otimes z) dy$$

und
$$B = 1/\tau \int \Phi' \Omega^{-1} \Phi dt = 1/\tau \int (I \otimes z) \Omega^{-1} (I \otimes z') dt = 1/\tau \int (\Omega^{-1} \otimes z z') dt$$

Daraus folgt für den ML-Schätzer

$$\begin{aligned} \hat{\alpha} &= \left[\Omega^{-1} \otimes \int z z' ds \right]^{-1} \left[\Omega^{-1} \otimes \int z \right] dy \\ &= \left[I \otimes \left(\int z z' ds \right)^{-1} \int z \right] dy \end{aligned}$$

Mit Hilfe von $\text{row}(ABC) = (\text{row } A \otimes C')$ $\text{row } B$ kann man diesen Ausdruck wieder "einrollen" und findet schließlich

$$(2.14) \quad \widehat{[A, B]} = \left[\int dy z' \right] \left[\int z z' ds \right]^{-1}$$

oder

$$\widehat{\begin{bmatrix} A' \\ B' \end{bmatrix}} = \left\{ \int \begin{bmatrix} yy' & yx' \\ xy' & xx' \end{bmatrix} ds \right\}^{-1} \left\{ \int \begin{bmatrix} y dy' \\ x dy' \end{bmatrix} \right\}$$

Interessant ist, daß die Diffusionsmatrix gar nicht mehr vorkommt (sie kann mit Hilfe von (2.10) unabhängig von der Drift bestimmt werden).

Die Voraussetzung kontinuierlicher Messung ist i.A. unrealistisch, sodaß die in (2.13) und (2.14) vorkommenden Integrale durch entsprechende Summen ersetzt werden müssen. Setzt man (in Anlehnung an **Le Breton** 1976)

$$\int yy' ds \rightarrow \sum_{i=0}^{T-1} y_i y_i' \Delta t_i \quad \text{und} \quad \int y dy' \rightarrow \sum_{i=0}^{T-1} y_i (y_{i+1} - y_i)'$$

so gilt für den entsprechenden diskretisierten Schätzer ("**diskretisiertes kontinuierliches Sampling**")

$$(2.15) \quad \widehat{\begin{bmatrix} A' \\ B' \end{bmatrix}}(\Delta t) = 1/\Delta t \begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}^{-1} \begin{bmatrix} Y \Delta Y' \\ X \Delta Y' \end{bmatrix}$$

wobei die aus den Integralen entstehenden Summen mit Hilfe der Daten-Matrizen $Y = [y_0, \dots, y_{T-1}]$, $Y_+ = [y_1, \dots, y_T]$ und $\Delta Y = Y_+ - Y$ geschrieben wurden.

Bemerkung: Analoge Resultate gelten für N unabhängige Trajektorien, wenn man y_i durch $[y_{1i}, \dots, y_{Ni}]$ ersetzt.

Le Breton zeigt (ohne exogene Variablen), daß

$$p\text{-lim}_{\Delta t \rightarrow 0} \hat{A}(\Delta t) = p\text{-lim}_{\Delta t \rightarrow 0} \hat{A}_{\text{EXAKT}} = \hat{A}$$

wobei A_{EXAKT} der ML-Schätzer für das sog. **exakte diskrete Modell** ist (d.h. für die Likelihood des Datensatzes $\{y_0, \dots, y_T\}$; vgl. Abs. 2.4).

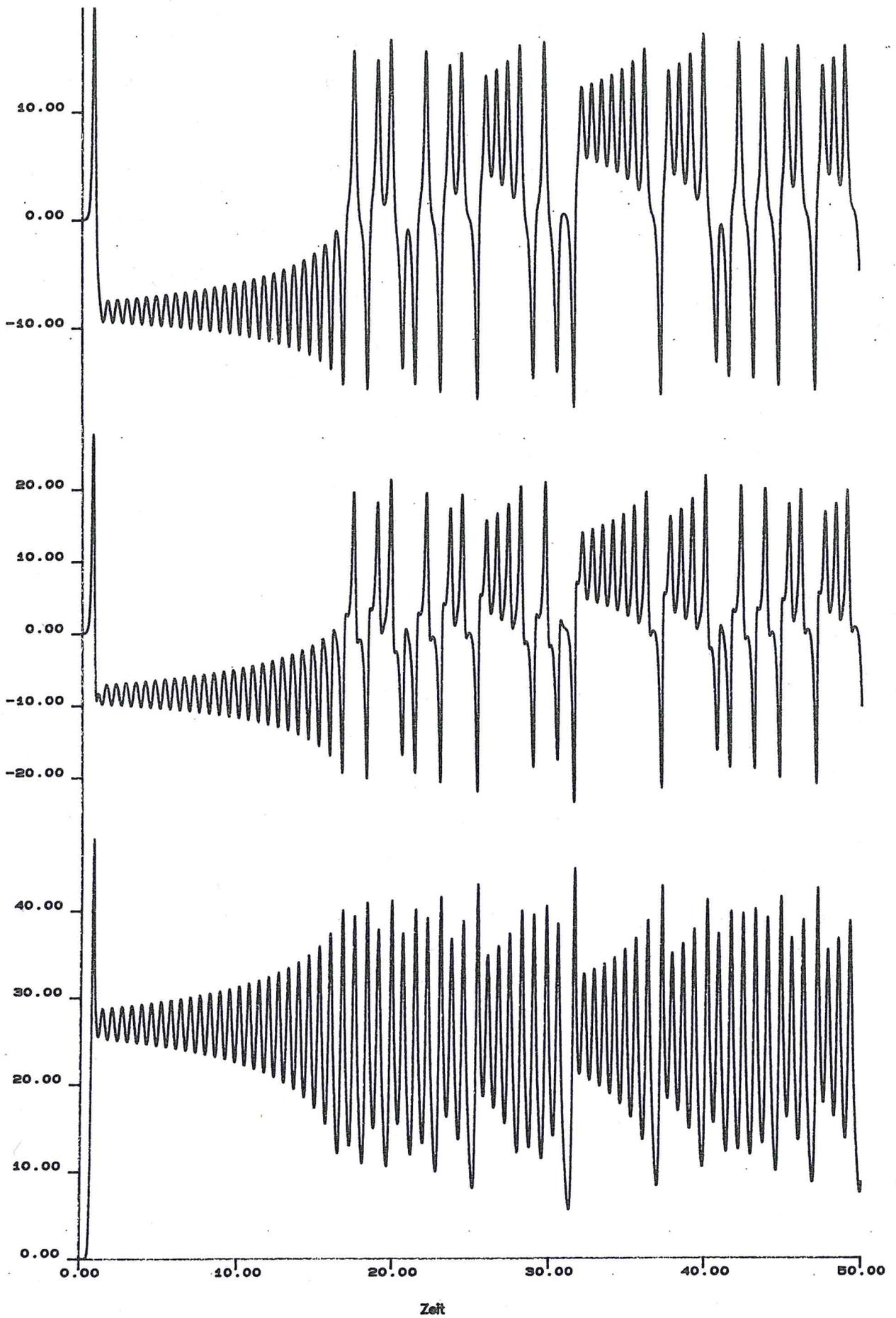


Abb.2.1. Trajektorien $x(t)$, $y(t)$, $z(t)$ des Lorenz-Attraktors (ohne Prozeß-Fehler)

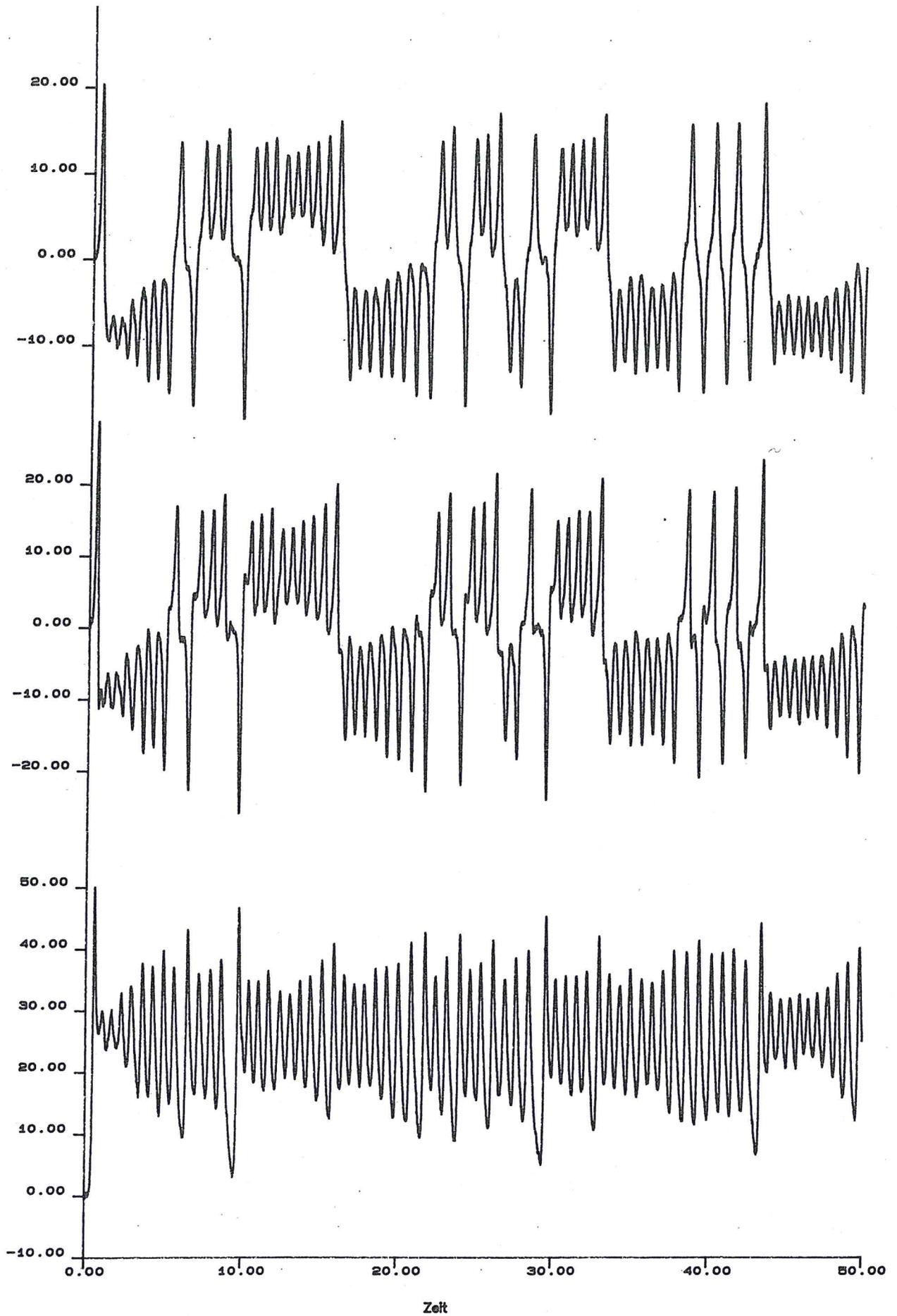


Abb. 2.2. Trajektorien des Lorenz-Modells (mit Prozeß-Fehler $G = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix}$)

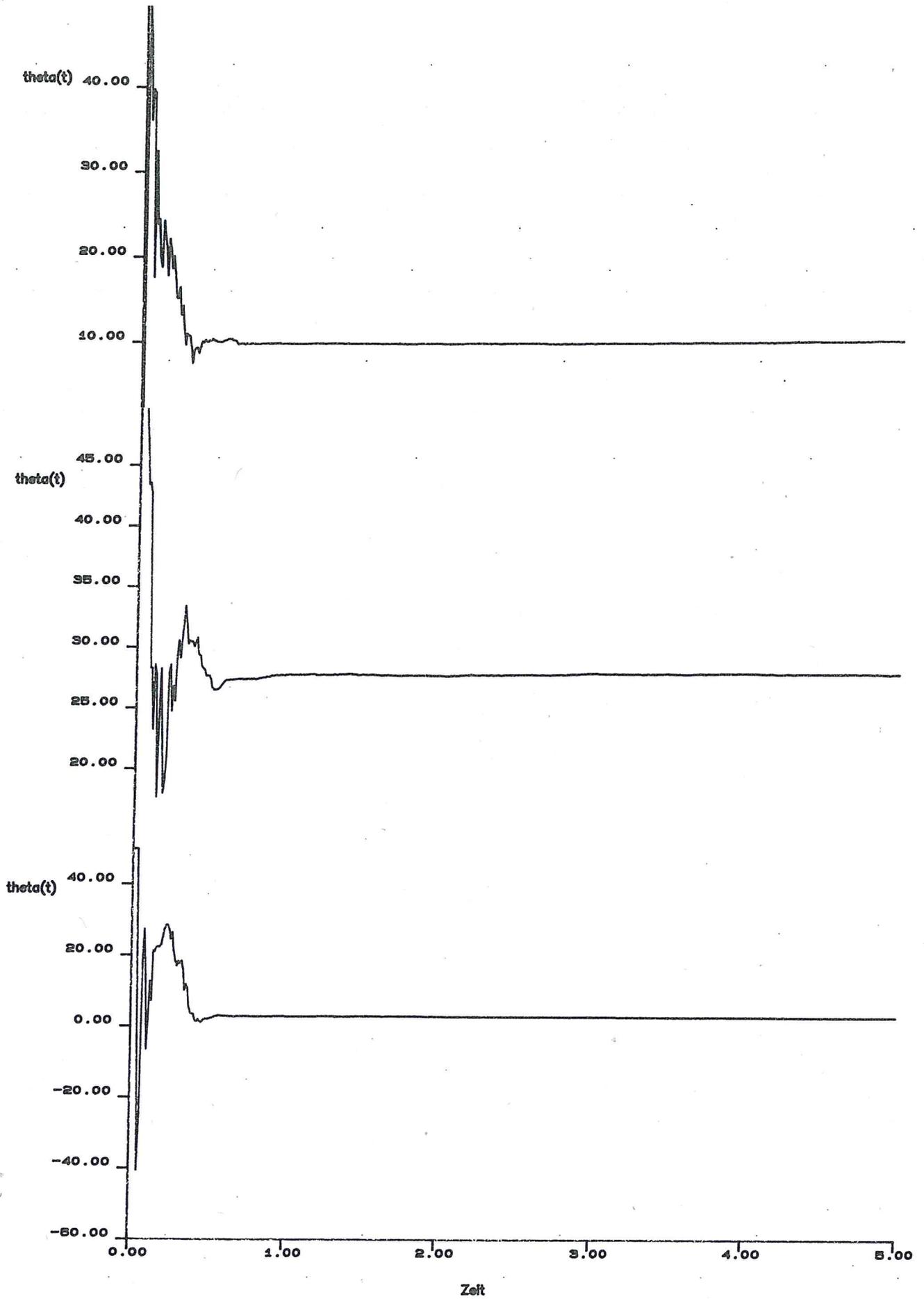


Abb. 2.3. ML-Schätzer von $\Theta = \{\sigma, r, b\}$ als Funktion der Zeit für die Trajektorie aus Abb. 2.2

sind die typischen Spikes am Beginn der Kurve, die wegen

$$\hat{\Theta} = \Theta + \left[\int \Phi' \Omega^{-1} \Phi \, ds \right]^{-1} \left[\int \Phi' \Omega^{-1} G dw(s) \right]$$

proportional zu $\Delta w / \Delta t = (w(t_1) - w(t_0)) / \Delta t$ (Weißes Rauschen) sind.

Insgesamt konvergiert der Schätzer schnell gegen die Parameter Θ . Das selbe Resultat ergibt sich auch für stärkere Prozeßfehleranteile, bei denen die deterministischen Trajektorien kaum noch erkennbar sind.

2.3. Die Trapez-Näherung

Eine in ökonomischen Anwendungen häufig benutzte Näherungsmethode ist das sog. **approximative diskrete Modell** (vgl. **Sargan 1976, Wymer 1976, Bergstrom 1984**). **Bergstroms** Idee war, die Integralform (2.4) der linearen stochastischen Differentialgleichung

$$y(t) - y(0) = A \int_{t_0}^t y(s) ds + B \int_{t_0}^t x(s) ds + \int_{t_0}^t G dw(s)$$

mit Hilfe einer Trapeznäherung für die Integrale als Funktion der Daten $[y_0, \dots, y_T]$ und $[x_0, \dots, x_T]$ auszudrücken. Schreibt man

$$\int_{t_i}^{t_{i+1}} y(s) ds \approx \Delta t / 2 [y_i + y_{i+1}] = \Delta t \bar{y}_i$$

(analog für x) und definiert die Datenmatrizen $\bar{Y} = [\bar{y}_0, \dots, \bar{y}_T]$, $\bar{X} = [\bar{x}_0, \dots, \bar{x}_T]$, $\Delta W = [w(t_1) - w(t_0), \dots, w(t_T) - w(t_{T-1})]$, so ergibt sich folgendes Schema:

$$(2.16) \quad Y_+ - Y - A\bar{Y}\Delta t - B\bar{X}\Delta t = G\Delta W$$

Da die Störgröße $G\Delta W$ die unabhängigen Zuwächse des **Wiener-Prozesses** enthält, ist deren Kovarianz durch $\text{Cov}[\text{col}(G\Delta W)] = I_T \otimes GG'\Delta t$ gegeben.

Üblicherweise wird das Schema (2.16) als simultanes Gleichungssystem umgeschrieben und mit Hilfe von Strukturgleichungsmodellen (z.B. LISREL) geschätzt (vgl. beispielsweise **Möbus u. Nagl 1983**).

Im folgenden soll jedoch der ML-Schätzer explizit berechnet werden, um einen direkten Vergleich mit den übrigen Methoden zu erhalten. Die log-Likelihood

lautet

$$(2.17) \quad l(A,B,\Omega) = \text{const.} - T/2 \log|\Omega\Delta t| - 1/2 \text{tr}[(\Omega\Delta t)^{-1}ZZ'] + \log J$$

wobei $Z = \Delta Y - A\bar{Y}\Delta t - B\bar{X}\Delta t$, $\Delta Y = Y_+ - Y$ und $J = |I - A\Delta t/2|$ die **Jacobi**-Determinante der Transformation (2.16) darstellt.

Mit Hilfe des **McDonald-Swaminathan**-Kalküls (Anhang 1) erhält man die ML-Gleichungen

$$(2.18) \quad \begin{aligned} \partial l / \partial \Omega &= -1/2 L \text{ row} \cdot [T\Omega^{-1} - 1/\Delta t \Omega^{-1} ZZ' \Omega^{-1}] = 0 \\ \partial l / \partial A &= \text{row} [(\Omega\Delta t)^{-1} Z \bar{Y}' \Delta t - T\Delta t/2(I - A'\Delta t/2)^{-1}] = 0 \\ \partial l / \partial B &= \text{row} [(\Omega\Delta t)^{-1} Z \bar{X}' \Delta t] = 0 \end{aligned}$$

Hierbei ist $L = (\partial S / \partial S)$ eine irreduzible Matrixableitung (S ist symmetrisch) (**McDonald u. Swaminathan** 1973). (2.18.1) ergibt die Lösung (zu beachten ist die Singularität von L)

$$(2.19) \quad \hat{\Omega} = 1/T\Delta t ZZ'$$

Setzt man dies in die 2. Gleichung ein, so findet sich

$$(I - A\Delta t/2) \bar{Y}' (\Delta Y - A\bar{Y}\Delta t - B\bar{X}\Delta t)' = 0$$

Der Term $Y\Delta\bar{Y}'$ ist eine Integralsumme des stochastischen Integrals im Sinne von **Stratonovich**. Beachtet man noch die 3. Gleichung, die zu $Z\bar{X}' = 0$ führt, ergibt sich schließlich

$$(2.20) \quad \begin{aligned} \bar{Y}' \Delta Y' - 1/2 \Delta Y \Delta Y' - \bar{Y}\bar{Y}' \Delta t A' - \bar{Y}\bar{X}' \Delta t B' &= 0 \\ \bar{X}' \Delta Y' - \bar{X}\bar{Y}' \Delta t A' - \bar{X}\bar{X}' \Delta t B' &= 0 \end{aligned}$$

Eine kurze Umrechnung zeigt, daß

$$(2.21) \quad \bar{Y}' \Delta Y' - 1/2 \Delta Y \Delta Y' = Y \Delta Y'$$

sodaß die **Stratonovich**-Summen in **Ito**-Summen transformiert werden (dies wird durch die **Jacobi**-Determinante bewirkt). Im Limes $\Delta t \rightarrow 0$ erhält man also einen Spezialfall der Transformationsformel (2.8)

$$(S) \int y dy' = (I) \int y dy' + 1/2 \Omega (t_T - t_0)$$

Auflösen der Normalgleichungen (2.20) nach A und B ergibt schließlich

$$\begin{bmatrix} \hat{A}' \\ \hat{B}' \end{bmatrix}_{\text{TRAPEZ}} = 1/\Delta t \begin{bmatrix} \overline{YY'} & \overline{YX'} \\ \overline{XY'} & \overline{XX'} \end{bmatrix}^{-1} \begin{bmatrix} \overline{Y\Delta Y'} \\ \overline{X\Delta Y'} \end{bmatrix}$$

Im Vergleich zum diskretisierten kontinuierlichen Sampling (2.15) erhält man hier eine etwas andere Näherung für die Zeitintegrale, wird jedoch wiederum auf das stochastische **Ito**-Integral geführt (ersetzt man $Y\Delta Y'$ durch $\overline{Y\Delta Y'}$, so erhält man einen Schätzer, der stochastisch gegen 0 konvergiert!).

Verwendet man statt der Trapez- eine Rechtecknäherung für die Integrale, so ergibt sich für die Drift-Koeffizienten Formel (2.15), also "diskretisiertes kontinuierliches Sampling", während die Lösung für die Diffusionsmatrix

$$\hat{\Omega} = 1/T\Delta t \text{ ZZ}'[\hat{A}, \hat{B}]$$

nicht mit (2.10) übereinstimmt. (im Falle des kontinuierlichen Samplings wird Ω als bekannt vorausgesetzt, während hier bzgl. Ω maximiert wurde; vgl. **Dacunha-Castelle u. Florens-Zmirou 1986**).

2.4 Das exakte diskrete Modell

Die lineare stochastische Differentialgleichung

$$dy(t) = [Ay(t) + Bx(t)]dt + Gdw(t)$$

kann explizit gelöst werden (siehe **Arnold 1973**, Kap. 8):

$$y(t) = e^{A(t-t_0)}y(t_0) + e^{At} \int_{t_0}^t e^{-As} [Bx(s) ds + Gdw(s)]$$

Für die Meßzeitpunkte ergibt sich also folgendes rekursives Schema (**exaktes diskretes Modell**):

$$y_{i+1} = e^{A\Delta t} y_i + e^{A\Delta t} \int_0^{\Delta t} e^{-As} Bx(t_i+s) ds + e^{A\Delta t} \int_0^{\Delta t} e^{-As} Gdw(t_i+s)$$

(2.23) $y_i = y(t_i)$

Abgekürzt hat man $y_{i+1} = A^* y_i + F_i(A, B, X) + u_i$; $i=0, \dots, T-1$

wobei $\text{cov}(u_i, u_j) = \delta_{ij} V$ und $V = \int \exp(As) \Omega \exp(A's) ds$.

Dieses Modell ist jedoch nichtlinear im Parameter A. Hieraus ergeben sich 2 Hauptprobleme, die einen großen Raum in dieser Arbeit einnehmen:

1. Die Matrix-Funktion $A^* = \exp(A\Delta t)$ ist nicht eindeutig; d.h. es gibt Matrizen $A_N \neq A$ mit $\exp(A_N \Delta t) = \exp(A \Delta t)$. Das daraus erwachsende **Identifikationsproblem** wird in Kap. 3 diskutiert.

2. Man erhält ein nichtlineares Schätzproblem, das numerisch gelöst werden muß (Kap. 4).

Von Autoren aus dem soziologischen und technischen Bereich (**Doreian u. Humm-on** 1976, **Tuma u. Hannan** 1984, **Armingier** 1986, **Isermann** 1988) wurde vorgeschlagen, das Modell (2.23) nach einer Reparametrisierung zu schätzen (mit Programmen wie LISREL etc.) und dann aus den reduzierten Parametern auf die ursprünglichen Strukturparameter zurückzuschließen. Dabei wird jedoch das Identifikationsproblem nicht beachtet (der Matrix-Logarithmus in $\hat{A} = 1/\Delta t \log(\hat{A}^*)$ ist mehrdeutig, wenn \hat{A}^* komplexe Eigenwerte aufweist und führt für negative Eigenwerte zu komplexem A. Weiterhin wird vernachlässigt, daß V eine Funktion der Strukturparameter A und Ω ist. Restriktionen auf A, B und Ω , die eine Identifikation ermöglichen, sind jedoch auf der Ebene des reparametrisierten Modells nicht implementierbar, da die reduzierten Parameter komplizierte nichtlineare Matrixfunktionen sind (ausführlich hierzu Kap. 3 und 4).

Für $\Delta t \rightarrow 0$ ist jedoch $A^*(A) = I + A\Delta t$, sodaß in diesem Grenzfall das Identifikationsproblem verschwindet und der ML-Schätzer explizit angegeben werden kann. Wie in Kap. 4 gezeigt (vgl. auch **Hamerle, Nagl u. Singer** 1988) gilt für die Likelihood der diskret gemessenen Daten $[y_0, \dots, y_T]$ und $[x_0, \dots, x_T]$ (bedingt auf den Anfangswert y_0)

$$(2.24) \quad l = -T/2 \left[\log|V| + \text{tr}[V^{-1} 1/T UU'] \right]$$

mit $U = Y_+ - A^* Y - B^* X$, $A^* = \exp(A\Delta t)$, $B^* = A^{-1}(A^* - I)B$.

Die Lösung der ML-Gleichungen liefert

$$(2.25) \quad \hat{V} = 1/T UU'[\hat{A}^*, \hat{B}^*]$$

$$\begin{bmatrix} \hat{A}^* \\ \hat{B}^* \end{bmatrix} = \begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}^{-1} \begin{bmatrix} YY_+ \\ XY_+ \end{bmatrix}$$

Im Grenzfall $\Delta t \rightarrow 0$ kann man A^* und B^* entwickeln und erhält so $A^* = I + A\Delta t + O(\Delta t^2)$ und $B^* = B\Delta t + O(\Delta t^2)$. Einsetzen in (2.25) führt dann zum Ergebnis

$$(2.26) \quad \begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix}_{\text{EXAKT}} = \begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}^{-1} \begin{bmatrix} Y\Delta Y' \\ X\Delta Y' \end{bmatrix} ; \Delta t \rightarrow 0$$

Dies stimmt aber mit Formel (2.15) für das diskretisierte kontinuierliche Sampling überein. Von **Le Breton** (1976) wurde gezeigt, daß $\hat{A}(\Delta t)$ und $\hat{A}_{\text{EXAKT}}(\Delta t)$ für $\Delta t \rightarrow 0$ mit dem zeitkontinuierlichen Schätzer (2.14) übereinstimmen. Entsprechend kann man nachweisen, daß der Likelihood-Quotient der Zufallsgrößen $[y_0, \dots, y_T]$ mit dem **Wiener**-Prozeß als Referenz:

$$L = \frac{|2\pi V|^{-T/2} \exp\{-1/2 \operatorname{tr}[V^{-1}UU']\}}{|2\pi\Omega\Delta t|^{-T/2} \exp\{-1/2 \operatorname{tr}[(\Omega\Delta t)^{-1}\Delta Y\Delta Y']\}}$$

mit $U = Y_+ - A^*Y - B^*X$ für $\Delta t \rightarrow 0$ gegen die **Radon-Nikodym**-Ableitung der kontinuierlich beobachteten Trajektorie strebt (vgl. **LeBreton** 1976).

Im univariaten Fall $p=1$ ist der Logarithmus eine injektive Funktion, sodaß Gleichung (2.25) für $\hat{A}^* > 0$ nach \hat{A} und \hat{B} aufgelöst werden kann :

$$\hat{A} = 1/\Delta t \log \hat{A}^* ; \quad \hat{B} = \hat{A}/(\hat{A}^* - 1) * \hat{B}^*$$

In den Arbeiten von **Jones** (1984) und **Jones u. Tryon** (1987) über Zeitreihenanalyse mit unregelmäßigen Zeitabständen wird die Maximierung der Likelihood (2.24) mit Hilfe eines numerischen Optimierungsverfahrens (Quasi-Newton-Algorithmus mit numerischen Ableitungen) vorgenommen. Obwohl einfach zu implementieren, ist es aus Gründen der Effizienz vorzuziehen, mit exakten analytischen Ableitungen zu arbeiten. **Jones** (1984) schreibt z.B.:

" The likelihood function is not well behaved and can have local maxima which are far below the global maximum. Local maxima can occur because of ripples caused by numerical roundoff.... The nonlinear optimization program spends a lot of effort searching for the minimum of $-\ln$ likelihood and the final result is estimates of the parameters for which approximate confidence intervals can not be obtained by numerical methods."

2.5. Minimum Distance-Verfahren

Ein von **Phillips** (1972) und **Hauptmann u. Schmid** (1980) vorgeschlagenes Ver-

fahren besteht darin, ausgehend vom exakten diskreten Modell (2.23) die Parameter durch Minimierung der Funktion

$$Q(A,B) = \sum_{i=0}^{T-1} \left\| y_{i+1} - A^* y_i + F_i(A,B,X) \right\|^2$$

zu schätzen. V wird dann als $\hat{V} = 1/T \sum u_i u_i'$ $[\hat{A}, \hat{B}]$ geschätzt. Die Norm $\|\cdot\|$ läßt sich mit Hilfe der positiv definiten Matrix S als quadratische Form $\|z\|^2 = z'S^{-1}z$ ausdrücken. Das Verfahren von **Phillips** besteht darin, die Minimierung mit $S=I$ zu beginnen und anschließend in einem zweiten Durchgang mit $S=\hat{V}$ den MD-Schätzer $A^{\|\|}$, $B^{\|\|}$ und $V^{\|\|}$ zu berechnen. Da die Momentenmatrix \hat{V} gegen V konvergiert ($T \rightarrow \infty$), ist der MD-Schätzer asymptotisch äquivalent zum ML-Schätzer bei festem V . Allerdings können bei diesem Verfahren keine Restriktionen für die Diffusionsmatrix Ω implementiert werden. Gegenüber den diskreten Approximationen hat der MD-Schätzer jedoch den Vorzug, die nicht-lineare Struktur von $A^* = \exp(A\Delta t)$ zu berücksichtigen, während die Abhängigkeit von V von A und Ω nur mit der vollen Likelihood-Methode (Kap. 4) in Betracht gezogen werden kann.

2.6. Asymptotische Eigenschaften

Bei der Approximation des stochastischen Ito-Integrals durch Integralsummen bzw. bei Anwendung der Rechteck- oder Trapezformel für die Zeit-Integrale werden Fehler gemacht, die sich in einer Verzerrung der Schätzer bemerkbar machen, dies umso mehr, je größer der Zeitabstand Δt der Messungen ist. Der Einfachheit halber wird im folgenden nur der Fall ohne Exogene x diskutiert,

2.6.1 Diskretisiertes kontinuierliches Sampling

Wie in Kap. 2.2 abgeleitet, gilt

$$\hat{A}(\Delta t) = \Delta Y Y' (Y Y' \Delta t)^{-1} = \sum_{i=0}^{T-1} (y_{i+1} - y_i) y_i' \left[\sum_{i=0}^{T-1} y_i y_i' \Delta t \right]^{-1}$$

Setzt man hier das exakte diskrete Modell $y_{i+1} = A^* y_i + u_i$ ein, ergibt sich

$$\hat{A}(\Delta t) = \left[\sum_{i=0}^{T-1} (A^* - I) y_i y_i' + \sum_{i=0}^{T-1} u_i y_i' \right] \left[\sum_{i=0}^{T-1} y_i y_i' \Delta t \right]^{-1}$$

$$= 1/\Delta t (A^* - I) + \sum_{i=0}^{T-1} u_i y_i' \left[\sum_{i=0}^{T-1} y_i y_i' \Delta t \right]^{-1}$$

Wenn man weiterhin berücksichtigt, daß (vgl. hierzu **Anderson** 1971, Kap. 5)

$$(2.27) \quad \text{qm-lim } 1/T \sum_{i=0}^{T-1} y_i y_i' = \Sigma \quad \text{und} \quad \text{qm-lim } 1/T \sum_{i=0}^{T-1} u_i y_i' = 0 \quad ,$$

wobei Σ die stationäre Kovarianz ist, die als Lösung der **Liapunov**-Gleichung $A\Sigma + \Sigma A' + \Omega = 0$ gewonnen werden kann (dazu muß vorausgesetzt werden, daß alle Eigenwerte von A negativen Realteil aufweisen), gilt

$$(2.28) \quad \underset{T \rightarrow \infty}{p\text{-lim}} \hat{A}(\Delta t) = 1/\Delta t (A^* - I) = A + A^2/2 \Delta t + O(\Delta t^2)$$

Der ML-Schätzer für das diskretisierte kontinuierliche Sampling ist also asymptotisch verzerrt. Für Parameterwerte A und Zeitabstände Δt , für die $\|\Delta t A\| \ll 1$ nicht erfüllt ist, können sich erhebliche Abweichungen vom wahren Wert ergeben (siehe 2.7.3).

2.6.2 Trapez-Methode

In Kap. 2.3 wurde gezeigt, daß die Trapez-Methode auf den ML-Schätzer

$$\hat{A}_{\text{TRAPEZ}} = \Delta Y Y' [\bar{Y} Y' \Delta t]^{-1} = \sum_{i=0}^{T-1} (y_{i+1} - y_i) y_i' \left[\sum_{i=0}^{T-1} (y_{i+1} + y_i) y_i' \Delta t / 2 \right]^{-1}$$

führt. Berücksichtigt man (2.27), so gilt

$$(2.29) \quad \underset{T \rightarrow \infty}{p\text{-lim}} \hat{A}_{\text{TRAPEZ}} = (2/\Delta t)(A^* - I)(A^* + I)^{-1} = (2/\Delta t) P [\tanh \lambda \Delta t / 2] P^{-1}$$

wobei die Eigenwertzerlegung $A = PAP^{-1}$ benutzt wurde. Eine **Taylor**-Entwicklung des Hyperbeltangens liefert dann:

$$\underset{T \rightarrow \infty}{p\text{-lim}} \hat{A}_{\text{TRAPEZ}} = A - 1/12 A^3 \Delta t^2 + O(\Delta t^4) ; \quad |\lambda \Delta t| < \pi$$

Für kleines Δt liefert also die Trapezmethode eine kleinere asymptotische Verzerrung ($T \rightarrow \infty$) als die Rechtecknäherung bzw. das diskretisierte kontinuierliche Sampling (cf. auch **Sargan** 1976).

2.6.3 Exaktes Diskretes Modell

Im Falle $p=1$ kann eine explizite Formel angegeben werden (siehe Kap. 2.4)

$$\hat{A}_{\text{Exakt}} = 1/\Delta t \log \hat{A}^*;$$

mit $\hat{A}^* = \sum_{i=0}^{T-1} y_i y_{i+1} / \sum_{i=0}^{T-1} y_i^2$. Mit Hilfe von $\Delta y_i = y_{i+1} - y_i$ kann man schreiben:

$$\begin{aligned} \hat{A}_{\text{Exakt}} &= 1/\Delta t \log \left[\sum_{i=0}^{T-1} y_i \Delta y_i / \sum_{i=0}^{T-1} y_i^2 + 1 \right] \\ &= 1/\Delta t \log \left[\Delta t * \hat{A}(\Delta t) + 1 \right] \end{aligned}$$

wobei $\hat{A}(\Delta t)$ der in Kap. 2.7.1 diskutierte Schätzer ist (diskretisiertes kontinuierliches Sampling). Dieser strebt jedoch, wie gezeigt, gegen $1/\Delta t(A^* - I)$, sodaß mit Hilfe des **Slutzkyschen** Theorems [$p\text{-lim } g(x_n) = g(x)$, wenn $p\text{-lim } x_n = x$, siehe **Wilks 1962**] sofort die Konsistenz $p\text{-lim } \hat{A}_{\text{Exakt}} = A$ folgt.

2.6.4 Numerische Beispiele

Um das eben abgeleitete asymptotische Verhalten der diskretisierten Schätzer zu illustrieren, werden im folgenden simulierte Datensätze betrachtet, die als Lösung der stochastischen Differentialgleichung

$$(2.30) \quad dy(t) = Ay(t)dt + Gdw(t)$$

mit der zufälligen Anfangsbedingung $y(0) \sim N(\mu, \Sigma)$ und den Werten $A=-2$, $G=2$, $\mu=0$ und $\Sigma=1$ entstehen. Die Daten wurden mit Hilfe von **Eulerschen** Polygonzügen simuliert, welche für $dt \rightarrow 0$ gegen die **Ito**-Lösung von (2.30) konvergieren (siehe **Gard 1988**, Kap. 7):

$$y_{i+1} = y_i + Ay_i dt + Gz_i dt^{1/2}; \quad y_0 \sim N(0,1) \quad ; \quad i=0, \dots, T-1$$

wobei z_i unabhängige und identisch normalverteilte Zufallsgrößen sind. Abb. 2.4 zeigt eine Trajektorie des Lösungsprozesses im Intervall $[0, 20]$, wobei ein Zeitraster von $dt=.01$ gewählt wurde. Dieser Pfad wurde nun für die Schätzung der Drift A benutzt.

