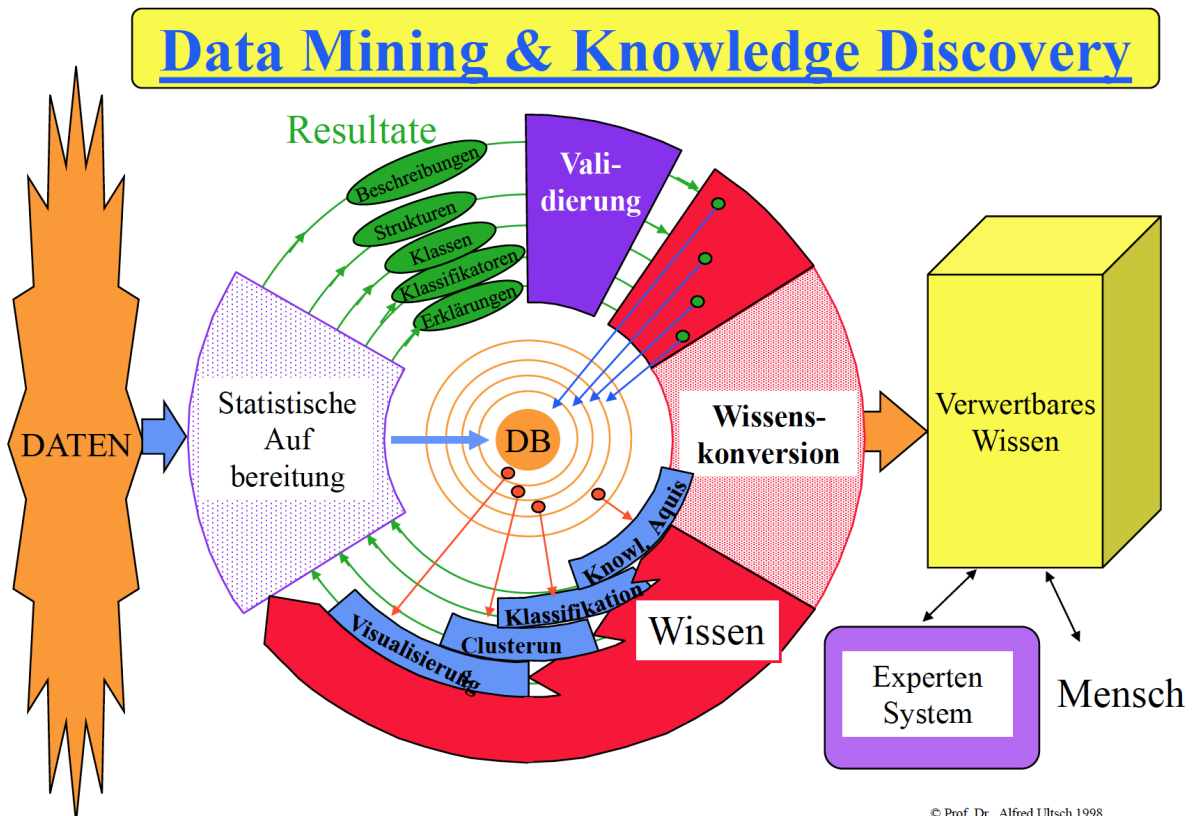
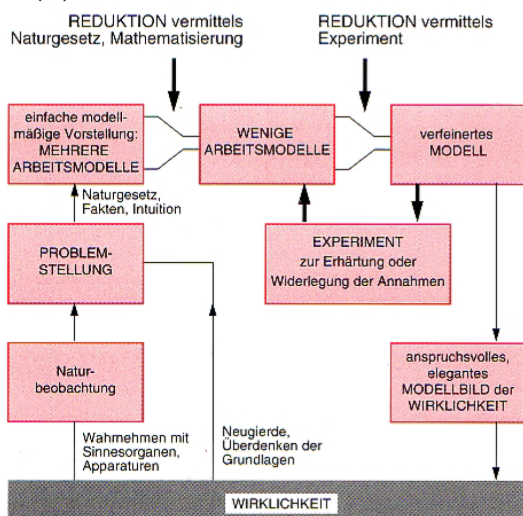


(a) Paradigma des Dataminings, in blau Methoden und in grün Gliederung

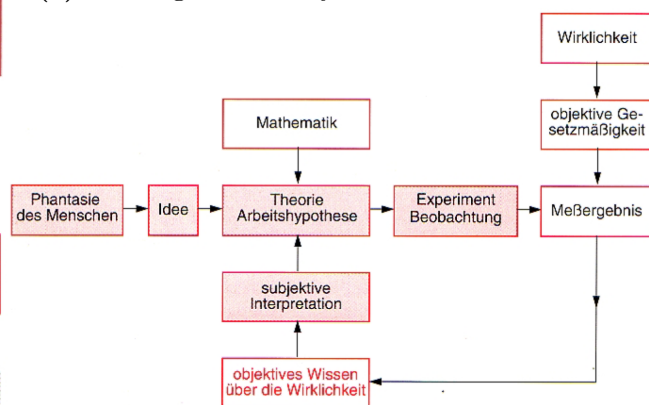


© Prof. Dr. Alfred Ultsch 1998

(b) Paradigma: Naturwissenschaften



(c) Paradigma der Physik



**Statistik:** gezielte Datensammlung bei Hypothese und Modell

**Data-Mining:** Erst aus Datensammlung folgt Wissen, kein Modell als Voraussetzung und keine Hypothese.

**Physik:** Beobachtung → Messung → mathematische Formulierung → Hypothese → Voraussagen → Experimentelle Bestätigung → Theorie → weitere Voraussage → keine Widersprüche zur Erfahrung → Naturgesetz. -

DB: Datenbank, Knowl.Aquis.: Knowledge Aquisition - Wissenserzeugung und -darstellung. B/ Bäume, falls für Menschen verständlich, Expertensystem: Automatischer Beweiser, s. KI-VI. (B/ Prolog)

# Inhaltsverzeichnis

<b>1</b>	<b>Grundlagen</b>	<b>1</b>
1.1	Beschreibung von Zeitreihen . . . . .	1
1.1.1	Renditenalternativen: Messung einer täglichen Änderung . . . . .	1
1.1.2	Grundlegende Filter . . . . .	2
1.2	Komponentenmodell . . . . .	3
1.2.1	Die glatte Komponente . . . . .	3
1.2.2	Die zyklische Komponente . . . . .	3
1.2.3	Restkomponente . . . . .	3
1.3	Vorverarbeitung von Zeitreihen . . . . .	4
<b>2</b>	<b>Zeitreihe als stochastischer Prozess</b>	<b>5</b>
2.1	Autokorrelation ACF . . . . .	5
2.2	Zeitreihenmodellierung: ARMA . . . . .	6
2.2.1	Zeitreihenmodellierung: ARIMA . . . . .	7
2.2.2	Modellfindung und Validierung . . . . .	7
2.3	Volatilitätsmodelle: ARCH und GARCH . . . . .	8
2.4	Markov-Ketten . . . . .	9
2.4.1	Hidden Markov Modelle . . . . .	9
<b>3</b>	<b>Zeitreihenanalyse mit dem Fourierprinzip</b>	<b>10</b>
3.1	Fourier Theorie . . . . .	10
3.2	Fouriertransformationen . . . . .	11
3.2.1	Spektralanalyse . . . . .	11
3.3	FT in der Praxis . . . . .	11
3.4	Anwendung der Faltung . . . . .	13
3.4.1	Beschreibung des Windowing-Effektes . . . . .	13
3.4.2	Bezug zur allgemeinen Fensterung . . . . .	13
3.4.3	Filterung . . . . .	13
3.4.4	Gabor Trafo - SFTF: Short time FT . . . . .	14
3.5	Wavelettrafo: Verallgemeinerung der Gabor-Trafo . . . . .	14
3.5.1	Arten der WT . . . . .	14
3.5.2	Haar-Wavelet . . . . .	16
3.5.3	Vergleich zu FT . . . . .	16
<b>4</b>	<b>temporale Knowledge Discovery</b>	<b>17</b>
4.1	Knowledge Discovery . . . . .	17
4.2	Dynamic Time Warping - DTW . . . . .	17
4.3	Wissensdarstellung von ZR . . . . .	18

# Kapitel 1

## Grundlagen

### 1.1. Beschreibung von Zeitreihen

Aus KD (DataMining I) ist Theorem von Bayes, Gesetz der großen Zahlen sowie Beschreibungen einer zeitunabhängigen Variablen insbesondere auch Korrelationen bekannt. Begriffsauflistung:

- Zeitreihe: Menge von zeitlich geordneten Vektoren, oft periodisch
- univariat: 1 Zeitreihe
- multivariat: mehrere Zeitreihen
- Skedastizität:  $\text{Varianz}(t) = \text{const}$
- Autokorrelation: die Korrelation einer Zeitreihe  $f(t)$  (Funktion/Signal) mit sich selbst zu einem verschobenen Zeitpunkt  $f(t + \tau)$ , Zshg weißes Rauschen?
- Abtastrate: Wiederholfrequenz der zeitlichen Messung
- kontinuierliches Abtasten: Zeitreihen bei Prozessbeschreibungen
- Monte-Carlo Ansatz: generieren von Daten und Vergleich mit theoretischen Verteilungen
- Cauchy-Verteilung: Quotient aus zwei Normalverteilungen. Folge: Es existiert kein Erwartungswert, Mittelwert oder Varianz, sondern nur ein Modus
- weißes Rauschen: Zufälliger Prozess, z.B. Brownsche Bewegung, Random Walk

Das eine Zeitreihe weißem Rauschen entspricht, ist generell schwierig zu zeigen. Von der Perspektive des Dataminings kann eine visuelle Strukturanalyse durchgeführt werden.

