

Mathematische Methoden in den Wirtschaftswissenschaften

Aufbaukurs

Josef Leydold

Department für Statistik und Mathematik – WU Wien

SS 2006

Übersicht

Finanzmathematik und *Quantitative Finance*

- Grundbegriffe
- Portfolio Management
- Derivative
- Das Binomialmodell
- Das stochastische Modell für Aktienkurse
- Stochastische Analysis
- Das Black-Scholes Modell

Prognose und Zeitreihenanalyse

- Daten und Indizes
- Beschreibung von Zeitreihen
- Zerlegung von Zeitreihen, Glättung und Saisonbereinigung
- Regression
- *Autoregressive moving average* (ARMA) Modelle

Literatur

- Cuthbertson, K. (1996): *Quantitative Financial Economics*, Wiley, Chichester. Kapitel 1
Renditen, Nutzenfunktion
- Spremann, K. (1996): *Wirtschaft, Investition und Finanzierung*, Oldenbourg, München.
deutsche Synonyme für die englischen Ausdrücke
- Bosch, Karl(1990): *Finanzmathematik*, Oldenbourg. Kapitel 7
Versicherungsmathematik
- Wilmott, P. (2001): *Wilmott Introduces Quantitative Finance*, Wiley, Chichester. Kapitelauswahl
Quantitative Finance
- Makridakis, Wheelwright and Hyndman (1998): *Forecasting: Methods and Applications* (3rd edition), Wiley. Kapitelauswahl
Zeitreihenanalyse

Kapitel 1

Multivariate Analysis

Lernziele

- Funktionen in mehreren Variablen
- Graph und Niveaulinien einer Funktion in zwei Variablen
- Partielle Ableitung und Gradient
- Lokale und globale Extrema
- Lagrange-Ansatz

Funktionen in mehreren Variablen

Eine **reelle Funktion in mehreren Variablen** ist eine Abbildung, die jedem Vektor x eine reelle Zahl zuordnet.

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto f(x) = f(x_1, x_2, \dots, x_n)$$

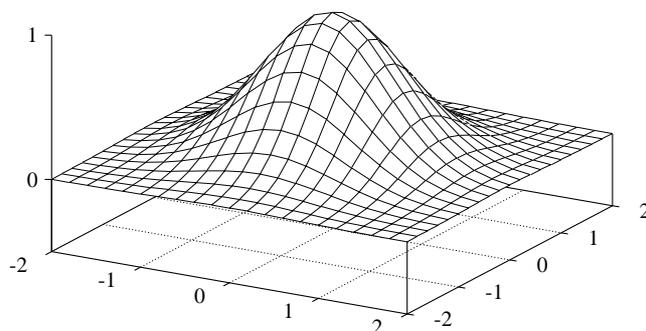
Die Komponenten x_i des Vektors x heißen die **Variablen** der Funktion f .

Funktionen in *zwei* Variablen lassen sich durch den **Graphen** (Funktionengebirge) veranschaulichen.

Dabei wird für jeden Punkt in der (x, y) -Ebene der Funktionswert $f(x, y)$ in die z -Richtung eingezeichnet.

Graph einer Funktion in zwei Variablen

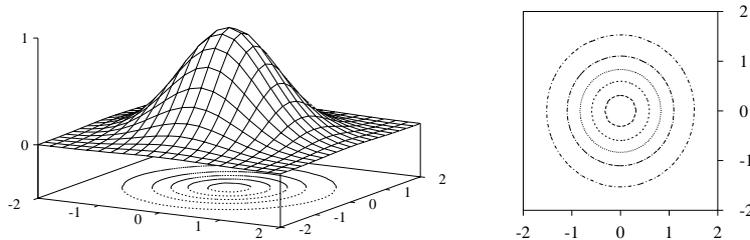
$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, (x, y) \mapsto f(x, y) = \exp(-x^2 - y^2)$$



Niveaulinien

Eine **Isoquante (Höhenlinie, Niveaulinie)** ist die Menge aller Punkte (x, y) mit $f(x, y) = c$ ($c \in \mathbb{R}$ konstant).

Die Funktion f hat daher auf einer Höhenlinie den gleichen Funktionswert.



Partielle Ableitung

Wir untersuchen die Änderung des Funktionswertes wenn wir eine Variable x_i variieren und alle anderen konstant lassen. Wir erhalten dadurch die (erste) **partielle Ableitung** von f nach x_i :

$$f_{x_i}(\mathbf{x}) = \frac{\partial f}{\partial x_i} = \lim_{\Delta x_i \rightarrow 0} \frac{F(\dots, x_i + \Delta x_i, \dots) - F(\dots, x_i, \dots)}{\Delta x_i}$$

Wir erhalten die partielle Ableitung nach x_i wenn wir alle anderen Variablen als Konstante auffassen und f nach den bekannten Regeln für Funktionen in einer Variable nach x_i ableiten.