Fig. 2.5 illustriert das zeitliche Verhalten des (quasi)kontinuierlichen **Ito**-Schät-

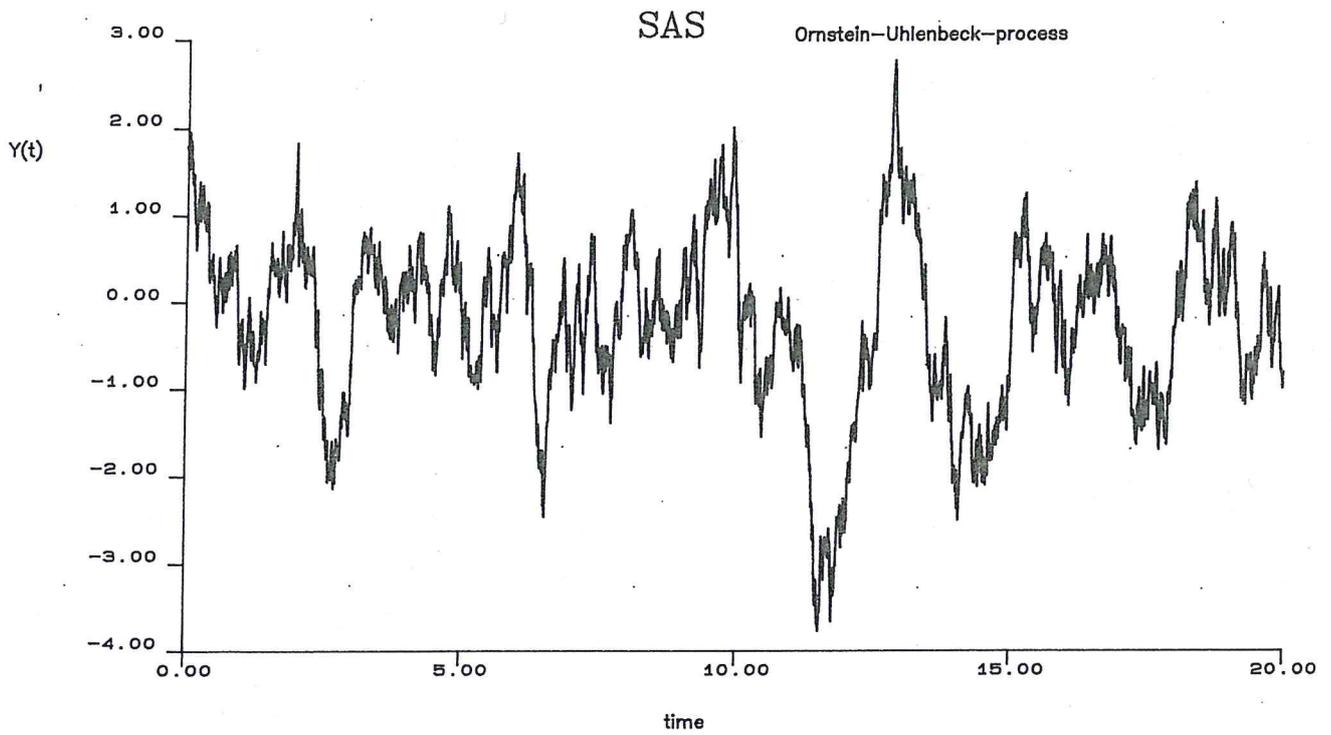


Abb. 2.4. Lösungstrajektorie des Ornstein-Uhlenbeck-Prozesses

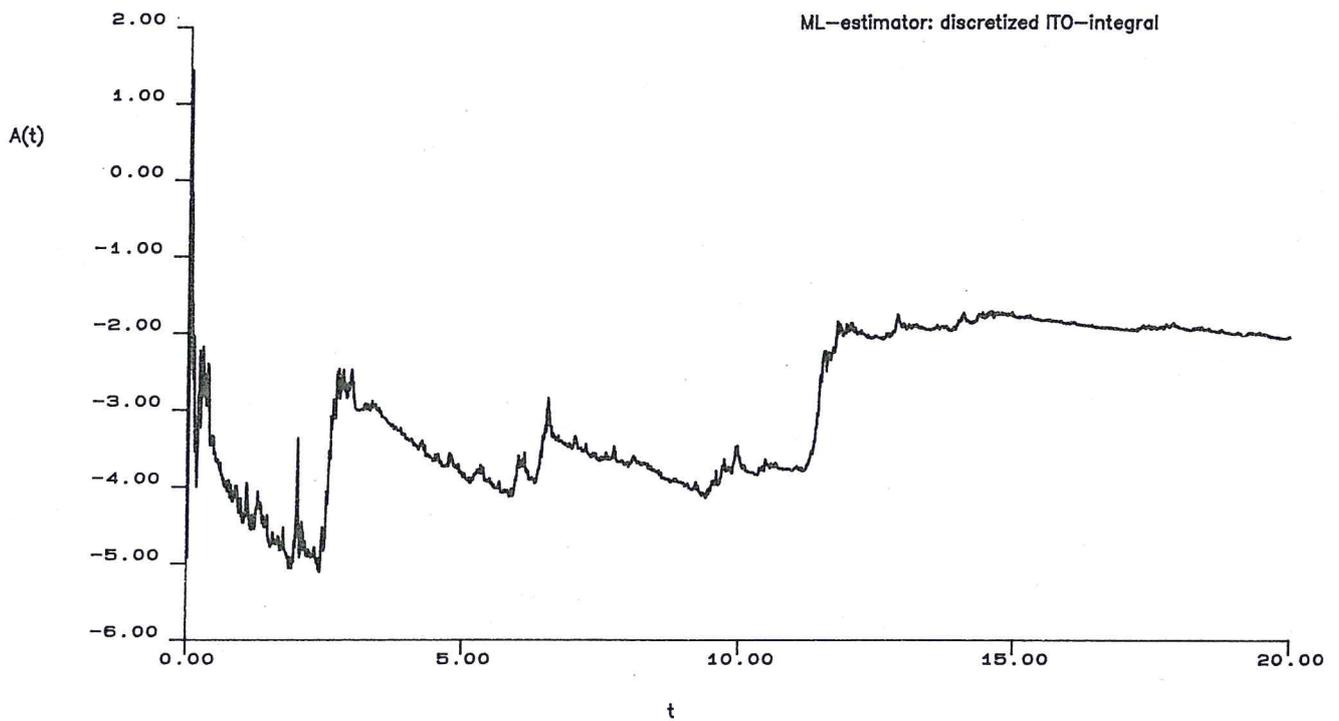


Abb. 2.5. zeitlicher Verlauf des ML-Schätzers (quasikontinuierliches Sampling)

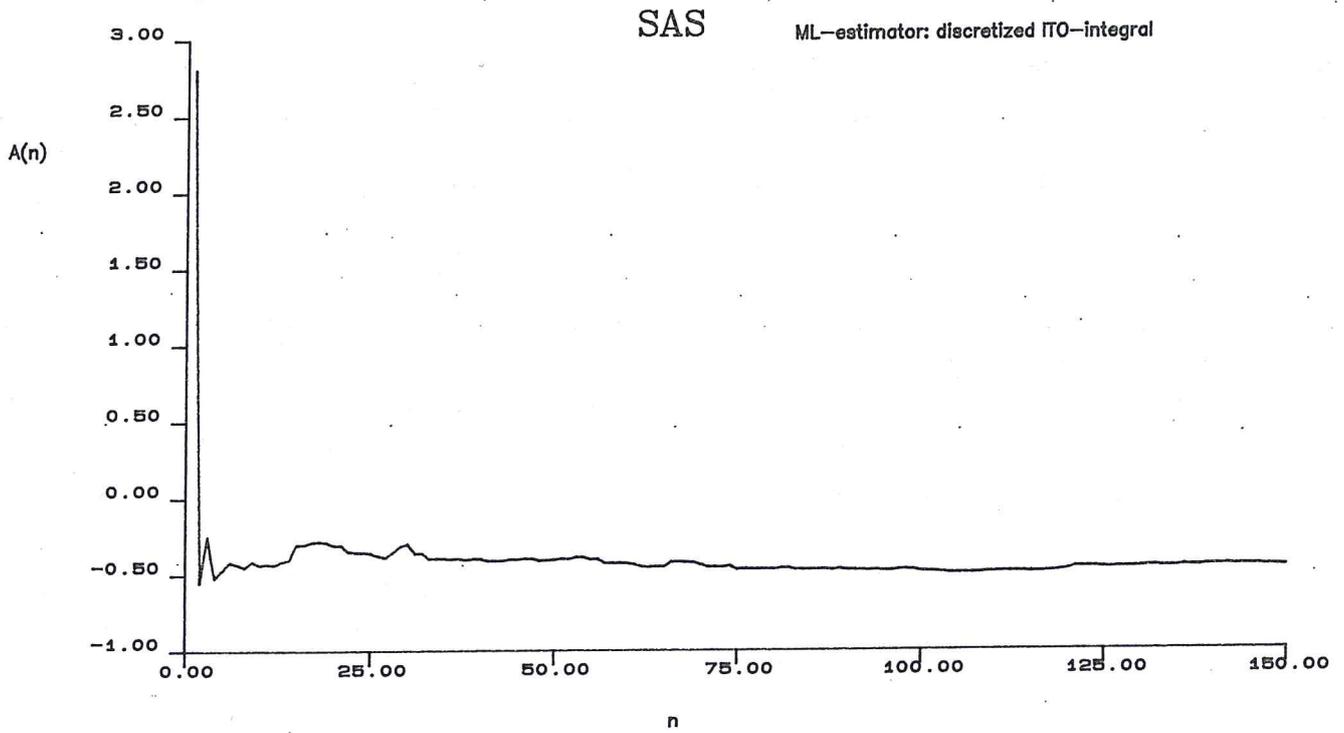


Abb. 2.6. Diskretisiertes kontinuierliches Sampling ($\Delta t=2$) als Funktion der Stichprobengröße n .

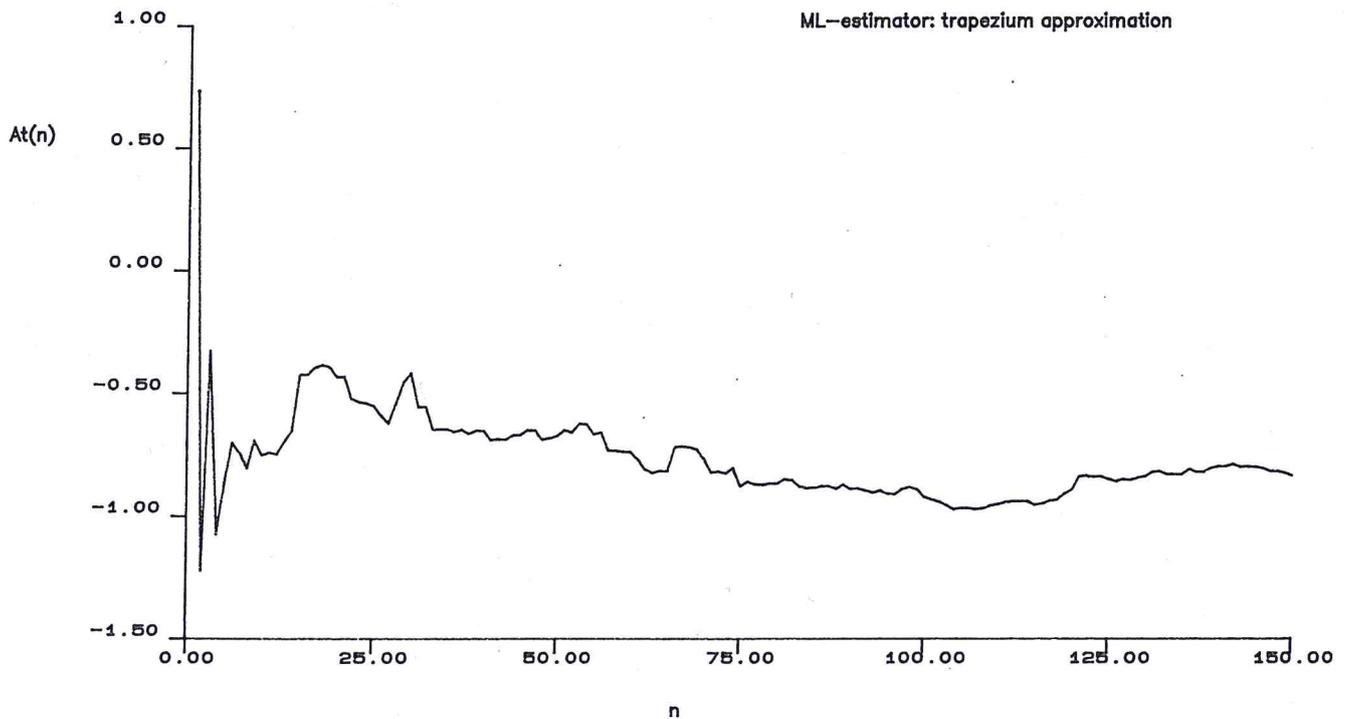


Abb. 2.7. Trapez-Schätzer als Funktion der Stichprobengröße

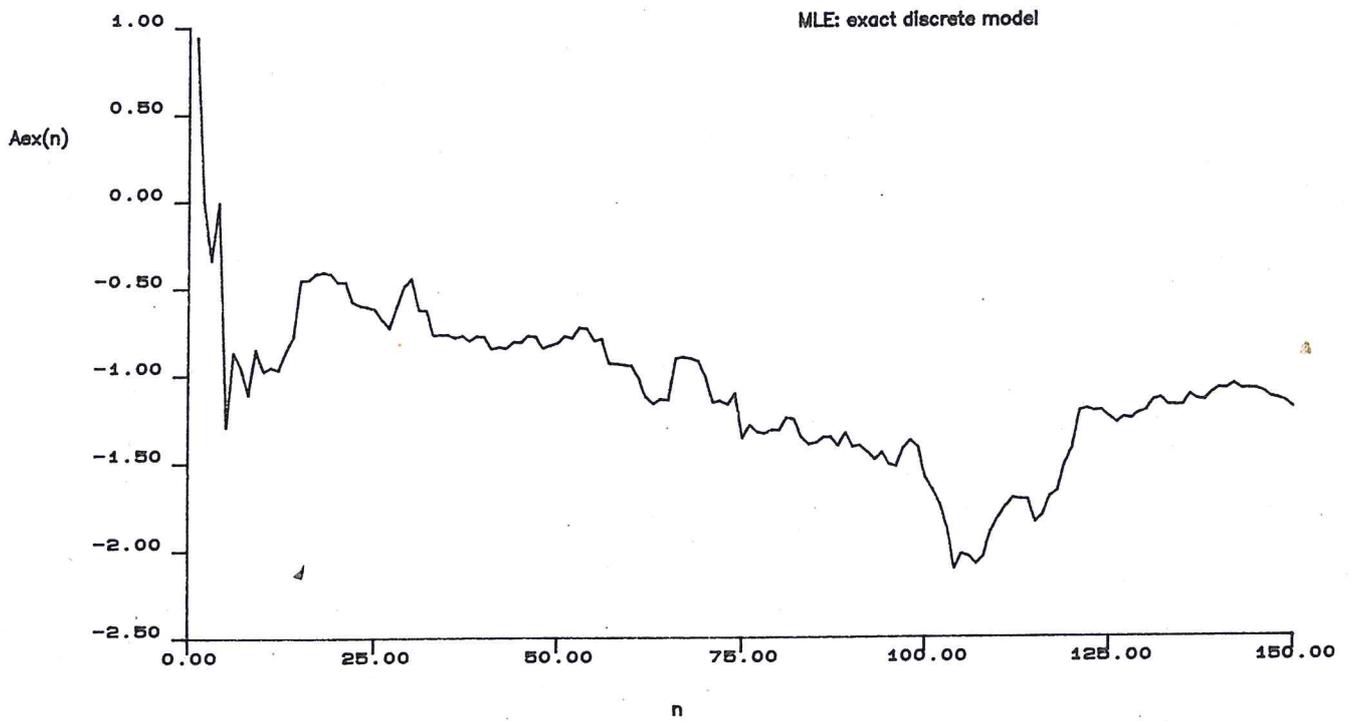


Abb. 2.8. exakter ML-Schätzer als Funktion der Stichprobengröße n

zers (2.15). Die starken Spikes für kleines t sind aus der Darstellung

$$\hat{A} = A + G \int_0^t y(s) dw(s) / \int_0^t y^2 ds$$

verständlich, wenn man für $t \rightarrow 0$ die Integrale durch $y(0)t$ und $y(0)^2 w(t)$ ersetzt, was zu $G/y(0) [w(t)/t]$, der formalen Ableitung des **Wiener**-Prozesses, d.h. "weißem Rauschen" führt. Für größere Zeiten ($t > 10$) nähert sich der Schätzer dem Wert $A = -2$ (\hat{A} ist jedoch in jedem endlichen Intervall verzerrt, vgl. **Liptser u. Shirayayev** 1987, Kap. 18). Die Trapeznäherung und das exakte diskrete Modell führen bei diesem kleinen Zeitraster zu Kurven, die mit dem Auge nicht unterscheidbar sind. Anders sind die Verhältnisse, wenn man ein großes Sampling-Intervall Δt betrachtet. Hierzu wurden $N=150$ Trajektorien mit $dt=.01$ im Intervall $[0, 2]$ simuliert, jedoch nur die Werte $y(t=0)$ und $y(t=2)$ als Daten verwendet (also ein "Zwei-Wellen-Panel"). Abb. 2.6 zeigt den diskretisierten **Ito**-Schätzer (2.15) als Funktion der Stichprobengröße n (die Anordnung der Einheiten ist natürlich beliebig). Man kann zeigen, daß analog zu (2.28), für $N \rightarrow \infty$ Konvergenz zum Wert $1/\Delta t(A^*-1) = -.49$ stattfindet (siehe hierzu Behauptung 1 und 2, Anhang 5).

Der Trapez-Schätzer geht gegen $2/\Delta t(A^*-1)(A^*+1) = -.96$ (Fig. 2.7), während nun das exakte diskrete Modell seine Vorzüge demonstriert (Abb. 2.8). Der Schätzer ist jedoch sehr empfindlich bei Fluktuationen von \hat{A}^* , da $\log x$ für sehr kleine x ausgewertet wird (in Panel-Studien ist Δt typischerweise groß, sodaß $A^* = .018$ im vorliegenden Fall). Weiterhin kann es vorkommen, daß $\hat{A}^* < 0$, was außerhalb des zulässigen Parameterraums liegt. Daher ist es auch im univariaten Fall empfehlenswert, die Maximierung bzgl. der Drift A durchzuführen.

2.7. Zusammenfassung

In Kapitel 2 wurden verschiedene in der Literatur vorgeschlagene Schätzverfahren diskutiert.

Bei kontinuierlichen Datensätzen ist das von **Liptser u. Shirayayev, Le Breton** und **Basawa u. Prakasa Rao** diskutierte ML-Verfahren mit stochastischen **Ito**-Integralen die Methode der Wahl. Insbesondere können nichtlineare Differentialgleichungen geschätzt werden, ohne daß die explizite Lösung bekannt sein muß.

Im Falle diskreter Datensätze kann eine Näherung verwendet werden, bei der man die **Ito**-Integrale (bzw. Zeit-Integrale) durch entsprechende Summen ersetzt. Damit verbunden ist jedoch ein Spezifikationsfehler, der zu einer asymptotischen Verzerrung der Schätzer führt, die bei einem größeren Zeitab-

stand Δt erheblich sein kann.

Analoge Überlegungen gelten für die Trapez-Approximation, bei der jedoch die asymptotische Verzerrung nur von Ordnung Δt^3 für $\Delta t \rightarrow 0$ ist. Auch hier sind bei großem Zeitabstand erhebliche Verzerrungen zu verzeichnen.

Die von einigen Autoren vorgeschlagene Schätzung des reparametrisierten Zeitreihenmodells und Zurückrechnung kann nicht empfohlen werden, da das Identifikationsproblem und die Nichtlinearitäten in den Parametern nicht adäquat gehandhabt werden können.

Das Minimum-Distance-Verfahren modelliert zwar die Nichtlinearität von $A^* = e^{A\Delta t}$ exakt, vernachlässigt jedoch die Struktur der Kovarianz V . Zu erwarten ist, daß das exakte Verfahren für endliche Stichproben kleinere Konfidenzintervalle ergibt.

3. Ein Identifikationsproblem

3.1 Einführung

Das exakte diskrete Modell (für eine Trajektorie)

$$y_{i+1} = A^* y_i + F_i(A, B, X) + u_i$$

mit $A^* = e^{A\Delta t}$, $F_i = e^{A\Delta t} \int_0^{\Delta t} e^{-As} Bx(t_i+s) ds$ und

$\text{Cov}(u_i, u_j) = \delta_{ij} V = \delta_{ij} \int_0^{\Delta t} \exp(As) \Omega \exp(A's) ds$ führt auf die Likelihood-Funktion

$$l = -T/2 [\log|V| + \text{tr}(V^{-1}W)]$$

mit $W = 1/T U U'$, $U = Y_+ - A^* Y - F$, $Y = [y_0, \dots, y_{T-1}]$, $Y_+ = [y_1, \dots, y_T]$ und $F = [F_0, \dots, F_{T-1}]$.

Es stellt sich nun die Frage, ob die Parameter $\Theta = \{A, B, \Omega\}$ eindeutig geschätzt werden können, d.h. ob die Likelihood eine eindeutige Funktion der Parameter ist. Wenn dies nicht der Fall ist, gibt es mehrere Parametersätze $\Theta_1, \Theta_2, \dots$, die auf die gleiche Likelihood führen (**beobachtungsäquivalente Strukturen**). Daß sich hier Schwierigkeiten ergeben, ist durch die Mehrdeutigkeit der Matrixfunktion $A^* = e^{A\Delta t}$ vorgezeichnet (vgl. Phillips 1973, Coddington u. Levinson 1955). Setzt man nämlich (wenn alle Eigenwerte von A als verschieden vorausgesetzt werden)

$$(3.1) \quad A_N = A + 2\pi i / \Delta t T M_N T^{-1}$$

wobei $N = \{n_1, \dots, n_c\}$ ein Vektor von ganzen Zahlen ist, die Matrix A c konjugiert komplexe Eigenwerte $\Lambda = \{\rho_1, \dots, \rho_r; \zeta_1, \zeta_1^*, \dots, \zeta_c, \zeta_c^*\}$ aufweist, T die Matrix aller Eigenvektoren und $M_N = \text{diag}\{0, \dots, 0, n_1, -n_1, \dots, n_c, -n_c\}$ ist, so gilt

$$\begin{aligned} \exp(A_N \Delta t) &= \exp(A + 2\pi i / \Delta t T M_N T^{-1} \Delta t) \\ &= \exp[T(\Lambda + 2\pi i M_N)T^{-1}] \\ &= T \exp(\Lambda + 2\pi i M_N) T^{-1} \end{aligned}$$

Λ und M_N sind als Diagonalmatrizen vertauschbar, sodaß $\exp(\Lambda + 2\pi i M_N) = \exp(\Lambda) \exp(2\pi i M_N)$ gilt. Da jedoch $e^{2\pi i k} = 1 \forall k \in \mathbb{Z}$, hat man $\exp(2\pi i M_N) = I$ und damit $\exp(A_N \Delta t) = \exp(A) = A^*$. Durch (3.1) sind abzählbar unendlich viele Matrizen gegeben, die alle auf das gleiche A^* führen (ähnliche

Probleme ergeben sich bei der Schätzung eines **Markoff-Sprungprozesses** aus Paneldaten (**Singer u. Spilerman 1976, Kalbfleisch u. Lawless 1985**).

Durch die Veränderung von $A \rightarrow A_N$ muß jedoch auch in $V=V(A,\Omega)$ und $F=F(A,B,X)$ B und Ω so reskaliert werden, daß V und F und damit die Likelihood $l(A^*,F,V)$ invariant bleiben. Es ist also zu fragen, ob Matrizen Ω_N und B_N existieren, die folgende Gleichungen erfüllen:

$$(3.2) \quad \begin{aligned} F(A_N, B_N, X) &= F(A, B, X) \\ V(A_N, \Omega_N) &= V(A, \Omega) \end{aligned}$$

Außerdem müssen die reskalierten Matrizen die an sie geforderten Restriktionen erfüllen (etwa muß Ω_N als Diffusionskoeffizient positiv semidefinit (p.s.d) sein - die Tragweite dieser Forderung wurde von **Hansen u. Sargent (1983)** erkannt. Um hier konkrete Resultate zu erzielen, muß das Funktional F explizit ausgewertet werden.

3.2 Exogene Variablen approximiert als Sprungfunktionen

Setzt man $x(t) = x(t_i) = x_i$ im Intervall $[t_i, t_{i+1})$, so gilt

$$F_i(A, B, X) = A^{-1}(e^{A\Delta t} - I)Bx_i = B^*x_i$$

bzw.

$$F = B^*X$$

Daher kann (3.2.1) erfüllt werden, wenn man $B_N = (A^* - I)^{-1}A_N B^*$ setzt, d.h. es gilt

$$F(A_N, B_N, X) = F(A, B, X)$$

Mit Hilfe der Gleichung

$$\begin{aligned} \text{row } V &= \left[\int_0^{\Delta t} e^{As} \otimes e^{As} \right] \text{row } \Omega \\ &= (A \otimes I + I \otimes A)^{-1} (e^{A\Delta t} \otimes e^{A\Delta t} - I) \text{row } \Omega \end{aligned}$$

sieht man, daß die Wahl

$$\text{row } \Omega_N = (A \otimes I + I \otimes A)^{-1} (A_N \otimes I + I \otimes A_N) \text{row } \Omega$$

(A_N und A sind vertauschbar) die Forderung $V(A_N, \Omega_N) = V(A, \Omega)$ erfüllt.

Ω_N ist zwar symmetrisch (wie man durch Multiplikation mit der Kommutationsmatrix zeigen kann, vgl. Anhang 1), jedoch nicht unbedingt p.s.d. Dieses Erkenntnis bildet den Ausgangspunkt der Resultate von **Hansen u. Sargent** (siehe

Kap. 3.4).

3.3 Exogene Variablen approximiert als Polygone und allgemeinere Funktionen

Approximiert man x im Intervall $[t_i, t_{i+1}]$ durch die Gerade $x(t) = x_i + t(x_{i+1} - x_i)/\Delta t$, so ergibt sich für F

$$F = B^*X + C^*\Delta X \quad , \text{ mit } C^* = A^{-1}(-B + 1/\Delta t B^*)$$

Die Invarianz von F kann nur gewährleistet werden, wenn $B^*(A_N, B_N) = B^*(A, B)$ und $C^*(A_N, B_N) = C^*(A, B)$. Die erste Forderung kann, wie in Kap. 3.2 gezeigt, durch

$$B_N = (A^* - I)^{-1} A_N B^*$$

erfüllt werden, jedoch gilt dann

$$\begin{aligned} C^*(A_N, B_N) &= A_N^{-1}(-B_N + 1/\Delta t B^*) \\ &= -(A^* - I)^{-1} B + A_N^{-1} 1/\Delta t B^* \\ &\neq C^*(A, B) = A^{-1}(-B + 1/\Delta t B^*) \end{aligned}$$

sodaß das Modell **ohne weitere Restriktionen identifizierbar** wird. Ähnliche Überlegungen gelten bei komplizierteren Interpolationsformeln für $x(t)$.

Für stückweise konstante, insbesondere verschwindende exogene Einflüsse bleibt das Identifikationsproblem jedoch erhalten. Hier bietet die Forderung von Restriktionen auf die Systemmatrizen einen Ansatzpunkt.

3.4 Identifikationsresultate von Hansen u. Sargent

Im folgenden wird angenommen, daß die System-Matrizen unrestringiert sind und nicht mit Hilfe der Exogenen identifiziert werden können (d.h. entweder sind die Exogenen konstant, Sprungfunktionen oder verschwinden ganz). Wie in 3.2. gezeigt, kann durch die Wahl

$$A_N = A + 2\pi i/\Delta t T M_N T^{-1}$$

und $\text{row } \Omega_N = (A \otimes I + I \otimes A)^{-1} (A_N \otimes I + I \otimes A_N) \text{ row } \Omega$

ein abzählbar unendlicher Satz $\{A_N, \Omega_N\}$ von beobachtungsäquivalenten Strukturen erzeugt werden. Es zeigt sich jedoch, daß die positive Definitheit der Diffusionsmatrizen Ω_N diese Klasse i.A. auf endlich viele einschränkt. **Hansen u. Sargent** (1983) geben (ohne Beweis) folgende Behauptungen an:

Theorem 2: Wenn es ein $A_k \dagger A$ gibt, sodaß

i) $\exp(A_k \Delta t) = A^*$

ii) $V(A_k, \Omega) = V(A, \Omega)$

dann existiert eine unendliche Folge unterschiedlicher A_N , die i) und ii) erfüllen.

Beweis

1. Aus i) folgt, daß A_k die Darstellung $A_k = A + 2\pi i / \Delta t T M_k T^{-1} = T(\Lambda + (2\pi i / \Delta t) M_k) T^{-1}$ hat.

2. $V = \int e^{As} \Omega e^{A^*s} ds$ kann mit Hilfe der Eigenwertzerlegung $A = T \Lambda T^{-1}$ als

$$V = T [\varepsilon \# T^{-1} \Omega T^{*-1}] T^+ := T [\varepsilon \# \Omega_T] T^+$$

geschrieben werden,

wobei $\varepsilon_{ik}[\lambda] = [\exp(\lambda_i + \lambda_k^*) \Delta t - 1] / (\lambda_i + \lambda_k^*); \quad i, j = 1, \dots, p,$

das **Hadamard**-Produkt (Anhang 1) und $T^+ = T^*$, die adjungierte Matrix (transponiert und konjugiert komplex) ist.

Bedingung ii) impliziert dann:

$$\varepsilon[\Lambda + (2\pi i / \Delta t) M_k] \# \Omega_T = \varepsilon[\Lambda] \# \Omega_T$$

in Komponenten $\frac{\exp[(\lambda_i + \lambda_j^*) \Delta t + 2\pi i (K_i - K_j)]}{\lambda_i + \lambda_j^* + (2\pi i / \Delta t) (K_i - K_j)} \Omega_{Tij} = \frac{\exp(\lambda_i + \lambda_j^*) \Delta t}{\lambda_i + \lambda_j^*} \Omega_{Tij} \quad i, j = 1, \dots, p$

Hierbei wurde $K = \{0, \dots, 0; k_1, -k_1, \dots, k_c, -k_c\} : 1 \times (r+2c)$ gesetzt. Daraus kann man dann

$$\frac{\Omega_{Tij}}{\lambda_i + \lambda_j^* + (2\pi i / \Delta t) (K_i - K_j)} = \frac{\Omega_{Tij}}{\lambda_i + \lambda_j^*}$$

folgern, was nur für Komponenten ij mit $\Omega_{Tij} = 0$ ($i \neq j$) oder $K_i - K_j = 0$ möglich ist. Diese Bedingung gilt dann jedoch auch für andere $k_i \in \mathbb{Z}$, sodaß unendlich viele A_k Bedingung i) und ii) erfüllen.

Q.e.d.

Theorem 3: Wenn es kein A_k gibt, das die Voraussetzungen von Theorem 2 erfüllt, existieren nur endlich viele Strukturen (A_k, Ω_k) , die

i) $\exp(A_k \Delta t) = \exp(A \Delta t) = A^*$

ii) $V(A_k, \Omega_k) = V(A, \Omega)$

iii) $\Omega_k \geq 0$ (p.s.d.) erfüllen .

Bemerkung: iii) ist der entscheidende Punkt (der von **Phillips** 1973 in seiner grundlegenden Arbeit übersehen wurde).

Beweis

Aus dem Beweis der vorigen Behauptung kann ersehen werden, daß die Existenz eines A_k mit $V(A_k, \Omega) = V(A, \Omega)$ äquivalent zum Verschwinden mindestens eines Elements von Ω_T ist. (etwa wenn man $j=i+1$ wählt). Wenn dies nicht erfüllt ist, muß auch Ω reskaliert werden und man hat die Darstellung

$$A_k = A + (2\pi i / \Delta t) T M_k T^{-1} = T(\Lambda + (2\pi i / \Delta t) M_k) T^{-1}$$

und $V(A_k, \Omega_k) = V(A, \Omega)$

Mit Hilfe der Eigenwertzerlegung gilt also wieder

$$\varepsilon[\Lambda + (2\pi i / \Delta t) M_k] * \Omega_{kT} = \varepsilon[\Lambda] * \Omega_T$$

oder

$$\Omega_{kT} = \left[\varepsilon[\Lambda] * \Omega_T \right] / \varepsilon[\Lambda + (2\pi i / \Delta t) M_k]$$

wobei / die elementweise Division bedeutet. Berücksichtigt man noch $\Omega_{kT} = T \Omega_k T^{+1}$, so gilt für Ω_k die Darstellung

$$\begin{aligned} \Omega_k &= T \left[\frac{\lambda_j + \lambda_1^* + (2\pi i / \Delta t)(K_j - K_1)}{\lambda_j + \lambda_1^*} \Omega_{T j1} \right] T^+ \\ &= \Omega + T \left[(2\pi i / \Delta t) \frac{(K_j - K_1)}{\lambda_j + \lambda_1^*} \Omega_{T j1} \right] T^+ \end{aligned}$$

Die Matrix $\chi = (2\pi i / \Delta t) \frac{(K_j - K_1)}{\lambda_j + \lambda_1^*}$ ist hermitesch, da

$$\chi^+ = -(2\pi i / \Delta t) \frac{(K_1 - K_j)}{(\lambda_1 + \lambda_j^*)^*} = (2\pi i / \Delta t) \frac{(K_j - K_1)}{\lambda_j + \lambda_1^*}$$

Das gleiche gilt für Ω_T , da $[T^{-1} \Omega T^{+1}]^+ = T^{-1} \Omega^+ T^{+1} = T^{-1} \Omega T^{+1}$ (Ω ist symmetrisch und reell). Daraus folgt aber, daß auch das **Hadamard**-Produkt $\chi * \Omega_T$ hermitesch ist.

Betrachten wir nun die hermitesche Form $\chi^+ T [\chi * \Omega_T] T^+ \chi$, wobei χ ein komplexer p- Vektor ist. Kürzt man $T^+ \chi$ als ψ ab, so hat man

$$\psi^+ [\chi * \Omega_T] \psi \geq 0 \text{ für p.s.d. } \chi * \Omega_T.$$

$\chi * \Omega_T$ hat jedoch verschwindende Diagonalelemente, sodaß für die Eigenwerte μ gilt:

$$\sum \mu_i = \text{tr}[\chi \# \Omega_T] = 0.$$

Da die Eigenwerte einer p.s.d. hermiteschen Matrix ≥ 0 und reell sind, ist $\chi \# \Omega_T$ entweder gleich der Null-Matrix oder indefinit. Ersteres ist n.V. ausgeschlossen. Damit ist aber $T[\chi \# \Omega_T]T^+$ indefinit und damit i.a. auch die Summe $\Omega_k = \Omega + T[\chi \# \Omega_T]T^+$.

Betrachten wir hierzu noch etwas genauer die Matrix χ :

$$\chi = (2\pi i / \Delta t) \frac{(K_j - K_1)}{\lambda_j + \lambda_1^*}$$

mit $\Lambda = \{\rho_1, \dots, \rho_r; \zeta_1, \zeta_1^*, \dots, \zeta_c, \zeta_c^*\}$ und $K = \{0, \dots, 0, k_1, -k_1, \dots, k_c, -k_c\}$

Betrachten wir nun eine unendliche Folge von Matrizen χ_k , so wird ab einem bestimmten K_0 das größte Element eine Schranke M überschreiten, sodaß der größte Eigenwert folgende Ungleichung erfüllt:

$$\max |\mu| = \|\chi\|_2 \geq \max |x_{ij}| \geq M.$$

Entsprechendes gilt für das Produkt $\chi \# \Omega_T$. Die hermitesche Form ($\langle \Phi, \Psi \rangle = \Phi^+ \Psi$)

$$S = \langle \Phi, [\Omega + T[\chi \# \Omega_T]T^+] \Phi \rangle = \langle \Phi, \Omega \Phi \rangle + \langle T^+ \Phi, \chi \# \Omega_T T^+ \Phi \rangle$$

wird also negativ werden, wenn $T^+ \Phi$ als Eigenvektor des kleinsten Eigenwerts von $\chi \# \Omega_T$ gewählt wird und K genügend groß ist. Daher können nur endlich viele Matrizen $\Omega_k = \Omega + T[\chi \# \Omega_T]T^+$ positiv semidefinit sein.

Q.e.d.

Theorem 4: Wenn $\Omega_T = T^{-1} \Omega T^{+1}$ keine Null-Elemente aufweist, gibt es höchstens endlich viele (A_N, Ω_N) , die i), ii) und iii) aus Theorem 3 erfüllen.

Beweis

folgt sofort aus dem Beweis von Theorem 2 und 3.

Theorem 5: Wenn Ω_T keine Null-Elemente aufweist, gibt es einen Zeitabstand Δt_0 , sodaß für $\Delta t < \Delta t_0$ das Modell (A, Ω) aus (A_N, Ω_N) identifizierbar ist.

Beweis: Analog zum Beweis von Theorem 3 und mit Hilfe der Beobachtung, daß

$$\chi = (2\pi i / \Delta t) \frac{(K_j - K_1)}{\lambda_j + \lambda_1^*}$$

umgekehrt proportional zu Δt ist, kann geschlossen werden, daß für gegebenes K und $\Delta t < \Delta t_0$ die Summe $\Omega_T + T\chi \# \Omega_T T^+$ indefinit ist. In diesem Falle können alle beobachtungsäquivalenten Strukturen ausgeschlossen werden, da Ω_k p.s.d.

nicht erfüllt wurde.

Q.e.d.

Bemerkung:

Die Zeitspanne Δt_0 , die bei der Messung unterschritten werden muß, um das Aliasing- Phänomen auszuschalten, ist abhängig vom Wert der Parameter A und Ω , die jedoch in realen Schätzproblemen unbekannt sind. Auch die genaue Anzahl von beobachtungsäquivalenten Strukturen ist (mit dem gleichen Argument) vom wahren Wert abhängig. Insofern sind die von **Hansen u. Sargent** angegebenen Theoreme **kein** Argument für eine Schätzung mit Hilfe des reparametrisierten Modells (siehe Kap. 2.4). Dies wäre nur möglich, wenn für alle zulässigen (A, Ω) und Sampling-Intervall Δt **a priori** nachgewiesen werden könnte, daß **alle** (A_N, Ω_N) aus dem zulässigen Bereich herausfallen.

Beispiel:

Betrachten wir ein Modell mit Drift $A = \begin{bmatrix} 0 & 1 \\ -16 & -4 \end{bmatrix}$, Diffusionskonstante $\Omega = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$ und Input-Matrix $B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ und das Modell der Sprungfunktionen ($\Delta t = 2$). Das Modell ist daher nicht identifizierbar und es ist von Interesse, die beobachtungsäquivalenten Strukturen zu bestimmen.

Man erhält in diesem Falle für $A_{-2} = A + (2\pi i / \Delta t) T \begin{bmatrix} 0 & -2 \\ -2 & 2 \end{bmatrix} T^{-1}$ und

$$A_{-1} = A + (2\pi i / \Delta t) T \begin{bmatrix} 0 & -1 \\ -1 & 1 \end{bmatrix} T^{-1} .$$

positiv definite Diffusionskoeffizienten. Die numerischen Werte lauten:

$$(A_{-2}, B_{-2}, \Omega_{-2}) = \left(\begin{bmatrix} -3.6276 & -.8138 \\ 13.0208 & -.3724 \end{bmatrix}, \begin{bmatrix} .2267 \\ -.8138 \end{bmatrix}, \begin{bmatrix} .2267 & 0 \\ 0 & .3724 \end{bmatrix} \right)$$

$$(A_{-1}, B_{-1}, \Omega_{-1}) = \left(\begin{bmatrix} -1.8131 & .0931 \\ -1.4896 & -.21862 \end{bmatrix}, \begin{bmatrix} .1134 \\ .0931 \end{bmatrix}, \begin{bmatrix} .1134 & 0 \\ 0 & 2.1862 \end{bmatrix} \right).$$

Die zugehörigen Zeitreihenparameter sind:

$$(A^*, B^*, V) = \left(\begin{bmatrix} .0210 & .0032 \\ -.0509 & .0083 \end{bmatrix}, \begin{bmatrix} .0612 \\ .00318 \end{bmatrix}, \begin{bmatrix} .0312 & .0002 \\ .00002 & .4999 \end{bmatrix} \right).$$

Man sieht insbesondere, daß die verschiedenen, zum gleichen (A^*, B^*, V) möglichen System-Matrizen, völlig verschiedene Kausalinterpretationen zulassen. Die Anzahl der beobachtungsäquivalenten Matrizen ist hier klein (=2). Wählt man $\Delta t = .5$, so sind alle Ω_N indefinit. Dies gilt jedoch nur für den gewählten Parameter (A, B, Ω) und nicht für einen Parameterbereich, in dem die ML-Schätzung ausgeführt werden könnte.

3.5 Lokale Identifizierbarkeit

Die Theoreme von **Hansen u. Sargent** schränken das Identifikationsproblem bei exogenen Sprungfunktionen auf einen endlichen Satz von beobachtungsäquivalenten Strukturen ein (andere Approximationen der Exogenen führen zur Identifikation, wie in 3.3. gezeigt). Ob jedoch und wie viele Aliasing-Matrizen existieren, hängt vom gewählten Parameterbereich (z.B. alle reellen $p \times p$ -Matrizen für A) und vom Zeitabstand Δt ab. Im allgemeinen müßte analytisch gezeigt werden, daß die entstehenden Aliasing-Matrizen (die ja auf die gleiche Likelihood führen) aus diesem Bereich herausfallen. Dies erscheint als schwierige Aufgabe, da die Eigenwerte von A und Ω unter den auferlegten Restriktionen analytisch zu bestimmen wären.