Zeitreihen sollten trendbereinigt sein: Der Verlauf sollte um Null schwanken, also weder im gesamten Verlauf steigen noch fallen.

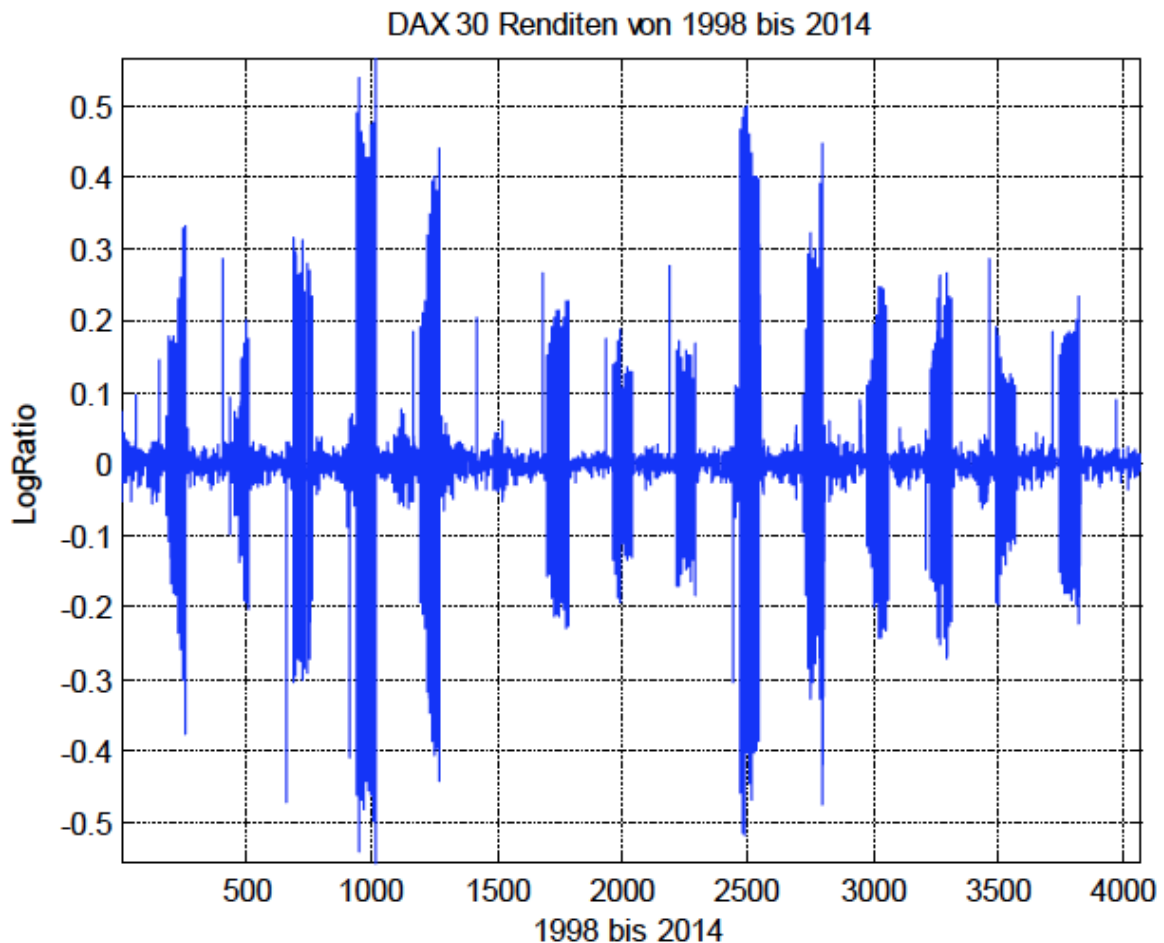
#### 1.1.1. Renditenalternativen: Messung einer täglichen Änderung

Prinzipiell gibt es mehrere Maße um Änderungen innerhalb der Zeitreihe abzubilden. Eine davon ist in Abb. 1.1-1 gezeigt. Der Begriff Volatilität ist sehr schwammig definiert, entspricht aber nicht der täglichen Änderung<sup>1</sup>. Es wird damit eher die temporale der zeitlichen Varianz angedeutet.

- %: Relative Änderung in der Wirtschaft als **Rendite** bezeichnet
- Log-Ratio
- Relative Differenz, Definition im nächsten Unterkapitel

Ein Hinweis für Cluster ist eine leptokurtische Verteilung der Wahrscheinlichkeitsdichte der Volatilität. Es ist möglich die Rendite in ein besseres Maß einer Änderung der Zeitreihe überzuführen. Aus Sicht des Knowledge Discovery ist Relative Differenz daher der Rendite bzw, der LogRatio und auch den Prozentsätzen vorzuziehen, weil so eine bessere Distanz definiert werden kann und die Möglichkeit einer Cauchy-Verteilung ausgeschlossen wird.

<sup>1</sup>vllt Standardabweichung der relativen Änderung des Parameters, welcher betrachtet wird.



**Abb. 1.1-1:** Heteroskedastizität: Volatilität der Zeitreihe bildet viele Cluster: Im Bild sind viele „Knubbel“/Schocks zu sehen, d.h. Zeitreihe unterliegt großer Schwankung

### 1.1.2. Grundlegende Filter

Ein **linearer Filter** ist eine Transformation der Daten mit:

$$y(t) = x(t) * L = \sum_{i=t-lag}^{t+lead} a_i * x_i(t) \quad (1.1)$$

Ein Beispiel ist der gleitende Durchschnitt, auch moving average genannt. Mittels eines linearen Filters lassen sich Ausreißer durch Glättung eliminieren.

Der **Differenzfilter** 1.Ordnung ist definiert mit

$$\Delta(x(t)) = x_t - x_{t-1} \quad (1.2)$$

Im Vergleich ist die relative Differenz definiert durch

$$\Delta(x(t), x(t+1)) = \frac{x_{t+1} - x_t}{\frac{1}{2}(x_{t+1} + x_t)} \quad (1.3)$$

Dadurch kann eine Rendite als normierter Differenzfilter 1.Ordnung angesehen werden. Der **Differenzfilter** 2.Ordnung ist definiert mit

$$\Delta^2(x(t)) = x_t - 2x_{t-1} + x_{t-2} \quad (1.4)$$

Dadurch ergibt sich die allgemeine Definition eines Differenzfilters der p-ten Ordnung:

$$\Delta^p(x(t)) = \Delta^{p-1}x_t - \Delta^{p-1}x_{t-1} \quad (1.5)$$

## 1.2. Komponentenmodell

Hier wird die Zeitreihe nur modelliert und nicht analysiert. Die Annahme ist, dass sie aus den folgenden vier Komponenten besteht.

1. Trend, z.B. linear oder quadratisch
2. Konjunktur: merjährige nicht notwendigerweise regelmäßige Schwankung
3. Saison: periodisch
4. Rest: Gaußförmig,  $\rightarrow \epsilon$

Diese Komponenten können additiert oder multipliziert werden. Eine solche Modellierung ist nicht eindeutig. Qualität des Modells hängt stark von den Annahmen ab, weil z.B. Trend von der Konjunktur nur schlecht abzugrenzen ist. Das Modell wird nach den Kriterien der Erklärbarkeit und Einfachheit ausgewählt.

### 1.2.1. Die glatte Komponente

Der Trend (und die Konjunktur) ist eine langfristige systematische Veränderung des mittleren Niveaus und wird über Polynome niedrigen Grades modelliert. Möglich ist dies über eine lineare Filterung oder eine Regressionsanalyse. Der Grad des Polynoms kann über einen Scree-Plot und die Qualität über die Residuen bestimmt werden<sup>2</sup>. Zum Schluss wird die Differenz von Zeitreihe und Trendmodellierung gebildet. Dadurch ist die Zeitreihe trendbereinigt. In späteren Kapiteln werden wavelets verwendet. Häufig werden auch Konjunktur und Trend zur glatten Komponente zusammengefasst.