Partielle Ableitung / Beispiel

Gesucht sind die ersten partiellen Ableitungen von

$$f(x_1, x_2) = \sin(2x_1) \cdot \cos(x_2)$$

$$f_{x_1} = 2 \cdot \cos(2x_1) \cdot \underbrace{\cos(x_2)}_{\text{als Konstante betrachtet}}$$

$$f_{x_2} = \underbrace{\sin(2x_1)}_{\text{als Konstante betrachtet}} \cdot (-\sin(x_2))$$

Partielle Ableitung / Beispiel

Gesucht sind die ersten partiellen Ableitungen von

$$f(x, y) = x^2 + 3xy$$

$$f_x = 2x + 3y$$

$$f_y = 0 + 3x$$

Der Gradient

Wir fassen die partiellen Ableitungen erster Ordnung zu einem Vektor, dem **Gradienten** an der Stelle x , zusammen.

$$\nabla f(x) = (f_{x_1}(x), \dots, f_{x_n}(x))$$

- ∇f heißt auch „nabla f “.
- Der Gradient wird oft als Zeilenvektor geschrieben.
- Der Gradient einer Funktion f zeigt in die Richtung des steilsten Anstieges von f . Seine Länge gibt diese Steigung an.
- Der Gradient steht immer normal auf die entsprechende Niveaulinie.
- Der Gradient „spielt“ die gleiche Rolle wie die erste Ableitung bei Funktionen in einer Variablen.

Der Gradient / Beispiel

Gesucht ist der Gradient von

$$f(x, y) = x^2 + 3xy$$

an der Stelle $x = (3, 2)$.

$$\nabla f(x) = (2x + 3y, 3x)$$

$$\nabla f(3, 2) = (12, 9)$$

Höhere partielle Ableitungen

Analog zu den Funktionen in einer Variablen können wir partielle Ableitungen nochmals ableiten und erhalten so **höhere partielle Ableitungen**.

$$f_{x_i x_k}(x) = \frac{\partial^2 f}{\partial x_k \partial x_i}(x) \quad \text{bzw.} \quad f_{x_i x_i}(x) = \frac{\partial^2 f}{\partial x_i^2}(x)$$

Es kommt auf die Reihenfolge beim Differenzieren nicht an (falls alle zweiten partiellen Ableitungen existieren und stetig sind):

$$\frac{\partial^2 f}{\partial x_k \partial x_i}(x) = \frac{\partial^2 f}{\partial x_i \partial x_k}(x)$$

Höhere partielle Ableitungen / Beispiel

Gesucht sind alle ersten und zweiten partiellen Ableitungen von

$$f(x, y) = x^2 + 3xy$$

Erste partielle Ableitungen:

$$f_x = 2x + 3y \quad f_y = 0 + 3x$$

Zweite partielle Ableitungen:

$$\begin{aligned} f_{xx} &= 2 & f_{xy} &= 3 \\ f_{yx} &= 3 & f_{yy} &= 0 \end{aligned}$$

Extrema (Optimierung)

x_0 heißt **globales Maximum** von f , falls für alle x gilt:

$$f(x_0) \geq f(x)$$

x_0 heißt **lokales Maximum** von f , falls für alle x in einer geeigneten Umgebung von x_0 (i.e. Menge aller Punkte in der Nähe von x_0) gilt:

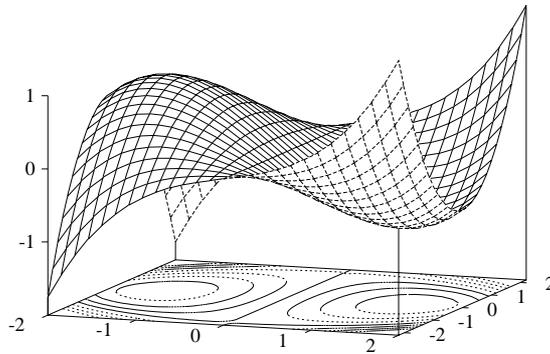
$$f(x_0) \geq f(x)$$

Analog werden **globales Minimum** und **lokales Minimum** definiert.

Minimum und Maximum werden als **Extremum** von f bezeichnet.

Aufgabe:

Suche Extrema einer gegebenen Funktionen.



$$f(x, y) = \frac{1}{6} x^3 - x + \frac{1}{4} x y^2$$

Stationärer Punkt

Ein Punkt x_0 , in dem der Gradient verschwindet, i.e.,

$$\nabla f(x_0) = 0$$

heißt **stationärer Punkt** von f .

Es gilt (*notwendige Bedingung*):

Jedes Extremum von f ist ein stationärer Punkt von f .