In einem praktischen Kontext erweist sich deshalb der Zusammenhang zwischen **lokaler Identifizierbarkeit und Nichtsingularität der Fisher-Information** als nützlich.

Hierzu einige Definitionen (vgl. **Rothenberg 1971**). Man geht davon aus, daß die Zufallsgröße Y , die die beobachteten Daten beschreibt (etwa eine Trajektorie $y(t)$), eine Wahrscheinlichkeitsverteilung $F(y, \Theta) = P_{\Theta}\{Y \leq y\}$ (bzw. Dichte $f(y, \Theta)$) aufweist, die durch den u -dimensionalen Vektor $\Theta \in T \subset \mathbb{R}^u$ parametrisiert ist. Verschiedene Strukturen lassen sich in diesem Spezialfall durch unterschiedliche Θ charakterisieren. Wenn für unterschiedliche Θ die gleiche Verteilung resultiert, sind die Modelle empirisch nicht unterscheidbar und man definiert:

Definition 1: Wenn $F(y, \Theta_1) = F(y, \Theta_2)$ für $\Theta_1 \neq \Theta_2$ und alle $y \in \mathbb{R}^n$, so heißen die Strukturen Θ_1 und Θ_2 **beobachtungsäquivalent**.

Definition 2: sei $\Theta_0 \in T$. Θ_0 heißt **global identifizierbar**, wenn kein anderes beobachtungsäquivalentes $\Theta \in T$ existiert. In diesem Fall ist die Verteilungsfunktion eine injektive Funktion von Θ .

Definition 3: $\Theta_0 \in T$ heißt **lokal identifizierbar**, wenn es eine offene Umgebung U_0 von Θ_0 gibt, sodaß kein $\Theta \in U_0$ beobachtungsäquivalent ist.

Definition 4: Sei $I(\Theta)$ eine in Θ stetige Matrix-Funktion. Sei Θ_0 ein Punkt, für den gilt: $I(\Theta)$ hat konstanten Rang für $\Theta \in U_0$. Dann heißt Θ_0 "**regulärer Punkt**".

Dann gilt folgendes **Theorem (Rothenberg 1971)** :

Sei $I(\Theta_0) = E_{\Theta_0} [\partial \log f / \partial \Theta (\partial \log f / \partial \Theta)']$ die **Fisher-Information** bei dem regulären Punkt Θ_0 .

Dann gilt: Θ_0 ist lokal identifizierbar $\Leftrightarrow I(\Theta_0)$ ist nichtsingulär.

Beweis (Rothenberg 1971, S.579 ff.)

4. Maximum Likelihood-Schätzung des exakten diskreten Modells mit analytischen Ableitungen

4.1. Modellspezifikation

Im folgenden wird ein wichtiger Spezialfall des State-space-Modells

$$(4.1) \quad dy_t = Ay_t dt + Bx_t dt + Gdw_t \quad ; \quad t \in [t_0, t_T] \quad ; \quad y_0: N(\mu, \Sigma) \quad ;$$

$$z_i = Hy_i + Dx_i + \varepsilon_i \quad ; \quad i=0, \dots, T \quad ,$$

das sogenannte "exakte diskrete Modell" diskutiert.

Wir nehmen an, daß alle Komponenten des Zustandsvektors $y(t)$ exakt gemessen werden können, d.h. $H=I$, $D=0$ und $\varepsilon_i=0$. Weiterhin sollen alle Wellen des Panels mit gleichem Zeitabstand Δt erhoben worden sein. Da die Daten $y(t)$ und $x(t)$ nur bei t_0, \dots, t_T erhältlich sind, ist es günstig, die Systemgleichung in ein Differenzenschema ("exaktes diskretes Modell") für diese Zeitpunkte umzuschreiben.

Man erhält (vgl. L. Arnold 1973, Kap. 8) als Lösung des Anfangswertproblems (4.1)

$$y_t = e^{A(t-t_0)} y_0 + e^{At} \int_{t_0}^t e^{-As} (Bx(s) ds + Gdw_s)$$

Daraus folgt für die Beobachtungszeitpunkte ($y_i = y(t_i)$)

$$y_{i+1} = e^{A\Delta t} y_i + e^{A\Delta t} \int_0^{\Delta t} e^{-As} Bx(t_i+s) ds + e^{A\Delta t} \int_0^{\Delta t} e^{-As} Gdw(t_i+s)$$

Abgekürzt hat man also

$$(4.2) \quad y_{i+1} = A^* y_i + F_i(A, B, X) + u_i$$

wobei u_i normalverteilte und unabhängige Zufallsvariable $N(0,V)$ mit

$$\dot{V} = \int_0^{\Delta t} \exp(As) \Omega \exp(A's) ds$$

sind (d.h. zeitdiskretes weißes Rauschen) und $F_i(A,B,X)$ ein Funktional der Exogenen darstellt. Zu beachten ist, daß der autoregressive Parameter A^* , der Einfluß der Exogenen (F) und die Kovarianz des Störterms u implizit und nichtlinear von den Strukturparametern A , B und G des ursprünglichen zeitstetigen Modells abhängen (vgl. hierzu die Ausführungen zur Parameteridentifikation in Kap. 3).

4.2 Maximum Likelihood-Schätzung

Die auf y_0 bedingte Log-Likelihood der **Gauß**-verteilten Daten $[y_0, \dots, y_T]$ für N unabhängige Trajektorien (N statistische Einheiten des Panels) läßt sich folgendermaßen schreiben :

$$l = \sum_{n=1, i=0}^{N, T-1} -1/2 \left\{ \log |V| + [y_{n,i+1} - A^* y_{ni} - F_{ni}(A,B,X)] V^{-1} [\dots] \right\}$$

Mit Hilfe der Abkürzungen $Y_t = [y_{1t}, \dots, y_{Nt}] : p \times N$ (analog für F)

$$Y = [Y_0, \dots, Y_{T-1}] : p \times NT$$

$$Y_+ = [Y_1, \dots, Y_T] : p \times NT$$

und $U = Y_+ - A^* Y - F(A,B,X)$; $W = \frac{1}{NT} U U'$

kann man dies kompakt als

$$l = - \frac{NT}{2} \left[\log |V| + \text{tr}(V^{-1} W) \right]$$

schreiben. Unter Einbeziehung der Dichte der Anfangswerte $y_{n0} : N(\mu, \Sigma)$

erhält man schließlich

$$(4.3) \quad I(A, B, G, \mu, \Sigma; Y_0, Y_+) = -\frac{NT}{2} \left[\log |V| + \text{tr}(V^{-1}W) \right] - \frac{N}{2} \left[\log |\Sigma| + \text{tr}(\Sigma^{-1}R) \right]$$

mit $R = \frac{1}{N} \sum_{i=1}^N (y_{i0} - \mu)(y_{i0} - \mu)'$

Bis jetzt wurde stillschweigend davon ausgegangen, daß der Einfluß der Exogenen $x(t)$ explizit berechnet werden kann. Das Funktional

$$F_t = e^{A\Delta t} \int_0^{\Delta t} e^{-As} Bx(t+s) ds$$

läßt sich jedoch nur bei kontinuierlichen Messungen exakt auswerten. Ansonsten muß eine Näherung durch die Datenpunkte $x(t_i)$ interpoliert werden (die Meßzeitpunkte müssen nicht unbedingt mit denen der Exogenen übereinstimmen).

Bei dem im folgenden diskutierten Algorithmus wird davon ausgegangen, daß $x(t)$ eine Sprungfunktion ist oder dadurch approximiert werden kann, d.h.

$$x(t) = \sum_{i=0}^{T-1} \{ t \in [t_i, t_{i+1}) \} \cdot x_i$$

Hierbei ist $\{ t \in [t_i, t_{i+1}) \} = \begin{cases} 1, & \text{wenn Aussage wahr} \\ 0 & \text{sonst} \end{cases}$

die Indikatorfunktion des Intervalls $[t_i, t_{i+1})$.

Damit läßt sich F explizit als

$$F_i(A, B, X) = A^{-1} (e^{A\Delta t} - I) B x_i$$

angeben, wobei die Nichtsingularität von A vorausgesetzt wurde.

Es ist klar, daß die Art der Approximation für $x(t)$ und die dabei gemachten Fehler sowohl von Zeitabstand, Kurvenverlauf (etwa langsam vs. schnell fluktuierend), als auch von der Größe des Drift-Koeffizienten abhängen (das "Superpositionsintegral" $\int \exp(-As)x(s+t)ds$ kann als Tiefpassfilter aufgefaßt werden, vgl. Lewis 1986, Kap.2 und die darin enthaltene Diskussion des "gesampelten" Modells).

Wie in Kap. 3 diskutiert, ändert sich der Identifikationsstatus des Modells je nach Approximation. Im Falle der Sprungfunktionen müssen daher weitere Re-

striktionen auf A, B und G bzw. genügend kleines Sampling-Intervall Δt (vgl. Hansen u. Sargent 1983) vorausgesetzt werden, um das Modell identifizierbar und schätzbar zu machen.

Zusammenfassend erhält man also bei Approximation der Exogenen durch Sprungfunktionen und Meßbarkeit von $y(t)$ bei t_i folgendes Modell

$$y_{n, i+1} = A^* y_{ni} + B^* x_{ni} + u_{ni}$$

$$y_{n0} : N(\mu, \Sigma) \quad ; \quad n=1, \dots, N \quad ; \quad i=0, \dots, T-1$$

und die Log-Likelihood

$$l(A, B, G, \mu, \Sigma; Y_0, Y_+) = - \frac{NT}{2} \left[\log |V| + \text{tr}(V^{-1}W) \right] - \frac{N}{2} \left[\log |\Sigma| + \text{tr}(\Sigma^{-1}R) \right] ;$$

4.3 Analytische Berechnung des Maximum-Likelihood-Schätzers $\hat{\Theta}$.

Im folgenden wird angenommen, daß die System-Matrizen A, B, G, Σ und μ Funktionen eines u -dimensionalen Parametervektors Θ sind. Damit lassen sich leicht Restriktionen für diese Matrizen implementieren (z.B. $A_{11} = \exp(\Theta) > 0$ etc.). Weiterhin können funktionale Abhängigkeiten, die etwa aus theoretischen Gründen bestehen, berücksichtigt werden (etwa ist der Diffusionskoeffizient eines Brownschen Teilchens eine Funktion von Viskosität, Temperatur und Masse (Einsteinsche Formel, vgl. Nelson 1967, Kap.4)).

Um den Maximum-Likelihood (ML) -Schätzer $\hat{\Theta}$ zu berechnen, muß das Maximum der Likelihood-Funktion (4.3) gefunden werden. Wenn l als differenzierbar vorausgesetzt wird, erhält man als notwendige Bedingung

$$(4.4) \quad s = \left(\frac{\partial l}{\partial \Theta} \right) = 0 \quad ; \quad u \times 1$$

wobei die erste Ableitung der Likelihood als **score-Funktion** und deren Kovarianzmatrix F als **Fisher-Informations-Matrix** bezeichnet wird (vgl. etwa Mardia

et al. 1979, Kap. 4 und Anhang 2). Anschaulich kann also F als inverse Breite des Maximums der Likelihood l aufgefaßt werden (d.h. wenig Information entspricht einem flachen Maximum). Für die **Maximum Likelihood-Gleichungen** (4.4.) erhält man mit Hilfe des **Mc Donald-Swaminathan-Kalküls** (siehe Anhang 1) folgende explizite Form

$$s(\Theta, Y_0, Y_+) = [\partial \Theta^* / \partial \Theta] * [\partial l / \partial \Theta^*] = 0$$

mit $\Theta^* = [V, A^*, B^*, \Sigma, \mu]$

$$(4.5) \quad [\partial l / \partial \Theta^*] = \begin{bmatrix} \frac{NT}{2} \text{row}[V^{-1} (W - V) V^{-1}] \\ \text{row}[V^{-1} U Y'] \\ \text{row}[V^{-1} U X'] \\ \frac{N}{2} \text{row}[\Sigma^{-1} (R - \Sigma) \Sigma^{-1}] \\ N \text{row}[\Sigma^{-1} (m - \mu)] \end{bmatrix}$$

und den Abkürzungen $U = Y_+ - A^* Y - B^* X$

$$W = \frac{1}{NT} U U'$$

$$R = \frac{1}{N} \sum_{i=1}^N (y_{i0} - \mu) (y_{i0} - \mu)'$$

und $m = \frac{1}{N} \sum_{i=1}^N y_{i0}$

Entscheidend für die Lösung der Likelihood-Gleichungen (4.4, 4.5) ist somit die Berechnung der Koeffizienten-Matrix $[\partial \Theta^* / \partial \Theta]$.

Sei im folgenden $\Theta = \text{ROW}(\Omega, A, B, \Sigma, \mu)$ gesetzt, d.h. alle Systemmatrizen sind unrestringiert (bis auf die Symmetrie von Ω und Σ). Dann gilt

$$[\partial\theta^*/\partial\theta] = \begin{bmatrix} V_{\Omega} & 0 & 0 & 0 & 0 \\ V_A & A_A^* & B_A^* & 0 & 0 \\ 0 & 0 & B_B^* & 0 & 0 \\ 0 & 0 & 0 & L & 0 \\ 0 & 0 & 0 & 0 & I \end{bmatrix}$$

wobei die **M-S**-Ableitungen durch $\partial Y/\partial X \equiv Y_X$ abgekürzt sind.
Explizit findet man

$$\begin{aligned} V_{\Omega} &= \partial/\partial\Omega \int_0^{\Delta t} \exp(As)\Omega\exp(A's) ds \\ &= L * \left(e^{\alpha \Delta t} - I(p^2) \right) \alpha^{-T} \end{aligned}$$

wobei $\alpha = A \otimes I + I \otimes A$ und $L = \partial\Omega/\partial\Omega$ eine irreduzible Matrix-Ableitung (vgl. **Mc Donald u. Swaminathan** 1973 und Anhang 1) darstellt.

Die erste Zeile von (4.5) impliziert also

$$(4.6) \quad V_{\Omega} \text{ row} \left[V^{-1} (W - V) V^{-1} \right] = 0$$

Daraus kann jedoch nicht geschlossen werden, daß $\text{row}(\cdot) = 0$, denn V_{Ω} ist singular. Bei der Ableitung der Likelihood-Gleichungen für die multivariate Normalverteilung stößt man auf vergleichbare Ausdrücke (vgl. **M/S** 1973), jedoch gilt dort

$$L * \text{row} \left[\Sigma^{-1} (R - \Sigma) \Sigma^{-1} \right] = 0$$

was $\hat{\Sigma} = R$ impliziert. Dies kann aus der Formel $L \text{ row}(X) = \text{row}(X+X'-\text{Diag}(X))$ abgeleitet werden (es ist interessant, daß an dieser Stelle verschiedene bekannte Autoren zu falschen Zwischenresultaten kommen, vgl. **M/S** 1973).

Im vorliegenden Falle steht jedoch zwischen L und $\text{row}(\cdot)$ eine Matrix-Funktion der Drift A , sodaß andere Lösungen außer $\hat{V}_{ML} = W$ möglich sind (diese Behauptung kann durch die numerische Berechnung der ML-Schätzer illustriert werden).

Tatsächlich kann man außer der OLS-Lösung weitere Maxima der Likelihood nachweisen, vgl. Kap. 3 und 4.9).

Hinreichend für (4.6) ist jedoch $\hat{V}_{ML} = W = 1/NT \ U U'$. Die weiteren in (4.5) enthaltenen Gleichungen liefern dann:

$$\begin{aligned} A_A^* \text{ row}[V^{-1} U Y'] + B_A^* \text{ row}[V^{-1} U X'] &= 0 \\ B_B^* \text{ row}[V^{-1} U X'] &= 0 \end{aligned} \quad (4.7)$$

$$\begin{aligned} L \text{ row}[\Sigma^{-1} (R - \Sigma) \Sigma^{-1}] &= 0 \\ \text{row}[\Sigma^{-1} (m - \mu)] &= 0 \end{aligned}$$

Unter der Annahme der Nichtsingularität von A_A^* , B_A^* und B_B^* und obigem Argument für $L \text{ row}(\cdot)$ erhält man folgende Normalgleichungen

$$\begin{bmatrix} U X' \\ U Y' \\ m - \mu \\ R - \Sigma \end{bmatrix} = 0$$

Aufgelöst nach den Parametern findet man also

$$\begin{bmatrix} \hat{A}^* \\ \hat{B}^* \end{bmatrix} = \begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}^{-1} \begin{bmatrix} Y Y_+ \\ X Y_+ \end{bmatrix}$$

$$\hat{V} = W[\hat{A}^*, \hat{B}^*]$$

$$\hat{\Sigma} = S \quad (= M - mm' = \frac{1}{N} Y_0 Y_0' - mm')$$

$$\hat{\mu} = m = \frac{1}{N} Y_0 * 1$$

In diesem speziellen Fall ist die ML-Lösung zur OLS-Lösung äquivalent (vgl. dazu auch die Ausführungen in **Hamerle, Nagl u. Singer 1988**). Man erhält jedoch nur Gleichungen für die Parameter Θ^* des EDM und nicht für die originalen Struktur-

parameter. Dies ist ein weiterer Ausdruck des Identifikationsproblems (siehe Kap. 3). Löst man mit Hilfe des Matrix-Logarithmus nach A, B und Ω auf, so kann der Fall eintreten, daß kein zulässiges \hat{A} existiert ($\hat{A} \in \mathbb{R}^{P \times P}$); weiterhin ist es möglich, daß zu einem gefundenen \hat{A} keine weiteren Aliasing-Lösungen existieren (wenn $\lambda(\hat{A}) \in \mathbb{R}$), jedoch der numerische Scoring-Algorithmus weitere Nullstellen der Score-Funktion findet, die oft wesentlich näher am wahren Wert liegen (vgl. Kap. 4.9). Diese Probleme treten schon im univariaten Fall auf, wenn nämlich $\hat{A}^* < 0$ ist (obwohl hier kein Identifikationsproblem vorliegt).

Nach der Einführung von Restriktionen, die eine Identifikation ermöglichen, erhält man komplizierte nichtlineare Gleichungen, deren Lösungen nicht mit OLS übereinstimmen.

Approximiert man die Exogenen mit Hilfe eines Polygonzugs, lautet das Funktional F explizit

$$F_i(A, B, X) = B^* X_i + C^* \Delta X_i \quad ; i=0, \dots, T-1$$

mit $C^* = A^{-1} (-B + 1/\Delta t B^*)$ und $\Delta X_i = X_{i+1} - X_i$

In diesem Fall erhält man die ML-Gleichungen durch die Ersetzung $B^* \rightarrow [B^*, C^*]$ und $X \rightarrow \begin{bmatrix} X \\ \Delta X \end{bmatrix}$.

$$(4.7.2) \text{ lautet jetzt } [B_B^*, C_B^*] \text{ row} \left[V^{-1} U \begin{bmatrix} X \\ \Delta X \end{bmatrix} \right] = 0$$

Daraus kann nun **nicht** mehr $U \begin{bmatrix} X \\ \Delta X \end{bmatrix} = 0$ geschlossen werden, sodaß die einfache OLS-Lösung im Falle der Polygonzüge und höherer Approximationen nicht mehr herauskommt.

Wichtiger ist jedoch, daß man nun (komplizierte) Gleichungen für die Systemparameter A, B und Ω selbst erhält (dies spiegelt die Identifizierbarkeit des Modells wider). In der soziologischen Literatur (vgl. Tuma u. Hannan 1984) wird an dieser Stelle in geradezu abenteuerlicher Weise versucht, den bei OLS entstehenden zusätzlichen Parameter C^* (d.h. \hat{A} und \hat{B} sollen aus A^* , B^* und C^* zurückgerechnet werden) durch Mittelwertbildung loszuwerden.

I.A. müssen die ML-Gleichungen numerisch aufgelöst werden. Dies wird im nächsten Abschnitt diskutiert.

4.4. Newton-Raphson und Scoring -Algorithmus zur numerischen Lösung der Likelihood-Gleichungen

Im vorigen Abschnitt wurde gezeigt, daß die Likelihood-Gleichungen i.a. nicht-linear und kompliziert sind und daher numerisch gelöst werden müssen. Alternativ kann das Maximum der Likelihood l mit einem nichtlinearen Optimierungsalgorithmus gefunden werden. Die im folgenden diskutierten Verfahren lehnen sich eng an die Newtonsche Methode an und werden deshalb als **Quasi-Newton-Methoden** bezeichnet (vgl. Dennis u. Schnabel 1983 (im folgenden D/S) und Quandt 1983).

Allgemein gesprochen wird versucht, ausgehend von einem Startwert Θ_0 iterativ das Maximum $\hat{\Theta}$ einer Funktion $l(\Theta)$ zu finden :

$$(4.9) \quad \Theta_{k+1} = \Theta_k - \lambda_k H_k^{-1} s_k \quad ; \quad k=0,1,\dots$$

Hierbei ist $s_k = \partial l / \partial \Theta_k$ der Gradient von l bei $\Theta_k \in \mathbb{R}^u$ und H_k eine Matrix, die die Suchrichtung bestimmt. Wählt man

$$H_k = \left. \partial^2 l / \partial \Theta^2 \right|_{\Theta_k}$$

so erhält man den **Newton-Raphson-Algorithmus**, während $H_k = -I$ die Methode des **steilsten Aufstiegs** liefert.

Der **Newton-Raphson-Algorithmus** (4.9) hat zwei Nachteile:

1. er ist nicht global konvergent (d.h. nicht für alle Startwerte $\in \mathbb{R}^u$)
2. H_k ist nicht unbedingt negativ definit (d.h. man findet auch Minima und Sattelpunkte) .

Daher sind folgende Modifikationen notwendig :

1. Wenn H_k nicht negativ definit ist, ersetze $H_k \rightarrow H_k - \mu I$, wobei $\mu > 0$ gerade so groß ist, daß $H_k - \mu I$ negativ definit ist. Wählt man μ als größten Eigenwert von H_k , so ergibt sich gerade der gewünschte Effekt.

2. Im Kontext der Maximum-Likelihood-Schätzung kann anstelle der 2. Ableitung deren Erwartungswert benutzt werden. Dies ergibt die sog. **Scoring-Methode** (Rao 1973). Aufgrund der Formel (vgl. Anhang 2)

$$F = E_{\Theta} [s s'] = - E_{\Theta} [H]$$

mit

$$s = \partial l / \partial \Theta \quad (\text{"score"})$$

ist leicht einzusehen, daß der Erwartungswert der Hesse-Matrix H proportional zur Kovarianz der Score-Funktion (**Fisher-Informations-Matrix**) und damit negativ definit ist.

3. Wenn der (Quasi) **Newton-Schritt** $p_k = -H_k^{-1} * s_k$ unbefriedigend ist, d.h. wenn $l(\theta_k + p_k) \leq l(\theta_k)$, muß eine "globale" Strategie benutzt werden. Im folgenden wird die Methode der **Schritt-Halbierung** verwandt, d.h. ausgehend vom vollen **Newton-Schritt** p_k wird dieser so lange halbiert, bis eine Zunahme der Likelihood erreicht ist. Dies ist aufgrund der **Taylor-Entwicklung**

$$l(\theta_k + \lambda p_k) = l(\theta_k) - \lambda (s_k' H_k^{-1} s_k) + O(\lambda^2)$$

für genügend kleines λ garantiert, wenn die Modell-Hesse-Matrix H_k negativ definit gewählt wurde (d.h. die **Newton-Richtung** ist eine "Aufstiegsrichtung")

Zusammenfassend wird folgender **Quasi-Newton-Algorithmus** verwandt (vgl. **D/S** Kap. 5 und 6):

1. Berechne den Gradienten $s_k = \partial l / \partial \theta_k$ und die Modell-Hesse-Matrix

$$H_k = \begin{cases} \partial^2 l / \partial \theta^2 \Big|_{\theta_k} & \text{(Newton-Raphson)} \\ - F_k = - E_{\theta_k} [s_k s_k'] = E_{\theta_k} [H] & \text{(Scoring)} \end{cases}$$

2. Wenn H_k nicht negativ definit ist, setze $H_k \rightarrow H_k - \mu I$ ($\mu > \max_i \mu_i$).

3. Löse $-H_k p_k = s_k$

4. Setze $\theta_{k+1} = \theta_k + p_k$

5. Wenn $l(\theta_{k+1}) > l(\theta_k) - \alpha (s_k' H^{-1} s_k)$ **nicht** erfüllt ist, dann

$p_k \rightarrow p_k / 2$ (**Schritt-Halbierung**)

Zurück zu 4.

6. Konvergenz-Kriterium: Abbruch, wenn $\|s_k\| < \varepsilon$ oder $\|p_k\| < \varepsilon$.

Dies bedeutet, daß entweder eine approximative Nullstelle der Score-Funktion (**Maximum**) gefunden oder extrem kleine Schrittweiten produziert werden.

In Anlehnung an **D/S**, Kap. 6, wird in 5. ein Wert von $\alpha = 10^{-4}$ benutzt, d.h. schon ein kleiner Zuwachs der Likelihood reicht aus, den Quasi-Newton-Schritt zu akzeptieren.

Bedingung 5 garantiert, daß die Methode global konvergent ist und in der Nähe eines Maximums, wo H negativ definit ist, erfüllt der volle Newton-Schritt Bedingung 5 (**D/S** Theorem 6.3.2 und 6.3.3).

Damit hat man die Kombination einer global konvergenten mit einer schnellen lokalen Strategie erreicht, d.h. wenn Θ genügend nahe an einem lokalen Maximum $\hat{\Theta}$ liegt, ist die Folge Θ_k q -quadratisch konvergent, d.h. $\|\Theta_{k+1} - \hat{\Theta}\| < c \|\Theta_k - \hat{\Theta}\|^2$ (vgl. **D/S** Kap. 2 und Theorem 5.2.1).

Mit dem Quasi-Newton-Algorithmus ist die Notwendigkeit verbunden, die 1. und 2. Ableitung der Likelihood zu berechnen. Da dies mit einem relativ hohen Aufwand verbunden ist, wird oft versucht, mit numerischen Ableitungen (Differenzen-Quotienten) auszukommen (vgl. hierzu die Arbeiten von **Jones** , siehe Kap. 2). Dies ist zwar konzeptuell einfach, jedoch aufgrund der endlichen Rechengenauigkeit mit Rundungsfehlern behaftet, sodaß die Länge des Inkrements $\Delta\Theta$ mit Bedacht gewählt werden muß. Außerdem sind für den Gradienten ∇l u zusätzliche Werte der Likelihood, für die 2. Ableitung zusätzlich $u(u+1)/2 + u$ zusätzliche Auswertungen erforderlich. Dies kann (zusätzlich muß die Likelihood bei der Schritthalbierung berechnet werden) zu einem hohen Aufwand an Rechenzeit führen.

Ernsthafter ist jedoch das Problem, daß bei numerischer Approximation des Gradienten die vorteilhaften Konvergenzeigenschaften des Newton-Algorithmus verlorengehen können: ∇l ist ja gerade die Größe, deren Nullstelle gesucht werden soll. Die Genauigkeit der Lösung $\hat{\Theta}$ wird also durch die Genauigkeit des Gradienten determiniert. Aus diesem Grund wird in der Literatur empfohlen, möglichst analytische Ausdrücke zumindest für den Gradienten zu verwenden. Relativ unproblematisch ist dagegen die Approximation der Hesse-Matrix durch numerische Ableitungen, bei der die Konvergenzeigenschaften nicht berührt werden.

Im folgenden werden exakte analytische Ausdrücke für die Score-Funktion und die Fisher-Informationen-Matrix abgeleitet (vgl. auch Kap. 5, wo ein Newton-Raphson-Algorithmus mit numerischen 2. Ableitungen (aus der exakten Score-Funktion) diskutiert wird).

4.5 Analytische Berechnung der Score-Funktion und der Fisher-Information

Die folgenden Ableitungen sind relativ technisch und können ohne weiteres überschlagen werden. Die Formeln sind jedoch essentiell für die praktische Implementierung des Scoring-Algorithmus.

Knüpfen wir an Kap. 4.3 an, so gilt für den Score:

$$s(\Theta, Y_0, Y_+) = \left[\partial\Theta^*/\partial\Theta \right] * \left[\partial l / \partial\Theta^* \right] = 0$$

mit $\partial\Theta^*/\partial\Theta = \left[V_\Theta, A_\Theta^*, B_\Theta^*, \Sigma_\Theta, \mu_\Theta \right]$

$$\left[\partial l / \partial\Theta^* \right] = \begin{bmatrix} -\frac{NT}{2} \text{row} \left[V^{-1} (W - V) V^{-1} \right] \\ \text{row} \left[V^{-1} U Y' \right] \\ \text{row} \left[V^{-1} U X' \right] \\ -\frac{N}{2} \text{row} \left[\Sigma^{-1} (R - \Sigma) \Sigma^{-1} \right] \\ N \text{row} \left[\Sigma^{-1} (m - \mu) \right] \end{bmatrix}$$

$\left[\partial l / \partial\Theta^* \right]$ ist der Ausdruck, den man auch im Falle der Zeitreihenanalyse erhalten hätte. Jedoch sind nun die Matrizen Θ^* aufgrund der zeitstetigen Struktur des Modells von spezieller Form, d.h. vorgegebene Funktionen der zeitstetigen Parameter $\{ \Omega, A, B, \Sigma, \mu \}$.

Es gilt:

$$\begin{aligned} V &= \int_0^{\Delta t} \exp(As) \Omega \exp(A's) ds \\ \text{row } V &= \alpha^{-1} (e^{\alpha \Delta t} - I) \text{row } \Omega \\ \text{mit } \alpha &= A \otimes I + I \otimes A ; \\ A^* &= e^{A\Delta t} \\ B^* &= A^{-1} (e^{A\Delta t} - I) B , \end{aligned} \tag{4.11}$$

wobei Ω, A, B, Σ und μ Funktionen des Parametervektors Θ sind.

Im folgenden wird davon ausgegangen, daß die Ableitungen $\{ \Omega_\Theta, A_\Theta, B_\Theta, \Sigma_\Theta, \mu_\Theta \}$ der zeitstetigen Parameter vom Benutzer bereitgestellt werden, was ohne Schwierigkeiten im Rahmen der Implementierung von Restriktionen möglich ist.

Für die Ableitungen der Parametermatrizen gilt (Beweis siehe Anhang 3)

$$(4.12) \quad \partial A^* / \partial \Theta = \partial A / \partial \Theta * \tau \left[\text{diag row } E_1 \right] \tau^{-1}$$

$$\text{mit } E_1 = \begin{cases} \Delta t e^{\lambda_i \Delta t} & i = j \\ \frac{e^{\lambda_i \Delta t} - e^{\lambda_j \Delta t}}{\lambda_i - \lambda_j} & i \neq j \end{cases}$$

$$\tau = T^{-T} \otimes T$$

$$\text{und } \Lambda = T^{-1} A T$$

Hierbei wird angenommen, daß alle Eigenwerte von A verschieden sind (vgl. hierzu die Diskussion in **Jennrich u. Bright 1976**).

$$(4.13) \quad \partial V / \partial \Theta = \partial A / \partial \Theta * \partial V / \partial A + \partial \Omega / \partial \Theta * \partial V / \partial \Omega$$

$$\begin{aligned} \text{mit } \partial V / \partial \Omega &= \partial / \partial \Omega \int_0^{\Delta t} \exp(A s) \Omega \exp(A' s) ds \\ &= \left(e^{\alpha \Delta t} - I(p^2) \right) \alpha^{-T} \end{aligned} ;$$

$$\alpha = A \otimes I + I \otimes A$$

und $\partial V / \partial A =$

$$\left[\partial \alpha / \partial A \right] * \left[\left(-\alpha^{-T} \otimes \alpha^{-1} \right) \left(I(p^2) \otimes \left[\left(e^{\alpha \Delta t} - I(p^2) \right) \text{row } \Omega \right] \right) + \partial e^{\alpha \Delta t} / \partial \alpha \right]$$

$$\partial \alpha / \partial A = \partial / \partial A \left(I(p) \otimes A + A \otimes I(p) \right)$$

$$= \left[I(p^2) \otimes \text{row } I(p) \right] \left[I(p) \otimes E(p,p) \otimes I(p) \right] \left[I(p^4) + E(p,p) \otimes E(p,p) \right]$$

$$\partial e^{\alpha \Delta t} / \partial \alpha = \sigma \left[\text{diag row } F_1 \right] \sigma^{-1}$$

$$\text{mit } F_1 = \begin{cases} \Delta t e^{x_i \Delta t} & i = j \\ \frac{e^{x_i \Delta t} - e^{x_j \Delta t}}{x_i - x_j} & i \neq j \end{cases}$$

$$\sigma = \tau^{-T} \otimes \tau$$

$$\text{und } \chi = \tau^{-1} \alpha \tau = I \otimes \Lambda + \Lambda \otimes I$$

Da jedoch Matrizen bis zur Größenordnung $p^4 \times p^4$ vorkommen, ist obiger Ausdruck, obwohl exakt, für Berechnungen auf dem Computer unbrauchbar.

Mit Hilfe der Ableitungen in Matrix-Anordnung (vgl. Anhang 1) können jedoch Formeln hergeleitet werden, die bezüglich des benötigten Speicherplatzes wesentlich effektiver sind. Man findet (Beweis siehe Anhang 3)

$$(4.14) \quad \begin{aligned} [\partial V / \partial \Theta_u] &= \int_0^{\Delta t} [A_{\Theta_u}^*(s)] \Omega A^{*'}(s) && (I) \\ &+ \int_0^{\Delta t} A^*(s) \Omega [A_{\Theta_u}^{*'}(s)] && (I') \\ &+ \int_0^{\Delta t} A^*(s) [\Omega_{\Theta_u}] A^{*'}(s) && (II) \end{aligned}$$

$u=1, \dots, U$

wobei
$$I_{bh}^u = \sum_{adf} [T^{-1} [A_{\Theta_u}] T]_{ad}^u [T_{ba} E_{3adf} T_{hf}]_{adf}^{bh} [T^{-1} \Omega T^{-T}]_{df}$$

$u=1, \dots, U; \quad b, h = 1, \dots, p$

$$E_{3adf} = \begin{cases} \frac{\Delta t}{\lambda_a + \lambda_f} e^{(\lambda_a + \lambda_f) \Delta t} - \frac{1}{(\lambda_a + \lambda_f)^2} [e^{(\lambda_a + \lambda_f) \Delta t} - 1] ; & a=d \\ \frac{1}{(\lambda_a - \lambda_d)} (E_{2af} - E_{2df}) & ; a \neq d \end{cases}$$

$$\Pi_{af}^u = \sum_{be} \left[T^{-1} \left[\Omega_{\Theta_u} \right] T^{-T} \right]_{be}^u \left[T_{ab} E_{2be} T_{fe} \right]_{be}^{af}$$

u=1.....U; a, f=1.....p

mit $E_{2be} = \frac{1}{\lambda_b + \lambda_e} \left[e^{(\lambda_b + \lambda_e)\Delta t} - 1 \right]$

Weiterhin gilt :

$$(4.15) \quad \left[B_{\Theta}^* \right] = -A^{-1} \left[A_{\Theta} \right] B^* + A^{-1} \left[A_{\Theta}^* \right] B + A^{-1} (A^* - I) \left[B_{\Theta} \right]$$

und für die Parameter des exakten diskreten Modells:

$$(4.16) \quad \begin{aligned} A^* &= e^{A\Delta t} = T e^{\Lambda\Delta t} T^{-1} \\ B^* &= A^{-1} (A^* - I) B \\ V &= T \left[E_2 * (T^{-1} \Omega T^{-T}) \right] T \end{aligned}$$

Mit den oben angegebenen Ableitungen der reduzierten Parameter Θ^* kann die **Fisher-Information-Matrix** berechnet werden . Man findet (Beweis siehe Anhang 4)

$$\begin{aligned} F &= E_{\Theta} \left[s s' \right] = -E_{\Theta} \left[\partial^2 l / \partial \Theta^2 \right] \\ &= - \left[\partial \Theta^* / \partial \Theta \right] F^* \left[\partial \Theta^* / \partial \Theta \right]' \end{aligned}$$

mit $F^* = -E_{\Theta} \left[\partial^2 l / \partial \Theta^{*2} \right] =$

$$\begin{bmatrix} NT/2 (V^{-1} \otimes V^{-1}) & 0 & 0 & 0 & 0 \\ 0 & V^{-1} \otimes E_{\Theta} [YY'] & V^{-1} \otimes E_{\Theta} [Y] X' & 0 & 0 \\ 0 & V^{-1} \otimes X E_{\Theta} [Y'] & V^{-1} \otimes XX' & 0 & 0 \\ 0 & 0 & 0 & N/2 (\Sigma^{-1} \otimes \Sigma^{-1}) & 0 \\ 0 & 0 & 0 & 0 & N \Sigma^{-1} \end{bmatrix}$$

F^* ist das Resultat, das man auch im Falle der Zeitreihenanalyse erhalten hätte. $\partial \Theta^* / \partial \Theta$ enthielte dann nur die Restriktionen der Θ^* - Matrizen (z.B. Symmetrie), während jetzt die spezielle Struktur des zeitstetigen Modells in den Matrizen $\partial \Theta^* / \partial \Theta$ enthalten ist. Insbesondere wird die Blockstruktur von F^* (keine Kovarianz zwischen \hat{A}^* (bzw. \hat{B}^*) und \hat{V} , wenn man die asymptotische Kovarianzmatrix der ML-Schätzer, $COV = F^{*-1}$ bildet), durch die Projektionsmatrizen zerstört, was die funktionale Abhängigkeit der Fehlerkovarianz V von den Parametern A und Ω ausdrückt.

Die Auswertung der Erwartungswerte $E[YY']$ und $E[Y]$ kann rekursiv erfolgen: Mit Hilfe des exakten diskreten Modells (4.2) erhält man für $\mu_{nt} = E[y_{nt}]$

$$\mu_{n,t+1} = A^* \mu_{nt} + B^* x_{nt} \quad , \quad t=0, \dots, T-1, \quad n=1, \dots, N$$

$$\mu_{n0} = \mu = E[y_0]$$

$$\begin{aligned} \text{Die zeitgleichen 2. Momente } M_{nt} = E[y_{nt} y'_{nt}] &= \text{cov} [y_{nt}, y'_{nt}] + \mu_{nt} \mu'_{nt} \\ &= \Sigma_t + \mu_{nt} \mu'_{nt} \end{aligned}$$

lassen sich dann mit folgender Rekursion für die Kovarianzmatrix auswerten

$$\Sigma_{t+1} = A^* \Sigma_t A^{*'} + V \quad , \quad \Sigma_0 = \Sigma \quad ,$$

da y_t unabhängig von den Zuwächsen $dW(t+s)$, $s > 0$ ist.

Auf der anderen Seite kann man sich fragen, ob es nicht (im Sinne eines modifizierten Scoring-Algorithmus) effizienter wäre, für die Maximierungs-Iteration

$E[YY']$ (bzw. $E[Y]$) durch YY' (Y) zu ersetzen. Das Stichproben-Moment

$$ZZ' = \begin{bmatrix} Y \\ X \end{bmatrix} [Y', X']$$

kann vor der Iteration berechnet werden und ist positiv definit, wenn mindestens $p+q$ der Vektoren z_{nt} linear unabhängig sind (dann ist die aus diesen Vektoren gebildete Summe in ZZ' nichtsingulär und damit positiv definit; die Summe einer p.d. und p.s.d-Matrix ist p.d.) .