### Erklärung der Analysemöglichkeiten

Verfahren der variaten Differenzen. Differenzenfilter 1. Ordnung= linearer Filter

Tipp: Nach Subtraktion von Trend aus Zeitreihe kann man das Schwanken um Null herum mit einem qqPlot prüfen.

### 1.2.2. Die zyklische Komponente

Die zyklische Komponente (Konjunktur+Saison) kann über eine Fourieranalyse erfolgen. Im Komponenten Modell wird jedoch oft der gleitende Durchschnitt mit einer Ordnung (lag) benutzt.

Die Restkomponente ergibt sich aus dem Residuum (=Daten-(Trend+Saison)).

### Ermittlung der zyklischen Komponente

- gleitender Durchschnitt(Moving Average): Durchschnitt von versetzten Zeitfenstern rückwärts
- Fourieranalyse

### 1.2.3. Restkomponente

Es werden die Residuen des gesamten Modells mit Methoden der Knowledge Discovery untersucht. Beachte: keine Autokorrelation, EW=0, keine Zeitliche Struktur, d.h. homoskedastizität, normalverteilt.

<sup>2</sup>RMS gegen Polynomgrad auftragen

Sind die Residuum um Null verteilt so ist nur noch weißes Rauschen vorhanden. Es Folgt daher einer ACF innerhalb der statistischen Grenzen.

### 1.3. Vorverarbeitung von Zeitreihen

Für eine mit den folgenden Modellen zu analysierende Zeitreihe sollte folgendes gegeben sein:

- mittelwertbereinigt: Erwartungswert/Mittelwert  $M(t)=const$
- trend und seasonsbereinigt
- homoskedastisch, Varianz  $Var(t) = const$ , d.h. Varianz von heute unabhängig von der Varianz von gestern
- : Achtung: weitere zeitliche Struktur ist möglich!

Bei nur noch weißem Rauschen ist nun  $M=0$ ,  $S=1$ ,  $W(t) = \epsilon(t)$ . Einfachste Modelle wären Random Walk und Brownscher Prozess.

## Kapitel 2

# Zeitreihe als stochastischer Prozess

Die Motivation ist, dass die Zeitreihe auch vom Zufall abhängig ist. Die Zeitreihe wurde *vorverarbeitet* und ist in den folgenden Modellen deswegen eine Realisierung einer Folge korrelierter Zufallsvariablen. Das bezeichnet einen stochastischen Prozess. Das Ziel dieses Kapitels ist es, die Zeitreihe so zu Verarbeiten, dass nur noch weißes Rauschen übrig bleibt (Abb. 2.1-1). Die Koeffizienten, die wir bei den Verarbeitungsschritten erhalten, beschreiben dann diese Zeitreihe. Begriffe:

- schwach stationär; Erwartungswert und Varianz sind zeitunabhängig
- weißes Rauschen: schwach stationär mit  $M=0$ ,  $S=1$
- heteroskedastisch:  $\text{Varianz}(t)$  ungleich 0
- mittelwertbereinig:  $M(t)=0$

gebräuchlichste Abkürzungen:

- AR(p): Autoregressive Prozesse
- MA(q): Moving Average Prozesse
- ACF: Autokorrelationsfunktion
- pACF: partielle ACF

## 2.1. Autokorrelation ACF

ACF  $\rightarrow$  heutiger Tag ist von Gestern abhängig. Sei  $M(x)$  der Mittelwert der Zeitreihe, so wird bei univariaten Zeitreihen der ACF bestimmt über

$$ACF(lag) = \frac{1}{n\sigma} \sum_t (x_t - M(x))(x_{t+lag} - M(x)) \quad (2.1)$$

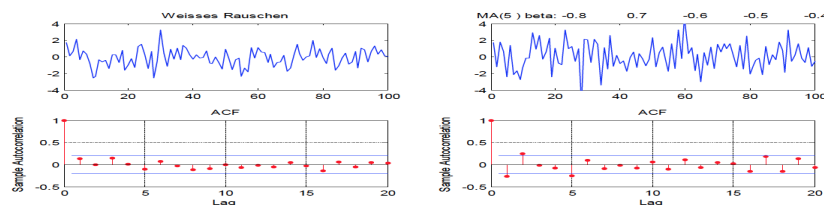
Weißes Rauschen hat die Gestalt

$$W(t) = c + rand \quad (2.2)$$

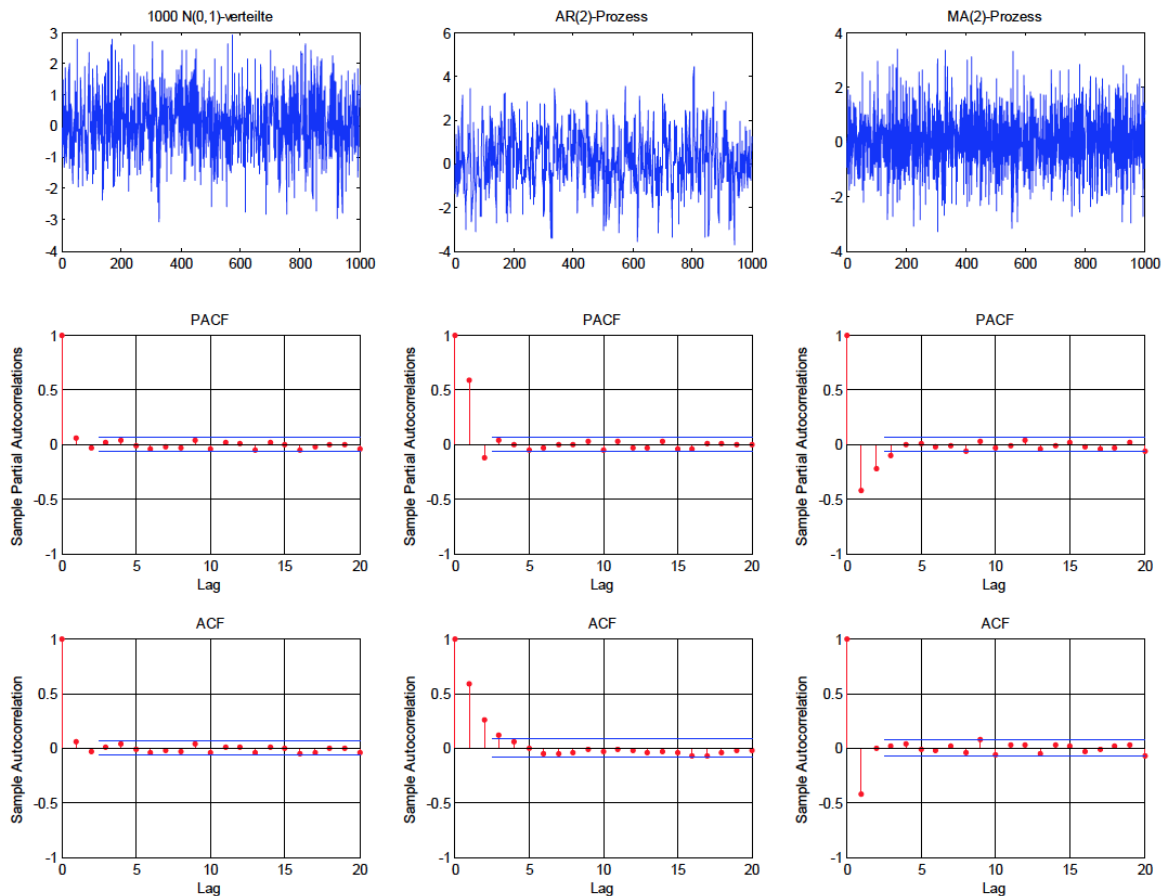
wobei  $c$  ein Mittelwert und  $rand$  eine standardnormalverteilte Zufallsvariable ist.

Folgerung: Wenn über 95% der Autokorrelationskoeffizienten  $ACF(Lag)$  einen Wert im Intervall  $[-2\sqrt{1/n}, 2\sqrt{1/n}]$  haben (farbige horizontale Linien im Korrelogram), so ist dies ein weißes Rauschen. (Statistischer Test: portmanteau test). Der Graph dieser Funktion heißt Korrelogramm. ACF strebt gegen Null, falls  $lag > q$

- Mit **ACF** wird empirisch entschieden, ob ein MA(q) Modell geeignet ist
- Mit **pACF** wird empirisch entschieden, ob ein AR(p) Modell geeignet ist



**Abb. 2.1-1:** Der Vergleich von weißem Rauschen und einem AR(5) Modell



**Abb. 2.1-2:** Die **pACF** besitzt im Gegensatz zur **ACF** ein Korrelogramm, dass bei einem  $AR(p)$  Prozess bei  $p$  abbricht; bei einem  $MA(q)$ -Prozess hingegen langsam ausklingt.