In Anhang 4 wurde hergeleitet, daß sich die 2. Ableitung der Likelihood (**Hesse-Matrix**) folgendermaßen darstellen läßt:

$$\partial^2 l / \partial \theta^2 = (\partial^2 \theta^* / \partial \theta^2) [I \otimes (\partial l / \partial \theta^*)] + [\partial \theta^* / \partial \theta] [\partial^2 l / \partial \theta^{*2}] [\partial \theta^* / \partial \theta]'$$

Weiterhin gilt $\partial^2 l / \partial A^{*2} = -V^{-1} \otimes YY'$ und $\partial^2 l / \partial A^* \partial B^* = -V^{-1} \otimes YX'$.

Weglassen der Erwartungswerte in der **Fisher-Informationen-Matrix** entspricht also einer Modifikation der **Hesse-Matrix**. Alle diese Betrachtungen sind asymptotisch gerechtfertigt durch folgende Grenzwerte (T fest, $N \rightarrow \infty$):

Behauptung 1:

$$\text{Sei } \frac{1}{N} \sum_n x_{nt} < \infty \text{ und } \frac{1}{N} \sum_n x_{nt} x'_{ns} < \infty \quad (*)$$

1. $\text{qm-lim}_{N \rightarrow \infty} \frac{1}{N} (\partial l / \partial \theta) = 0$
2. $\text{qm-lim}_{N \rightarrow \infty} \frac{1}{N} YY' = \lim_{N \rightarrow \infty} \frac{1}{N} E[YY']$
3. $\text{qm-lim}_{N \rightarrow \infty} \frac{1}{N} YX' = \lim_{N \rightarrow \infty} \frac{1}{N} E[Y] X'$

Beweis: (Anhang 5)

Analoge Bedingungen für die Exogenen müssen auch bei Konsistenzbeweisen gefordert werden (vgl. hierzu **Anderson 1971**, S. 203 ff.).

Mit Hilfe der obigen Grenzwerte findet man also:

$$\begin{aligned} \text{qm-lim}_{N \rightarrow \infty} 1/N [\partial^2 l / \partial \theta^2] &= \lim_{N \rightarrow \infty} 1/N [\partial \theta^* / \partial \theta] E [\partial^2 l / \partial \theta^{*2}] [\partial \theta^* / \partial \theta]' \\ &= - \lim_{N \rightarrow \infty} 1/N * F \end{aligned}$$

Damit ist auch im Falle **trajektorienweise verschiedener exogener Einflüsse** gezeigt,

daß **Hesse**-Matrix, **Fisher**-Informations-Matrix und die modifizierte **Fisher**-Information (Erwartungswerte weggelassen) asymptotisch äquivalent sind .

Für trajektorienweise gleiche Exogene gelten selbstverständlich die üblichen Resultate (insbesondere ist hier $1/N \sum \mu_{nt} = \mu_t$ und $1/N \sum \mu_{nt} \mu'_{ns} = \mu_t \mu'_s$).

Mit Hilfe des **starken Gesetzes der großen Zahlen** läßt sich unter etwas anderen Bedingungen an die Exogenen sogar fast sichere Konvergenz nachweisen:

Behauptung 2. Sei $\sum_n x_{nt}/n^2 < \infty$ und $\sum_n x_{nt} x'_{ns} /n^2 < \infty$ (**)

1. $1/N (\partial l / \partial \theta) \longrightarrow 0$ (fast sicher)
2. $1/N YY' - 1/N E[YY'] \longrightarrow 0$ (f.s)
3. $1/N YX' - 1/N E[Y]X' \longrightarrow 0$ (f.s.)

Beweis: (Anhang 5)

Analog zum Fall der Quadratmittel-Konvergenz ist damit gezeigt, daß **Hesse**-Matrix, **Fisher**-Informations-Matrix und modifizierte **Fisher**-Info asymptotisch äquivalent sind, dies sogar im Sinne fast sicherer Konvergenz.

4.6. Konsistenz des Maximum-Likelihood-Schätzers

Wie in Abs. 4.2 gezeigt wurde, läßt sich die (auf die Anfangswerte bedingte) Log-Likelihood als Funktion der reduzierten Parameter $\theta^* = [A^*, B^*, V]$ schreiben

$$l^*(A^*, B^*, V; M_{NT}) = - \frac{NT}{2} \left[\log |V| + \text{tr}(V^{-1} W) \right]$$

wobei A^* , B^* und V Funktionen des Parametervektors θ sind.

Damit hat man

$$l(\theta; M_{NT}) = l^*(\theta^*(\theta); M_{NT})$$

wobei M_{NT} als Abkürzung für die Momentenmatrix der Stichprobe steht:

$$M_{NT} = 1/NT \begin{bmatrix} Y_+ Y_+' & Y_+ Y' & Y_+ X' \\ & YY' & YX' \\ & & XX' \end{bmatrix}$$

Definition: sei $l(\Theta_{NT}; M_{NT}) \geq l(\Theta; M_{NT}) \forall \Theta \in \mathbb{R}^u$.

Dann heißt Θ_{NT} **Maximum-Likelihood-Folge**.

Damit gilt die für den Konsistenzbeweis zentrale Ungleichung (Θ ist der wahre Parameter)

$$l(\Theta; M_{NT}) \leq l(\Theta_{NT}; M_{NT}) \leq l^*(\hat{\Theta}^*; M_{NT})$$

wobei $\hat{\Theta}^* = \underset{\Theta^*}{\operatorname{argmax}} l^*(\Theta^*; M_{NT})$

Insbesondere ist es möglich, daß $\hat{\Theta}^*$ gar nicht als Funktion des Parametervektors Θ darstellbar ist.

Wie in Abs. 4.3 hergeleitet, gilt für das Maximum der Funktion l^*

$$\begin{aligned} \begin{bmatrix} \hat{A}^* \\ \hat{B}^* \end{bmatrix} &= \begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}^{-1} \begin{bmatrix} Y Y_+' \\ X Y_+' \end{bmatrix} \\ &= \begin{bmatrix} A^* \\ B^* \end{bmatrix} + \begin{bmatrix} YY' & YX' \\ XY' & XX' \end{bmatrix}^{-1} \begin{bmatrix} Y U' \\ X U' \end{bmatrix} \end{aligned}$$

$$\hat{V} = W[\hat{A}^*, \hat{B}^*] = \frac{1}{NT} (Y_+ - \hat{A}^* Y - \hat{B}^* X)(Y_+ - \hat{A}^* Y - \hat{B}^* X)'$$

Im vorigen Abschnitt (Behauptung 2) wurde jedoch hergeleitet, daß unter der Bedingung

$$\sum_n x_{nt}/n^2 < \infty \text{ und } \sum_n x_{nt} x'_{ns} /n^2 < \infty \quad (**)$$

gilt:

$$\begin{aligned} 1/NT \text{ UU}' &\longrightarrow V \\ 1/N \text{ UY}' &\longrightarrow 0 \\ 1/N \text{ UX}' &\longrightarrow 0 \\ 1/N \text{ YY}' - 1/N \text{ E}[\text{YY}'] &\longrightarrow 0 \\ 1/N \text{ YX}' - 1/N \text{ E}[\text{Y}] \text{X}' &\longrightarrow 0 \end{aligned} \quad \text{(fast sicher)}$$

Wenn weiterhin $1/NT \text{ E} \left[\begin{bmatrix} \text{Y} \\ \text{X} \end{bmatrix} [\text{Y}', \text{X}'] \right] \rightarrow Q$ für $N \rightarrow \infty$, was unter der Bedingung

$$1/N \sum x_{nt} < \infty \text{ und } 1/N \sum x_{nt} x'_{ns} < \infty \quad (*)$$

gewährleistet ist (siehe Beweis von Behauptung 1 in Anhang 5), und unter der Annahme, daß Q nichtsingulär ist, folgt die **starke Konsistenz** der ML-Schätzer des zeitdiskreten Modells:

$$\begin{bmatrix} \hat{A}^* \\ \hat{B}^* \end{bmatrix} \longrightarrow \begin{bmatrix} A^* \\ B^* \end{bmatrix} \quad \text{für } N \rightarrow \infty \quad \text{(f.s.)}$$

und

$$\begin{aligned} \hat{V} &= 1/NT \left[(A^* - \hat{A}^*) Y + (B^* - \hat{B}^*) X + U \right] \left[\dots \dots \dots \right]' \\ &= \left[\Delta A^*, \Delta B^*, I \right] \left\{ 1/NT \begin{bmatrix} Y \\ X \\ U \end{bmatrix} \left[Y', X', U' \right] \right\} \begin{bmatrix} \Delta A^* \\ \Delta B^* \\ I \end{bmatrix} \end{aligned}$$

$$\longrightarrow V \quad \text{für } N \rightarrow \infty \quad \text{(fast sicher)}$$

wobei $\Delta A^* = A^* - \hat{A}^*$ und $\Delta B^* = B^* - \hat{B}^*$ gesetzt wurde.
Damit ist gezeigt:

Die ML-Schätzer der Parameter Θ^* sind **stark konsistent**, falls

$$1/N \sum_n x_{nt} < \infty \text{ und } 1/N \sum_n x_{nt} x'_{ns} < \infty \quad (*)$$

$$\sum_n x_{nt}/n^2 < \infty \text{ und } \sum_n x_{nt} x'_{ns} /n^2 < \infty \quad (**)$$

d.h. $[\hat{A}^*, \hat{B}^*, \hat{V}] \longrightarrow [A^*, B^*, V] \quad \text{für } N \longrightarrow \infty \quad (\text{f.s.})$

Damit gilt : $l(\Theta, M_{\infty T}) \leq l(\hat{\Theta}, M_{\infty T}) \leq l^*(\Theta^*(\Theta), M_{\infty T}) = l(\Theta, M_{\infty T})$

und somit die **starke Konsistenz der ML-Schätzer des zeitkontinuierlichen Modells**

$$\hat{\Theta} = \lim_{N \rightarrow \infty} \hat{\Theta}_{NT} = \Theta \quad (\text{fast sicher})$$

wenn noch zusätzlich die Identifikationsbedingung (vgl. Kap. 2 und 3)

$$l(\Theta_1, M_{NT}) = l(\Theta_2, M_{NT}) \iff \Theta_1 = \Theta_2$$

erfüllt ist.

4.7 Asymptotische Normalität der ML-Schätzer

Da unter den in (4.6) formulierten Bedingungen $\hat{\Theta}_N \longrightarrow \Theta$ (fast sicher), kann man ein $N_0(\omega)$ finden, sodaß der bei einer **Taylor**-Entwicklung von $s = \partial l / \partial \Theta$ um Θ gemachte Fehler δ_N fast immer kleiner als eine vorgegebene Schranke wird , d.h.

$$\partial l / \partial \Theta |_{\Theta+h} = \partial l / \partial \Theta |_{\Theta} + \partial^2 l / \partial \Theta^2 |_{\Theta} h + \delta_N$$

sodaß man näherungsweise schreiben kann ($h = \hat{\Theta}_N - \Theta$)

$$N^{-1/2} \partial l / \partial \Theta = \left[-1/N \left(\partial^2 l / \partial \Theta^2 \right) |_{\Theta} \right] N^{1/2} (\hat{\Theta}_N - \Theta)$$

Nun gilt (wie gezeigt) für die skalierte Hesse-Matrix

$$-1/N \left(\partial^2 l / \partial \Theta^2 \right) \longrightarrow J \quad (\text{fast sicher})$$

wobei $J = \lim_{N \rightarrow \infty} 1/N E \left[-\partial^2 l / \partial \Theta^2 \right] = \lim_{N \rightarrow \infty} (1/N) F$

$$\text{und } N^{-1/2} \partial l / \partial \theta = N^{-1/2} \left[\partial \theta^* / \partial \theta \right] \begin{bmatrix} -\frac{NT}{2} \text{ row} \left[V^{-1} (W - V) V^{-1} \right] \\ \text{row} \left[V^{-1} U Y' \right] \\ \text{row} \left[V^{-1} U X' \right] \\ -\frac{N}{2} \text{ row} \left[\Sigma^{-1} (R - \Sigma) \Sigma^{-1} \right] \\ N \text{ row} \left[\Sigma^{-1} (m - \mu) \right] \end{bmatrix}$$

Die Komponenten des Scores setzen sich aus N unabhängigen Summanden zusammen, die, falls auf jede Trajektorie identische exogene Einflüsse $x_{nt} = x_t$ wirken, sogar identisch verteilt sind. Berücksichtigt man noch $E[s] = 0$ und $E[ss'] = F$, so gilt der **Zentrale Grenzwertsatz**

$$N^{-1/2} \partial l / \partial \theta \sim N(0, J) \quad \text{für } N \rightarrow \infty$$

$$\text{sodaß } N^{1/2} (\hat{\theta}_N - \theta) \sim \left[-1/N \left(\partial^2 l / \partial \theta^2 \right) \Big|_{\theta} \right]^{-1} N^{-1/2} \partial l / \partial \theta$$

$$\xrightarrow{D} N(0, J^{-1}) \quad \text{für } N \rightarrow \infty$$

(vgl. Basawa u. Prakasa Rao 1980, S.221).

Im Falle trajektorienweise verschiedener Exogener muß eine allgemeinere Form des zentralen Grenzwertsatzes benutzt werden. Insbesondere wäre zu prüfen, ob die **Lindeberg**-Bedingung oder allgemeinere Bedingungen (vgl. Shirayev 1984, S. 326 ff) erfüllt sind.

4.8. Praktische Implementierung des Scoring-Algorithmus: Das Programmpaket LSD

Um die in dieser Arbeit betrachteten Methoden auf praktische Probleme anwenden zu können und einem weiteren Kreis von Interessenten zugänglich zu machen, wurde das Programmpaket **LSD** (Lineare Stochastische Differentialgleichungen) mit Hilfe der Matrix-Sprache SAS/IML (vgl. SAS/IML User's Guide, 1985) erstellt (eine Programmbeschreibung ist in Vorbereitung). Vorteilhaft hierbei ist, daß einerseits die übliche SAS-Umgebung voll zur Verfügung steht (insbesondere können bestehende SAS-Datensätze von IML bearbeitet werden), andererseits Matrix-Ausdrücke in einer der Mathematik ähnlichen Form programmiert werden können.

Hauptbestandteile des Programms zur Schätzung des exakten diskreten Modells sind ein Scoring-Algorithmus (Kap. 4.4) und Routinen (sog. **Module**) zur Berechnung der Ableitungen der Zeitreihenparameter θ^* nach dem Parameter-

vektor Θ . Da die Driftmatrix i.A. nichtsymmetrisch ist, mußte eine komplexe Arithmetik zur Berechnung von Eigenwerten und Eigenvektoren implementiert werden (siehe hierzu auch **Jennrich u. Bright 1976**). Zu diesem Zweck wurde ein QR-Algorithmus mit anschließender inverser Iteration geschrieben (vgl. **Golub u. Van Loan 1983**).

Im folgenden wird die Arbeitsweise der wichtigsten Module kurz beschrieben:

NEWTON (THETA, COV, THETA0, KMAX, EPS, M0, M00, M11, M12, M13, M22, M23, M33, DT, N, T, OPTION)

Zweck: Quasi-Newton- (Scoring)-Algorithmus) zur Bestimmung des ML-Schätzers $\hat{\Theta}$ (Kap. 4.4).

Ausgabe-Parameter:

THETA: $u \times 1$ -Vektor, der am Ende der Iteration den ML-Schätzer $\hat{\Theta}$ enthält (oder den letzten Wert $\Theta_{k_{\max}}$, falls keine Konvergenz erzielt wurde).

COV: $p \times p$. Asymptotische Kovarianzmatrix der geschätzten Parameter. Wird als $COV = GINV(F)$ berechnet .

Eingabe-Parameter:

THETA0: $u \times 1$. Startwert der Scoring-Iteration. Muß vom Benutzer bereitgestellt werden.

KMAX: Beschränkung der Maximalzahl der Iterationen.

EPS: Wenn der Score s und der Newton-Schritt $p < EPS$ ist, wird die Iteration abgebrochen, da eine approximative Nullstelle des Scores s gefunden oder extrem kleine Schrittlängen p produziert wurden.

M={M0,M00,.....M33}: Stichprobenmomente.

$M0=1/N Y_0 * 1'$; $M00=1/N Y_0 Y_0'$; $M11=1/NT Y_+ Y_+'$; $M12=1/NT Y_+ Y_+'$; $M13= 1/NT Y_+ X'$;
 $M22=1/NT YX'$; $M23=1/NT Y X'$;
 $M33=1/NT XX'$;

Diese können zuvor mit dem Modul MOMENT1 aus den Datenmatrizen YY und XX berechnet werden.

DT: Abstand der Meßzeitpunkte ($=\Delta t$).

N: Anzahl der statistischen Einheiten (Replikationen).

T: (T+1) ist die Anzahl der gemessenen Zeitpunkte t_0, \dots, t_T .

OPTION: Wenn=1, werden in jedem Iterationsschritt die Eigenwerte der Fisher-Informationsmatrix F berechnet. Dies ist sinnvoll, wenn bei unklarem Identifikationsstatus die Iteration überwacht werden soll.

NEWLIN (THETA, COV, THETA0, KMAX, EPS, M0, M00, M11, M12, M13, M22, M23, M33, DT, N, T)

Zweck: Quasi-Newton- (Scoring)-Algorithmus) zur Bestimmung des ML-Schätzers $\hat{\theta}$ für das linearisierte Modell unter Restriktionen (diskretisiertes kontinuierliches Sampling bzw. Rechtecknäherung, siehe Kap. 2.2 und 2.3)

Ausgabe-Parameter und **Eingabe-Parameter:** siehe **NEWTON**

MOMENT1 (M0,.....M33, YY, XX, P, Q, N, T)

Zweck: Berechnung der Stichprobenmomente aus den Datenmatrizen.

Ausgabeparameter:

M={M0,M00,.....M33}: Stichprobenmomente (siehe oben)

Eingabeparameter:

YY: $P \times N(T+1)$ -Datenmatrix : $YY=[Y_0, Y_1, \dots, Y_T]$

XX: $Q \times N(T+1)$ -Datenmatrix : $XX=[X_0, X_1, \dots, X_T]$

P: Dimension des Systemzustands y

Q: Dimension der Exogenen x.

N, T: siehe oben

EXAKT (YY, A, B, G, Y0, XX, P, Q, N, T, DT, SEED, PRINT)

Zweck: Simulation eines Datensatzes YY, der N Trajektorien an den Meßzeitpunkten t_0, \dots, t_T enthält, d.h. $YY=[Y_0, Y_1, \dots, Y_T]$. Benutzt wurde der Zufallsgenerator RANNOR (normalverteilte Zufallszahlen, siehe SAS User's Guide: Basics 1985)) und die Rekursionsbeziehungen aus Kap. 4.1.

Ausgabeparameter:

YY: $P \times N(T+1)$. Simulierte Datenmatrix.

Eingabeparameter:

A: $P \times P$ -Driftmatrix .

B: $P \times Q$ -Input-Matrix

G: $P \times R$ -Diffusionskoeffizient.

Y0: $P \times N$. Anfangswerte für N Trajektorien. In **NEWTON** wird angenommen, daß Y0 Gauß-verteilt ist.

XX: $Q \times N(T+1)$. Exogene an den Zeitpunkten t_0, \dots, t_T , d.h. $XX=[X_0, X_1, \dots, X_T]$

P, Q, N, T: siehe oben.

PAR (OMEGA, A, B, SIGMA, MUE, THETA)

Zweck: Modellspezifikation. Systemmatrizen als Funktion des Parametervektors THETA. Muß vom Benutzer bereitgestellt werden.

Ausgabeparameter:

OMEGA: $P \times P$ -Diffusionskoeffizient (OMEGA=G*G')

A: Drift.

B: Input-Matrix.

SIGMA: Kovarianzmatrix der Anfangsbedingung y_0 .

MUE: Erwartungswert der Anfangsbedingung y_0 : $N(\mu, \Sigma)$.

Eingabeparameter:

THETA: $u \times 1$

PARD (DOMEGA, DA, DB, DSIGMA, DMUE, THETA)

Zweck: Modellspezifikation. **McDonald-Swaminathan**-Ableitungen der Systemmatrizen als Funktion des Parametervektors THETA. Muß vom Benutzer bereitgestellt werden.

Ausgabeparameter:

DOMEGA: $\partial\Omega/\partial\theta$: $u \times p^2$. **M/S**-Ableitung der Diffusionsmatrix.

DA: $\partial A/\partial\theta$: $u \times p^2$. Ableitung des Drift-Koeffizienten

DSIGMA: $\partial\Sigma/\partial\theta$: $u \times p^2$. Ableitung der Kovarianzmatrix der Anfangsbedingung y_0 .

DMUE: $\partial\mu/\partial\theta$: $u \times p$. Ableitung des Erwartungswerts der Anfangsbedingung y_0 : $N(\mu, \Sigma)$.

Eingabeparameter:

THETA: $u \times 1$

ALIASING (AN, BN, OMEGAN, A, B, OMEGA, N, DT)

Zweck: Berechnet zu vorgegebenen Systemmatrizen A, B und Ω die "Aliasing-Matrizen" A_N , B_N und Ω_N , wobei N: $a \times NC$ (a=Zahl der zu berechnenden Aliasing-Matrizen, $NC \geq$ Anzahl der c.c. Eigenwerte) ganze Zahlen enthält. Weiterhin werden die Eigenwerte von Ω_N berechnet, um die positive Definitheit überprüfen zu können (vgl. Kap. 3)

Ausgabeparameter:

AN, BN, OMEGAN

Eingabeparameter:

A, B, OMEGA, N, DT

4.9. Beispiele

4.9.1. Ornstein-Uhlenbeck-Prozeß

Im folgenden wird die Maximum-Likelihood-Schätzung des exakten diskreten Modells an Hand eines simulierten Datensatzes verdeutlicht. Es handelt sich um einen gedämpften harmonischen Oszillator, der von einer stochastischen Kraft angeregt wird (zeitkontinuierlicher autoregressiver Prozeß 2. Ordnung):

$$\ddot{y}(t) + \gamma \dot{y}(t) + \omega_0^2 y(t) = g\zeta(t) + bx(t) = F(t)$$

Die stochastische Kraft F hat in diesem Falle die Besonderheit, daß ihr Erwartungswert $E[F]=bx(t)$ nicht verschwindet ($\zeta(t)$ ist weißes Rauschen), d.h. im Mittel wird der Oszillator aus seiner stabilen Nulllage abgelenkt. Modelle dieser Art sind essentiell zur Beschreibung von periodischen Vorgängen, die durch Kopplung an externe Einflüsse gestört sind.

Abb 4.1 zeigt den simulierten Verlauf der Trajektorien $y(t)$ und $\dot{y}(t)$ unter dem Einfluß der konstanten Kraft $x(t)=1$ und den Parameterwerten $\omega_0^2=16$, $\gamma=4$, $b=1$, $g=2$ und der zufälligen Anfangsbedingung $[y(0), \dot{y}(0)]^T \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$. Die Trajektorien wurden mit Hilfe des Moduls EXAKT (siehe Kap. 4.8) und einem Zeitschritt von $dt=10^{-2}$ erzeugt. Klar erkennbar ist der glatte Verlauf von $y(t)$, während $\dot{y}(t)$ die charakteristische Rauigkeit des **Wiener**-Prozesses aufweist. Schreibt man obige Gleichung als Dgl. 1. Ordnung im 2-dimensionalen Phasenraum (mechanische Systeme können durch Ort und Geschwindigkeit vollständig charakterisiert werden):

$$\begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b \end{bmatrix} x + \begin{bmatrix} 0 & 0 \\ 0 & g \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix}$$

oder $\dot{y} = A y + B x + G \zeta$

bzw. $dy = Aydt + Bxdx + Gdw$

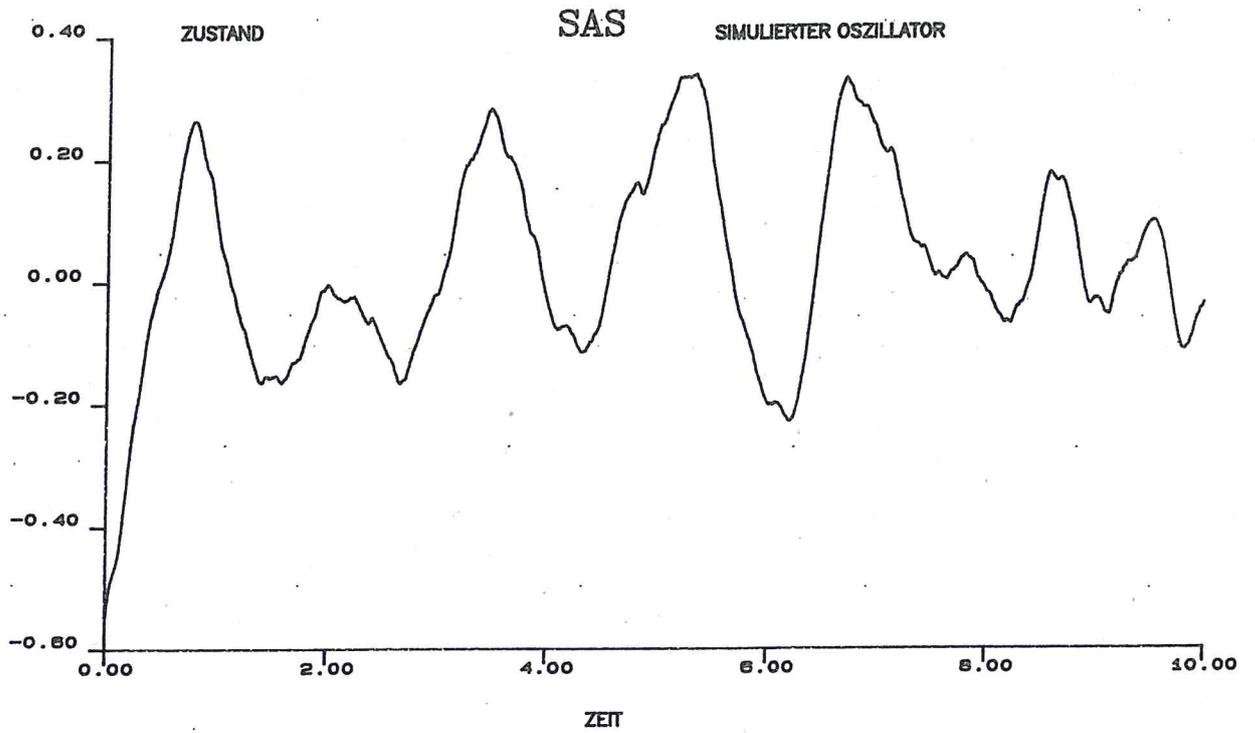
so wird klar, daß $y_1(t) = \int_{t_0}^t y_2(s) ds$ geschrieben werden kann (die Behandlung der **Brownschen** Bewegung mit differenzierbarem Ortsprozeß geht auf **Uhlenbeck u. Ornstein** 1930 zurück). Die Kreisfrequenz $\omega_0 = 2\pi\nu_0 = 4$ bzw. die Periodenlänge $T_0 = 2\pi/\omega_0 = 1.57$ des ungedämpften Oszillators wird durch die Reibung γ zu

$$\lambda_{1,2} = -\gamma/2 \pm \sqrt{\gamma^2/4 - \omega_0^2} = -2 \pm \sqrt{-12} = -2 \pm i\sqrt{12} = -2 \pm 3.464 i$$

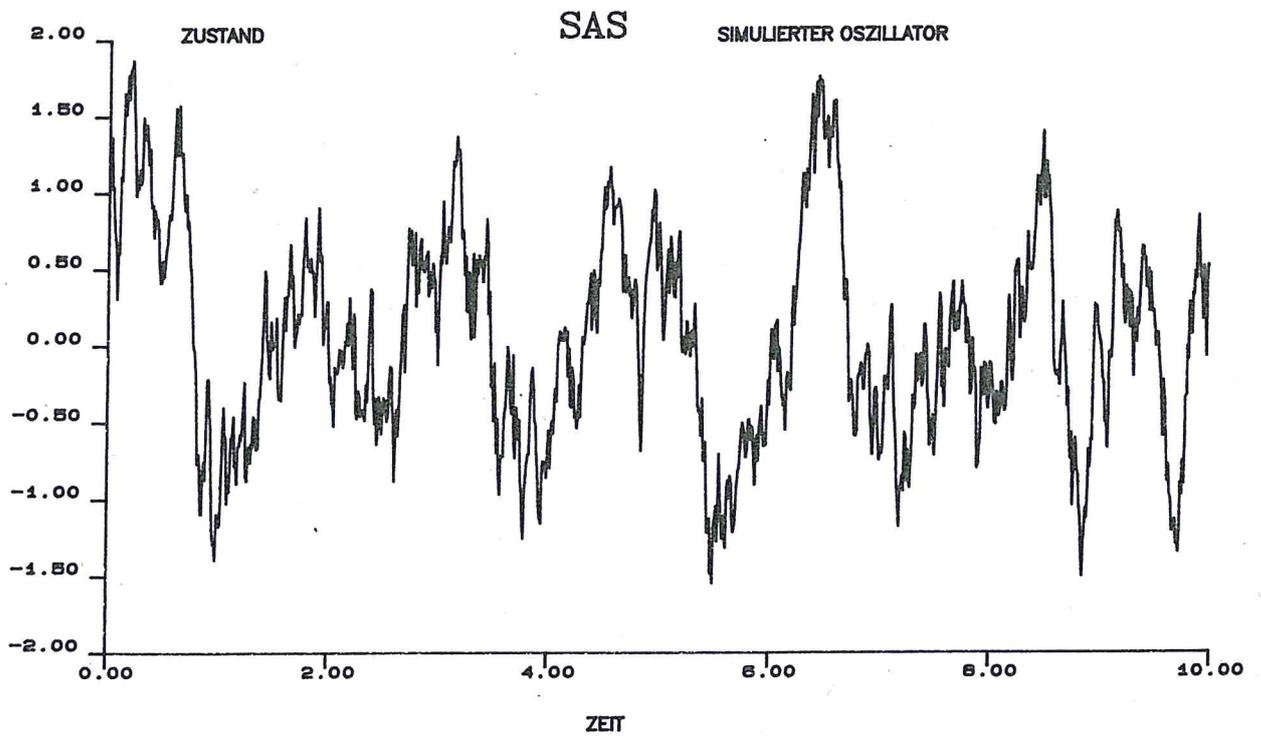
verändert (Dämpfung und verringerte Oszillationsfrequenz $\omega = \sqrt{-\gamma^2/4 + \omega_0^2}$). Dies kann man anhand der Eigenwerte von A leicht einsehen. Da der Realteil von λ negativ ist, stellt sich nach einiger Zeit ein stationärer Zustand ein, der durch konstanten Erwartungswert und Kovarianzmatrix charakterisiert ist. Es gilt:

$$\dot{\mu} = A\mu + Bx = 0 \Rightarrow \mu_s = -A^{-1}B = \begin{bmatrix} bx/\omega_0^2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/16 \\ 0 \end{bmatrix}$$

d.h. eine durch die konstante Kraft x verschobene Ruhelage und



a) erste Komponente (Ortsprozeß)



b) zweite Komponente (Geschwindigkeitsprozeß)

Abb. 4.1. Simulierte Trajektorie eines stochastischen Oszillators mit zufälliger Anfangsbedingung $N(0, I_2)$.

$$\dot{\Sigma} = A\Sigma + \Sigma A' + \Omega = 0$$

$$\Rightarrow \Sigma_S = g^2/2\gamma \begin{bmatrix} 1/\omega_0^2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/32 & 0 \\ 0 & 1/2 \end{bmatrix} \text{ bzw. } \Sigma_S^{1/2} = \begin{bmatrix} .177 & 0 \\ 0 & .7071 \end{bmatrix}$$

Diese Größen lassen sich durch optische Inspektion qualitativ bestätigen.

Identifikation

Bevor das Modell geschätzt werden kann, muß der Identifikationsstatus geklärt werden. Da A im schwach gedämpften Fall ($\omega_0^2 > \gamma^2/4$) konjugiert komplexe Eigenwerte $\lambda_{1,2}$ aufweist, sind (vgl. Kap 3.) Aliasing-Matrizen der Form $A_N = A + (2\pi i/\Delta t)P \begin{bmatrix} n & -n \end{bmatrix} P^{-1}$ mit $A = PAP^{-1}$ zu erwarten. Diese müssen jedoch mit den geforderten Restriktionen $A_{11} = 0$, $A_{12} = 1$ übereinstimmen. Im vorliegenden Fall läßt sich P noch leicht berechnen und man findet:

$$P = \begin{bmatrix} \Phi_1 & \Phi_1^* \\ \Phi_2 & \Phi_2^* \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix} ; P^{-1} = \begin{bmatrix} \psi_1 & \psi_2 \\ \psi_1^* & \psi_2^* \end{bmatrix} = -(1/2i\omega) \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}$$

wobei * konjugiert-komplex bedeutet. Die Erfüllung der Restriktionen ergibt dann die Bedingungen:

$$0 = \text{Im}(\Phi_1\psi_1) \text{ und } 0 = \text{Im}(\Phi_1\psi_2)$$

bzw. $0 = -\gamma/4\omega$ und $0 = -(2\omega)^{-1}$.

Dies läßt sich jedoch nicht erfüllen, sodaß das Modell identifizierbar ist.

Schätzung der System-Parameter:

4.9.1.1. Exakte ML-Schätzung

Im folgenden wird angenommen, daß N=50 Trajektorien im Zeitintervall [0, 10] vorliegen, jedoch nur zu 6 diskreten Zeitpunkten mit Abstand $\Delta t=2$ beobachtet werden konnten (d.h. T=5, $\Delta t=2$). Aus diesen Daten sollen nun ML-Schätzwerte der Parameter $\Theta = \{-\omega_0^2, -\gamma, b, g, \mu, \Sigma\}$ und die geschätzte Kovarianzmatrix von $\hat{\Theta}$ bestimmt werden.

Dazu wurden 50 Trajektorien mit Hilfe des Moduls EXAKT simuliert und aus den Datenmatrizen $YY = [Y_0, Y_1, \dots, Y_T]$ und $XX = [X_0, X_1, \dots, X_T]$ mit dem Modul MOMENT1 die Stichprobenmomente M0, M33 berechnet. Zur Modellspezifikation müssen die Module PAR und PARD definiert werden. Setzt man $\Theta = \{\Theta_1, \dots, \Theta_6\}$, so gilt:

$$A(\Theta) = \begin{bmatrix} 0 & 1 \\ \Theta_1 & \Theta_2 \end{bmatrix} \quad B(\Theta) = \begin{bmatrix} 0 \\ \Theta_3 \end{bmatrix} \quad G(\Theta) = \begin{bmatrix} 0 & 0 \\ 0 & \Theta_4 \end{bmatrix} ; \Omega = GG'$$

$$\mu(\Theta) = \begin{bmatrix} \Theta_5 \\ \Theta_6 \end{bmatrix} ; \quad \Sigma(\Theta) = \begin{bmatrix} \Theta_7 & \Theta_8 \\ \Theta_8 & \Theta_9 \end{bmatrix}$$

Die Parametrisierung von Ω über GG' hat den Vorteil, daß Ω positiv semidefinit ist (zur Vermeidung weiterer Identifikationsprobleme wird die **Cholesky**-Zerlegung gewählt, d.h. G ist eine Dreiecksmatrix). Maximierung über unrestringiertes Ω führt häufig in Parametergebiete, wo Ω indefinit ist. Im Bedarfsfall kann man für Σ analog vorgehen.

Die **McDonald-Swaminathan**-Ableitungen der Parametermatrizen (Modul PARD) sind durch folgende Ausdrücke gegeben:

$$\frac{\partial A}{\partial \Theta} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \frac{\partial B}{\partial \Theta} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \frac{\partial \Omega}{\partial \Theta} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\Theta_4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\frac{\partial \mu}{\partial \Theta} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \frac{\partial \Sigma}{\partial \Theta} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Die ML-Schätzung des Modells wird mit dem Modul NEWTON durchgeführt. Dazu müssen Startwerte für die Scoring-Iteration festgelegt werden. Da es sich um ein nichtlineares Schätzproblem handelt, müssen die Startwerte relativ gut sein. Des weiteren empfiehlt es sich, mit mehreren Startwerten zu arbeiten, um die Konvergenz des Algorithmus zu Nebenmaxima abzutesten. Im vorliegenden Beispiel wurde ein Startwert von

$$\Theta_0 = \{-16.3, -4.3, .7, 1.7, -3, -3, .7, -3, .7\}$$

(zum Vergleich $\Theta = \{-16, -4, 1, 2, 0, 0, 1, 0, 1\}$) benutzt.

Bei einem Abbruchkriterium von $\varepsilon = 10^{-4}$ (vgl. Kap. 4.4) konvergiert der Scoring-Algorithmus nach 9 Iterationen zum ML-Schätzwert (siehe Tab. 4.2) mit der

exakt		linearisiert	
$\hat{\Theta}$	Std	$\hat{\Theta}$	Std
-16.8229	1.9937	-.0375	.0459
-4.1320	.9274	-.4655	.0288
.8609	.2107	.0268	.0223
2.0228	.2558	.4985	.0223
-.0683	.1454	-.0683	.1454
.0578	.1441	.0578	.1441
1.0572	.2114	1.0572	.2114
.0248	.1482	.0248	.1482
1.0381	.2076	1.0381	.2076

Tab . 4.2. Maximum-Likelihood-Schätzer und geschätzte Varianz (exaktes Modell und linearisiertes Modell)

Log-Likelihood (Konstante weggelassen) $l=226.3492$. Andere Startwerte (Θ +Zufallsvektor) führen zum gleichen Resultat. Die Standardabweichungen sind relativ klein. Approximative 95%-Konfidenzintervalle $\hat{\Theta} \pm 2 \text{std}$ ($\text{std}(i) = \sqrt{(F^{-1})_{ii}}$) überdecken den wahren Parameter - eigentlich müßten simultane Konfidenzintervalle konstruiert werden.

4.9.1.2. Linearisiertes Modell

Äußerst aufschlußreich ist der Vergleich des exakten (nichtlinearen) ML-Verfahrens mit dem linearisierten Modell (Rechtecknäherung bzw. diskretisiertes kontinuierliches Sampling). Ausgehend vom gleichen Startwert Θ_0 konvergiert das Modul NEWLIN in 12 Iterationen zum ML-Schätzwert (siehe Tab. 4.2). Die Schätzungen sind sehr stark verzerrt, obwohl die geschätzten Standardabweichungen gute Schätzwerte suggerieren. Wie in Kap. 2 bemerkt, ist die Rechteck- bzw. Trapeznäherung nur gut, wenn $A^* = \exp(A\Delta t)$ durch $I + A\Delta t$ approximiert werden kann. Im vorliegenden Falle gilt jedoch

$$A^* = \exp(A\Delta t) = \begin{bmatrix} .0210 & .0032 \\ -.0509 & .0083 \end{bmatrix}; \quad I + A\Delta t = \begin{bmatrix} 1 & 2 \\ -32 & -7 \end{bmatrix}$$

I.A. ist jedoch der wahre Wert des Parameters unbekannt, sodaß derartige Abschätzungen unmöglich sind. Daher ist die Linearisierung eine ad-hoc-Näherung, deren Qualität nur im Rahmen des exakten ML-Verfahrens beurteilt werden kann.

4.9.1.3 Eine Simulationsstudie

Um die exakte und linearisierte ML-Methode genauer vergleichen zu können, wurden $M=100$ Stichproben mit jeweils $T=5$ und $N=50$ erzeugt. Die Mittelwerte und Stichproben-Kovarianzmatrizen $\bar{\Theta}$ und S sind in Tab. (4.3) aufgeführt. Wiederum bestätigt sich die asymptotische Verzerrung der Schätzer auch in kleinen Stichproben. Eine univariate Analyse der Schätzungen mit SAS/Univariate (siehe SAS user's guide: Basics 1985) zeigt eine approximative Normalverteilung (**Kolmogoroff-Smirnoff-Test**).

Insgesamt kann gesagt werden, daß im untersuchten Fall die linearisierte Methode völlig unrealistische Schätzungen ergibt, während die exakte ML-Methode zu recht genauen Aussagen über die unbekannt Parameter führt. Die gesamte Rechenzeit für alle 100 Replikationen betrug 673.88 Sekunden (Simulation, linearisierte und exakte ML-Schätzung). Damit kann das Verfahren als sehr effektiv bezeichnet werden.

4.9.1.4 Reparametrisierung

In Kap. 2 und 3 wurde gezeigt, daß die Methode der Schätzung des exakten diskreten Modells und anschließende Zurückrechnung (**Tuma u. Hannan, Arming**) nicht zu empfehlen ist. Im folgenden werden die Probleme am eben diskutierten Modell und 2 Datensätzen demonstriert.

Im 2-dimensionalen Fall lassen sich die Matrizen des Zeitreihenmodells in expliziter Form angeben. Man findet (vgl. auch **Arató** 1982, Kap. 1.7)

$$A^* = \exp \left\{ \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix} \Delta t \right\}$$

$$= e^{-\gamma/2 \Delta t} \begin{bmatrix} \gamma/2\omega \sin \omega\Delta t + \cos \omega\Delta t & 1/\omega \sin \omega\Delta t \\ -(\gamma^2/4 + \omega^2)/\omega \sin \omega\Delta t & \cos \omega\Delta t - \gamma/2\omega \sin \omega\Delta t \end{bmatrix}$$

mit $\omega = \sqrt{-\gamma^2/4 + \omega_0^2}$.

An dieser Formel wird unmittelbar ersichtlich, daß sich einfache Restriktionen auf A in komplizierte nichtlineare Restriktionen bei A^* umwandeln, die auf der Zeitreihenebene nicht handhabbar sind. Die unrestringierte Schätzung von A^*

THLQ	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
ROW1	-0.0286	-0.4990	.0029675	0.4988	-.005101	-2.0E-04	0.9613	-.004506	0.9958
THLS	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
ROW1	.0020487	1.9E-04	-3.3E-04	4.6E-05	-2.4E-06	-6.4E-04	-4.5E-04	-5.4E-05	-4.5E-04
ROW2	1.9E-04	8.7E-04	4.9E-06	8.2E-05	-7.4E-05	-4.6E-05	8.4E-04	3.8E-04	-2.5E-04
ROW3	-3.3E-04	4.9E-06	5.4E-04	-9.2E-07	4.4E-04	2.8E-04	-4.8E-04	-1.0E-04	-2.2E-04
ROW4	4.6E-05	8.2E-05	-9.2E-07	3.9E-04	-2.3E-04	1.8E-04	7.2E-04	-1.1E-04	-1.5E-04
ROW5	-2.4E-06	-7.4E-05	4.4E-04	-2.3E-04	0.0245	-.004251	-.003687	.0015355	-9.8E-04
ROW6	-6.4E-04	-4.6E-05	2.8E-04	1.8E-04	-.004251	0.0247	.0026103	0.002052	.0010264
ROW7	4.5E-04	8.4E-04	-4.8E-04	7.2E-04	-.003697	.0025103	0.0357	.0011079	-.006278
ROW8	-5.4E-05	3.8E-04	-1.0E-04	-1.1E-04	.0015355	0.002052	.0011079	0.0211	.0014997
ROW9	-4.5E-04	-2.5E-04	-2.2E-04	-1.5E-04	-9.8E-04	.0010264	-.006278	.0014997	0.0331
THEXQ	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
ROW1	-16.0799	-4.1346	0.9930	2.0197	-.005101	-2.0E-04	0.9613	-.004501	0.9958
THEXS	COL1	COL2	COL3	COL4	COL5	COL6	COL7	COL8	COL9
ROW1	3.4435	0.2888	-0.1849	-0.1555	0.0257	-0.0243	-0.0157	-0.0195	-0.0126
ROW2	0.2888	1.2722	-0.00734	-0.2879	.0032925	-0.0134	0.0143	0.0101	0.0191
ROW3	-0.1849	-0.00734	0.0398	.0055933	-3.3E-04	.0035013	.0023133	.0040983	-.002636
ROW4	-0.1555	-0.2879	.0055933	0.0718	-.002445	.0043149	-4.1E-04	-.003053	-.004791
ROW5	0.0257	.0032925	-3.3E-04	-.002445	0.0245	-.004251	-.003687	.0015355	-9.8E-04
ROW6	-0.0243	-0.0134	.0035013	.0043149	-.004251	0.0247	.0026103	0.002052	.0010264
ROW7	-0.0157	0.0143	.0023133	-4.1E-04	-.003687	.0026103	0.0357	.0011071	-.006279
ROW8	-0.0195	0.0101	.0040983	-.003063	.0015355	0.002052	.0011071	0.0211	0.001498
ROW9	-0.0126	0.0191	-.002636	-.004791	-9.8E-04	.0010264	-.006279	0.001498	0.0331

Tab. 4.3. Mittelwerte und Stichproben-Kovarianz-Matrizen des ML-Schätzers für M=100 simulierte Datensätze des Oszillatormodells (linearisiertes und exaktes ML-Verfahren)

und anschließende Zurückrechnung auf A führt zu einem beobachtungsäquivalenten Satz von Matrizen A_N , die jedoch die geforderten Restriktionen ($A_{11}=0$, $A_{12}=1$) nicht erfüllen. Da häufig in sozialwissenschaftlichen Anwendungen keine Struktur für A bekannt ist, könnte man trotzdem versuchen, von A^* zurückzurechnen:

Datensatz 1: (identisch mit dem oben diskutierten)

Wie in Kap. 4.3 gezeigt, erhält man als ML-Schätzer für A^* , B^* , V den OLS-Schätzer (hinreichend), wenn die Exogenen durch Sprungfunktionen approximiert werden ($x(t)$ ist hier konstant):

$$\hat{A}^* = \begin{bmatrix} .0223 & -.0088 \\ -.0749 & .0690 \end{bmatrix}; \hat{B}^* = \begin{bmatrix} .0500 \\ .0535 \end{bmatrix}; \hat{V} = \begin{bmatrix} .0289 & -.0036 \\ -.0036 & .4970 \end{bmatrix}$$

und die Likelihood 228.19 > 227.69 (exakte Restriktionen).

Mit Hilfe des Matrix-Logarithmus (Hauptwert; vgl. **Coddington u. Levinson** 1955, Kap. 3)

$$A = 1/\Delta t P \left[\log \Lambda \right] P^{-1}, \text{ wobei } P^{-1} A^* P = \Lambda$$

gilt:

$$\hat{A} = \begin{bmatrix} -2.0945 & -.1270 \\ -1.0761 & -1.4241 \end{bmatrix}; \hat{B} = \begin{bmatrix} .1130 \\ .1307 \end{bmatrix}; \hat{\Omega} = \begin{bmatrix} .1203 & .0806 \\ .0806 & 1.4144 \end{bmatrix}$$

Da jedoch der Logarithmus einer komplexen Zahl nicht eindeutig ist, existieren "Aliasing-Matrizen" der Form (siehe Kap.3, **Phillips** 1972)

$$\hat{A}_N = \hat{A} + (2\pi i/\Delta t) P \begin{bmatrix} 0 & & \\ & N & \\ & & -N \end{bmatrix} P^{-1}$$

Im vorliegenden Fall sind die Eigenwerte von \hat{A} reell ($\lambda(\hat{A}) = \{-1.2602, -2.2583\}$), sodaß zu diesem Datensatz keine Aliasing-Matrizen existieren. Es wäre jedoch ein Trugschluß, daraus die Identifizierbarkeit des unrestringierten 2-dimensionalen Modells zu folgern. Für andere Datensätze sind die Eigenwerte von \hat{A} konjugiert komplex, sodaß beobachtungsäquivalente Strukturen existieren. Zu bemerken ist weiterhin, daß die Schätzwerte fernab der wahren Parameter liegen und die Restriktionen nicht erfüllt sind.

Von Interesse ist nun das Verhalten des Scoring-Algorithmus im vorliegenden Fall, wenn alle Systemmatrizen unrestringiert sind (14 Parameter). Nach 51 Iterationen mit fast singulärer **Fisher**-Informations-Matrix (was die Nicht-Identifizierbarkeit des Modells widerspiegelt; Satz von **Rothenberg**, Kap. 3.5) findet der Algorithmus eine approximative Nullstelle der Score-Funktion, die relativ nahe am wahren Parameter liegt. Die Likelihood liegt zwischen dem exakt restringierten Modell und der OLS-Lösung.

Datensatz 2:

Ein anderer Datensatz zeigt eine weitere Schwachstelle der Reparametrisie-

rungs-Methode:

$$\hat{A}^* = \begin{bmatrix} .0440 & .0227 \\ .0041 & -.0302 \end{bmatrix}; \hat{B}^* = \begin{bmatrix} .0752 \\ -.0559 \end{bmatrix}; \hat{V} = \begin{bmatrix} .0303 & -.0016 \\ -.0016 & .5260 \end{bmatrix}; l=214.57$$

Hier ist ein Eigenwert von \hat{A}^* negativ ($\lambda(\hat{A}^*) = \{.0452, -.0314\}$), sodaß der Matrix-Logarithmus auf eine komplexe Lösung für \hat{A} führt (dies tritt schon im univariaten Fall auf, wenn $\hat{A}^* < 0$ ist!). Die Scoring-Iteration führt nach 38 Iterationen auf eine Nullstelle der Score-Funktion mit $l=213.28$.

Auch in diesem Fall, wo keine Rückrechnung auf (reelles) A möglich ist, findet man also weitere Nullstellen der Score-Funktion, die von der OLS-Lösung abweichen. Zum Schätzwert \hat{A}_{ML} , \hat{B}_{ML} und $\hat{\Omega}_{ML}$ existieren weitere beobachtungsäquivalente Strukturen mit positiv definiten Diffusionsmatrizen $\hat{\Omega}_N$, von denen ein \hat{A}_N mit dem Realteil des Hauptwerts von $\log \hat{A}^*$ ungefähr übereinstimmt. Die Verhältnisse sind also relativ kompliziert.

Um **Mißverständnissen** vorzubeugen: Die vollen Modelle sind **nicht identifizierbar** und sollten daher gar nicht geschätzt werden. Die obige Diskussion dient lediglich dem Zweck, die verschiedenen Pathologien bei naiver Rückrechnung aus dem exakten diskreten Modell zu demonstrieren.

4.9.2. Das Phillips-Modell

Im Rahmen einer Simulationsstudie wurde von **Phillips** (1972) das Differentialgleichungsmodell

$$d/dt \begin{bmatrix} C(t) \\ Y(t) \\ K(t) \end{bmatrix} = \begin{bmatrix} -\alpha & \alpha(1-s) & 0 \\ \lambda & \lambda(\gamma v - 1) & -\lambda\gamma \\ 0 & \gamma v & -\gamma \end{bmatrix} \begin{bmatrix} C(t) \\ Y(t) \\ K(t) \end{bmatrix} + \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} F + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix}$$

mit Hilfe der Minimum-Distance-Methode und der Trapez-Approximation geschätzt.

Die System-Variablen können als C ="real consumption", Y ="net national income", K ="fixed capital" und F ="autonomous consumption" interpretiert werden. Die Daten wurden mit dem Parameterwert $\Theta = \{\alpha, \lambda, \gamma, v, s\} = \{.6, 4.0, .4, 2.0, .25\}$ und $F=5$ simuliert. In diesem Fall ist die Drift-Matrix eine Funktion von Θ , was im Rahmen von LSD leicht implementiert werden kann.

Identifikation:

Man kann zeigen (**Phillips** 1973), daß ein 3-dimensionales System unter der Restriktion $A_{13}=0$ identifiziert werden kann. Da weiterhin Θ aus A identifiziert ist, existieren keine beobachtungsäquivalenten Strukturen und das Modell kann geschätzt werden.

Parameterschätzung:

Im Rahmen der Minimum-Distance-Methode ist die Diffusionsmatrix kein frei wählbarer (fixer oder restringierter) Parameter, sondern eine aus den Residuen zurückgerechnete Größe. Dagegen kann bei der ML-Methode Ω wahlweise (je nach Informationstand) als fixe ($\Omega=I$) oder diagonale bzw. völlig freie Größe modelliert werden (die entprechenden Parameter sind zu Θ hinzuzufügen).

In einer Publikation von **Möbus** u. **Nagl** (1983) wurde das obige Modell mit Hilfe eines von **Phillips** (1972) angegebenen Datensatzes und der Trapez-Näherung mit dem Programm LISREL geschätzt (die Diffusionsmatrix war hier restringiert auf Diagonalform mit $\Omega_{23} \neq 0$). Im folgenden werden die Schätzwerte der Trapezmethode mit der exakten ML-Methode und den Resultaten des linearisierten Modells (Rechtecknäherung) verglichen:

	exakt	Rechteck	Trapez	exakte ML-Methode
$\hat{\alpha}$.6	.5783	.635	.6080
$\hat{\lambda}$	4.0	2.1524	3.201	3.2662
$\hat{\gamma}$.4	.4055	.407	.4103
\hat{v}	2.0	2.0096	2.012	2.0148
\hat{s}	.25	.2394	.235	.2362
$\hat{\Omega}_{11}$	1	1.3526	1.321	1.4635
$\hat{\Omega}_{22}$	1	3.4899	31.200	2.0160
$\hat{\Omega}_{23}$	0	1.5778	-11.02	2.6072
$\hat{\Omega}_{33}$	1	3.2779	4.178	3.9426
Lik		-82.6069		-67.5763

Bei der Schätzung der Drift-Parameter sind die Methoden in etwa vergleichbar, wobei die exakte Methode am besten abschneidet. Auffällig verzerrt sind jedoch die Schätzungen des Diffusionskoeffizienten im Falle der Trapez-Näherung. Dies müßte jedoch im Rahmen einer Simulationsstudie überprüft werden.

Weiterhin wurden $M=100$ Datensätze mit $T=25$, $N=1$ und der deterministischen Anfangsbedingung $y_0 = -A^{-1}B$ (vgl. **Phillips** 1972) simuliert (Gesamtrechenzeit 1737.53 Sekunden für Simulation, exakte und linearisierte ML-Schätzung). Um die Daten mit der Arbeit von **Phillips** vergleichen zu können, wurde angenommen, daß Ω diagonal ist.

Man erhält (in Klammern Standardabweichungen):

	exakt	Rechteck	Trapez	Minimum-Distance	exakte ML-Methode
$\bar{\alpha}$.6	.4998 (.1529)	.6652 (.1800)	.5734 (.1410)	.5723 (.1539)
$\bar{\lambda}$	4.0	2.6746 (.7374)	2.7444 (.8015)	4.0709 (.7077)	4.1283 (.7631)
$\bar{\gamma}$.4	.3786 (.0273)	.4182 (.0241)	.4016 (.0153)	.4020 (.0202)
\bar{v}	2.0	1.9962 (.0140)	1.9995 (.0311)	2.0021 (.0149)	2.0008 (.0115)
\bar{s}	.25	.2547 (.0251)	.2767 (.0937)	.2537 (.0259)	.2509 (.0151)
$\bar{\Omega}_{11}$	1	.9520 (.2889)			.8789 (.2405)
$\bar{\Omega}_{22}$	1	3.1514 (1.0151)			.9088 (.8638)
$\bar{\Omega}_{33}$	1	1.3358 (.4002)			1.060 (.2821)

Die Trapez-Schätzung wurde mit Hilfe der 3-stufigen kleinste-Quadrate-Methode gewonnen. Auffällig ist wiederum die starke Verzerrung von λ in den linearisierten Modellen. Die Resultate von Minimum-Distance (MD) und exaktem Modell (ML) sind in etwa gleich gut.

Die Ergebnisse beruhen jedoch auf unterschiedlichen simulierten Datensätzen (die Daten von **Phillips** waren mir nicht zugänglich). Außerdem wurden im ML-Verfahren die Diffusions-Parameter Ω mitgeschätzt, was bei MD gar nicht möglich ist (hier ergibt sich nur \hat{V} aus den Residuen, vgl. Kap. 2.6).

5. ML-Schätzung des diskret/kontinuierlichen Zustandsraum-Modells mit EM, GEM und Newton-Raphson-Algorithmen.

5.1 Modellspezifikation

Im folgenden werden Schätzalgorithmen für das allgemeine diskret/kontinuierliche Zustandsraum-Modell (System-Modell + Meß-Modell) diskutiert:

$$(S) \quad dy_n(t) = Ay_n(t)dt + Bx_n(t)dt + Gdw_n(t); \quad t \in [t_0, t_T], n=1, \dots, N;$$

$$(5.1) \quad y_n(0) \sim N(\mu, \Sigma) \text{ i.i.d.}$$

$$(M) \quad z_{ni} = Hy_{ni} + Dx_{ni} + \varepsilon_{ni}; \quad \varepsilon_{ni} \sim N(0, R) \text{ i.i.d.}; \quad n=1, \dots, N, i=0, \dots, T$$

Wie in Kap. 4 sind N Trajektorien zu diskreten Zeitpunkten t_0, \dots, t_T meßbar, jedoch zusätzlich durch Meßfehler ε_{ni} und deterministische Trends Dx_{ni} überlagert. Weiterhin wird angenommen, daß nur Linearkombinationen Hy_{ni} des Systemzustands der Messung zugänglich sind. Wiederum ist es günstig, die Systemgleichung in ein Differenzenschema für die Meßzeitpunkte t_0, \dots, t_T umzuschreiben:

$$(S') \quad y_{ni+1} = A^*y_{ni} + B^*x_{ni} + u_{ni}$$

wobei $A^* = e^{A\Delta t}$, $B^* = A^{-1}(A^* - I)$ und $u_{ni} \sim N(0, V)$ mit $V = \int_0^{\Delta t} e^{As} \Omega e^{A's} ds$.

Hierbei wurde angenommen, daß die Exogenen x zwischen den Messungen durch konstante Werte approximiert werden können (vgl. Kap. 3).

Obiges State-Space-Modell stellt eine sehr allgemein und flexibel verwendbare Spezifikation dar, die Anwendungen von Differentialgleichungen höherer Ordnung bis zu nichtlinearen Regressionsmodellen mit autokorrelierter Fehlerstruktur, dynamischer Faktorenanalyse etc. zuläßt.

5.2 Parameterschätzung

Zur ML-Schätzung der Systemparameter stehen prinzipiell mehrere Methoden zur Verfügung.

Konzeptuell am einfachsten wäre es, die Likelihood für die **Gauß**-verteilten Daten z_{ni} mit Hilfe des **Kalman**-Filter-Algorithmus zu berechnen (Vorhersagefehler-Zerlegung, vgl. Kap. 5.5) und die Maximierung mit Hilfe eines numerischen Optimierungsverfahrens durchzuführen (mit numerischen Ableitungen, vgl. **Jones** 1984, **Jones u. Tryon** 1987). Da jedoch der Score, dessen Nullstelle gesucht wird, nur in einer numerischen Approximation vorliegt und zusätzlich die Likelihood zur Berechnung der Ableitungen häufig ausgewertet werden muß, erscheint dieses Verfahren als nicht besonders effizient.

Alternativ können Rekursionsgleichungen für die Score-Funktion und die **Fisher**-Informations-Matrix abgeleitet werden, die in jedem Schritt der Scoring-Iteration zusammen mit den üblichen **Kalman**-Filter-Gleichungen ausgewertet werden müssen (im zeitdiskreten Fall wurde dies von **Goodrich u. Caines** 1979, 1980 durchgeführt).

Ein weiteres Verfahren besteht darin, die zu schätzenden Parameter als zeitunabhängige Komponenten des Systemzustands aufzufassen und ein optimales Filter für den erweiterten Systemzustand $[y_n(t), \Theta]$ zu betrachten. Dieses Filter ist jedoch nichtlinear, sodaß das erweiterte **Kalman**-Filter (EKF) oder verwandte Approximationsmethoden benützt werden müssen (vgl. hierzu **Krebs, 1980**). Die im EKF benutzte Linearisierung (**Taylor**-Entwicklung) kann jedoch bei stärkeren Nichtlinearitäten zu einer Verzerrung der Schätzer führen.

Da im vorigen Kapitel ein Algorithmus zur Schätzung des System-Modells ohne Meßfehler abgeleitet und erprobt wurde, bietet es sich an, den EM-Algorithmus zu benutzen, der eine Anpassung der Methoden auf die unvollständige Meßbarkeit der Daten zuläßt.

5.3. Der EM-Algorithmus

Wenn sowohl $Y = \{y_{ni}, n=1, \dots, N, i=0, \dots, T\}$ als auch $Z = \{z_{ni}, n=1, \dots, N, i=0, \dots, T\}$ meßbar wären, hätte man einen vollständigen Datensatz für das Modell (5.1) zur Verfügung (die Meßbarkeit der deterministischen Kontrollvariablen $x_n(t)$ wird stillschweigend unterstellt). Da jedoch nur Z zur Verfügung steht, hat man es mit einem Problem unvollständiger Daten zu tun. Die Maximierung der Likelihood $L_{\Theta}(Z, Y)$ kann daher nicht durchgeführt werden. Eine in solchen Situationen seit langem benutzte Methode besteht darin, zuerst unter Annahme eines Parametervektors Θ_0 die fehlenden Daten zu schätzen (aus Z) und dann einen neuen Schätzwert Θ_1 zu berechnen (Iteration bis zur Konvergenz). Dieses heuristische Verfahren wurde von **Dempster, Laird u. Rubin** (1977) unter dem Titel EM (Estimation-Maximization) formalisiert. Die Idee besteht darin, die unbekannte

Likelihood $l_{\theta}(Z,Y)=\log L_{\theta}(Z,Y)$ durch die Daten Z zu schätzen (E-Schritt), d.h anstelle von $l_{\theta}(Z,Y)$ benutzt man eine Pseudolikelihood $Q(\theta,\theta_k)=E_{\theta_k}[l_{\theta}(Y,Z)|Z]$, deren Maximum den neuen Schätzwert θ_{k+1} liefert (Iteration). Unter bestimmten Voraussetzungen (vgl **Wu** 1983) konvergiert die Sequenz θ_k zum ML-Schätzwert $\hat{\theta}=\operatorname{argmax} l_{\theta}(Z)$. Die folgende Darstellung orientiert sich an der Arbeit von **Dembo** u. **Zeitouni** (1985), in der ein EM-Algorithmus für den kontinuierlich/kontinuierlichen Fall diskutiert wird (das diskret/diskrete Problem wird von **Watson** u. **Engle** 1983 behandelt).

Betrachten wir zuerst den Likelihood-Quotienten der beobachtbaren Daten. Es gilt:

$$(5.2) \quad L_{\theta}(Z)/L_{\theta'}(Z) = E_{\theta'} [L_{\theta}(Z,Y)/L_{\theta'}(Z,Y) | Z]$$

$$\begin{aligned} \text{denn } \int [L_{\theta}(Z,Y)/L_{\theta'}(Z,Y)] L_{\theta'}(Z|Y) dY &= \int [L_{\theta}(Z,Y)/L_{\theta'}(Z,Y)] [L_{\theta'}(Z,Y)/L_{\theta'}(Z)] dY \\ &= \int [L_{\theta}(Z,Y)/L_{\theta'}(Z)] dY = L_{\theta}(Z)/L_{\theta'}(Z) \end{aligned}$$

(der Einfachheit halber wurde der gleiche Buchstaben L für bedingte bzw. gemeinsame Likelihood benutzt). Der zu maximierende Likelihood-Quotient der gemessenen Daten läßt sich also als bedingte Erwartung des vollständigen Likelihood-Quotienten darstellen. Entsprechendes gilt für die Score-Funktion (vgl. **Louis** 1981):

$$(5.3) \quad s_{\theta}(Z) = \partial/\partial\theta \log L_{\theta}(Z) = \partial/\partial\theta l_{\theta}(Z) = E_{\theta}[s_{\theta}(Z,Y)|Z]$$

wobei $s_{\theta}(Z,Y)=\partial/\partial\theta l_{\theta}(Z,Y)$ der Score der vollständigen Daten ist.

Die Gültigkeit von (5.3) ist leicht aus den Relationen

$$\partial/\partial\theta l_{\theta}(Z) = \partial/\partial\theta \log \int L_{\theta}(Z,Y) dY = (1/\int L_{\theta}(Z,Y) dY) * \int \partial L_{\theta}(Z,Y)/\partial\theta dY$$

$$= (1/L_{\theta}(Z)) \int [\partial l(Z,Y)/\partial\theta] L_{\theta}(Z,Y) dY = \int s_{\theta}(Z,Y) L_{\theta}(Y|Z) dY$$

einzu sehen (benutzt wurde $\partial l/\partial\theta=1/L \partial L/\partial\theta$).

Für die **Fisher**-Information ist dieser einfache Zusammenhang jedoch nicht gegeben. Es gilt (**Louis** 1981) für die sog. **beobachtete Fisher-Information**:

$$\begin{aligned} (5.3') \quad I(Z) &= -\partial^2 l_{\theta}(Z)/\partial\theta^2 \\ &= E_{\theta} [-\partial^2 l_{\theta}(Z,Y)/\partial\theta^2 | Z] - E_{\theta} [s(Z,Y)s'(Z,Y)|Z] + s_{\theta}(Z)s_{\theta}'(Z) \end{aligned}$$

Aus Glg. (5.2) folgt nun mit Hilfe der **Jensen**-Ungleichung (log ist konvex):

$$(5.4) \quad \begin{aligned} \log L_{\Theta}(Z)/L_{\Theta'}(Z) &= \log E_{\Theta'} [L_{\Theta}(Z,Y)/L_{\Theta'}(Z,Y) | Z] \\ &\geq E_{\Theta'} [\log L_{\Theta}(Z,Y)/L_{\Theta'}(Z,Y) | Z] := Q(\Theta, \Theta'). \end{aligned}$$

Es gilt also: $l_{\Theta}(Z) - l_{\Theta'}(Z) \geq Q(\Theta, \Theta')$ mit $Q(\Theta, \Theta) = 0$.

Setzt man $\Theta' = \Theta_k$ als Schätzwert bei der k-ten Iteration, so kann ein neuer Schätzwert $\Theta_{k+1} = \underset{\Theta}{\operatorname{argmax}} Q(\Theta, \Theta_k)$ gewonnen werden, der zu einer Likelihood

$$l_{\Theta_{k+1}}(Z) \geq l_{\Theta_k}(Z) + Q(\Theta_{k+1}, \Theta_k) \geq l_{\Theta_k}(Z)$$

führt. Damit ist der **EM-Algorithmus** definiert:

$$(5.5) \quad \begin{aligned} \text{E-Schritt:} \quad & Q(\Theta, \Theta_k) = E_{\Theta_k} [l_{\Theta}(Z,Y) - l_{\Theta_k}(Z,Y) | Z] \\ \text{M-Schritt:} \quad & \Theta_{k+1} = \operatorname{argmax}_{\Theta} Q(\Theta, \Theta_k) \end{aligned}$$

Dabei genügt es, den ersten Term von Q zu maximieren (der zweite hängt nicht von Θ ab).

Wählt man dagegen irgendeinen neuen Schätzwert Θ_{k+1} , für den gilt $Q(\Theta_{k+1}, \Theta_k) \geq 0$, so erhält man einen verallgemeinerten EM - einen **GEM-Algorithmus**:

$$(5.5) \quad \begin{aligned} \text{E-Schritt:} \quad & Q(\Theta, \Theta_k) = E_{\Theta_k} [l_{\Theta}(Z,Y) - l_{\Theta_k}(Z,Y) | Z] \\ \text{M-Schritt:} \quad & \Theta_{k+1} \in \{ \Theta \mid Q(\Theta, \Theta_k) \geq 0 \} \end{aligned}$$

Nimmt man beispielsweise die Maximierung von Q mit Hilfe eines Quasi-Newton-Algorithmus vor, so kann es günstig sein, diese Iteration schon nach einem Schritt abubrechen und mit dem E-Schritt fortzufahren.

5.4 Praktische Implementierung des EM-Algorithmus.

Zerlegt man die Likelihood der vollständigen Daten in

$$l_{\Theta}(Z,Y) = l_{\Theta}(Z|Y) + l_{\Theta}(Y),$$

mit
$$l_{\Theta}(Z|Y) = \text{const.} - \frac{N(T+1)}{2} [\log|R| + \text{tr}[R^{-1}S]]$$

(bedingte Likelihood von Z auf Y)

$$\begin{aligned}
 E &= Z-HY-DX \ ; \ S = \frac{1}{N(T+1)} EE' \\
 Z &=[Z_0, \dots, Z_T], \ Y=[Y_0, \dots, Y_T], \ X=[X_0, \dots, X_T] \\
 Z_i &=[z_{1i}, \dots, z_{ni}] \text{ etc.}
 \end{aligned}$$

(5.6)

$$l_{\Theta}(Y) = \text{const.} - \frac{NT}{2} \left[\log|V| + \text{tr}(V^{-1}W) \right] - \frac{N}{2} \left[\log|\Sigma| + \text{tr}[\Sigma^{-1}S_0] \right]$$

mit $W = \frac{1}{NT} UU'$; $U = Y_+ - A^* Y_{\sim} - B^* X_{\sim}$; $S_0 = \frac{1}{N} \sum (y_{n0}^{-\mu})(y_{n0}^{-\mu})'$

$$Y_+ = [Y_1, \dots, Y_T], \ Y_{\sim} = [Y_0, \dots, Y_{T-1}], \ X_{\sim} = [X_0, \dots, X_{T-1}]$$

so ergibt sich die Aufgabe, im E-Schritt die bedingten Erwartungswerte der Momentenmatrizen des Systemmodells

$$M_s = \left\{ \frac{1}{NT} \begin{bmatrix} Y_+ Y_+' & Y_+ Y_{\sim}' & Y_+ X_{\sim}' \\ & Y_{\sim} Y_{\sim}' & Y_{\sim} X_{\sim}' \\ & & X_{\sim} X_{\sim}' \end{bmatrix}, \frac{1}{N} Y_0 Y_0', \frac{1}{N} Y_0 1' \right\}$$

und des Meßmodells

$$M_m = \frac{1}{N(T+1)} \begin{bmatrix} ZZ' & ZY' & ZX' \\ & YY' & YX' \\ & & XX' \end{bmatrix}$$

zu berechnen. Dies kann, wie im nächsten Abschnitt gezeigt, mit Hilfe eines Kalman- Glätters erfolgen. Die zu maximierende Q-Funktion setzt sich also aus

$$\begin{aligned}
 Q(\Theta, \Theta_k) &= E_{\Theta_k} [l_{\Theta}(Z|Y) | Z] + E_{\Theta_k} [l_{\Theta}(Y) | Z] + h(\Theta_k) \\
 &:= Q_m(\Theta, \hat{M}_m) + Q_s(\Theta, \hat{M}_s) + h(\Theta_k)
 \end{aligned}$$

zusammen, wobei Q_m und Q_s aus den Likelihoods (5.6) durch die Ersetzung $M \rightarrow \hat{M} = E_{\Theta} [M|Z]$ (bedingte Momente) entsteht, $h(\Theta_k)$ nicht von Θ abhängt und deshalb gar nicht berechnet werden muß. Die Maximierung von Q_m und Q_s kann mit dem im 4. Kapitel diskutierten Quasi-Newton-Algorithmus erfolgen. Der Gradient $g(\Theta) = \partial/\partial\Theta Q(\Theta, \Theta_k)$ läßt sich dann aus den Score-Funktionen der Likelihoods (5.6) durch Einsetzen der geschätzten Momente gewinnen. Aufgrund der Formel (5.3) von Louis und der Linearität der Likelihood in den Momenten ist dieser Gradient sogar mit dem Score $s_{\Theta}(Z)$ bei $\Theta = \Theta_k$ identisch. Ersetzt man auch in den modifizierten Fisher-Informationsmatrizen (siehe Kap. 4.5) die Momente M durch \hat{M} , erhält man eine Modell-Hesse-Matrix, die positiv semi-

definit ist. Diese ist jedoch, wie (5.3) zeigt, keineswegs mit der Fisher-Info der beobachteten Daten identisch (vgl. Kap. 5.7).

Weiterhin wird angenommen, daß die Matrizen $\Theta^* = [V, A^*, B^*, \Sigma, \mu]$ des Systemmodells von einem Parametervektor Θ , die des Meßmodells $\Phi^* = [R, H, D]$ von einem Vektor Φ abhängen. Damit lassen sich die Maximierungsaufgaben für Q_m und Q_s getrennt durchführen, was zu einem Zuwachs an Effizienz durch kleinere Matrizen führt. Von Nachteil ist allerdings, daß Restriktionen **zwischen** den Parametern des Systems und der Messung nicht implementierbar sind. Bei Bedarf kann jedoch eine gemeinsame Maximierung leicht durchgeführt werden.

Zusammenfassend erhält man also folgenden **EM-Algorithmus** (5.7):

sei $\Psi_k = [\Theta_k, \Phi_k]$ der Schätzwert im k-ten Schritt:

E-Schritt: Pseudolikelihood

$$Q(\Psi, \Psi_k) = E_{\Psi_k} [l_{\Phi}(Z|Y) | Z] + E_{\Psi_k} [l_{\Theta}(Y) | Z] + h(\Psi_k)$$

$$:= Q_m(\Phi, \hat{M}_m) + Q_s(\Theta, \hat{M}_s) + h(\Psi_k)$$

$$\hat{M}_m = E_{\Psi_k} [M_m | Z]; \quad \hat{M}_s = E_{\Psi_k} [M_s | Z]$$

M-Schritt: $\Phi_{k+1} = \underset{\Phi}{\operatorname{argmax}} Q_m(\Phi, \hat{M}_m)$,

$$\Theta_{k+1} = \underset{\Theta}{\operatorname{argmax}} Q_s(\Theta, \hat{M}_s)$$

wobei die Maximierung von Q mit dem Quasi-Newton-Verfahren (Kap.4)

$$\Theta^{i+1} = \Theta^i + F_s(\Theta^i)^{-1} g_s(\Theta^i) ; \quad \Theta^0 = \Theta_k \quad ; \quad i=0, \dots$$

$$\Phi^{i+1} = \Phi^i + F_m(\Phi^i)^{-1} g_m(\Phi^i) ; \quad \Phi^0 = \Phi_k \quad \text{vorgenommen wird, mit}$$

$$g_s(\Theta^i) = \partial Q_s(\Theta, \hat{M}_s) / \partial \Theta |_{\Theta^i}$$

$$g_m(\Phi^i) = \partial Q_m(\Phi, \hat{M}_m) / \partial \Phi |_{\Phi^i}$$

$$F_s = [\partial \Theta^* / \partial \Theta] \begin{bmatrix} NT/2(V^{-1} \otimes V^{-1}) & 0 & 0 & 0 & 0 \\ 0 & V^{-1} \otimes \hat{Y} \hat{Y}^T & V^{-1} \otimes \hat{Y} \hat{X}^T & 0 & 0 \\ 0 & V^{-1} \otimes \hat{X} \hat{Y}^T & V^{-1} \otimes \hat{X} \hat{X}^T & 0 & 0 \\ 0 & 0 & 0 & N/2(\Sigma^{-1} \otimes \Sigma^{-1}) & 0 \\ 0 & 0 & 0 & 0 & N \Sigma^{-1} \end{bmatrix} [\partial \Theta^* / \partial \Theta]^T$$

$$F_m = [\partial\Phi^*/\partial\Phi] \begin{bmatrix} \frac{N(T+1)}{2} (R^{-1} \otimes R^{-1}) & 0 & 0 \\ 0 & R^{-1} \otimes \widehat{Y\dot{Y}}' & R^{-1} \otimes \widehat{Y\dot{X}}' \\ 0 & R^{-1} \otimes \widehat{X\dot{Y}}' & R^{-1} \otimes \widehat{X\dot{X}}' \end{bmatrix} [\partial\Phi^*/\partial\Phi]'$$

Bemerkung: Bricht man die Quasi-Newton-Iteration nach einem Schritt ab, erhält man einen verallgemeinerten EM (GEM)-Algorithmus. Wie sich zeigt, ist dies effektiver als die Ausführung der Maximierung der Pseudo-Likelihood Q.

5.5 Rekursive Berechnung der bedingten Erwartungswerte der Momentenmatrizen

Wie im vorigen Abschnitt gezeigt, werden zur Berechnung der Pseudolikelihood Q bedingte Erwartungswerte der Momentenmatrizen benötigt, d.h. es müssen Ausdrücke der Form $E[y_i|Z]$, $E[y_i y_i' | Z]$ und $E[y_i y_{i+1}' | Z]$ berechnet werden. Da es sich bei y und z um **Gauß**-verteilte Größen handelt, können diese Erwartungswerte mit Hilfe des **Theorems über die Normalkorrelation** (Liptser u. Shirayev 1978, Kap. 13) durch die Regressionsfunktion

$$\hat{y}_i = E[y_i|Z] = E[y_i|z_0, \dots, z_T] = E[y_i] + \text{Cov}(y_i, Z) \text{Cov}(Z, Z)^{-1} [Z - E[Z]],$$

$$(5.8) \quad \begin{aligned} \text{wobei } \text{Cov}(y_i, Z) &= [\text{Cov}(y_i, z_0) H', \dots, \text{Cov}(y_i, z_T) H']', \\ \text{Cov}(z_i, z_j) &= H \text{Cov}(y_i, y_j) H' + R \delta_{ij}; \quad i, j = 0, \dots, T \end{aligned}$$

$$\text{und } \hat{P}_i = \text{Var}[y_i|Z] = E[(y_i - \hat{y}_i)(y_i - \hat{y}_i)' | Z] = \text{Cov}(y_i, y_i) - \text{Cov}(y_i, Z) \text{Cov}(Z, Z)^{-1} \text{Cov}(Z, y_i)$$

berechnet werden, wenn man noch die Relation

$$\text{Var}[y_i|Z] = E[y_i y_i' | Z] - E[y_i|Z] E[y_i|Z]'$$

berücksichtigt.

Der um eine Zeiteinheit verschobene Term $E[y_{i+1} y_i' | Z]$ kann mit Hilfe des Vektors $\Psi_i = [y_{i+1}', y_i']'$ gewonnen werden.

Obige Formeln erfordern jedoch die Auswertung von großen Matrizen und deren Inversion, sodaß eine rekursive Auswertung mit dem **Kalman-Filter-Algorithmus** als der günstigere Weg erscheint.

Bei der Größe $E[y_i|Z]$ handelt es sich um eine **geglättete (interpolierte)** Schätzung für y_i , da auch Daten aus der Zukunft $[z_{i+1}, \dots, z_T]$ benutzt wurden. Kombiniert man einen vorwärts laufenden (0.....i) mit einem rückwärts laufenden (T,.....i+1) Kalman-Filter, so kann $\hat{y}_i = E[y_i|Z]$ als Linearkombination der Filterwerte $y_{\leq} = E[y_i|z_0, \dots, z_i]$ und $y_{>} = E[y_i|z_{i+1}, \dots, z_T]$ dargestellt werden.