## 2.2. Zeitreihenmodellierung: ARMA

Die Folgenden Modelle besitzt Voraussetzungen, welche kaum geprüft werden, weswegen man von einer Modellierung unter bestimmten Annahmen spricht, aber nicht von einer Analyse der Zeitreihe.  $AR(p)$  alleine gibt nur eine kurzfristige Prognose, da keine Periode betrachtet wird. Wenn das Ergebnis der Modellierung innerhalb von den Schranken ist (also statistisch Null), wurde der Prozess als  $AR$ -Prozess erfolgreich modelliert. Der Wert auf der x-Achse entspricht  $p$  (Abb. 2.1-2). Definition: Ein Prozess  $X(t)$  ist ein  $AR(p)$   $\Leftrightarrow$

$$X(t) = \sum_{\tau=1}^q \alpha_{\tau} X(t - \tau) + \epsilon(\tau) \quad (2.3)$$

- $\alpha_{\tau}$  sind die Parameter des Modells
- Für einen  $AR(p)$  Prozess gilt, dass  $pACF(p) = \alpha_p$

$MA(q)$  ist die gewichtete Summe der Vergangenheit und kann als Filter angesehen werden. Definition: Ein Prozess  $X(t)$  ist ein  $MA(q)$   $\Leftrightarrow$

$$X(t) = \sum_{\tau=1}^q \beta_{\tau} \epsilon(t - \tau) + \epsilon(t) \quad (2.4)$$

- $\beta_{\tau}$  sind die Parameter des Modells
- $\epsilon(\tau)$  ist weißes Rauschen



- $\langle X(t) \rangle = 0$
- $Var(X(t)) = \sum_{\tau=1}^q \beta_{\tau}^2$
- Jeder Wert zum Zeitpunkt  $t$  ist ein gewichtetes Mittel seiner Vorgänger verbunden mit einem weißen Rauschen

Beachte:

- Voraussetzung: Zeitreihe stationär
- Bei Invertierbarkeit: Eindeutigkeit des MA-Prozesses zu gegebener **ACF**
- gewöhnlich  $q < 5$
- MA  $\rightarrow$  In dieser Modellierung wird weißes Rauschen gefiltert.
- AR  $\rightarrow$  Filter zum Zeitpunkt davor, also aus sich selber  $\rightarrow$  Autoregression

Box-Jenkins Modelle sind Modelle, die Sowohl **MA** als auch **AR** benutzen:

### 2.2.1. Zeitreihenmodellierung: ARIMA

Für nicht stationäre Zeitreihen kann die differenzierte Zeitreihe auf ARMA verwendet werden. Dies heißt dann ARIMA. Im Klartext ARMA( $p, q$ ) mithilfe eines  $i$ -fachen Differenzfilters ergibt ARIMA( $i, p, q$ ). ARIMA-Modelle bestehen aus folgenden drei Teilen:

- gemittelte Summe der Vorgänger
- gemittelte Fehlersumme der Vorgänger
- Dem Entfernen von Trends und Nicht-Stationarität

In der Praxis<sup>1</sup> werden vor allem ARMA Modelle mit  $p+q < 3$  verwendet. Generell gilt: Ist kein ARMA-Modell möglich, wird von einem ARIMA Modell ausgegangen. Hierfür werden Daten nach einer  $i$ -ten Differenzierung verwendet. Durch Ausprobieren vereinfacht dies die Entscheidungsfindung.

### 2.2.2. Modellfindung und Validierung

1. Option: Verändere Daten, so dass Die Varianz stationär ist
2. Option: Sind die Daten nicht stationär, differenziere sie solange, bis sie stationär werden.

Wenn ACF oder pACF rapide gegen Null innerhalb der statistischen Grenzen fallen, so ist die Zeitreihe stationär. Die Differenz der Zeitreihe und des Modells ergibt das Residuum. Ist dieses ein weißes Rauschen, so war die Modellierung erfolgreich. Die Güte eines Modells lässt sich auch mithilfe eines Hold-Out-Set-Prinzips prüfen.

<sup>1</sup>SCHLITTEGEN, RAINER und BERND H. J. STREITBERG: Zeitreihenanalyse

**Tab. 2.2-1:** Box-Jenkins Modelle. - pACF verhält sich umgekehrt(dual) zur ACF, s. aa Abb. 2.1-2

•	ACF	pACF
AR( $p$ )	unendlich	pACF( $i$ )=0, $i > p$
MA( $q$ )	ACF( $i$ )=0, $i > q$	unendlich
ARMA( $p, q$ )	Sinus oder gedämpfte Exp	Sinus oder gedämpfte Exp

## 2.3. Volatilitätsmodelle: ARCH und GARCH

Das Folgende Modell besitzt Voraussetzungen, welche kaum geprüft werden, weswegen man von einer Modellierung unter bestimmten Annahmen spricht, aber nicht von einer Analyse der Zeitreihe.

- ARCH(p)=autoregressive conditional heteroskedasticity
- p= Anzahl der Tage, welche man zurückblicken muss
- GARCH(p,q)=Generalisiertes ARCH
- Voraussetzung ist eine berechnete Volatilität der Zeitreihe  $Y(t)$ , welche leptokurtisch verteilt ist
- der wesentliche Unterschied zu ARMA besteht in der Persistenz von Schocks

ARCH(p) ist eine Art von einer gewichteten Summe. Die Annahme ist also das die Varianz von der Vergangenheit abhängt. In der Zeitreihe sind dann viele „Knubbel“ bzw. Schocks zu sehen s. Abb. 1.1-1. Präziser ausgedrückt ist die heutige Varianz abhängig von Zufallsfehlern der Zeitreihe einer vorherigen Periode. Große und kleine Fehlern tendieren dabei in Gruppen aufzutreten.

GARCH-Modelle gehen von einer autoregressiv Volatilität aus, daher werden GARCH-Modelle auf Renditen angewendet. Sei  $p=1$  so ist definiert

$$Y(t) = \epsilon_t \sqrt{\alpha_0 + \alpha_1 Y_{t-1}^2} \quad (2.5)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 \quad (2.6)$$

Die Varianz  $\sigma$  über die komplette Zeitreihe lautet dann

$$\sigma = \frac{\alpha_0}{1 - \alpha_0} \quad (2.7)$$

Das ARCH(p) Modell wird durch Erweiterung um autoregressive Terme der Volatilität (Abb. 1.1-1) zu einem **GARCH(p,q)** verallgemeinert werden („Arma auf Varianz“). Nun hängt die Varianz nicht nur von der „Geschichte“ der Zeitreihe ab sondern auch von der vorherigen Varianz einer Periode. Es wird dadurch die heutige Varianz aus Linearkombination vorheriger Varianzen modelliert. Praxis:

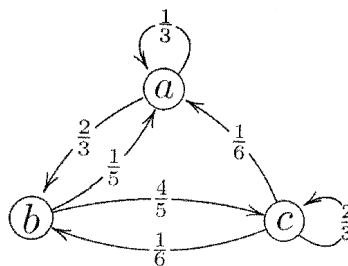
- en vague Finanzökonomie
- Volatilität bzgl. Vorzeichen: Asymetrie der Varianz, also vergangener Schocks<sup>2</sup> nicht abbildbar: Gute Nachrichten haben den selben Einfluss wie schlechte Nachrichten
- GARCH(1,1) häufig ausreichend
- Parameter p und q von Garch(p,q) lassen sich mit ACF und pACF begründen

Bemerkung:

- geschätzte Modellparameter  $\alpha_i$  durch iterative Maximum-Likelihood-Methode mit dem Startwert des Yule-Walker-Schätzers
- $p=14$  für Marktmodelle sinnvoll => Nachteil: komplizierte Modelle

<sup>2</sup>Serie aufeinanderfolgender großer Ausschläge

## 2.4. Markov-Ketten



**Abb. 2.4-1:** Beispiel einer Markov-Kette

- endlicher Automat aus Zuständen in Kreisen und Übergängen als Kanten
- Übergänge zwischen Zuständen  $(i,j)$  wahrscheinlichkeitsbasierend mit  $p(i,j) \in P$
- Die *Ordnung* des Modells ist über die zeitliche Abhängigkeit zwischen den Zuständen definiert.
- Markov-Eigenschaft: Kette vergisst ihre Vergangenheit
- Übergangswahrscheinlichkeit kann zeitlich unabhängig sein: zeitlich homogene Markov-Kette
- Übergangsmatrix  $P$ : stochastische Matrix mit Übergangswahrscheinlichkeiten als Elemente
- Zeilensumme der Matrix  $P$  ist immer eins
- Existiert ein Übergang nicht, so ist das Element gleich null

Praxis: Die Kette aus Zuständen ist, abgesehen von letztvorherigen Zustand, unabhängig von der Vergangenheit.  $\Rightarrow$  Modell *1.Ordnung*.