Die Gleichungen für das Vorwärts- und Rückwärts-Filter lauten (siehe etwa **Lewis** 1986, Kap. 2.8, **Gelb** 1974, Kap. 5):

Vorwärts-Filter:

System-Modell: $Y_{i+1} = A^* Y_i + B^* X_i + U_i$; $i=0 \dots T-1$; $Y_i = [Y_{i1} \dots Y_{iN}]$
 (5.9) $Z_i = H Y_i + D X_i + \varepsilon_i$; $i=0, \dots, T$

(5.10):

Anfangsbedingung: $Y_0 = E[Y_0 | Z] = \mu_N + \Sigma H' (H \Sigma H' + R)^{-1} (Z_0 - H \mu_N - D X_0)$; $\mu_N = 1_N \otimes \mu$;
 $P_0 = (1/N) \text{Var}[Y_0 | Z_0] = \Sigma - \Sigma H' (H \Sigma H' + R)^{-1} H \Sigma$;

Rekursion $i=0 \dots T-1$:

Zeit-Entwicklung (a priori-Schätzungen):

$$Y_{i+1}^- = A^* Y_i^- + B^* X_i$$

$$P_{i+1}^- = A^* P_i^- A^{*'} + V$$

Messung:

$$K_{i+1} = P_{i+1}^- H' (H P_{i+1}^- H' + R)^{-1} \quad (\text{Kalman-Gain})$$

$$Y_{i+1}^- = Y_{i+1}^- + K_{i+1} [Z_{i+1} - H Y_{i+1}^- - D X_{i+1}]$$

$$P_{i+1} = (I - K_{i+1} H) P_{i+1}^-$$

Likelihood: $-2 \log L = \sum_{i=0}^{T-1} \text{tr}[\Gamma_{i+1}^- v_{i+1} v_{i+1}'] + N \log |\Gamma_{i+1}| + \text{tr}[\Gamma_0^- v_0 v_0'] + N \log |\Gamma_0|$

mit $v_{i+1} = Z_{i+1} - H Y_{i+1}^- - D X_{i+1}$ (Innovation)
 $\Gamma_{i+1} = 1/N \text{Var}(v_{i+1}) = H P_{i+1}^- H' + R$ (Innovations-Kovarianz)
 und $v_0 = Z_0 - H \mu_N - D X_0$; $\Gamma_0 = H \Sigma H' + R$ (Likelihood der Anfangswerte Z_0)

Obiges Filter hat die Besonderheit, daß N Trajektorien simultan behandelt werden. Da jedoch die Formeln für die Kovarianzen gleich sind ($n=1 \dots N$) werden diese nur einmal ausgewertet.

Rückwärtsfilter:

Dazu muß die Systemgleichung (5.9) in eine zeitlich rückwärts laufende Rekursion umgeschrieben werden. Unter der Annahme nichtsingulären A^* s erhält man:

(5.11) $Y_i = A^{*-1} Y_{i+1} - A^{*-1} B^* X_i - A^{*-1} U_i$; $i=0 \dots T-1$

In analoger Weise erhält man dann die Filtergleichungen für $Y_i^b = E[Y_i | Z_{i+1} \dots Z_T]$ und $Y_i^b = E[Y_i | Z_i \dots Z_T]$:

Iteration: $i=T \dots 1$

Messung: $Y_i^b = P_i^b [(P_i^b)^{-1} Y_i^b + H' R^{-1} Z_i]$
 $P_i^b = [(P_i^b)^{-1} + H' R H]^{-1}$

(5.12)

Zeit-Entwicklung: $Y_{i-1}^b = A^{*-1} Y_i^b - A^{*-1} B^* X_{i-1}$
 $P_{i-1}^b = A^{*-1} (P_i^b + V) A^{*-T}$

Die etwas ungewöhnliche Form der Gleichungen (5.12.1, 5.12.2) läßt sich mit Hilfe des Matrix-Inversions-Lemmas aus der üblichen Form ableiten (siehe hierzu **Lewis** 1986, S. 53).

Die geglättete Trajektorie kann nun als gewichtete Linearkombination der Filterlösungen Y_i und Y_i^{b-} geschrieben werden:

$$(5.13.1) \quad \hat{Y}_i = \hat{P}_i^{-1} [P_i^{-1} Y_i + (P_i^{b-})^{-1} Y_i^{b-}]^{-1}$$

wobei der Glättfehler durch

$$(5.13.2) \quad \hat{P}_i = [P_i^{-1} + (P_i^{b-})^{-1}]^{-1}$$

gegeben ist.

Aus der Bedingung $\hat{P}_T = P_T$ läßt sich ersehen, daß das Rückwärts-Filter mit $(P_i^{b-})^{-1} = 0$ initialisiert werden muß. Entsprechend ist der Anfangswert Y_T^{b-} beliebig (eine Formulierung mit den "Informationsmatrizen" $S_i = (P_i^b)^{-1}$ und $S_i^- = (P_i^{b-})^{-1}$ und transformierten Zuständen $\eta_i = S_i Y_i$ und $\eta_i^- = S_i^- Y_i^{b-}$ ist ebenfalls möglich).

Anstelle der Benutzung von 2 Filtern kann im Anschluß an die Vorwärts-Filterung eine Rekursion angeschlossen werden, die nur die gefilterten a priori und a posteriori-Größen verwendet (**Rauch-Tung-Striebel-Glätter, R/T/S** 1965). Dazu werden die Variablen des Rückwärtsfilters eliminiert (für eine ausführliche Herleitung siehe **Lewis** 1986, Kap. 2.8). Beispielsweise kann der Glättfehler

$$\hat{P}_i = [P_i^{-1} + (P_i^{b-})^{-1}]^{-1} = P_i - P_i (P_i + P_i^{b-})^{-1} P_i$$

mit Hilfe der Formeln $P_i^{b-} = A^{*-1} (P_{i+1}^b + V) A^{*-T}$ (5.12.4) und $P_{i+1}^b = [P_{i+1}^{-1} - (P_{i+1}^{f-})^{-1}]^{-1}$ in eine Rekursion für \hat{P}_i , P_i und P_i^- umgeschrieben werden:

$$(5.14.1) \quad \hat{P}_i = P_i + F_i [\hat{P}_{i+1}^- - P_{i+1}^-] F_i' ; \quad i=T-1, \dots, 0 ; \quad \hat{P}_T = P_T$$

wobei der Glätter-Gain als $F_i = P_i A^{*'} (P_{i+1}^-)^{-1}$ gegeben ist.

Analog folgt für \hat{Y}_i :

$$(5.14.2) \quad \hat{Y}_i = Y_i + F_i [\hat{Y}_{i+1}^- - Y_{i+1}^-] ; \quad \hat{Y}_T = Y_T$$

Interessant ist, daß in der RTS-Formulierung die ursprünglichen Daten Z und die Exogenen X bei der Glättungs-Iteration nicht mehr vorkommen.

Mit Hilfe der geglätteten Daten \hat{Y}_i und dem Glättfehler \hat{P}_i können vermöge der Relation

$$\hat{Y}_i \hat{Y}_i' = N \hat{P}_i + \hat{Y}_i \hat{Y}_i'$$

(der Faktor N kommt durch die spezielle Form des Vorwärts-Filters (5.12) zustande) die Momentenmatrizen mit Termen ohne Zeitverschiebung berechnet werden.

Zur Behandlung des Ausdrucks $E[Y_{i+1} Y_i' | Z]$ betrachtet man einen erweiterten

Zustand der Form $\Psi_i = [Y_{i+1}, Y_i]'$. Die System-Gleichung lautet dann:

$$(5.15) \quad \Psi_{i+1} = \begin{bmatrix} A^* & 0 \\ I & 0 \end{bmatrix} \Psi_i + \begin{bmatrix} B^* \\ 0 \end{bmatrix} X_{i+1} + \begin{bmatrix} U_{i+1} \\ 0 \end{bmatrix} \quad ; i=0, \dots, T-1$$

$$Z_i = [0, H] \Psi_i + DX_i + \varepsilon_i \quad ; i=0, \dots, T$$

Bezeichnet man den Glätt-Fehler des erweiterten Modells (5.15) als $\hat{\Pi}_i = \text{Var}[\Psi_i | Z]$, so gilt:

$$E[Y_{i+1} Y_i | Z] = \hat{\Pi}_i[1,2] + \hat{Y}_{i+1} \hat{Y}_i' \quad ; \quad \hat{\Pi}_i[1,2] = \text{Cov}[Y_{i+1}, Y_i | Z]$$

Es genügt also, den RTS-Glätter für das erweiterte Modell nur bzgl. des Glättfehlers zu implementieren.

5.6. Newton-Raphson-Algorithmus

Da, im Gegensatz zum **Newton**-Verfahren, die Konvergenz des EM-Algorithmus nur linear ist, ist es oft vorteilhaft, in der Nähe des Maximums der Likelihood-Funktion einen schnell konvergenten Algorithmus zur Verfügung zu haben (Nachiteration, vgl. **Watson u. Engle** 1983). Wie sich zeigt (vgl. Kap. 5.7, 5.8), stellt eine Kombination aus EM und **Newton-Raphson** (zuverlässiges Auffinden einer approximativen Nullstelle + schnelle Konvergenz in deren Nähe) eine brauchbare Lösung des Schätzproblems dar.

Wie von **Louis** (1981) gezeigt, gilt für den Score und die gemessene **Fisher**-Informationsmatrix (Glg. 5.3. 5.3')

$$s_\psi(Z) = \partial / \partial \psi \log L(Z) = E_\psi[s_\psi(Z, Y) | Z]$$

$$\text{und} \quad I_\psi(Z) = E[-\partial^2 / \partial \psi^2 \log L(Z)] = E[s_\psi(Z, Y) s_\psi'(Z, Y) | Z] + s_\psi(Z) s_\psi'(Z)$$

Während sich vermöge der Linearität von $S_\psi(Z, Y)$ der Score $s_\psi(Z)$ durch die Ersetzung der Momentenmatrizen M durch bedingte Momente berechnen läßt (vgl. Kap. 5.4), ist die Formel für die Fisher-Information doch recht kompliziert. Stattdessen werden numerische Ableitungen benutzt, die als Vorwärts-Differenzenquotienten gebildet wurden. Die **Hesse**-Matrix ist also näherungsweise durch

$$H = [H_1, \dots, H_u] \quad \text{mit} \quad H_i = 1/h(s_{\psi_i} - s_{\psi'}) \quad \text{und} \quad \Psi_i = [\psi_1, \dots, \psi_i + h, \dots, \psi_u]'$$

gegeben, wobei h in der Größenordnung der Wurzel der Maschinengenauigkeit gewählt werden sollte (vgl. hierzu **Dennis** u. **Schnabel** 1983). Im vorliegenden Programmpaket wurde ein Wert von $h=10^{-8}$ benutzt. Der **Newton-Raphson**-Algorithmus ist analog zu den in Kap. 4.4 diskutierten **Quasi-Newton**-Verfahren implementiert.

Da der Score $s_\psi(Z)$ und der Gradient der Pseudolikelihood $Q(\psi, \psi_k)$ bei $\psi=\psi_k$ übereinstimmen, kann der GEM-Algorithmus (5.7), bei dem die Maximierung von Q nach einem Schritt abgebrochen wird, als **Quasi-Newton**-Verfahren mit **Modell-Hesse**-Matrix

$$H = - \begin{bmatrix} F_s & 0 \\ 0 & F_m \end{bmatrix}$$

angesehen werden.

5.7. IML-Module zur Schätzung des Zustandsraum-Modells

Die oben entwickelte Methode wurde wieder im Rahmen des Programmpakets LSD in Software-Prozeduren umgesetzt:

PAR und **PARD** (Spezifikation des System-Modells, siehe Kap. 4.8)

PARM (R, H, D, PHI)

Zweck: Modellspezifikation (Meßmodell). R, H, D als Funktionen des Parametervektors PHI. **Muß vom Benutzer bereitgestellt werden.**

Eingabeparameter:

PHI: $u_2 \times 1$ -Parametervektor

Ausgabe-Parameter:

R: $k \times k$; Kovarianzmatrix des Meßfehlers

H: $k \times p$; Output-Matrix

D: $k \times q$; Input-Matrix (Meßmodell)

PARMD (DR, DH, DD, PHI)

Zweck: Modellspezifikation (Meßmodell). **McDonald-Swaminathan**-Ableitungen DR, DH, DD als Funktionen des Parametervektors PHI. **Muß vom Benutzer bereitgestellt werden.**

Eingabeparameter:

PHI: $u_2 \times 1$ -Parametervektor

Ausgabe-Parameter:

DR: $u_2 \times k^2$; Ableitungen der Kovarianzmatrix des Meßfehlers

DH: $u_2 \times kp$; Ableitungen der Output-Matrix

DD: $u_2 \times kq$; Ableitungen der Input-Matrix (Meßmodell)

EM (THETA, PHI, Z, X, K, P, Q, P, N, T, DT, THETA0, PHI0, IMAX, KMAX, EPS)

Zweck: EM und GEM-Algorithmus zur Schätzung des State-Space-Modells (5.1).

Eingabeparameter:

Z: $k \times N(T+1)$ -Datenmatrix: $Z=[Z_0, \dots, Z_T]$

X: $K \times N(T+1)$ -Datenmatrix (Exogene Größen): $X=[X_0, \dots, X_T]$

K: Dimension der gemessenen Größen ($z_{nt}: k \times 1$)

P: Dimension des Systemzustands ($y_{nt}: p \times 1$)

Q: Dimension der Exogenen ($x_{nt}: q \times 1$)

N: Anzahl der statistischen Einheiten

T: T+1 ist die Anzahl der Meßzeitpunkte

DT: Abstand der Meßzeitpunkte

THETA0: $u_1 \times 1$; Startwert für EM-Iteration (Parameter des Systemmodells)

PHI0: $u_2 \times 1$; Startwert für EM-Iteration (Parameter des Meßmodells)

IMAX: Maximalzahl der EM-Iterationen

KMAX: Maximalzahl der Scoring-Iterationen (M-Schritt). $KMAX < K_0$ entspricht einem GEM-Algorithmus, wobei K_0 =Zahl der Iterationen bis zum Maximum von Q.

EPS: Abbruchkriterium: Wenn $\max \text{abs}(\Theta_{i+1} - \Theta_i) < \epsilon$ und $\max \text{abs}(\Phi_{i+1} - \Phi_i) < \epsilon$, Abbruch der EM-Iteration.

Ausgabe-Parameter:

THETA: ML-Schätzer $\hat{\Theta}$

PHI: ML-Schätzer $\hat{\Phi}$

NEWGRAPH(PSI, COV, PSI0, KMAX, EPS, Z, X, K, P, Q, U1, N, T, DT, OPTION)

Eingabe-Parameter:

PSI0: $u \times 1$; Startwert der Newton-Raphson-Iteration

KMAX: Maximale Zahl der Iterationen.

EPS: Abbruchkriterium (siehe NEWTON, Kap. 4.8)

Z, X, K, P : siehe EM.

U1: Dimension von Θ : $u_1 \times 1$

N, T, DT: siehe EM.

OPTION: Berechnung der Eigenwerte der Hesse-Matrix in jedem Iterationsschritt

und Transformation auf negative Definitheit (Kap. 4.4)

Ausgabe-Parameter:

PSI: $u \times 1$; ML-Schätzwert

COV: $u \times u$; geschätzte asymptotische Kovarianzmatrix ($=F^{-1}$)

STATESPA (Z, Y, H, D, R, A, B, G, Y0, X, P, Q, N, T, DT, SEED, PRINT).

Zweck: Simulation eines Datensatzes Z, der N Trajektorien an den Zeitpunkten t_0, \dots, t_T umfaßt (State-Space-Modell 5.1)

Eingabe-Parameter:

H: $k \times p$; Output-Matrix

D: $k \times q$; Input-Matrix des Meßmodells

R: $k \times k$; Kovarianz des Meßfehlers

A: $p \times p$; Drift-Matrix

B: $k \times q$; Input-Matrix des System-Modells

G: $p \times r$; Diffusions-Koeffizient

Y0: $p \times N(T+1)$; Anfangsbedingung

P, Q, N, T: siehe oben

SEED: Seed des Zufallsgenerators RANNOR (siehe SAS User's Guide: Basics (1985), Kap. 6).

PRINT: Druck-Parameter (=0: wenig Output)

Ausgabe-Parameter:

Z: $k \times N(T+1)$; Simulierte beobachtete Daten.

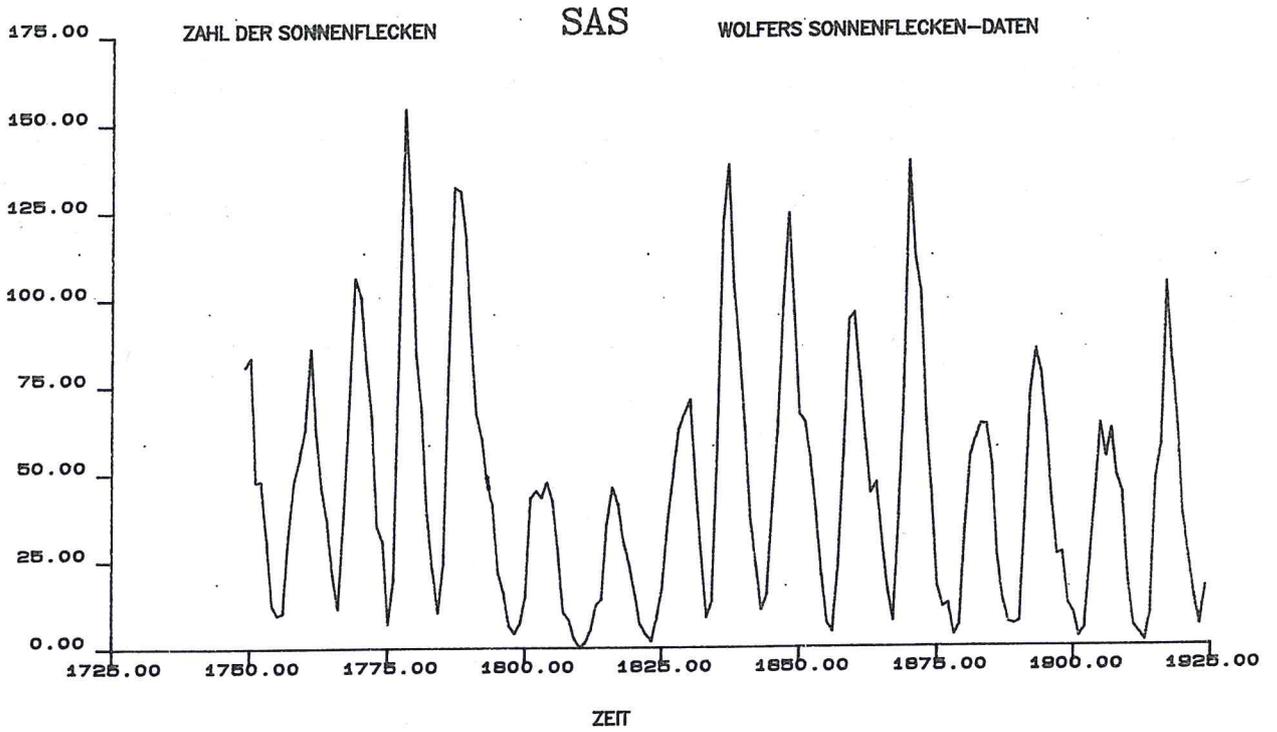
Y: $k \times N(T+1)$; Simulierte unbeobachtete Daten

5.8 Beispiel: Die Wolferschen Sonnenflecken-Daten

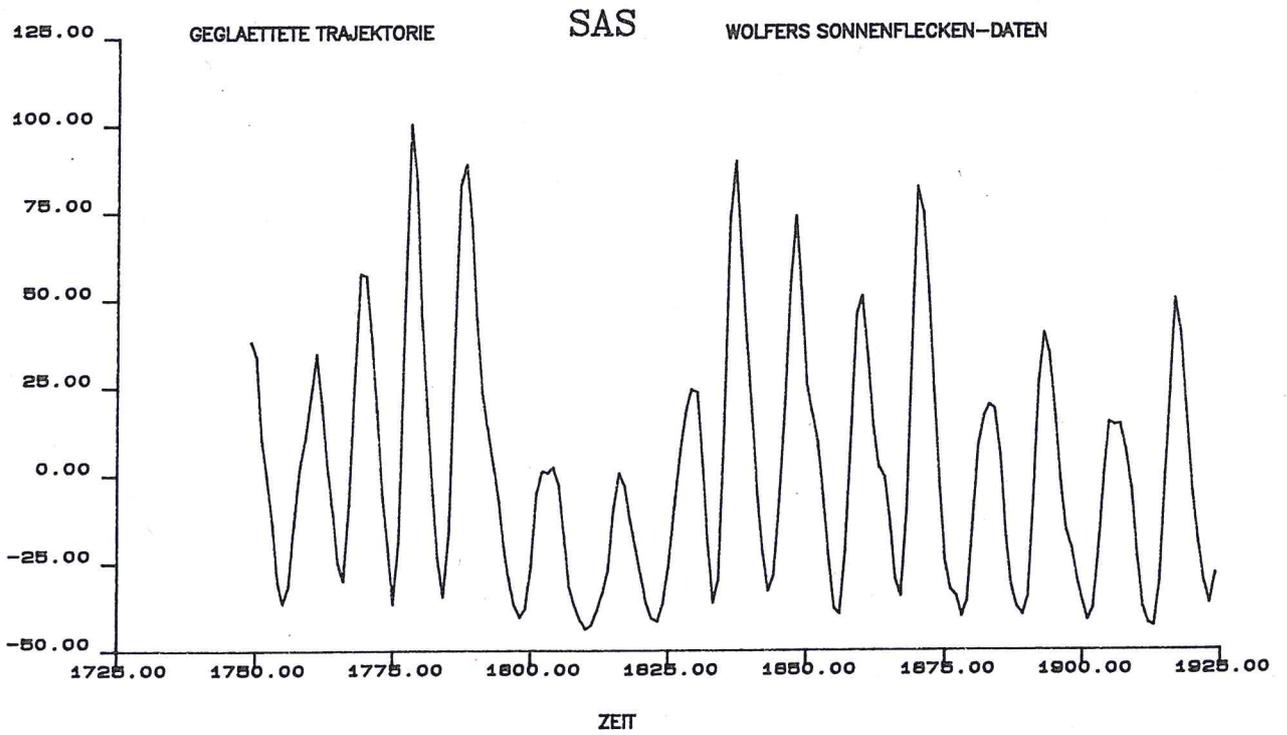
Als Beispiel für die eben entwickelte Methode soll eine Zeitreihe analysiert werden, die die jährliche Anzahl von Sonnenflecken vom Jahr 1749 bis 1924 umfaßt (Quelle T.W. Anderson 1971). Die Sonnenfleckendaten wurden von verschiedenen bekannten Autoren (Yule, Bartlett, Anderson, Box u. Jenkins etc.) untersucht, jedoch vorwiegend mit zeitdiskreten Modellen. Der Verlauf der Daten ist in Abb. 5.1 aufgezeichnet. Der Kurvenverlauf ähnelt einer Sinuskurve mit einer Periode von ca. 10 Jahren, die um einen konstanten Wert von etwa 50 schwankt. In der Tat wurden in der Literatur vorwiegend diskrete AR(2)-Modelle benutzt, während die Koeffizienten höherer Modelle nicht signifikant sind (vgl. hierzu T.W. Anderson 1971).

5.8.1. Modellspezifikation

Im folgenden wird davon ausgegangen, daß es sich um einen harmonischen Os-

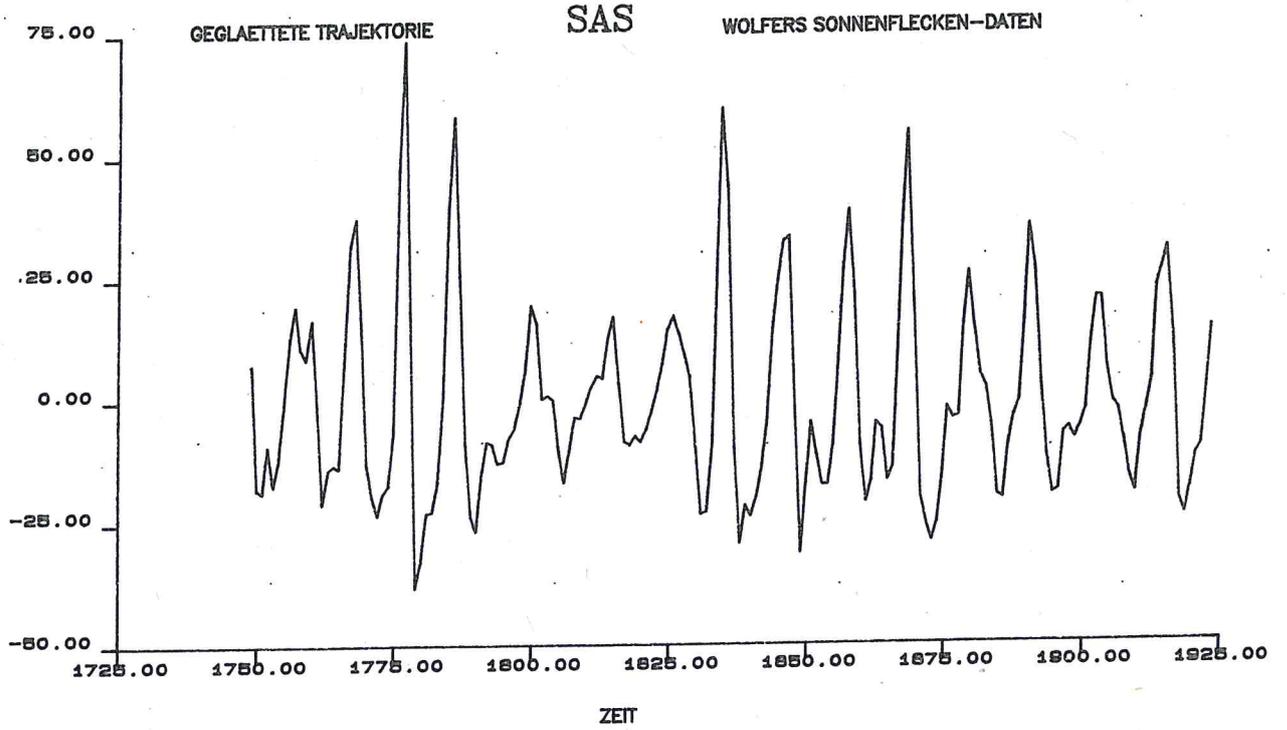


a) Trajektorie von 1749-1924

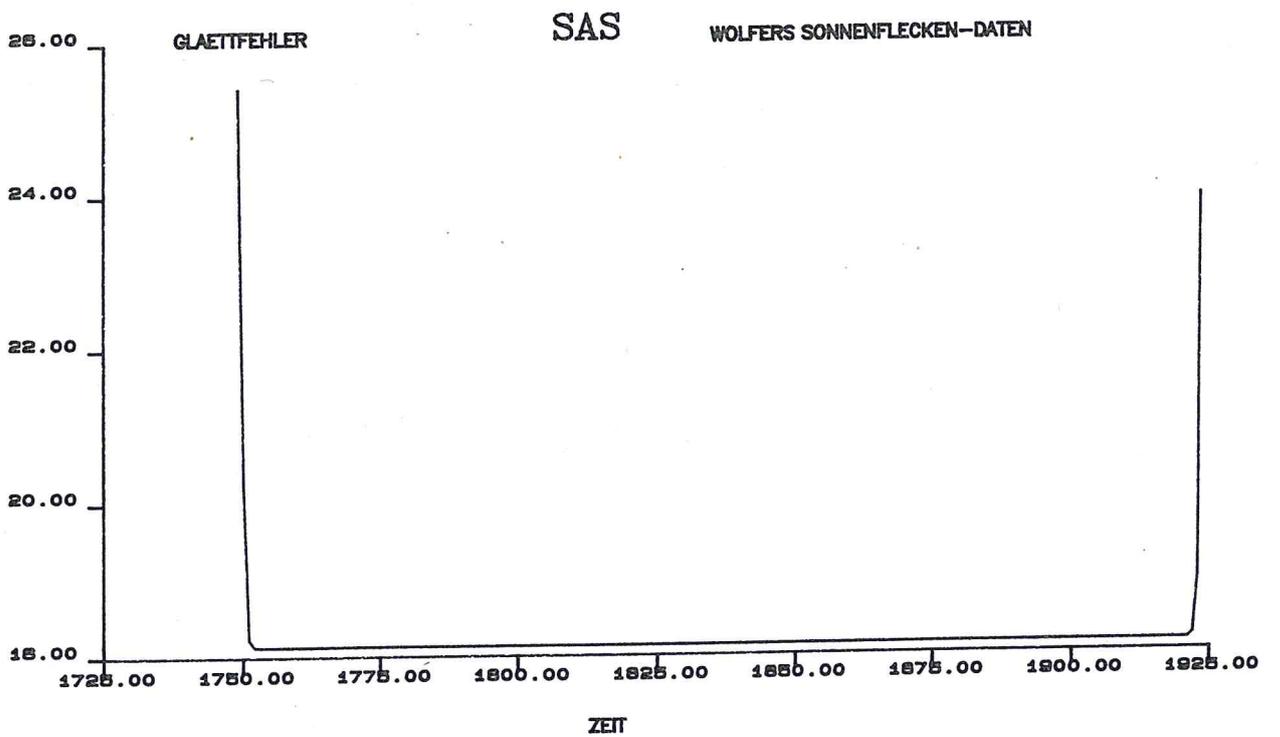


b) geglättete Trajektorie $\hat{y}_1(t) = E[y_1(t)|Z]$ (Modell II)

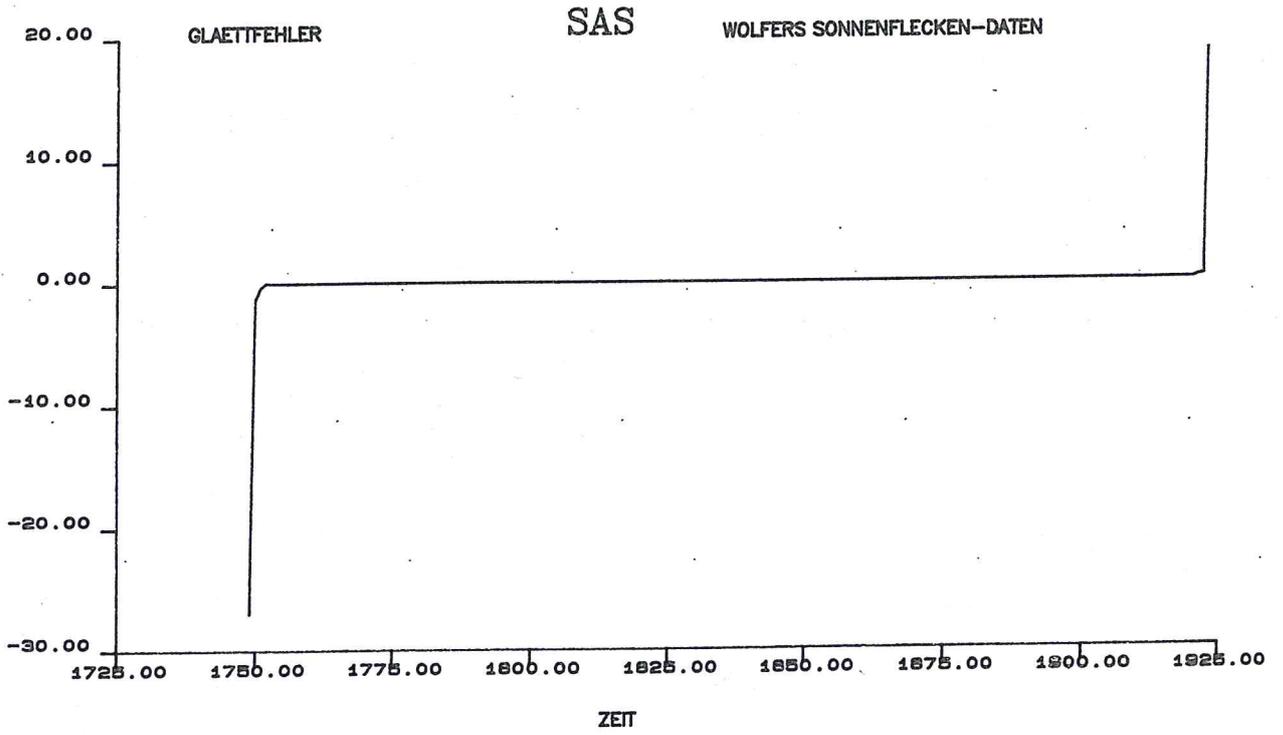
Abb. 5.1. Wolfersche Sonnenflecken-Daten.



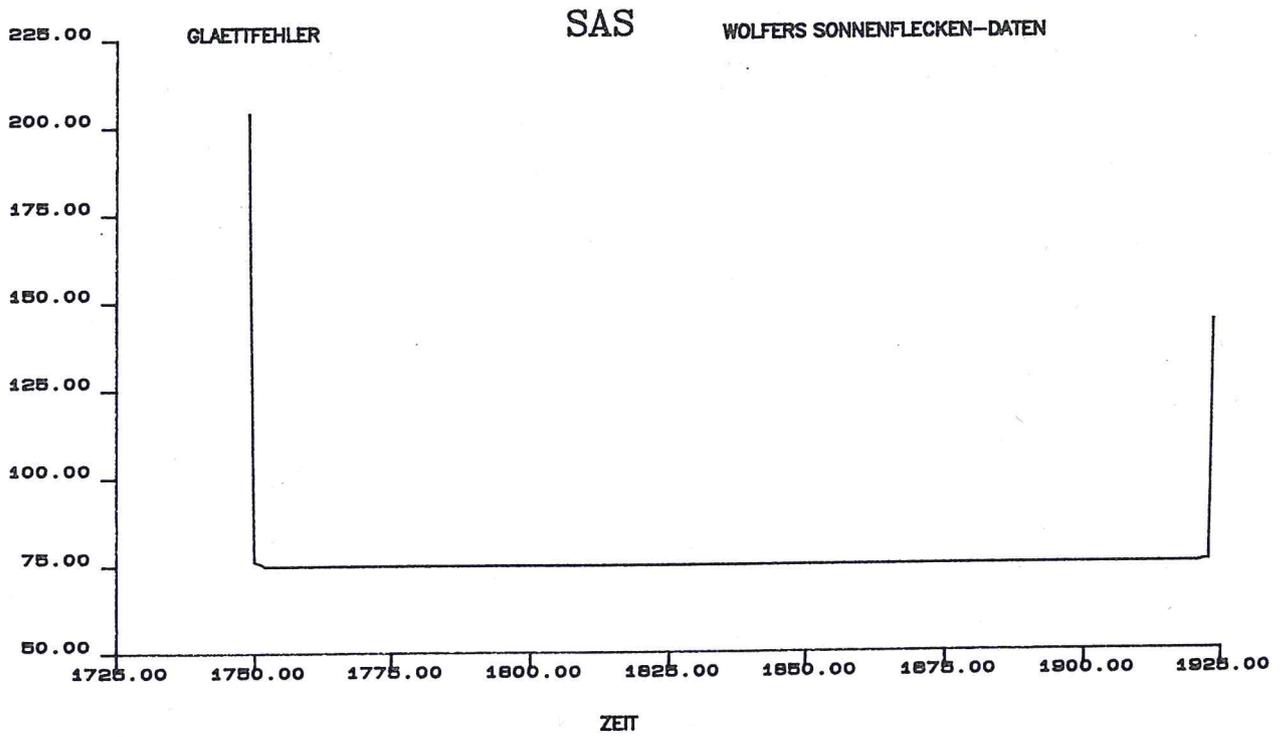
c) geglättete Trajektorie $\hat{y}_2 = E[y_2|Z]$ (Modell II)



d) Glättfehler $\text{Var}(y_1|Z)$ (Modell II)



e) Glättfehler $\text{Cov}(y_1, y_2 | Z)$ (Modell II)



f) Glättfehler $\text{Var}(y_2 | Z)$ (Modell II)

zillator (AR(2)-Prozeß) handelt, der um eine Ruhelage D schwingt. Diese Annahme ist rein phänomenologisch, da keine physikalische Theorie der Sonnenflecken zugrundegelegt wurde. Abweichungen von der idealen Sinusform sind durch stochastische Fluktuationen verursacht, die aber genauso gut als Approximation einer nichtlinearen Systemdynamik angesehen werden können (vgl. Kap. 1). Weiterhin ist davon auszugehen, daß die Daten Meßfehler enthalten. Man erhält dann folgendes Modell

$$\begin{aligned} \text{Systemmodell:} & \quad \ddot{y} + \gamma \dot{y} + \omega_0^2 y = g \zeta_t \\ \text{Meßmodell:} & \quad z_i = y_i + D + \varepsilon_i \quad ; \quad i=0, \dots, T \quad ; \quad y_i = y(i\Delta t) \quad \text{mit} \quad \Delta t = 1 \end{aligned}$$

Umschreiben in eine Ito-Differentialgleichung ergibt:

$$\begin{aligned} d \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} dt + \begin{bmatrix} 0 & 0 \\ 0 & g \end{bmatrix} \begin{bmatrix} dw_1(t) \\ dw_2(t) \end{bmatrix} \\ z_i &= [1 \ 0] \begin{bmatrix} y_1(t_i) \\ y_2(t_i) \end{bmatrix} + D x_i + \varepsilon_i \quad ; \quad x_i = 1; \quad i=0, \dots, T \end{aligned}$$

Die deterministische Anfangsbedingung $y(0)$ kann durch die Wahl $\mu=0$, $\Sigma \rightarrow \infty$ modelliert werden (Gleichverteilung).

Damit verbleiben als freie Parameter $\Theta = \{-\omega_0^2, -\gamma, g\}$, $\Phi = \{D\}$ (**Modell I** - ohne Meßfehler) bzw. $\Phi = \{D, R\}$ (**Modell II** - mit Meßfehler). Häufig wird vor der Analyse der Daten der konstante Mittelwert abgezogen (d.h. $\tilde{z}_i = z_i - \bar{z}$), jedoch erscheint es im Rahmen einer systemtheoretischen Betrachtungsweise logischer, D mit den anderen Parametern simultan zu schätzen.

5.8.2. Identifikation

In Kap. 4.9.1 wurde gezeigt, daß die Parameter des Oszillator-Modells aus denen des diskreten Modells, d.h. A^* , B^* und V erschlossen werden können. Im vorliegenden Fall ist nur die erste Komponente von y einer Messung zugänglich, sodaß sich neue Probleme ergeben.

Man erhält folgende Identifikationsgleichungen:

$$v_i = E[z_i] = HE[y_i] + D, \quad \text{mit} \quad y_i = A^{*i} y_0 + \sum_{l=0}^{i-1} A^{*i-l-1} B^* u_l$$

$$\begin{aligned} \text{d.h.} \quad v_i &= HA^{*i} \mu + D = D, \quad \text{d.h. } D \text{ ist identifiziert.} \\ \Gamma_{ij} &= HCov(y_i, y_j)H' + R\delta_{ij}; \quad Cov(y_i, y_j) = A^{*i} \Sigma A^{*j} + \sum_{l=0}^{i \wedge j - 1} A^{*i-l-1} V A^{*j-l-1} \end{aligned}$$

Aus $\Gamma_{00} = H \Sigma H' + R$ ist also R identifiziert.