Der Übergang zur Zeitreihe erfolgt mit dem state-emission Modell: Kommen wir in Zustand  $a$ , so ziehen wir zufällig eine Zahl aus einer Verteilung, d.h. Ausgabesymbol entspricht der Zeitreihe.

### 2.4.1. Hidden Markov Modelle

„Hidden“, da nur Zeitreihe gegeben, aber keine eigentlichen Zustände. Zustände können nicht beobachtet werden. Die Ausgabe in einem Zustand ist also probabilistisch, da weder Übergangswahrscheinlichkeit noch Zustände bekannt sind. Die Anzahl muss über ein Clusterverfahren oder aus Expertenwissen vorgegeben werden. Die Zustände sollten eine Semantik besitzen. Voraussetzung für Verständnis sind aus der KD:

- kmeans: Zustandsanzahlwahl
- GMM: Zustandsanzahlwahl
- EM-Algorithmus: Bestimmt die posteriori Wahrscheinlichkeit heuristisch iterativ

Verfahren:

1. Wahl der Anzahl von Zuständen (KD)
2. Der Baum-Welch-Algorithmus (generalisierter EM) bestimmt die Übergangsmatrix  $P$  und Outputwahrscheinlichkeit  $Q$ : Ein Trick ist die Position der Zentroide von kmeans über GMM zu wählen.
3. Viterbi-Algorithmus: Bestimmung der wahrscheinlichen Zustandsfolge bei gegebener Beobachtung, d.h. Clustering wird als Output wiedergegeben

Nachteil: Durch EM starke Abhängigkeit von Anfangsparametern und Anzahl an Zuständen.

## Kapitel 3

# Zeitreihenanalyse mit dem Fourierprinzip

Begriffe:

1. Fourierreihe
2. FT Berechnungszeit  $N^2$
3. DFT:  $f(t_n)$  nur zu diskreten Zeitpunkten bekannt äquidistante Abtastrate
4. IDFT: inverse DFT
5. FFT: DFT für eine Faktorisierung rekursiver durch DFTs halber Länge zu berechnen  $\rightarrow$  FAST FT mit  $N = 2^k$  in Berechnungszeit:  $N \log_2 N$

### 3.1. Fourier Theorie

Bekannt sollten sein

- Periode  $T$ : temporäres Teilintervall einer Zeitreihe, welches sich wiederholt. falls nicht erkennbar, erfolgt die Annahme, dass die Periode der Zeitdauer der Länge der kompletten gemessenen Zeitreihe entspricht.
- $T$  Periode  $\Leftrightarrow f(t) = f(t + T)$
- (Grund-)Frequenz  $f = \frac{1}{T}$
- $\Rightarrow$  höchste darstellbare Frequenz:  $0.5 * N * f$  bei  $N$  Abtastwerten ist **Nyquist-Frequenz**
- Kreisfrequenz  $\omega = 2\pi f$
- Phase  $\phi$ : Verschiebungsfaktor eines periodischen Vorgangs
- Amplitude: Höhe einer Schwingung, Vorfaktor von Sinus/Kosinus
- Sin, Cos sowie ihre Additionstheoreme
- komplexe Zahlen und eulersche Gleichung:  $e^{ix} = \cos(x) + i\sin(x)$

Satz: Jede beliebige Zeitreihe lässt sich als Summe von Sinus und Kosinusanteilen darstellen:

$$X_n(t) = \frac{a_0}{2} + \sum_{k=1}^n a_k \cos(k * \omega t) + b_k \sin(k * \omega t) \quad (3.1)$$

In Gl. 3.1 wird  $a_0$  als DC oder Offset definiert und ist der Mittelwert der Zeitreihe.

Satz: Jede beliebige stetige Funktion lässt sich beliebig genau als Fourierreihe darstellen

Die Komplexe Fourier-Reihe in Gl. 3.2 ist eine andere Form von Gl. 3.1, bei welcher die Eulersche Formel benutzt wurde ( $n \rightarrow \infty$ ):

$$f(t) = \sum_{k=-n}^n c_k e^{ik\omega t} \quad (3.2)$$

Die Fourier-Analyse ist dann die Berechnung der Fourier-Koeffizienten.

## 3.2. Fouriertransformationen

Die allgemeine Fouriertransformation lauten, wie folgt:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega \quad (3.3)$$

Und die Rücktransformation

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt \quad (3.4)$$

In der Praxis entstehen zum Einem Quantisierungsfehler: Durch Endlichkeit der Zahlendarstellung entsteht eine Diskretisierung, d.h analoge Signale werden digitalisiert.

Zudem wird die Zeitreihe künstlich/erzwungen periodisch gemacht. Das führt zu einem **Windowing-Effekt**:

- größte mögliche Periode=Dauer des Fensters
- tiefste Frequenz: Grundfrequenz des Fensters
- höchste Frequenz: höchster Wert unterhalb des Abtastspektrums
- Verzerrungsfehler bei Fourierkoeffizienten
- Prognose deswegen nur eingeschränkt möglich
- **Trick**: Spiegelung, und danach Punktspiegelung der Zeitreihe, um Periodizität zu erreichen

### 3.2.1. Spektralanalyse

Ist die Analyse der Energie, *dabeigilt*  $E = hf$ , folglich: Bestimmung von Frequenzen, welche mit möglichst wenigen Fourierkoeffizienten die Zeitreihe rekonstruieren. Defakto wird die Zeitreihe der der Fouriertransformation geplottet. Def. Power-Spektrum:

$$P(\omega) = |F(\omega)|^2 \quad (3.5)$$

Selektion der Frequenzen erfolgt über:

- Power-Spektrum P: Überproportional viele tiefe Frequenzen
- LogPower-Spektrum  $\ln(P)$ : relativiert den Einfluss tiefer Frequenzen
- Power(db)-Spektrum  $\log_{10}(\frac{P}{f_0})$ : starkes Erhöhen der Amplituden mit wenig Power

## 3.3. FT in der Praxis

Zeitreihe wird mit DFT zu Fourierkoeffizienten überführt, wenn  $n = 2^k$ . Dadurch lässt sich jede Zeitreihe über Grundfrequenz<sup>1</sup> und ihre ganzen Vielfachen darstellen. Die Grundfrequenz entspricht gleichzeitig dem zeitlichen Mittelwert der Amplitude?

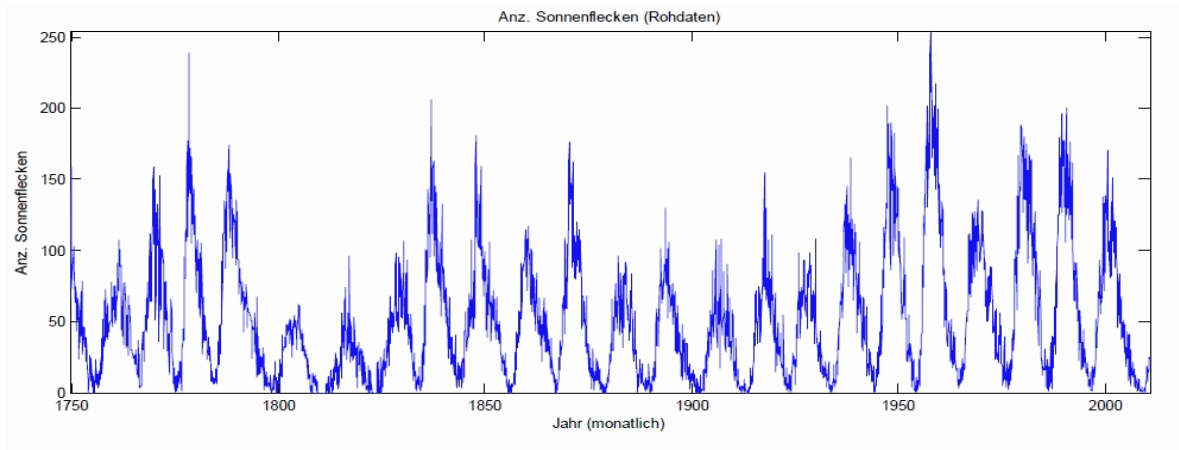
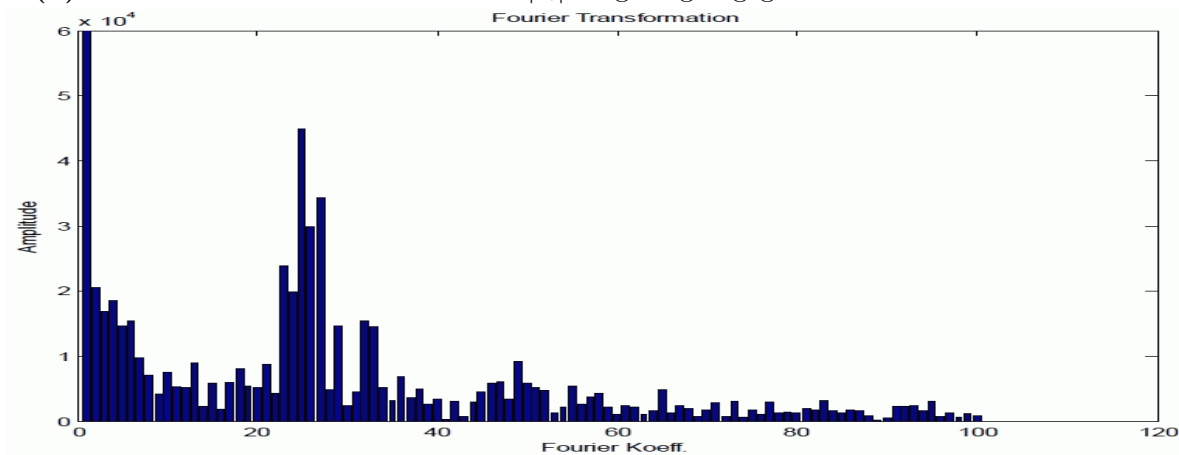
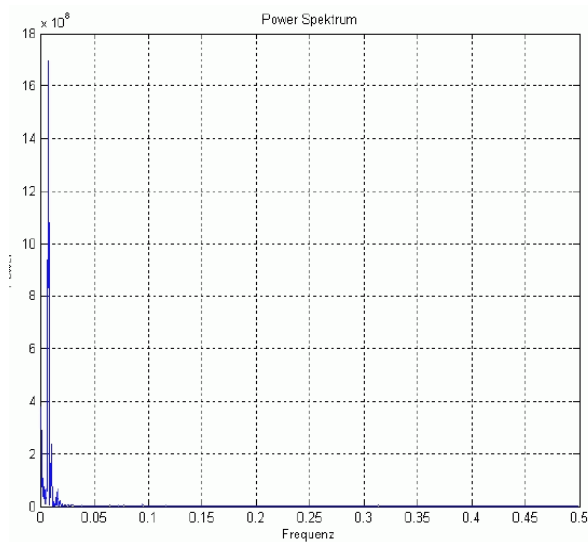
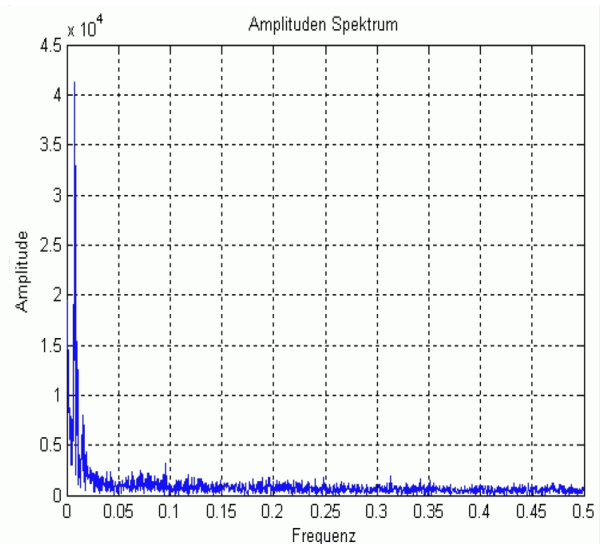
- weißes Rauschen ist im Power-Spektrum gleichverteilt
- Durch die Nyquist Grenze besitzen relevante Frequenzen Indizes  $i < \frac{1}{\text{Nyquistzahl}}$

Die Amplitude  $A_i$  der iten Frequenz ist definiert über Gl. 3.2 mit

$$A_i = \sqrt{RE(c_i)^2 + IM(c_i)^2} \quad (3.6)$$

<sup>1</sup>auch nullte Oberschwingung

(a) Beispiel einer Zeitreihe

(b) Die Größe der Fourierkoeffizienten  $|c_i|$  aufgetragen gegen i(c) Powerspektrum der ZR:  $P(\omega) = |F(\omega)|^2$ (d) Amplitudenspektrum der ZR:  $A_i = |c_i|$ 

**Abb. 3.3-1:** Bemerke die Unterschiede in der X-Achsenauftragung: Abb. 3.3-1a zeigt den Zeitverlauf. Abb. 3.3-1b Indiziert die Koeffizienten. Abb. 3.3-1c und Abb. 3.3-1d ist gegen die Frequenzen aufgetragen bis zur Nyquist-Grenze.

### 3.4. Anwendung der Faltung

Windowing-Effekt in 3.2: Durch die FT wird die endliche Zeitreihe durch einfaches Anstückeln periodisch gemacht. Oft Passen diese Endstücke nicht zueinander. Dadurch entstehen erhebliche Verzerrungen der Fourierkoeffizienten

#### 3.4.1. Beschreibung des Windowing-Effektes

Die Faltung zweier Zeitsignale im Zeitbereich entspricht der Multiplikation ihrer FT im Frequenzbereich:

$$f(t) * g(t) = IFT(FT(f(t))FT(g(t))) \quad \text{label: faltung} \quad (3.7)$$

Dabei war die Faltung im reelem Intervall definiert

$$f(t) * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad (3.8)$$

Die Faltung kann als Skalarprodukt und dadurch auch als Distanz gedeutet werden.

#### 3.4.2. Bezug zur allgemeinen Fensterung

Schneidet man aus einer Funktion ein endliches Stück aus, so kann dies als eine Multiplikation mit einer Fensterfunktion  $g$  verstanden werden, welche als zwei  $\theta$  Funktionen  $n$  (Rechteckfenster) definiert ist. Kurz: Frequenzbereich gefaltet  $\Leftrightarrow$  Zeitreihe gefenstert.

Aus dem Faltungssatz in Gl. ?? folgt

$$FT(f(t)g(t)) = FT(f(t))FT(g(t)) \quad (3.9)$$

Das bedeutet: Im Spektrum der FT der gefensterten Zeitreihe findet sich Frequenzanteile der Zeitreihe als auch der Fensterfunktion. Denn die rechteckige Fensterfunktion entspricht letztlich der Messung der Zeitreihe, d.h. die Beobachtung hat einen Einfluss auf die Messung  $\Rightarrow$  Heisenbergsche Unschärferelation.

Problem: rechteckige Fenster besitzen ungünstige Eigenschaften und die FT Koeffizienten werden deutlich beeinflusst. Die Lösung ist es sanft abklingende Fenster zu wählen:

- Gauss  $\rightarrow$  Gabor
- Parzen
- Welch
- cosinus

#### 3.4.3. Filterung

Eine Filterung kann immer als eine Faltung eines Signals  $f$  mit einer Filterfunktion  $g$  beschrieben werden. Das bedeutet das Frequenzspektrum des Signals wird mit der Übertragungsfunktion des Filters gewichtet.

- Tiefpass: hohe Frequenzen werden unterdrückt  $\Rightarrow$  Moving Average MA
- Bandpass: Frequenzen außerhalb eines bestimmten Bereiches werden unterdrückt
- Hochpass: tiefe Frequenzen werden unterdrückt
- Bandsperre: Frequenzen innerhalb eines bestimmten Bereiches werden unterdrückt

Ziel: selektive Auswahl von Frequenzen zum Zweck der Identifizierung verschiedener Komponenten einer Zeitreihe wie Grundfrequenz und Rauschen.

### 3.4.4. Gabor Trafo - STFT: Short time FT

Nachteile der FT sind, dass kein Abklingen des Signals sowie keine Berücksichtigung des lokalen Charakters der Zeitreihe existiert. (Beispiel: Fensterfunktion gewählt => Lokalisierung zeitlich und Frequenz fiert => kurzlebige Phänomene, wie Singularitäten, werden nicht erkannt.)

Deswegen wird nun eine spezielle gefensterte Fourier-Transformation eingeführt, welche eine optimale Zeit-Frequenz-Auflösung besitzt. Diese STFT kann lokale Änderungen in der Frequenz darstellen.

Hintergrund: Mit dem Fenster wird die Periode und somit die höchste Frequenz (Nyquist) festgelegt. Es ist eine Analogie zur Heisenbergschen Unschärferelation zwischen Energie und Zeit:

- viele hohe Frequenzen also großes Fenster  $\Leftrightarrow$  ungenaue/geringe zeitliche Lokalisation, d.h. wann etwas stattfindet
- nur kleines Fenster, also wenige hohe Funktionen  $\Leftrightarrow$  scharfe zeitliche Lokalisation
- optimale Fenstergröße wird mit Gaußfunktionen erreicht

Der beste Kompromiss ist die Gabor-Transformation definiert über eine Gaußfunktion bei einer  $\sigma$  Auflösung und der Fläche  $= \sigma_t \sigma_\omega$ . Dies wird auch als Short-time FT von der Zeitreihe mit Gaußfenster beschrieben.