Setzt man $\Sigma = \varepsilon I_2$ (ε groß wegen Gleichverteilung), so hat man weiterhin:

$$\Gamma_{i0} = HA^{*i} \Sigma H' = \varepsilon e^{\gamma/2 t_i} [\gamma/2 \sin \omega t_i + \cos \omega t_i] \quad (\text{siehe Kap. 4.9.1}).$$

Damit sind T nichtlineare Gleichungen für γ und $\omega = \sqrt{\omega_0^2 - \gamma^2/4}$ gegeben.

Gleichungen für $V = \int \exp(As) G G' \exp(A's) ds$ sind aus den anderen Γ_{ij} zu entnehmen.

men. Leider sind die Verhältnisse sehr unübersichtlich, schon in diesem relativ einfachen Modell, sodaß der pragmatische Weg eingeschlagen wurde, die Identifizierbarkeit über die Fisher-Informations-Matrix zu beurteilen [für das zeitdiskrete AR(p)-Modell werden in **Caines** (1988), Kap. 9 hinreichende Bedingungen angegeben. Die einfache Struktur der zeitdiskreten Übergangsmatrix ist jedoch hier nicht gegeben - einfache Restriktionen auf A spiegeln sich **nicht** in A^* wider!].

5.8.3. Maximum-Likelihood-Schätzung

Modell I (ohne Meßfehler):

Um zu vermeiden, daß die Likelihood $L(Z|Y)$ des Meßmodells singulär ist ($L = \prod \delta(z_i - Hy_i - Dx_i)$ für $R \rightarrow 0$), wurde R auf den kleinen, fixen Wert $R=10^{-2}$ gesetzt. Analysiert wurden zwei Datensätze (A: 1770-1889 und B: 1749-1924; der zweite liegt den Berechnungen von **Yule** (1927) zugrunde). Ausgehend von einem Startwert von $\Theta_0 = \{-1, -1, 2\}$ und $\Phi_0 = \{46.0\}$ findet der EM-Algorithmus nach 30 Iterationen einen Punkt in der Nähe des Maximums der Likelihood. Da die Konvergenz dort sehr langsam stattfindet, wurde eine Newton-Raphson-Nachiteration angeschlossen, die in 4 bzw. 6 Schritten zum ML-Schätzwert konvergiert (Tab. 5.2). Die gesamte Rechenzeit (pro Datensatz) beläuft sich auf ca. 135 Sekunden (BASF 7/73 unter VM/CMS).

Modell II (mit Meßfehler):

Schon in den frühen Arbeiten von Yule (1927) wurde die Möglichkeit diskutiert, daß in den Daten Beobachtungsfehler enthalten sind, was zur Benutzung einer geglätteten Trajektorie (graduated series) führte. Im folgenden wird deshalb ein Meßfehler ε mit Kovarianz R zugelassen. Ausgehend von den Startwerten $\Theta_0 = \{-1, -1, 2\}$ und $\Phi_0 = \{46, 1\}$ konvergiert der kombinierte EM/Newton-Raphson-Algorithmus nach 30+3 (Datensatz A) bzw. 30+5 Iterationen (Datensatz B) zum ML-Schätzwert (Tab. 5.2).

		$-\omega_0^2$	$-\gamma$	g	D	R
I	1770-1889	- .4578	- .6540	28.6962	45.8880	-
		(.0722)	(.1469)	(2.6349)	(5.7562)	
	1749-1924	- .5064	- .8000	30.9014	44.1756	-
		(.0691)	(.1454)	(2.5292)	(4.7823)	
II	1770-1889	- .4153	- .4667	22.9461	46.0587	13.5181
		(.0617)	(.1397)	(3.3120)	(5.0901)	(8.2712)
	1749-1924	- .4008	- .3782	18.7669	44.5706	27.0055
		(.0466)	(.1032)	(2.4397)	(3.5703)	(8.0596)

Tabelle 5.2. ML-Schätzwerte für die **Wolferschen** Sonnenfleck-Daten mit (II) und ohne Meßfehler (I). (in Klammern: geschätzte Standardabweichungen).

Vergleicht man Modell I und II, so fällt auf, daß die Einführung eines Meßfehlers zu einer Verringerung des Prozeßfehleranteils und der Dämpfung γ führt. Eine Betrachtung der Formeln für die stationäre Kovarianz (siehe Kap. 4.9.1)

$$\Sigma_s = g^2/2\gamma \begin{bmatrix} 1/\omega_0^2 & 0 \\ 0 & 1 \end{bmatrix}$$

zeigt, daß - konstantes Σ_s und ω_0^2 vorausgesetzt- γ und g^2 proportional sind. Die Unterschiede zwischen Datensatz A und B werden aus einer Betrachtung des Kurvenverlaufs in den Jahren 1749-1770 und 1889-1924 verständlich: die Schwankung der Sonnenflecken ist in diesen Jahren geringer, sodaß die stationäre Kovarianz für Datensatz B geringer ausfallen sollte. Man erhält in der Tat:

$$1770-1889: \Sigma_s^{1/2} = \begin{bmatrix} 37.08 & 0 \\ 0 & 25.09 \end{bmatrix} ; 1749-1924: \Sigma_s^{1/2} = \begin{bmatrix} 34.33 & 0 \\ 0 & 24.43 \end{bmatrix}$$

Von Interesse ist weiterhin die geschätzte Periodendauer $T=2\pi/\omega=2\pi/(\omega_0^2-\gamma^2/4)^{1/2}$ der Oszillationen. Man erhält:

Autor	ω_0^2	γ	a	b	ω	T
Bartlett	.3631	.3186	-1.4255	.7272	.5812	10.8111
Yule (R)	.3961	.4231	-1.3425	.6550	.5928	10.5998
Yule (G)	.3289	.2200	-1.5153	.8025	.5628	11.1639
IA	.4578	.6540	-1.1965	.5200	.5923	10.6073
IIA	.4153	.4667	-1.3065	.6271	.6007	10.4597
IB	.5064	.8000	-1.1151	.4493	.5886	10.6756
IIB	.4088	.3782	-1.3623	.6851	.6042	10.3994

Tab. 5.3: Schätzwerte verschiedener Autoren für die Sonnenflecken-Daten (Zeitraum 1749-1924; 1770-1889 für IA und IB). Die Schätzungen von **Yule(G)** und **Bartlett** beruhen auf geglätteten Daten (graduierte Reihe).

Zu beachten ist, daß die Resultate von **Yule** (1927) mit einem zeitdiskreten AR(2)-Prozeß der Form

$$y_{i+2} + ay_{i+1} + by_i = u_i$$

gewonnen wurden. Eine Umrechnung der Parameter in das kontinuierliche Modell

$$\ddot{y} + \gamma\dot{y} + \omega_0^2 y = g\zeta$$

ist näherungsweise möglich, wenn man die Lösungsansätze $y(t) \sim \exp(\lambda t)$ und $y_i \sim \exp(\lambda i \Delta t)$ einsetzt (Störterme weglassen) und dann die Koeffizienten vergleicht. Man findet (vgl. **Bartlett** 1955, S.315):

$$a = -2\exp(\gamma\Delta t/2) \cos(\omega\Delta t) ; b = \exp(-\gamma)$$

Umschreiben der Differentialgleichung auf eine Differenzgleichung führt auf

$$y_{i+2} + ay_{i+1} + by_i = v_i$$

wobei es sich jedoch bei v_i um einen Moving-average-Prozeß handelt (v_i ist mit v_{i+1} korreliert, während u_i in Yules Gleichung weißes Rauschen darstellt (hierzu ausführlich **Bartlett** (1946)). Die Schätzwerte von **Bartlett** beruhen auf den Korrelationen, die mit Yules gradierter (geglätteter) Reihe gewonnen wurden (**Bartlett** 1955, S. 315).

Auffällig ist, daß trotz größerer Unterschiede in den Schätzungen für die autoregressiven Parameter die Resultate für die Periodendauer T wenig schwanken. Die zeitstetigen Resultate von **Bartlett** sind ähnlich zu Modell IIB (mit Meßfehler).

In Anbetracht der Unterschiede in den Schätzungen zwischen Modell I und II ist zu fragen, welches der beiden (bzw. weitere konkurrierende Modelle, z.B. mit höheren Ableitungen, Trends etc.) das bessere ist. Zur Wahl der Modell-Ordnung (**Modell-Identifikation**) kann das Konzept der **Kompensierten Likelihood** herangezogen werden (vgl. **Azencott u. Dacunha-Castelle** 1986), bei dem die Anzahl u der Parameter und deren quantitative Ausprägung simultan als Maximum

$$[\hat{u}, \hat{\Theta}_u] = \operatorname{argmax}_{u, \Theta_u \in \mathbb{R}^u} \lambda(u, \Theta_u, Z)$$

der kompensierten log-Likelihood $\lambda = l(\Theta_u, Z) - \delta(u, M)$ berechnet werden, wobei δ als **Kompensator** bezeichnet wird (M=Anzahl der Beobachtungen). Anschaulich gesprochen muß der Zuwachs an Likelihood bei größerer Parameterzahl (besserer Fit) kompensiert werden, um eine Überparametrisierung und entsprechende große simultane Konfidenzintervalle zu vermeiden. Am bekanntesten ist hierbei das **Akaike-Informationskriterium** (AIC), bei dem als Kompensator die Anzahl der Parameter gewählt wird. Es führt jedoch zu einer Überparametrisierung, die mit endlicher Wahrscheinlichkeit eintritt (d.h. \hat{u} ist nicht konsistent für $M \rightarrow \infty$). Im folgenden werden 4 Kriterien (siehe **Azencott u. Dacunha-Castelle** 1986, **Koehler u. Murphree** 1988) ausgewertet:

	Kompensator
Akaike	u
Schwarz	u/2 log M
Azencott u. Dacunha-Castelle	1/2 u(1+u)(1+ε)loglog M (ε klein, hier=10 ⁻³)
Rissanen	u log M

Für Modell I und II ergibt sich:

	I	II	I	II
Lik	-394.5	-393.5 *	-578.6	-571.8 *
AIC	-398.5	-398.0 *	-582.6	-576.8 *
SIC	-404.1 *	-405.0	-588.9	-584.7 *
AD	-410.2 *	-416.5	-595.1 *	-596.5
R	-413.7 *	-417.0	-599.3	-597.7 *
u	4	5	4	5
M	120	120	175	175
Zeitraum	1770-1889	1770-1889	1749-1924	1749-1924

Im Zeitraum 1770-1889 favorisieren die Kriterien SIC, AD und R das Modell ohne Meßfehler, während für den größeren Datensatz mit 175 Beobachtungen eher Modell II bevorzugt wird.

6. Abschließende Bemerkungen

Hauptziel der vorliegenden Arbeit war es, die Maximum-Likelihood-Parameterschätzung zeitstetiger dynamischer Systeme im Rahmen kontinuierlicher und diskreter Messungen durchzuführen.

Die Methodik, implementiert im **Softwarepaket LSD**, ist nun auch für mathematisch weniger interessierte Anwender eine reale Alternative zu Programmen wie LISREL geworden, die nur eine statische oder approximativ dynamische Modellierung (Trapez-Methode) zulassen. Zusätzlich zur Parameterschätzung (exaktes diskretes Modell, diskret/kontinuierliches State-Space-Modell) werden Module zur Simulation und graphischen Darstellung der untersuchten Modelle angeboten. Die verschiedenen Möglichkeiten werden in einer speziellen Programmbeschreibung genauer diskutiert.

Aufgrund der Komplexität des Modells sind eine Vielzahl von Erweiterungen möglich. Dies betrifft einerseits die Möglichkeit, unterschiedliche Zeitintervalle Δt_i zuzulassen. In diesem Fall müssen die Formeln auf zeitabhängige Parameter wie z.B. $A_i^* = \exp(A\Delta t_i)$ etc. erweitert werden. Man erkennt jedoch wieder, daß die Parameter des zeitstetigen Modells davon unberührt bleiben.

Eng damit verknüpft ist das Problem von fehlenden Daten (missing data), die mit Hilfe des in Kap. 5 diskutierten EM-Algorithmus aus den erhältlichen Daten geschätzt werden können. Dazu ist es jedoch erforderlich, im Meßmodell (5.1.1) eine zeitabhängige Output-Matrix H_i einzuführen.

Die Approximation der exogenen Variablen stellt einen wichtigen Punkt dar. Wenn diese nicht kontinuierlich erhältlich sind und keine weiteren Annahmen über die funktionale Form vorliegen, enthält das exakte diskrete Modell (trotz der Bezeichnung!) immer einen Spezifikationsfehler, der zu Verzerrungen führt, die sich auf alle Parameter auswirken (die Wirkung von $x(t)$ ist ja über ein Tiefpassfilter $\int \exp(-As)Bx(t_i+s)ds$ vermittelt). Für die Interpolation der exogenen Variablen zwischen den Messungen ist eine Vielzahl von Methoden möglich, die auf jeweils unterschiedliche Funktionen der Systemparameter A und B und deren Ableitungen führen. Eine Alternative bestünde darin, alle Größen als endogene Variablen zu behandeln. Das entstehende Identifikationsproblem muß dann aber durch Restriktionen oder unregelmäßige Zeitabstände ausgeschaltet werden.

In Panel-Studien kann man davon ausgehen, daß die verschiedenen statistischen Einheiten eine gewisse Heterogenität aufweisen, die sich z.B. in einer unterschiedlichen Dynamik ausdrücken kann. Man wird so auf Modelle mit Personeneffekten geführt, die als fixe oder zufällige Parameter behandelt werden. Im additiven Fall erlaubt das Zustandsraum-Modell (5.1) nach Erweiterung des

Systemzustands y im Prinzip eine Darstellung, jedoch sind die entsprechenden Systemmatrizen dann singulär. Weiterhin führt eine Heterogenität der Drift auf multiplikative Terme, sodaß eine spezielle Behandlung des Problems angebracht erscheint.

Weiterhin ist zu erwähnen, daß sich mit Hilfe des Zustandsraum-Modells auch integrierte Daten (flow data) behandeln lassen.

Eine sehr interessante Erweiterung stellt schließlich die Behandlung raum-zeitlicher Modelle dar. Häufig, wie z.B. in der Meteorologie, stehen nur Messungen an diskreten Orten (Meßstationen) und zu diskreten Zeitpunkten zur Verfügung. Es stellt sich daher die Aufgabe, aus den partiellen Differentialgleichungen der Hydrodynamik exakte diskrete Modelle für die Meßpunkte abzuleiten, deren Parameter zu schätzen und optimale Interpolationsfunktionen und Prognosen für beliebige raum-zeitliche Gebiete abzuleiten.

Anhänge:

Anhang 1: Das McDonald-Swaminathan-Kalkül

Im folgenden soll eine kurze Zusammenstellung der wichtigsten Formeln des Matrix-Kalküls von McDonald u. Swaminathan (M/S 1973, M 1976, S 1976; siehe auch Magnus u. Neudecker 1988) und der wichtigsten Matrix-Operationen gegeben werden.

A1.1 Matrix-Produkte

Sei $A: I \times J$ und $B: I \times J$; Die $I \times J$ -Matrix $A * B = A_{ij}B_{ij}$, $i=1\dots I$, $j=1\dots J$, heißt **Hadamard-Produkt** (elementweises Produkt).

Sei $A: I \times J$ und $B: K \times L$. Die $IK \times JL$ -Matrix $A \otimes B = A_{ij}B$, $i=1\dots I$, $j=1\dots J$, heißt (rechtshändiges) **Kronecker-Produkt**.

Sei $A = [a_1, \dots, a_j]$ eine $i \times j$ -Matrix. Dann ist die **spaltenweise Vektor-Operation** als

$$\text{col}(A) = \begin{bmatrix} a_1 \\ \vdots \\ a_j \end{bmatrix} \quad ij \times 1$$

definiert.

Sei $A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_i \end{bmatrix}$ eine $i \times j$ -Matrix. Die **zeilenweise Vektor-Operation** lautet:

$$\text{row}(A) = \begin{bmatrix} \alpha_1' \\ \vdots \\ \alpha_i' \end{bmatrix} \quad ij \times 1$$

Es gilt: $\text{col } A = \text{row } A' \quad (A'_{ij} = A_{ji})$

$$\text{row } ABC = (A \otimes C') \text{row } B$$

$$\text{col } ABC = \text{row } C'B'A' = (C' \otimes A) \text{row } B' = (C' \otimes A) \text{col } B$$

A 1.2 Matrix-Ableitungen

Sei $Y(X) : t \times u$ eine Matrix-Funktion von $X: r \times s$.

Dann heißt

$$\partial Y / \partial X := \partial (\text{row } Y) / \partial \text{row } X : rs \times tu$$

McDonald-Swaminathan-Ableitung von $Y(X)$ nach X .

In Komponenten gilt:

$$\partial Y / \partial X = \partial Y_{kl} / \partial X_{ij} ; i=1, \dots, r, j=1, \dots, s, k=1, \dots, t, l=1, \dots, u$$

Da die Matrizen Y und X vektorisiert werden, spricht man auch von Ableitungen in **Vektor-Anordnung**. Mit Hilfe des Differentialoperators

$$\partial / \partial X = [\partial / \partial x_{11}, \partial / \partial x_{12}, \dots, \partial / \partial x_{rs}]'$$

kann man auch schreiben

$$\begin{aligned} \partial Y / \partial X &= \partial / \partial X (\text{row } Y)' \\ &= \partial / \partial X \otimes (\text{row } Y)' \end{aligned}$$

Definiert man dagegen $\partial / \partial [X]$ als $r \times s$ -Matrix mit den Komponenten $\partial / \partial x_{ij}$, so ergibt sich die Ableitung in **Matrix-Anordnung**

$$[\partial Y / \partial X] = \partial / \partial [X] \otimes Y$$

(siehe hierzu die Diskussion in McDonald 1976 . Wenn X speziell ein Vektor Θ ist, erhält man also

$$[\partial Y / \partial \Theta] = \begin{bmatrix} \partial Y / \partial \theta_1 \\ \dots \\ \partial Y / \partial \theta_u \end{bmatrix}$$

wobei $\partial Y / \partial \theta_i$ eine $t \times u$ -Matrix darstellt.

Sei $Z=Z(Y(X))$ eine Matrix-Funktion ($Z: r \times s$, $Y: t \times u$, $X: v \times w$).

Dann gilt die **Kettenregel**:

$$\begin{aligned} (\partial Z / \partial X) &= (\partial Y / \partial X) (\partial Z / \partial Y) \\ vw \times rs & \quad vw \times tu \quad tu \times rs \end{aligned}$$

Produktregel: $\partial/\partial X [Y(X)Z(X)] = (\partial Y/\partial X) (I_u \otimes Z) + (\partial Z/\partial X)(Y \otimes I_w)$

wobei I_u eine Einheitsmatrix der Ordnung u ist ($X: m \times n, Y: u \times v, Z: v \times w$).

Folgerung: sei A und B konstant $\Rightarrow \partial/\partial X [AY(X)B] = \partial Y/\partial X (A \otimes B)$.

Ableitung des Logarithmus einer Determinante:

$$\partial/\partial X \log |X| = \text{row } X^{-T}$$

wobei $|X|$ =Determinante von X und $X^{-T}=X^{-1}$.

Ableitung der Spur:

$$\partial/\partial X \text{tr } X = \text{row } I_p \quad ; X: p \times p$$

A1.3 Irreduzible Matrix-Ableitungen

Sei S eine symmetrische Matrix. Dann ist die irreduzible Matrix-Ableitung L folgendermaßen definiert:

$$L = \partial S / \partial S$$

Bsp.: $S = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \Rightarrow L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Man erkennt, daß L singular ist.

Weiterhin hat man: $L \text{ row } A = \text{row} [A + A' - \text{Diag } A]$.

Sei $X: m \times n$. Dann heißt $E_{m,n} = \partial X' / \partial X$ **Kommutationsmatrix**.

Dieser Name ist aus folgenden Relationen verständlich ($A: n \times m$):

$$\partial X' / \partial X \text{ row } A = E_{m,n} \text{ row } A = \text{row } A'$$

Daher gilt ($A: m \times n, B: n \times o, C: o \times p$)

$$\begin{aligned} E_{p,m} \text{ row } ABC &= \text{row } C'B'A' \\ \Rightarrow E_{p,m} (A \otimes C') \text{ row } B &= (C' \otimes A) \text{ row } B' \\ &= (C' \otimes A) E_{o,n} \text{ row } B \end{aligned}$$

Da B beliebig ist, gilt

$$E_{p,m} (A \otimes C') = (C' \otimes A) E_{o,n}$$

A1.4. Partitionierte Matrizen

Bisher wurde angenommen, daß die Elemente der Matrizen skalare reelle oder komplexe Zahlen sind. Häufig sind Matrizen jedoch in Blöcke unterteilt, sodaß im folgenden die Matrix-Elemente wieder Matrizen sein können:

$$A = \{ A_{ij} \} ; A_{ij} : \alpha_i \times \beta_j$$

Bsp.: $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ wobei $A_{11} = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$ etc.

In Komponenten gilt also: $A = A_{ij}^{\alpha\beta}$; $i=1\dots I, j=1\dots J; \alpha=1\dots\alpha_i, \beta=1\dots\beta_j$.

Um einen unendlichen Regreß zu vermeiden, müssen auf der untersten Ebene die Elemente Skalare sein.

Partitioniertes Matrix-Produkt

Für "normale" Matrizen gilt: $AB = \sum_j A_{ij} B_{jk}$.

Für partitionierte Matrizen definiert man analog

$$AB = \sum_j A_{ij} B_{jk} ,$$

wobei das Matrixprodukt als

$$A_{ij} B_{jk} = \sum_{\beta} A_{ij}^{\alpha\beta} B_{jk}^{\beta\gamma}$$

definiert ist (die Blöcke müssen selbverständlich zusammenpassen).

Daher gilt

$$AB = \sum_i \sum_{\beta} A_{ij}^{\alpha\beta} B_{jk}^{\beta\gamma} .$$

Dies stimmt mit dem unpartitionierten Produkt überein. Tiefer verschachtelte Produkte werden analog definiert.

Partitioniertes Kronecker-Produkt

Seien A und B partitioniert. Dann heißt

$$A \otimes B = A_{ij} \otimes B_{kl} = A_{ij}^{\alpha\beta} B_{kl}^{\gamma\delta}$$

Partitioniertes Kronecker-Produkt. Das Kronecker-Produkt \otimes wird also auf die Elemente von A und B (die wieder Matrizen sein können) angewandt, bis man auf der Ebene der Skalare angelangt ist ($a \otimes b = ab$, falls a und b skalar).

Partitionierte Vektor-Operation

Sei $A = A_{ij}$; $i=1\dots I, j=1\dots J$ partitioniert. Dann heißt

$$\text{ROW } A = \begin{bmatrix} \text{row } A_{11} \\ \text{row } A_{12} \\ \dots \\ \text{row } A_{IJ} \end{bmatrix} = A_{ij}^{\alpha\beta}$$

partitionierte zeilenweise Vektor-Operation (analog für COL).

Es gilt

$$\text{ROW } ABC = (A \otimes\otimes C') \text{ROW } B,$$

wobei $(C')_{ij} = (C_{ji})'$ der ij -te Block von C' ist.

Partitionierte M/S-Ableitung

Sei $Y(X)$ eine partitionierte Matrix-Funktion der partitionierten Matrix X .

Dann heißt

$$\partial[Y]/\partial[X] = \partial[\text{ROW } Y]' / \partial[\text{ROW } X]$$

Partitionierte M/S-Ableitung. Die Rechenregeln aus A1.2. gelten analog, wenn man \otimes durch $\otimes\otimes$ und row durch ROW ersetzt.

Da nichtpartitionierte als trivial partitionierte Matrizen betrachtet werden können (A_{ij} =skalar), wird folgende **Vereinbarung** getroffen:

Alle Matrix-Operationen sind in partitionierter Form zu verstehen und werden rekursiv auf allen Ebenen der Partitionierung angewandt, bis man auf der skalaren Ebene angelangt ist. Zur Vereinfachung werden die in A.1.2. definierten einfachen Symbole verwandt.

Bsp.: i) $A \otimes B = A_{ij} \otimes B_{kl} = A_{ij}^{\alpha\beta} \otimes B_{kl}^{\gamma\delta}$. Wenn $A_{ij}^{\alpha\beta}$ und $B_{kl}^{\gamma\delta}$ Skalare sind, ist man am Ende der Rechnung angelangt.

ii) $\text{row } A = \text{row } A_{ij} = A_{ij}^{\alpha\beta}$

iii) $\partial Y / \partial X = \partial Y_{ij} / \partial X_{kl} = \partial(\text{row } Y_{ij})' / \partial \text{row } X_{kl} = \partial Y_{ij}^{\alpha\beta} / \partial X_{kl}^{\gamma\delta}$

Anhang 2 : Maximum-Likelihood-Schätzung

Sei $X = \{X_1, \dots, X_N\}$ eine Zufallsstichprobe mit der gemeinsamen Dichte $f(x_1, \dots, x_N; \Theta)$. Dann heißt f , betrachtet als Funktion des Parametervektors $\Theta : u \times 1$, **Likelihood-Funktion**

$$L(\Theta; X) = f(x_1, \dots, x_N; \Theta)$$

$l = \log L$ heißt **Log-Likelihood-Funktion**. Da \log eine monotone Transformation darstellt, kann das Maximum von L auch bzgl. l gesucht werden.

$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} L(\Theta; X) = \underset{\Theta}{\operatorname{argmax}} l(\Theta; X)$ heißt **Maximum-Likelihood (ML)-Schätzer**.

Wenn l differenzierbar ist, kann $\hat{\Theta}$ als Nullstelle des Gradienten $s = \partial l / \partial \Theta$ gefunden werden. Die Zufallsvariable $s(\Theta; X) : u \times 1$, heißt **score-Funktion**.

Theorem A2.1.

Der Erwartungswert der score-Funktion verschwindet, d.h.

$$E[s] = 0.$$

Beweis: (Mardia et al. 1979, Kap. 4).

Die Kovarianzmatrix von s wird **Fisher-Informations-Matrix** genannt, d.h. mit Theorem A2.1 gilt

$$F = \operatorname{Cov}(s, s) = E[ss'] : u \times u$$

Theorem A2.2

Sei $s(\Theta; X)$ der score und t eine Funktion von X und Θ . Dann gilt

$$E[st'] = \partial / \partial \Theta E[t] - E[\partial t / \partial \Theta]$$

Beweis: (Mardia et al. 1979, Kap. 4)

Corollar 1: wähle $t = \text{const.}$ $\Rightarrow E[st'] = 0 \forall t \Rightarrow E[s] = 0$.

Corollar 2: sei $t = s$. $\Rightarrow F = [ss'] = -E[\partial s / \partial \Theta] = -E[\partial^2 l / \partial \Theta^2]$, d.h. die Fisher-Information kann als negativer Erwartungswert der Hesse-Matrix gewonnen werden.

Theorem A2.3. (Cramer-Rao-Ungleichung)

Sei $t(X)$ ein unverzerrter Schätzer von Θ (d.h. $E[t] = \Theta$). Dann gilt

$$\operatorname{Cov}(t) \geq F^{-1} \text{ (d.h. } \operatorname{Cov}(t) - F^{-1} \text{ ist positiv semidefinit)}$$

Beweis: (Mardia et al., a.a.o)

Theorem A2.4.

Unter diversen Regularitäts- und Identifizierbarkeitsbedingungen (siehe Caines 1988, Theorem 3.2, S. 317) gilt:

Sei X_1, \dots, X_N eine Folge unabhängiger und identisch verteilter Zufallsvariablen.

i) Dann ist der ML-Schätzer $\hat{\Theta}_N$ **stark konsistent**, d.h.

$$\lim_{N \rightarrow \infty} \hat{\Theta}_N = \Theta \quad (\text{fast sicher})$$

ii) $\sqrt{N}(\hat{\Theta}_N - \Theta) \sim N(0, J^{-1})$ für $N \rightarrow \infty$, wobei $J = \lim_{N \rightarrow \infty} 1/N F = \lim_{N \rightarrow \infty} 1/N E_{\Theta}[ss']$ die skalierte Fisher-Information ist.

(Asymptotische Normalität des ML-Schätzers).

Anhang 3 : Analytische Ableitungen der reduzierten Parameter

Behauptung 1:

$$(4.12) \quad \partial A^* / \partial \Theta = \partial A / \partial \Theta * \tau \left[\text{diag row } E_1 \right] \tau^{-1}$$

$$\text{mit } E_1 = \begin{cases} \Delta t e^{\lambda_i \Delta t} & i = j \\ \frac{e^{\lambda_i \Delta t} - e^{\lambda_j \Delta t}}{\lambda_i - \lambda_j} & i \neq j \end{cases}$$

$$\tau = T^{-T} \otimes T$$

$$\text{und } \Lambda = T^{-1} A T$$

Herleitung:

Sei $\Phi(s) = e^{As}$. Dann ist $A^* = \Phi(\Delta t)$ Lösung der Matrix-Differentialgleichung

$$(A3.1) \quad \dot{\Phi} = A\Phi ; \quad \Phi(0) = I$$

Sei weiterhin $\psi(s) = \partial \Phi(s) / \partial A = \Phi_A$. Dann erfüllt ψ die Differentialgleichung:

$$\begin{aligned} \dot{\psi} &= \partial \dot{\Phi} / \partial A = \partial / \partial A (A\Phi) \\ &= \partial A / \partial A (I \otimes \Phi) + \partial \Phi / \partial A (A' \otimes I) \\ &= I \otimes e^{As} + \psi (A' \otimes I) \end{aligned}$$

Mit den Abkürzungen $H_2 = I \otimes A$ und $H_1 = A' \otimes I$ gilt:

$$(A3.2) \quad \dot{\psi} = e^{H_2 s} + \psi H_1$$

Lösung der homogenen Gleichung: $\dot{\psi} = \psi H_1 \Rightarrow \psi(s) = \psi(0) e^{H_1 s}$

wobei $\psi(0) = \partial \Phi(0) / \partial A = 0$

Allgemeine Lösung von (A3.2) :

$$\psi(s) = \left[\int_0^s e^{H_2 u} e^{-H_1 u} du \right] e^{H_1 s}$$

Da weiterhin H_1 und H_2 vertauschbar sind ($[H_1, H_2] = A' \otimes A - A' \otimes A = 0$)

kann man schreiben

$$(A3.3) \quad \psi(s) = \left[\int_0^s e^{(H_2 - H_1) u} du \right] e^{H_1 s}$$

Falls nun $H_2 - H_1$ nichtsingulär wäre, hätte man das Endresultat

$$\psi(\Delta t) = \partial A^* / \partial A = (I \otimes A - A' \otimes I)^{-1} \left[e^{(I \otimes A) \Delta t} - e^{(A' \otimes I) \Delta t} \right]$$

Dies ist jedoch nicht der Fall, wie man mit Hilfe der Eigenwertzerlegung

$$\Lambda = T^{-1} A T$$

leicht zeigen kann. Man findet nämlich

$$\begin{aligned} \mu &= \tau^{-1} (H_2 - H_1) \tau = I \otimes \Lambda - \Lambda \otimes I = \mu_2 - \mu_1 \\ &= \text{Diag} (0, \lambda_2 - \lambda_1, \lambda_3 - \lambda_1, \dots; \lambda_1 - \lambda_2, 0, \lambda_3 - \lambda_2, \lambda_4 - \lambda_2, \dots) \end{aligned}$$

(A.3.3) läßt sich nun explizit auswerten :

$$\begin{aligned} \psi(s) &= \tau \left[\int_0^s e^{\mu u} du \right] e^{\mu_1 s} \tau^{-1} \\ &= \tau \left[\text{diag row } E_1 \right] \tau^{-1} \end{aligned}$$

$$\text{mit } E_{1 \text{ ij}} = \begin{cases} \Delta t e^{\lambda_i \Delta t} & i = j \\ \frac{e^{\lambda_i \Delta t} - e^{\lambda_j \Delta t}}{\lambda_i - \lambda_j} & i \neq j \end{cases}$$

Mit Hilfe der Kettenregel $\partial A^* / \partial \Theta = (\partial A / \partial \Theta) (\partial A^* / \partial A)$ findet man schließlich

$$\partial A^* / \partial \Theta = \partial A / \partial \Theta * \tau \left[\text{diag row } E_1 \right] \tau^{-1}$$

q.e.d.

Formel (4.13) kann auch mit Hilfe der Taylor-Entwicklung von $\exp(A\Delta t)$ und der Eigenwertzerlegung hergeleitet werden (vgl. Jennrich u. Bright 1973, Kalbfleisch u. Lawless 1985). Etwas kompakter geschrieben gilt dann

$$\left[A^*_{\Theta_i} \right] = T \left[(T^{-1} [A_{\Theta_i}] T) \# E_1 \right] T^{-1} ; i = 1 \dots u$$

$$\text{und } \partial A^* / \partial \Theta = \begin{bmatrix} \text{row } (A^*_{\Theta_1}) \\ \vdots \\ \text{row } (A^*_{\Theta_u}) \end{bmatrix}$$

Die Äquivalenz der Formeln kann unmittelbar mit Hilfe von

$$\left[\text{diag row } X \right] \text{row } Y = \text{row } (X * Y)$$

und

$$(\text{row } B) (A \otimes C) = \text{row } ABC$$

eingesehen werden.

Behauptung 2

$$(4.13) \quad \partial V / \partial \Theta = \partial A / \partial \Theta * \partial V / \partial A + \partial \Omega / \partial \Theta * \partial V / \partial \Omega$$

$$\begin{aligned} \text{mit } \partial V / \partial \Omega &= \partial / \partial \Omega \int_0^{\Delta t} \exp(As) \Omega \exp(A's) ds \\ &= \left(e^{\alpha \Delta t} - I(p^2) \right) \alpha^{-T} \quad ; \end{aligned}$$

$$\alpha = A \otimes I + I \otimes A$$

$$\begin{aligned} \text{und } \partial V / \partial A &= \left[\partial \alpha / \partial A \right] * \left[\left(-\alpha^{-T} \otimes \alpha^{-1} \right) \left(I(p^2) \otimes \left[\left(e^{\alpha \Delta t} - I(p^2) \right) \text{row } \Omega \right] \right) \right. \\ &\quad \left. + \left(\partial e^{\alpha \Delta t} / \partial \alpha \right) \left(\alpha^{-T} \otimes \text{row } \Omega \right) \right] \end{aligned}$$

$$\begin{aligned} \partial \alpha / \partial A &= \partial / \partial A \left(I(p) \otimes A + A \otimes I(p) \right) \\ &= \left[I(p^2) \otimes \text{row } I(p) \right] \left[I(p) \otimes E(p,p) \otimes I(p) \right] \left[I(p^4) + E(p,p) \otimes E(p,p) \right] \end{aligned}$$

$$\partial e^{\alpha \Delta t} / \partial \alpha = \sigma \left[\text{diag row } F_1 \right] \sigma^{-1}$$

$$\text{mit } F_1 = \begin{cases} \Delta t e^{x_i \Delta t} & i = j \\ \frac{e^{x_i \Delta t} - e^{x_j \Delta t}}{x_i - x_j} & i \neq j \end{cases}$$

$$\sigma = \tau^{-T} \otimes \tau$$

$$\text{und } \chi = \tau^{-1} \alpha \tau = I \otimes \Lambda + \Lambda \otimes I$$

Herleitung :

Mit Hilfe der Produktregel findet man (die Symmetrie von Ω ist in $\partial \Omega / \partial \Theta$ berücksichtigt)

$$\begin{aligned} \partial V / \partial \Omega &= \partial / \partial \Omega \int_0^{\Delta t} \exp(A s) \Omega \exp(A' s) ds \\ &= \int_0^{\Delta t} \exp(A' s) \otimes \exp(A' s) ds \\ &= \int_0^{\Delta t} \exp(\alpha' s) ds \\ &= \left(e^{\alpha' \Delta t} - I(p^2) \right) \alpha^{-T} ; \end{aligned}$$

$$\begin{aligned} \partial V / \partial A &= \left[\partial \alpha / \partial A \right] * \left[\partial V / \partial \alpha \right] \\ &= \left[\partial \alpha / \partial A \right] \left[\left(\partial \alpha^{-1} / \partial \alpha \right) \left(I(p^2) \otimes \left[\left(e^{\alpha \Delta t} - I(p^2) \right) \text{row } \Omega \right] \right) \right. \\ &\quad \left. + \left(\partial e^{\alpha \Delta t} / \partial \alpha \right) \left(\alpha^{-T} \otimes \text{row } \Omega \right) \right] \end{aligned}$$

Weiterhin gilt mit Hilfe der Ableitungs-Regeln für **Kronecker-Produkte** (vgl. **Magnus u. Neudecker 1988, S. 184**)

$$\begin{aligned} \partial / \partial A \left[A \otimes I(p) \right] &= \left[I(p^2) \otimes \text{row}' I(p) \right] \left[I(p) \otimes E(p,p) \otimes I(p) \right] \\ \partial / \partial A \left[I(p) \otimes A \right] &= \partial / \partial A \left[A \otimes I(p) \right] \left[E(p,p) \otimes E(p,p) \right] \end{aligned}$$

Schließlich kann $\partial e^{\alpha \Delta t} / \partial \alpha$ mit Hilfe von Formel (4.12) und der Ersetzung $A^* \rightarrow \alpha$ als

$$\sigma \left[\text{diag row } F_1 \right] \sigma^{-1}$$

geschrieben werden,

$$\text{mit } F_1 = \begin{cases} \Delta t e^{\chi_i \Delta t} & i = j \\ \frac{e^{\chi_i \Delta t} - e^{\chi_j \Delta t}}{\chi_i - \chi_j} & i \neq j \end{cases}$$

$$\sigma = \tau^{-T} \otimes \tau$$

$$\text{und } \chi = \tau^{-1} \alpha \tau = I \otimes \Lambda + \Lambda \otimes I$$

q.e.d.

Behauptung 3:

$$\left[\partial V / \partial \Theta_u \right] = \int_0^{\Delta t} \left[A_{\Theta_u}^*(s) \right] \Omega A^{*'}(s) \quad (I)$$

$$+ \int_0^{\Delta t} A^*(s) \Omega \left[A_{\Theta_u}^{*'}(s) \right] \quad (I')$$

$$+ \int_0^{\Delta t} A^*(s) \left[\Omega_{\Theta_u} \right] A^{*'}(s) \quad (II)$$

$u=1, \dots, U$

$$\text{wobei } I_{bh}^u = \sum_{adf} \left[T^{-1} \left[A_{\Theta_u} \right] T \right]_{ad}^u \left[T_{ba} E_{3adf} T_{hf} \right]_{adf}^{bh} \left[T^{-1} \Omega T^{-T} \right]_{df}$$

$u=1, \dots, U; \quad b, h = 1, \dots, p$

$$E_{3adf} = \begin{cases} \frac{\Delta t}{\lambda_a + \lambda_f} e^{(\lambda_a + \lambda_f) \Delta t} - \frac{1}{(\lambda_a + \lambda_f)^2} \left[e^{(\lambda_a + \lambda_f) \Delta t} - 1 \right] ; & a=d \\ \frac{1}{(\lambda_a - \lambda_d)} \left(E_{2af} - E_{2df} \right) ; & a \neq d \end{cases}$$

$$\Pi_{af}^u = \sum_{be} \left[T^{-1} \left[\Omega_{\Theta u} \right] T^{-T} \right]_{be}^u \left[T_{ab} E_{2be} T_{fe} \right]_{be}^{af}$$

$$u=1,\dots,U; \quad a, f=1,\dots,p$$

$$\text{mit } E_{2be} = \frac{1}{\lambda_b + \lambda_e} \left[e^{(\lambda_b + \lambda_e)\Delta t} - 1 \right]$$

Beweis:

(I)

$$\int_0^{\Delta t} \left[A_{\Theta}^*(s) \right] \Omega A^{*'}(s) ds = \int_0^{\Delta t} \left[A_{\Theta} \right]_{ij} T_{ba} T_{ai}^{-1} E_{1ad} T_{jd} T_{dc}^{-1} \Omega_{ce} T_{ef}^{-T} e^{\lambda_f s} \delta_{fg} T_{gh} ds$$

$$\text{Für das Integral } E_{3adf} = \int_0^{\Delta t} E_{1ad}(s) e^{\lambda_f s} ds$$

findet man

$$E_{3adf} = \begin{cases} \frac{\Delta t}{\lambda_a + \lambda_f} e^{(\lambda_a + \lambda_f)\Delta t} - \frac{1}{(\lambda_a + \lambda_f)^2} \left[e^{(\lambda_a + \lambda_f)\Delta t} - 1 \right] & ; a=d \\ \frac{1}{(\lambda_a - \lambda_d)} \left(E_{2af} - E_{2df} \right) & ; a \neq d \end{cases}$$

$$\text{mit } E_{2be} = \frac{1}{\lambda_b + \lambda_e} \left[e^{(\lambda_b + \lambda_e)\Delta t} - 1 \right]$$

Damit ist (I) gezeigt.