## 3.5. Wavelettrafo: Verallgemeinerung der Gabor-Trafo

Wähle eine beliebige Welle(schen)(Funktion) mit Eigenschaften:

- quadratintegrabel  $\Leftrightarrow$  „Energieerhaltung“
- Breite des Waveletts=Skala und außerhalb konvergiert schnell gegen null
- skalierbar
- verschiebbar

und der Transformationsdefinition

$$Wt f(m, s) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-m}{s}\right) dt \quad (3.10)$$

Das Wavelet in Abb. 3.5-1a wird solange gestaucht und verschoben bis es der Zeitreihe entspricht. Wie bei der STFT wird das zu untersuchende Signal also mit der Fensterfunktion verglichen. Aber es wird nicht die Zeitreihe verschoben und moduliert (Verschiebung im Frequenzbereich) sondern das Fenster. Die Bedingung lautet dabei, wie gut korreliert das Wavelet mit der darzustellenden Funktion/Zeitreihe.

In der Praxis eignet sich diese Trafo gegenüber FT besser wenn Wissen über die Zeitreihe besteht

### 3.5.1. Arten der WT

Sei das Wavelet mit  $\Psi$  definiert, so existieren es folgende Transformationen:

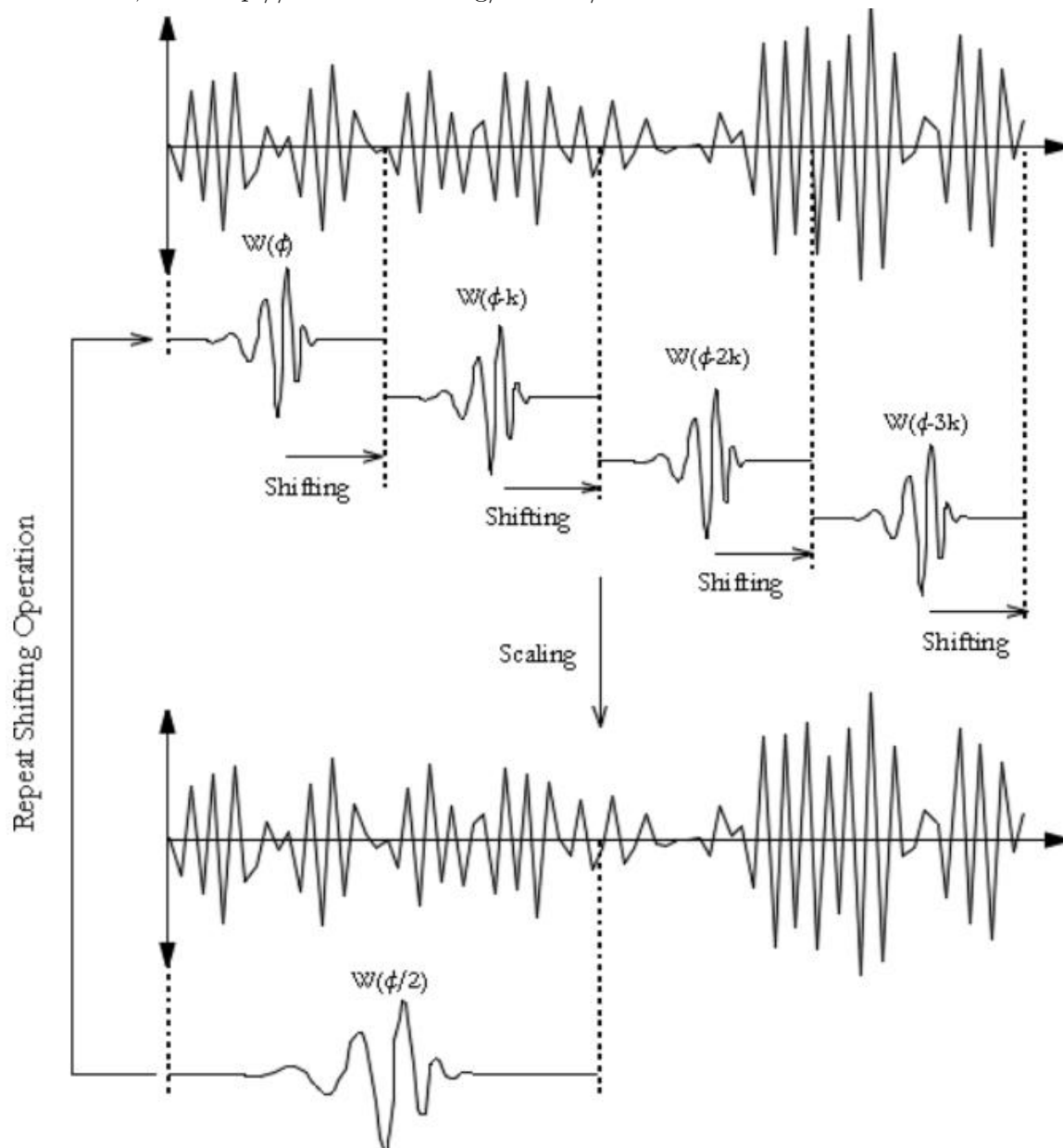
- CWT - kontinuierliche WT: Berechnet Energie statt Frequenz bzgl Position, d.h.  $\Psi(\omega)$
- DWT: diskrete WT  $\rightarrow$  dyadisches Gitter<sup>2</sup>  $\rightarrow$  Berechnungszeit  $O(n)$
- MODWT: Maximal Overlap DWT, Aufwand  $O(n \log(n))$ , für Varianzanalyse<sup>3</sup>
- WTMM: Wavelet-Transformation Modulus Maxim
- 

<sup>2</sup>y- und -Achse halbieren

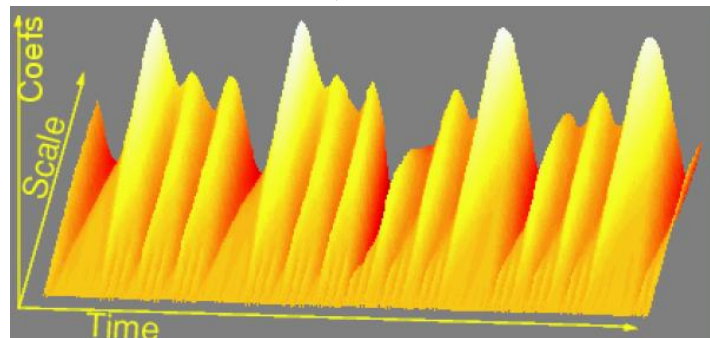
<sup>3</sup>ANOVA



(a) Skalierung und Verschiebung eines Daubechies-Waveletts für die Anpassung an eine Zeitreihe, aus: <http://www.wavelet.org/tutorial/wbasic.htm>



(b) Der Graph der Koeffizienten von  $W$  (Abb. 3.5-1a Verschiebung  $k$  und Skalierung  $\frac{c}{2}$ )



Eigenschaften:

- $\Psi$  enthält viele Frequenzen, deswegen wird von Skalierung gesprochen
- $\Psi$  ist genau dann analytisch, wenn Frequenzen definiert, d.h. Trafo Power WT  $\rightarrow$  Power FT möglich
- Die Power der WT ist  $|\Psi(f(,s))|^2$ : Welcher Anteil ist in der Frequenz der Fall?  $\rightarrow$  Amplitude = Stärke der Skalierung
- Haar-Wavelett: Abfrage von größer/kleiner

Vorteil: keine Fensterfunktion nötig, da Wavelet per Definition abklingt. Beschreibungen:

- Je breiter Wavelet skaliert, desto grober werden die Signaleigenschaften wiedergegeben.
- WT bei tiefen Frequenzen f bessere Frequenzauflösung
- WT bei hohen Frequenzen f bessere Zeitauflösung
- ist f hoch so laufen viele Schwingungen in kurzer Zeit ab, also hohe zeitliche Auflösung zur Modellierung nötig
- ist f niedrig, wirken sich kleine Änderungen stark aus  $\Rightarrow$  hohe Frequenzauflösung nötig

### 3.5.2. Haar-Wavelet

Wavelet aus der Kombination zweier Rechteckfunktionen. Der Vorteil ist die Ausnutzung der schneller Wavelet Transformation bei der Berechnungszeit ( $O(n)$ ) und die gute Approximation der Koeffizienten. Der Nachteil ist, daß hohe Frequenzen nicht exakt abgebildet werden können.