(II)

$$\begin{aligned} \int_0^{\Delta t} A^*(s) \left[\Omega_{\Theta u} \right] A^{*'}(s) ds &= \int_0^{\Delta t} T_{ab} e^{\lambda_b s} T_{bc}^{-1} \Omega_{\Theta cd} T_{de}^{-T} e^{\lambda_e s} T_{ef} ds \\ &= T_{ab} T_{bc}^{-1} \Omega_{\Theta cd} T_{ed}^{-1} T_{fe} E_{2be} \end{aligned}$$

q.e.d

Anhang 4: Fisher-Informations-Matrix des exakten diskreten Modells

Für die Fisher-Informations-Matrix gilt:

$$F = E_{\Theta} [s s'] = -E_{\Theta} [\partial^2 l / \partial \Theta^2]$$

$$= - [\partial \Theta^* / \partial \Theta] F^* [\partial \Theta^* / \partial \Theta]'$$

mit $F^* = -E_{\Theta} [\partial^2 l / \partial \Theta^{*2}] =$

$$\begin{bmatrix} NT/2 (V^{-1} \otimes V^{-1}) & 0 & 0 & 0 & 0 \\ 0 & V^{-1} \otimes E_{\Theta} [Y Y'] & V^{-1} \otimes E_{\Theta} [Y] X' & 0 & 0 \\ 0 & V^{-1} \otimes X E_{\Theta} [Y'] & V^{-1} \otimes X X' & 0 & 0 \\ 0 & 0 & 0 & N/2 (\Sigma^{-1} \otimes \Sigma^{-1}) & 0 \\ 0 & 0 & 0 & 0 & N \Sigma^{-1} \end{bmatrix}$$

Beweis:

Knüpfen wir an Kap. 4.3 an, so gilt für den Score:

$$E_{\Theta} [s(\Theta, Y_0, Y_+)] = [\partial \Theta^* / \partial \Theta] * E_{\Theta} [\partial l / \partial \Theta^*] = 0$$

mit $\partial \Theta^* / \partial \Theta = [V_{\Theta}, A_{\Theta}^*, B_{\Theta}^*, \Sigma_{\Theta}, \mu_{\Theta}]$

und $[\partial l / \partial \Theta^*] =$

$$\begin{bmatrix} -\frac{NT}{2} \text{row}[V^{-1} (W - V) V^{-1}] \\ \text{row}[V^{-1} U Y'] \\ \text{row}[V^{-1} U X'] \\ -\frac{N}{2} \text{row}[\Sigma^{-1} (R - \Sigma) \Sigma^{-1}] \\ N \text{row}[\Sigma^{-1} (m - \mu)] \end{bmatrix}$$

Mit Hilfe der Kettenregel kann man dann für die 2. Ableitung schreiben

$$\begin{aligned} \partial^2 1 / \partial \theta^2 &= \partial / \partial \theta \left\{ \left[\partial \theta^* / \partial \theta \right] * \left[\partial 1 / \partial \theta^* \right] \right\} \\ &= \left(\partial^2 \theta^* / \partial \theta^2 \right) \left(I \otimes \partial 1 / \partial \theta^* \right) + \left(\partial^2 1 / \partial \theta \partial \theta^* \right) \left(\left[\partial \theta^* / \partial \theta \right]' \otimes I_1 \right) \\ &= \left(\partial^2 \theta^* / \partial \theta^2 \right) \left(I \otimes \partial 1 / \partial \theta^* \right) + \left[\partial \theta^* / \partial \theta \right] \left[\partial^2 1 / \partial \theta^*{}^2 \right] \left[\partial \theta^* / \partial \theta \right]' \end{aligned}$$

$$\begin{aligned} \text{Damit gilt zunächst} \quad F &= E_{\theta} \left[s s' \right] = -E_{\theta} \left[\partial^2 1 / \partial \theta^2 \right] \\ &= - \left[\partial \theta^* / \partial \theta \right] F^* \left[\partial \theta^* / \partial \theta \right]' \end{aligned}$$

$$\text{mit} \quad F^* = -E_{\theta} \left[\partial^2 1 / \partial \theta^*{}^2 \right].$$

Es genügt also, die Ableitungen der Score-Funktion und deren Erwartungswert zu berechnen. Man findet (alle Matrizen frei, Restriktionen erscheinen in $\partial \theta^* / \partial \theta$)

Ableitung nach V :

$$\begin{aligned} \partial / \partial V \left(\text{row} \left[V^{-1} (W - V) V^{-1} \right] \right) &= \left(\partial V^{-1} / \partial V \right) \left[I \otimes (W - V) V^{-1} \right] \\ &\quad + \left(\partial (W - V) / \partial V \right) \left(V^{-1} \otimes V^{-1} \right) \\ &\quad + \left(\partial V^{-1} / \partial V \right) \left[(W - V) V^{-1} \otimes I \right] \end{aligned}$$

Da jedoch $E[W]=V$ gilt und W nur von A^* und B^* abhängt, findet man einfach

$$E \left\{ \partial / \partial V \left(\text{row} \left[V^{-1} (W - V) V^{-1} \right] \right) \right\} = - \left(V^{-1} \otimes V^{-1} \right)$$

Die Erwartungswerte der Ableitungen der anderen Komponenten des Scores verschwinden, da $E[UY']=0$ und $E[U]=0$ bzw. die Komponenten gar nicht von V abhängen.

Ableitungen nach A^* :

$$\begin{aligned} \partial / \partial A^* \left(\text{row} \left[V^{-1} (W - V) V^{-1} \right] \right) &= \left(\partial W / \partial A^* \right) \left(V^{-1} \otimes V^{-1} \right) \\ &= 1/NT \partial / \partial A^* (U U') \left(V^{-1} \otimes V^{-1} \right) \\ &= 1/NT \left[\begin{array}{l} \left(\partial U / \partial A^* \right) (I \otimes U') \\ \left(\partial U' / \partial A^* \right) (U' \otimes I) \end{array} \right] \left(V^{-1} \otimes V^{-1} \right) \end{aligned}$$

Nun ist $(\partial U/\partial A^*) = -I \otimes Y$ und $(\partial U'/\partial A^*) = -(I \otimes Y) E_{pp}$ (vgl. Anhang 1).
Da jedoch $E[UY']$ verschwindet (die Zuwächse u_i sind von den Zuständen y_i unabhängig, verschwindet obiger Ausdruck.

Weiterhin hat man

$$\partial/\partial A^* \left(\text{row} \left[V^{-1} U Y' \right] \right) = \left(\partial U/\partial A^* \right) (V^{-1} \otimes Y') = -V^{-1} \otimes YY'$$

$$\partial/\partial A^* \left(\text{row} \left[V^{-1} U X' \right] \right) = \left(\partial U/\partial A^* \right) (V^{-1} \otimes X') = -V^{-1} \otimes YX'$$

Die anderen Komponenten des Scores hängen nicht von A^* ab und verschwinden.

Ableitungen nach B:

$$\begin{aligned} \partial/\partial B^* \left(\text{row} \left[V^{-1} (W - V) V^{-1} \right] \right) &= 1/NT \partial/\partial B^* (UU') (V^{-1} \otimes V^{-1}) \\ &= 1/NT \left[\left(\partial U/\partial B^* \right) (I \otimes U') \right. \\ &\quad \left. \left(\partial U'/\partial B^* \right) (U' \otimes I) \right] (V^{-1} \otimes V^{-1}) \end{aligned}$$

Nun ist $(\partial U/\partial B^*) = -I \otimes X$ und $(\partial U'/\partial B^*) = -(I \otimes X) E_{pp}$ (vgl. Anhang 1).
Da jedoch $E[XU']$ verschwindet (die Zuwächse haben Erwartungswert 0), ist obiger Ausdruck 0.

$$\partial/\partial B^* \left(\text{row} \left[V^{-1} U Y' \right] \right) = \left(\partial U/\partial B^* \right) (V^{-1} \otimes Y') = -V^{-1} \otimes XY'$$

$$\partial/\partial B^* \left(\text{row} \left[V^{-1} U X' \right] \right) = \left(\partial U/\partial B^* \right) (V^{-1} \otimes X') = -V^{-1} \otimes XX'$$

Die restlichen Ableitungen verschwinden.

Ableitung nach Σ :

$$E \left\{ \partial/\partial \Sigma \left(\text{row} \left[\Sigma^{-1} (R - \Sigma) \Sigma^{-1} \right] \right) \right\} = - \left(\Sigma^{-1} \otimes \Sigma^{-1} \right)$$

da $E[R] = \Sigma$ und analog zur Ableitung nach V.

$$E \left\{ \partial/\partial \Sigma \text{ row} \left[\Sigma^{-1} (m - \mu) \right] \right\} = E \left\{ \partial \Sigma^{-1} / \partial \Sigma \left(I \otimes (m - \mu) \right) \right\} = 0$$

da $E[m] = \mu$.

Ableitung nach μ :

$$\begin{aligned} \partial/\partial\mu \text{ row } \Sigma^{-1}(R-\Sigma)\Sigma^{-1} &= (\partial R/\partial\mu) \Sigma^{-1} \otimes \Sigma^{-1} \\ &= 1/N \sum_{i=1}^N \left[-(\partial\mu/\partial\mu) I \otimes (y_{0i}-\mu)' - (\partial\mu'/\partial\mu) \left[(y_{0i}-\mu)' \otimes I \right] \right] \Sigma^{-1} \otimes \Sigma^{-1} \end{aligned}$$

Da jedoch $E[y_{0i}] = \mu$, verschwindet der Erwartungswert.

$$\partial/\partial\mu \text{ row } \Sigma^{-1}(m-\mu) = -(\partial\mu/\partial\mu)(\Sigma^{-1} \otimes I_1) = -\Sigma^{-1}.$$

Wiederum sind die anderen Ableitungen = 0.

Zusammenfassend erhält man

$$F^* = -E[\partial^2 l / \partial \theta^{*2}] =$$

$$\begin{bmatrix} NT/2 (V^{-1} \otimes V^{-1}) & 0 & 0 & 0 & 0 \\ 0 & V^{-1} \otimes E_{\theta}[YY'] & V^{-1} \otimes E_{\theta}[Y]X' & 0 & 0 \\ 0 & V^{-1} \otimes X E_{\theta}[Y] & V^{-1} \otimes XX' & 0 & 0 \\ 0 & 0 & 0 & N/2 (\Sigma^{-1} \otimes \Sigma^{-1}) & 0 \\ 0 & 0 & 0 & 0 & N \Sigma^{-1} \end{bmatrix}$$

q.e.d.

Anhang 5: Asymptotische Eigenschaften der Score-Funktion und der Hesse-Matrix

Behauptung 1:

Sei $1/N \sum x_{nt} < \infty$ und $1/N \sum x_{nt} x'_{ns} < \infty$ (*)

1. $\text{qm-lim}_{N \rightarrow \infty} 1/N (\partial l / \partial \theta) = 0$
2. $\text{qm-lim}_{N \rightarrow \infty} 1/N YY' = \lim_{N \rightarrow \infty} 1/N E[YY']$
3. $\text{qm-lim}_{N \rightarrow \infty} 1/N YX' = \lim_{N \rightarrow \infty} 1/N E[Y] X'$

Beweis:

Zuvor ist folgendes **Lemma** erforderlich :

$$x_n \xrightarrow{2} c \text{ (deterministisch)} \Leftrightarrow E[x_n] \rightarrow c \text{ und } V[x_n] \rightarrow 0$$

Beweis: $E | x_n - c |^2 = V[x_n - c] + E^2[x_n - c]$
 $= V[x_n] + [E(x_n) - c]^2$

Da beide Terme auf der rechten Seite positiv sind, ist die Behauptung gezeigt.
q.e.d.

1. Der Score des diskreten Modells ist durch folgenden Ausdruck gegeben (vgl. Abs 4.3):

$$\left[\frac{\partial l}{\partial \theta^*} \right] = \begin{bmatrix} -\frac{NT}{2} \text{row}[V^{-1} (W - V)V^{-1}] \\ \text{row}[V^{-1} U Y'] \\ \text{row}[V^{-1} U X'] \\ -\frac{N}{2} \text{row}[\Sigma^{-1} (R - \Sigma)\Sigma^{-1}] \\ N \text{row}[\Sigma^{-1} (m - \mu)] \end{bmatrix}$$

- 1.1. Es gilt $W = (1/NT) UU' = (1/NT) \sum u_{nt} u'_{nt}$, mit $E[u_{nt} u'_{nt}] = V$ und für die Komponenten von u ($u^a = u^a_{nt}$, $a=1, \dots, p$):

$$\text{Cov}[u^a u^b, u^c u^d] = E[u^a u^b u^c u^d] - E[u^a u^b]E[u^c u^d]$$

Mit Hilfe der Formel

$$E[ijkl] = E[ij]E[kl] + E[ik]E[jl] + E[il]E[jk] - 2 E[i]E[j]E[k]E[l]$$

wobei $E[ij] \equiv E[u^i u^j]$ etc., die leicht mit Hilfe der charakteristischen Funktion für die Normalverteilung abgeleitet werden kann, gilt

$$\text{Cov}[u^a u^b, u^c u^d] = v_{ac}v_{bd} + v_{ad}v_{bc}$$

und somit $\text{Cov}[W^{ab}, W^{cd}] = (1/N)^2 \sum v_{ac}v_{bd} + v_{ad}v_{bc} \rightarrow 0$ für $N \rightarrow \infty$

Damit ist die Quadratmittelkonvergenz von W gegen V gezeigt (elementweise Konvergenz impliziert die Konvergenz der Norm).

1.2. $(1/N) UY' \xrightarrow{2} 0$,

denn $E[(1/N) UY'] = 0$

$$\begin{aligned} \text{und } \text{Cov}[u^a y^b, u^c y^d] &= E[u^a y^b u^c y^d] - E[u^a y^b]E[u^c y^d] \\ &= v_{ac} E[y^b y^d] \end{aligned}$$

Weiterhin ist

$$\begin{aligned} E[y_{nt} y'_{nt}] &= \text{cov}[y_{nt}, y'_{nt}] + \mu_{nt} \mu'_{nt} \\ &= \Sigma_t + \mu_{nt} \mu'_{nt} \end{aligned}$$

sodaß für das Verschwinden der Kovarianz von $(1/N) UY'$ die Konvergenz von

$$1/N \sum_{nt} \mu_{nt} \mu'_{nt}$$

ausreicht.

Aufgrund der Formel $\mu_{nt} = A^{*t} \mu + \sum_{s=0}^{t-1} A^{*t-1-s} B^* x_{ns}$

ist hierfür $1/N \sum x_{nt} < \infty$ und $1/N \sum x_{nt} x'_{ns} < \infty$ hinreichend. **q.e.d.**

1.3. $(1/N) UX' \xrightarrow{2} 0$,

denn $E[(1/N) UX'] = 0$ und

$$\begin{aligned} \text{Cov}[u^a x^b, u^c x^d] &= E[u^a x^b u^c x^d] - E[u^a x^b] E[u^c x^d] \\ &= \text{vac } x^b x^d \end{aligned}$$

Wiederum genügt die Konvergenz von $1/N \sum x_{nt} x'_{nt}$, um die Kovarianz von $(1/N) UX'$ im Limes $N \rightarrow \infty$ zum Verschwinden zu bringen.

1.4. Auch für die Komponenten des Score, die zur Anfangsbedingung Y_0 gehören, läßt sich die Behauptung leicht nachweisen:

$$R = 1/N \sum (y_{n0} - \mu)(y_{n0} - \mu)' \xrightarrow{2} \Sigma$$

denn $E[R] = \Sigma$

und $\text{Cov}[(y_{n0} - \mu)^a (y_{n0} - \mu)^b, (y_{n0} - \mu)^c (y_{n0} - \mu)^d] = \Sigma^{ac} \Sigma^{bd} + \Sigma^{ad} \Sigma^{bc}$

$E[m] = E[1/N \sum y_{0t}]$ und $\text{Cov}[m, m] = 1/N \Sigma \rightarrow 0$ für $N \rightarrow \infty$.

Damit ist Teil 1 der Behauptung gezeigt.

2. $\text{qm-lim}_{N \rightarrow \infty} 1/N YY' = \lim_{N \rightarrow \infty} 1/N E[YY']$, denn

$$\begin{aligned} E[1/N YY'] &= 1/N \sum E[y_{nt} y'_{nt}] = 1/N \sum \{ \text{cov}[y_{nt}, y'_{nt}] + \mu_{nt} \mu'_{nt} \} \\ &= 1/N \sum (\Sigma_t + \mu_{nt} \mu'_{nt}) \end{aligned}$$

Dieser Ausdruck konvergiert wiederum, wenn die in der Behauptung geforderten Bedingungen an die Exogenen x erfüllt sind.

Weiterhin kann man für die Kovarianz von $1/N YY'$ folgenden Ausdruck ableiten :

$$\begin{aligned} 1/N^2 \sum_{\substack{nm \\ ts}} \text{Cov}[y_{nt}^a y_{nt}^b, y_{ms}^c y_{ms}^d] &= 1/N^2 \sum_{nts} E[y_{nt}^a y_{ns}^c] E[y_{nt}^b y_{ns}^d] \\ &\quad + E[y_{nt}^a y_{ns}^d] E[y_{nt}^b y_{ns}^c] \\ &\quad - 2E[y_{nt}^a] E[y_{nt}^b] E[y_{ns}^c] E[y_{ns}^d] \\ &= 1/N^2 \sum_{nts} \Sigma_{ts}^{ac} \Sigma_{ts}^{bd} + \Sigma_{ts}^{bd} \mu_{nt}^a \mu_{ns}^c + \Sigma_{ts}^{ac} \mu_{nt}^b \mu_{ns}^d \\ &\quad + \Sigma_{ts}^{ad} \Sigma_{ts}^{bc} + \Sigma_{ts}^{bc} \mu_{nt}^a \mu_{ns}^d + \Sigma_{ts}^{ad} \mu_{nt}^b \mu_{ns}^c \end{aligned}$$

wobei $\Sigma_{ts} = \text{Cov}[y_{nt}, y_{ns}]$ und $\mu_{nt} = E[y_{nt}]$

Für die Konvergenz dieses Ausdruckes zu 0 im Limes $N \rightarrow \infty$ genügt wiederum Bedingung (*) aus der Voraussetzung.

3. Die Behauptung für $1/N YX'$ folgt mit einer analogen Argumentation.

Q.E.D.

Analoge Bedingungen für die Exogenen müssen auch bei Konsistenzbeweisen gefordert werden (vgl. hierzu **Anderson 1971**, S. 203 ff.).

Mit Hilfe der obigen Grenzwerte findet man also:

$$\begin{aligned} \text{qm-lim}_{N \rightarrow \infty} 1/N \left[\partial^2 l / \partial \theta^2 \right] &= \lim_{N \rightarrow \infty} 1/N \left[\partial \theta^* / \partial \theta \right] E \left[\partial^2 l / \partial \theta^{*2} \right] \left[\partial \theta^* / \partial \theta \right]' \\ &= - \lim_{N \rightarrow \infty} (1/N) F \end{aligned}$$

Damit ist auch im Falle trajektorienweise verschiedener exogener Einflüsse gezeigt, daß **Hesse-Matrix**, **Fisher-Informationen-Matrix** und die modifizierte **Fisher-Information** (Erwartungswerte weggelassen) asymptotisch äquivalent sind.

Für trajektorienweise gleiche Exogene gelten selbstverständlich die üblichen Resultate (insbesondere ist hier $1/N \sum \mu_{nt} = \mu_t$ und $1/N \sum \mu_{nt} \mu'_{ns} = \mu_t \mu'_s$).

Mit Hilfe des **starken Gesetzes der großen Zahlen** läßt sich unter etwas anderen Bedingungen an die Exogenen sogar fast sichere Konvergenz nachweisen:

Behauptung 2: sei $\sum_n x_{nt} / n^2 < \infty$ und $\sum_n x_{nt} x'_{ns} / n^2 < \infty$ (**)

1. $1/N (\partial l / \partial \theta) \rightarrow 0$ (fast sicher)
2. $1/N YY' - 1/N E[YY'] \rightarrow 0$ (f.s.)
3. $1/N YX' - 1/N E[Y]X' \rightarrow 0$ (f.s.)

Beweis:

Obige Behauptungen folgen aus dem **starken Gesetz der großen Zahlen** von **Kolmogoroff**.

Dies läßt sich wie folgt formulieren (siehe **Shiryayev 1984**, S 364 ff.)

Sei ξ_n eine Folge unabhängiger Zufallsvariablen mit endlichen 2. Momenten.
Seien b_n positive Zahlen mit $b_n \rightarrow \infty$ und

$$\sum_n V[\xi_n] / b_n^2 < \infty$$

Dann gilt $(S_N - E[S_N]) / b_N \rightarrow 0$ (fast sicher), wobei $S_N = \sum \xi_n$.
Insbesondere, wenn $b_n = n$, gilt

$$(S_N - E[S_N]) / N \rightarrow 0 \text{ (fast sicher)}$$

1.1 Wie im Beweis der vorigen Behauptung gezeigt wurde, gilt

$$V^{abcd} = \text{Cov}[u^a u^b, u^c u^d] = V^{ac} V^{bd} + V^{ad} V^{bc} \Rightarrow \sum V^{abcd} / n^2 < \infty$$

und damit $1/N \sum_n u_{nt} u'_{nt} - 1/N E \sum_n u_{nt} u'_{nt} \rightarrow 0$ (f.s.)

Dies impliziert $W - V = 1/NT \sum_n u_{nt} u'_{nt} - V \rightarrow 0$ (f.s.)

1.2. $UY' = \sum u_{nt} y_{nt}$: Die einzelnen Summanden sind unabhängig, da

a) alle Trajektorien unabhängig sind und

b) für $t < s$ gilt $E[u_t y_t u_s y_s] = E[u_t y_t] E[u_s y_s] = 0$ und $E[u_t y_t] = 0$. (entsprechend für $t > s$).

Weiterhin ist $\text{Cov}[u_{nt}^a y_{nt}^b, u_{nt}^c y_{nt}^d] = V^{ac} (\sum_t^{bd} + \mu_{nt}^b \mu_{nt}^d)$

Daher genügt

$$\sum_n x_{nt} / n^2 < \infty \text{ und } \sum_n x_{nt} x'_{nt} / n^2 < \infty \quad (**)$$

um die Voraussetzungen des Gesetzes der großen Zahlen zu erfüllen.

1.3. $UX' = \sum_{nt} u_{nt} x'_{nt}$. $E[UX'] = 0$ und die einzelnen Summanden sind unabhängig.

Weiterhin hat man $\text{Cov}[u_{nt}^a x_{nt}^b, u_{nt}^c x_{nt}^d] = V^{ac} x_{nt}^b x_{nt}^d$. Wieder sind mit (**) die Voraussetzungen erfüllt und es folgt

$$1/N UX' - 1/N E[UX'] \rightarrow 0 \text{ (f.s.)} \Rightarrow 1/N UX' \rightarrow 0 \text{ (f.s.)}$$

2. $YY' = \sum_n (\sum_t y_{nt} y'_{nt}) = \sum_n \zeta_n$. Die einzelnen Summanden ζ_n sind unabhängig und für die Kovarianz der Komponenten gilt

$$\begin{aligned}
 \text{Cov}[\zeta_n, \zeta_n] &= \sum_{ts} \text{Cov}[y_{nt}^a y_{nt}^b, y_{ns}^c y_{ns}^d] \\
 &= \sum_{ts} \left\{ E[y_{nt}^a y_{ns}^c] E[y_{nt}^b y_{ns}^d] \right. \\
 &\quad + E[y_{nt}^a y_{ns}^d] E[y_{nt}^b y_{ns}^c] \\
 &\quad \left. - 2E[y_{nt}^a] E[y_{nt}^b] E[y_{ns}^c] E[y_{ns}^d] \right\} \\
 &= \sum_{ts} \left\{ \Sigma_{ts}^{ac} \Sigma_{ts}^{bd} + \Sigma_{ts}^{bd} \mu_{nt}^a \mu_{ns}^c + \Sigma_{ts}^{ac} \mu_{nt}^b \mu_{ns}^d \right. \\
 &\quad \left. \Sigma_{ts}^{ad} \Sigma_{ts}^{bc} + \Sigma_{ts}^{bc} \mu_{nt}^a \mu_{ns}^d + \Sigma_{ts}^{ad} \mu_{nt}^b \mu_{ns}^c \right\}
 \end{aligned}$$

wobei $\Sigma_{ts} = \text{Cov}[y_{nt}, y_{ns}]$ und $\mu_{nt} = E[y_{nt}]$

$\sum_n x_{nt}/n^2 < \infty$ und $\sum_n x_{nt} x'_{ns}/n^2 < \infty$ (***) garantiert also wiederum die Konvergenz der Summe $\sum_n \text{Var}[\zeta_n^{ab}]$ **q.e.d.**

$$3. \quad YX' = \sum_n \left(\sum_t y_{nt} x'_{nt} \right) = \sum_n \zeta_n.$$

Für die Kovarianz der Komponenten hat man

$$\begin{aligned}
 \text{Cov}[\zeta_n, \zeta_n] &= \sum_{ts} \text{Cov}[y_{nt}^a, y_{ns}^c] x_{nt}^b x_{ns}^d \\
 &= \sum_{ts} \Sigma_{ts}^{ac} x_{nt}^b x_{ns}^d
 \end{aligned}$$

Damit ist $\sum_n \text{Var}[\zeta_n^{ab}]$ konvergent, wenn $\sum_n x_{nt} x'_{ns}/n^2 < \infty$ **Q.E.D.**

Literatur

- Anderson, T.W.** (1971), *The Statistical Analysis of Time Series*, Wiley, New York
- Arató, M.** (1982), *Linear Stochastic Systems with Constant Coefficients*, Springer, Berlin, Heidelberg, New York
- Arminger, G.** (1986), *Linear Stochastic Differential Equation Models for Panel Data with Unobserved Variables*, in: **Tuma, N.B.** (ed.), *Sociological Methodology*, Washington, 187-212
- Arnold, L.** (1973), *Stochastische Differentialgleichungen*, Oldenbourg, München, Wien
- Arnold, V.I.** (1978), *Mathematical Methods of Classical Mechanics*, Springer, Heidelberg, New York
- Arnold, V.I.** (1973), *Ordinary Differential Equations*, MIT Press, Cambridge (Mass.), London
- Azencott, R., Dacunha-Castelle, D.** (1986), *Series of Irregular Observations*, Springer, New York
- Bartlett, M.S.** (1946), *On the theoretical specification and sampling properties of autocorrelated time-series*, *Journal of the Royal Statistical Society (Supplement)*, 7, 27-41
- Bartlett, M.S.** (1955, 1978), *An Introduction to Stochastic Processes*, Cambridge University Press, Cambridge
- Basawa, I.V., Prakasa Rao, B.L.S.** (1980), *Statistical Inference for Stochastic Processes*, Academic Press, London
- Bergstrom, A.R.** (1966), *Non Recursive Models as Discrete Approximations to Systems of Stochastic Differential Equations*, in: **Bergstrom** (1976).
- Bergstrom, A.R.** (ed.) (1976), *Statistical Inference in Continuous Time Models*, North Holland Amsterdam
- Bergstrom, A.R.** (1984), *Continuous Time Stochastic Models and Issues of Aggregation over Time*, in: **Griliches u. Intriligator** (1984).
- Bergstrom, A.R.** (1988), *The History of Continuous-Time Econometric Models*, *Econometric Theory*, 4, 365-383
- Caines, P.E.** (1988), *Linear Stochastic Systems*, Wiley, New York
- Coddington, E.A., Levinson, N.** (1955), *Theory of ordinary differential equations*, *McGraw-Hill*, New York
- Courant, R., Hilbert, D.** (1924), *Methoden der Mathematischen Physik*, Band I und II, Springer, Berlin, Heidelberg, New-York
- Dacunha-Castelle, D., Florens-Zmirou** (1986), *Estimation of the Coefficients of a Diffusion from Discrete Observations*, *Stochastics*, 19, 263-284

- Dembo, A., Zeitouni, O.** (1986), Parameter Estimation of Partially Observed Continuous Time Stochastic Processes Via the EM Algorithm, *Stochastic Processes and their Applications*, 23, 91-113
- Dempster, A.P., Laird, N.M., Rubin, D.B.** (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm, *J. Royal Statistical Society B*, 39, 1-38
- Dennis, J.E. Jr, Schnabel, R.B.** (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs
- Deppe, W.**(1977), *Formale Modelle in der Psychologie*, Kohlhammer, Stuttgart
- Doreian, P., Hummon, N.P.** (1976), *Modelling Social Processes*, Elsevier, New York , Oxford, Amsterdam
- Gard, T.C.** (1988), *Introduction to stochastic differential equations*, Dekker, New York
- Gelb, A.** (ed.)(1974), *Applied Optimal Estimation*, MIT Press, Cambridge, Mass., London
- Gelfand, I.M., Wilenkin, N.J.** (1964), *Verallgemeinerte Funktionen, Band 4*, VEB Deutscher Verlag der Wissenschaften, Berlin
- Golub, G.H., Van Loan, C.** (1983), *Matrix Computations*, John Hopkins University Press, Baltimore
- Goodrich, R.L., Caines, P.E.** (1979), Linear system identification from non-stationary cross sectional data, *IEEE Trans. Autom. Control*, **AC-24**(1), 125-127
- Goodrich, R.L., Caines, P.E.** (1980), Identification of time-invariant linear systems from nonstationary cross-sectional data, in: **Jacobs, O.L.R. et al.** (eds.)(1980), *Analysis and Optimization of Stochastic Systems*, Academic Press, New York, 491-507
- Griliches, Z., Intriligator, M.D.** (1983,1984), *Handbook of Econometrics*, North Holland, Amsterdam
- Gulasu, S., Shenitzer, A.** (1985), The Principle of Maximum Entropy, *Mathematical Intelligencer*, 7, 1, 42-48
- Hamerle, A., Nagl, W., Singer, H.** (1988), Parameter Estimation in Continuous Time Dynamic Systems using Panel Data, *Diskussionsbeiträge 109*, Universität Konstanz
- Hansen, L.P., Sargent, T.J.** (1983), The dimensionality of the aliasing problem in models with rational spectral density, *Econometrica*, 51, 377-387
- Hauptmann, H., Schmid, F.** (1980), Spezifikationen und Schätzprobleme bei dynamischen Systemen, in: **Hauptmann, H., Schenk, K.E.** (Hrsg.), *Anwendungen der Systemtheorie und Kybernetik in Wirtschaft und Verwaltung*, Dunker und Humblot
- Isermann, R.** (1988), *Identifikation dynamischer Systeme, Band I und II*, Springer, Berlin Heidelberg New York
- Itô, K.** (1951), On stochastic differential equations, *American Mathematical Society, Memoirs 4*, New York

- Jaynes, E.T.** (1978), Where Do We Stand on Maximum Entropy ?, in: **Levine, R. D., Tribus, M.** (eds.) (1979), *The Maximum Entropy Formalism*, MIT Press, Cambridge, Mass.
- Jennrich, R.I., Bright, P.B.** (1976), Fitting Systems of Linear Differential Equations Using Computer Generated Exact Derivatives, *Technometrics*, 18, 4, 385-392
- Jones, R.H.** (1984), Fitting Multivariate Models to Unequally Spaced Data, in: **Parzen, E.** (ed.) (1984), *Time Series Analysis of Irregularly Observed Data*, Springer, New York, 158-188
- Jones, R.H., Tryon, P.V.** (1987), Continuous time series models for unequally spaced data applied to modeling atomic clocks, *SIAM J. Sci. Stat. Comput.*, 8, 1, 71-81
- Kalbfleisch, J.D., Lawless, J.F.** (1985), The analysis of panel data under a Markoff assumption, *JASA*, 80, 863-871
- Koehler, A.B., Murphree, E.S.** (1988), A Comparison of the Akaike and Schwarz Criteria for selecting Model Order, *Applied Statistics*, 37, 2, 187-195
- Krebs, V.** (1980), *Nichtlineare Filterung*, Oldenbourg, München, Wien
- Kushner, H.J.** (1984), *Approximation and Weak Convergence Methods for Random Processes, with Applications to Stochastic Systems Theory*, MIT Press, Cambridge, Mass., London
- Le Breton, A.** (1976), On continuous and discrete sampling for parameter estimation in diffusion type processes, *Mathematical Programming Study* 5, 124-144
- Lewis, F.L.** (1986), *Optimal Estimation*, Wiley, New York
- Lighthill, M.J.** (1966), *Einführung in die Theorie der Fourier-Analyse und der verallgemeinerten Funktionen*, Bibliographisches Institut, Mannheim, Wien, Zürich
- Liptser, R.S., Shiriyayev, A.N.** (1977, 1978), *Statistics of Random Processes, Band I und II*, Springer, New York, Heidelberg, Berlin
- Louis, T.A.** (1982), Finding the Observed Information Matrix when Using the EM Algorithm, *Journal of the Royal Statistical Association B*, 44, 2, 226-233
- Magnus, J.R., Neudecker, H.** (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester
- Mardia, K.V., Kent, J.T., Bibby, J.M.** (1979), *Multivariate Analysis*, Academic Press, London
- McDonald, R.P., Swaminathan, H.** (1973), A simple matrix calculus with applications to multivariate statistics, *General Systems*, XVIII, 37-54
- McDonald, R.P.** (1976), The McDonald-Swaminathan Matrix Calculus: Clarifications, Extensions, and Illustrations, *General Systems*, XXI, 87-94
- Möbus, C. Nagl, W.** (1983), Messung, Analyse und Prognose von Veränderungen. In: *Hypothesenprüfung, Band 5 der Serie Forschungsmethoden der Psychologie der Enzyklopädie der Psychologie*, Hogrefe, Göttingen

- Nelson, E.** (1967), *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton
- Phillips, P.C.B.** (1972), The Structural Estimation of a Stochastic Differential Equation System, in: **Bergstrom** (1976).
- Phillips, P.C.B.** (1973), The Problem of Identification in Finite Parameter Continuous Time Models, in: **Bergstrom** (1976).
- Quandt, R.E.** (1983), Computational Problems and Methods, in: **Griliches u. Intriligator** (1983)
- Rao, C.R.** (1973), *Linear Statistical Inference and its Applications*, 2. Auflage, Wiley, New York
- Rauch, H.E., Tung, F., Striebel, C.T.** (1965), Maximum Likelihood Estimates of Linear Dynamic Systems, *AIAA Journal*, 3, 8, 1445-1450
- Rothenberg, T.J.** (1971), Identification in Parametric Models, *Econometrica*, 39, 3, 577-591
- Ruelle, D.** (1980), Strange Attractors, *Mathematical Intelligencer*, 2, 126-137
- Ruymgaart, P.A., Soong, T.T.** (1971), A Sample Treatment of Langevin-Type Stochastic Differential Equations, *Journal of Mathematical Analysis and Applications*, 34, 325-338
- Sargan, J.D.** (1976), Some Discrete Approximations to Continuous Time Stochastic Models, in: **Bergstrom** (1976).
- SAS User's Guide: Basics** (1985), SAS Institute, Cary, N.C.
- SAS/IML Users Guide** (1985), SAS Institute, Cary, N.C.
- Schuster, H.G.** (1984), *Deterministic Chaos*, Physik-Verlag, Weinheim
- Shiryayev, A.N.** (1984), *Probability*, Springer, New York
- Singer, B., Spilerman, S.** (1976), Representation of social processes by Markov models, *American Journal of Sociology*, 82, 1-53
- Singer, H.** (1988), Parameter Estimation in Stochastic Differential Equations: Comparison of Some Estimators, *Diskussionsbeiträge 112*, Universität Konstanz
- Singer, H., Hautzinger, M.** (1988), Kausalattributionen, Depressivität und kritische Lebensereignisse als Stochastischer Prozeß, in: **Kammer, D., Hautzinger, M.**(Hrsg.)(1988), *Kognitive Depressionsforschung*, Huber, Bern, Stuttgart, Toronto
- Soong, T.T.** (1973), *Random Differential Equations in Science and Engineering*, Academic Press, New York, London
- Swaminathan, H.** (1976), Matrix Calculus for Functions of Partitioned Matrices, *General Systems*, XXI, 95-99
- Tigelaar, H.H.** (1982), *Identification and Informative Sample Size*, Mathematisch Centrum, Amsterdam
- Tuma, N.B., Hannan, M.** (1984), *Social Dynamics*, Academic Press, New York
- Uhlenbeck, L.S., Ornstein, L.S.** (1930), On the Theory of Brownian Motion, in: **Wax, N.** (ed.)(1954), *Selected Papers on Noise and Stochastic Processes*, Dover, New York

- Urbanik, K.** (1958), Generalized Stochastic Processes, *Studia Mathematica* **16**, 268-334
- Van Kampen, N.G.**(1981), *Stochastic Processes in Physics and Chemistry*, North Holland, Amsterdam, New York Oxford
- Van Kampen, N.G.**(1987), How do stochastic processes enter into physics , in: **Albeverio, S., Blanchard, Ph., Streit, L.** (eds.)(1987), *Stochastic Processes - Mathematics and Physics II*, Springer, Berlin, Heidelberg, New York
- Watson, M.W., Engle, R.F.** (1983), Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and varying Coefficient Regression Models, *Journal of Econometrics*, **23**, 385-400
- Wilks, S.S.** (1962), *Mathematical Statistics*, Wiley, New York, London
- Wong E., Zakai, M.** (1965), On the convergence of ordinary integrals to stochastic integrals, *Ann. Math. Statistics*, **36**, 1560-1564
- Wong, E., Hajek, B.** (1985), *Stochastic Processes in Engineering Systems*, Springer, New York
- Wu, C.F.J.** (1983), On the Convergence Properties of the EM Algorithm, *Annals of Statistics* **11**, 1, 95-103
- Wymer, C.R.** (1972), Econometric Estimation of Stochastic Differential Equation Systems, in: **Bergstrom** (1976)
- Yule, G.U.** (1927), On a method for investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, *Philos. Trans. Roy. Soc. London, Ser. A*, **226**, 267-298