- bewahrt Euklidische Distanz
- exakt umkehrbar (Rücktrafo)

### 3.5.3. Vergleich zu FT

FT: Erhält gut globale Frequenzen  $\rightarrow$  Konjunktur WT: Erhält gut Trend über Jahre hinweg

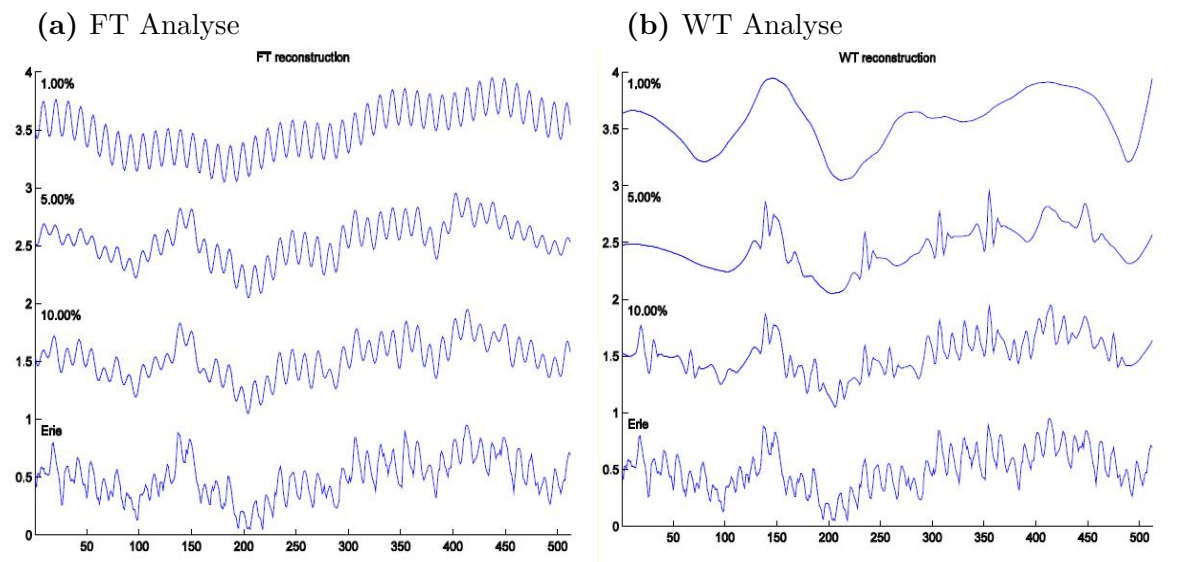


Abb. 3.5-2: Vergleich der Analysen

## Kapitel 4

# temporale Knowledge Discovery

Ein Vektor beschreibt die Zeitreihe. Dieser entsteht aus Koeffizienten entweder der ARIMA oder GARCH Modelle oder der Analysen von Waveletts bzw. Fouriertrafos. Nun werden Strukturen in der Menge der Daten gesucht

### 4.1. Knowledge Discovery

Zeitreihen werden auf Ähnlichkeit überprüft. Die Definition des Distanz oder Ähnlichkeitsmaßes entsteht aus den Koeffizienten und ist schwierig.

Details s. Zusammenfassung zur Knowledge Discovery. Zusätzlich sollte man im multivariaten Fall folgende Fragen fuer eine Distanzbildung beantworten:

- Amplituden Verschiebung relevant? Achtung, da euklid nicht Skaleninvariant ergibt eine Veränderung der Amplitude eine unterschiedliche Ähnlichkeit z.B. Lautstärke eines Musikstückes
- DC verschiebung (1.Frequenz nach FFT)?
- zeitliche Verschiebung? (Ist Imaginärteil nach FFT und wird somit rausgerechnet.  $f(t+a) \rightarrow F(w) * \exp(iaw)$ )
- Gewichtung verschiedener ZR?
- Fourier-Analyse sinnvoll? Frequenzspektrum (Power) bevorzugt lange Frequenzen in Euklid, z.B. Musik mit ähnlichem Bass
- $\log(\text{Power})$  mit Euklid  $\Rightarrow$  kurze Frequenzen relevant  $\Rightarrow$  Rauschen in Musik
- Normierung? (bei FFT:  $f(at) \rightarrow 1/a F(w/a)$ ); Ist relevant, falls Amplitude relevant fuer Ähnlichkeitsbildung ist.

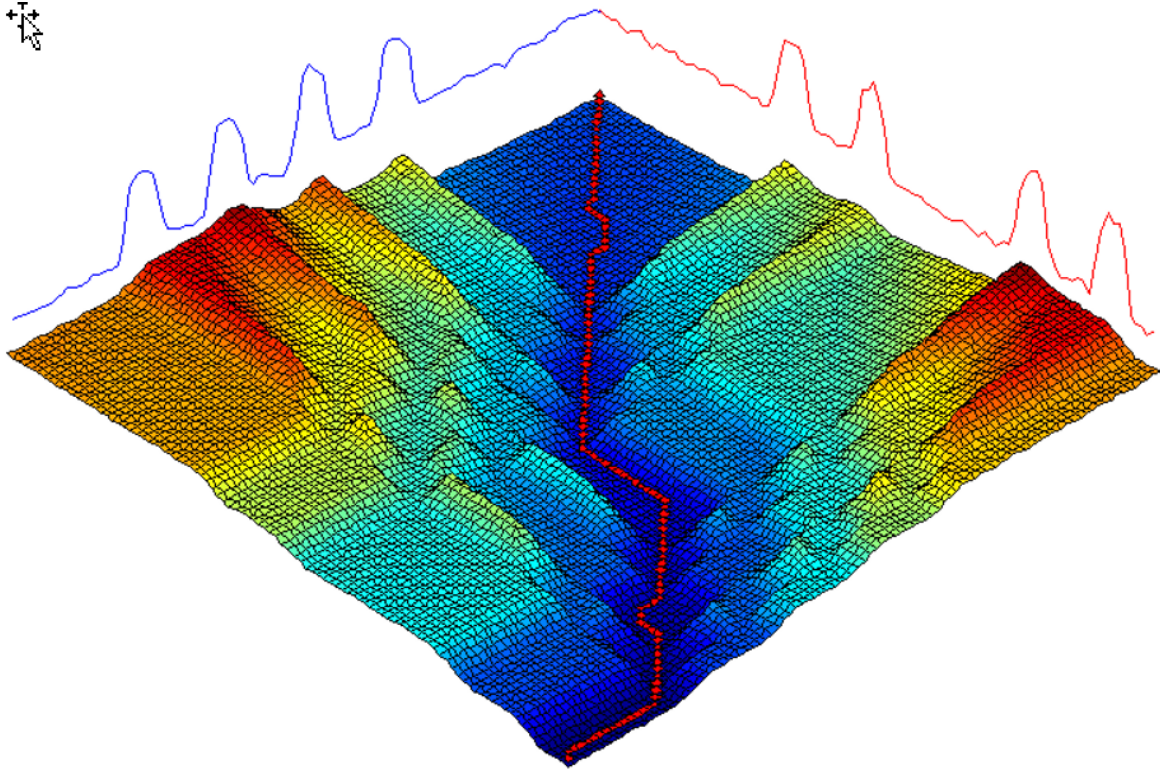
Eine oft genutzte Ähnlichkeit bei ZR ist die KullbackLeibler Divergenz die im symmetrischen Fall aber immernoch nicht die Dreiecksungleichung erfüllt.

### 4.2. Dynamic Time Warping - DTW

Häufig treten Verzerrung auf der Zeitachse auf, die objektiv nur schwer zu normieren sind und bei denen die üblichen Distanzmaße versagen. DTW ist die Lösung dafür:

Zeitreihe1 und Zeitreihe2 werden  $n^2$  mal verschoben, danach wird der kürzeste Pfad berechnet, siehe Abb. 4.2-1. Ein Pfad muss immer bei  $(0,0)$  anfangen und bei  $(n,n)$  enden. Voraussetzungen:

- ZR sind gleich lang.
- Ein Distanzmaß fuer einen Punkt  $(i,j)$  ist sinnvoll definiert.



**Abb. 4.2-1:** KuerzesterPfad ist Distanz zweier ZR bei  $n^2$  moeglichen Verschiebungen. Das ergibt als Nachteil einen hohen Rechenaufwand.

### 4.3. Wissensdarstellung von ZR

Dissertation von Mörchen in 2006

1. primitive patterns: cluster von features, atomic time units
2. succession: abfolge gleicher primitive patterns, Dauer
3. events: Gleichzeitige Mustervorkommen verschiedener ZR, synchronicity
4. sequence: order, zeitliche abfolge von events
5. temp. patterns: verschiedene typen von 4)

->Prädikatenlogik 1. Stufe -> Prolog -> Fuzzy Kalkül