Stationärer Punkt / Beispiel

Suche alle stationären Punkte von

$$f(x, y) = \frac{1}{6} x^3 - x + \frac{1}{4} x y^2$$

Bilde alle ersten partiellen Ableitungen und setze diese Null:

$$(I) \quad f_x = \frac{1}{2} x^2 - 1 + \frac{1}{4} y^2 = 0$$

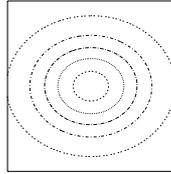
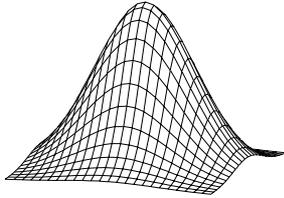
$$(II) \quad f_y = \frac{1}{2} x y = 0$$

Die stationären Punkte von f sind nun die Lösungen dieses Gleichungssystems:

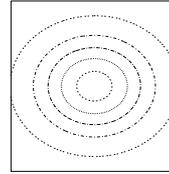
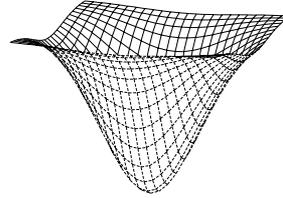
$$x_1 = (0, 2) \qquad x_2 = (0, -2)$$

$$x_3 = (\sqrt{2}, 0) \qquad x_4 = (-\sqrt{2}, 0)$$

Stationäre Punkte: Lokale Extrema

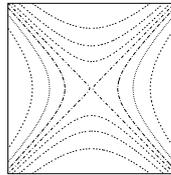
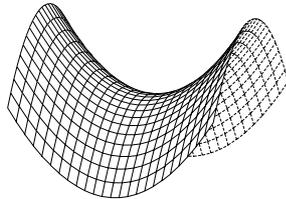


Lokales Maximum

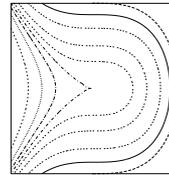
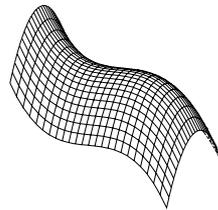


Lokales Minimum

Stationäre Punkte: Sattelpunkte und andere



Sattelpunkt



Beispiel für höhere Ordnung

Die Hesse-Matrix

Die Matrix aller zweiten partiellen Ableitungen einer Funktion f an der Stelle x wird als **Hesse-Matrix** $H_f(x)$ bezeichnet.

$$H_f(x) = \begin{pmatrix} f_{x_1x_1}(x) & f_{x_1x_2}(x) & \cdots & f_{x_1x_n}(x) \\ f_{x_2x_1}(x) & f_{x_2x_2}(x) & \cdots & f_{x_2x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1}(x) & f_{x_nx_2}(x) & \cdots & f_{x_nx_n}(x) \end{pmatrix}$$

Die Hesse-Matrix bestimmt ob die Funktion f in der Nähe von x konvex oder konkav ist (oder nicht).

Sie „spielt“ die gleiche Rolle, wie die zweite Ableitung von Funktionen in einer Variablen.

Die Hesse-Matrix / Beispiel

Berechne die Hesse-Matrix an den Stellen $x_1 = (0, 2)$ und $x_4 = (-\sqrt{2}, 0)$ von

$$f(x, y) = \frac{1}{6}x^3 - x + \frac{1}{4}xy^2$$

$$\begin{aligned} f_x &= \frac{1}{2}x^2 - 1 + \frac{1}{4}y^2 & f_{xx} &= x & f_{xy} &= \frac{1}{2}y \\ f_y &= \frac{1}{2}xy & f_{yx} &= \frac{1}{2}y & f_{yy} &= \frac{1}{2}x \end{aligned}$$

$$H_f(\mathbf{x}) = \begin{pmatrix} f_{xx}(\mathbf{x}) & f_{xy}(\mathbf{x}) \\ f_{yx}(\mathbf{x}) & f_{yy}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} x & \frac{1}{2}y \\ \frac{1}{2}y & \frac{1}{2}x \end{pmatrix}$$

$$H_f(\mathbf{x}_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad H_f(\mathbf{x}_4) = \begin{pmatrix} -\sqrt{2} & 0 \\ 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

Hauptminoren

Wir benötigen das „Vorzeichen“ (die **Definitheit**) der Hesse-Matrix.

Die Definitheit der Hesse-Matrix kann mit Hilfe der sogenannten **Hauptminoren** festgestellt werden.

Die Determinante der linken oberen $k \times k$ -Untermatrix heißt der **k -te (führende) Hauptminor**:

$$M_k = \begin{vmatrix} f_{x_1x_1}(\mathbf{x}) & f_{x_1x_2}(\mathbf{x}) & \dots & f_{x_1x_k}(\mathbf{x}) \\ f_{x_2x_1}(\mathbf{x}) & f_{x_2x_2}(\mathbf{x}) & \dots & f_{x_2x_k}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_kx_1}(\mathbf{x}) & f_{x_kx_2}(\mathbf{x}) & \dots & f_{x_kx_k}(\mathbf{x}) \end{vmatrix}$$

Hauptminoren / Beispiel

Berechne alle Hauptminoren der Hesse-Matrix an den Stellen $x_1 = (0, 2)$ und $x_4 = (-\sqrt{2}, 0)$ von

$$f(x, y) = \frac{1}{6}x^3 - x + \frac{1}{4}xy^2$$

$$H_f(\mathbf{x}_1) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$M_1 = 0$$

$$M_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$$

$$H_f(\mathbf{x}_4) = \begin{pmatrix} -\sqrt{2} & 0 \\ 0 & -\frac{\sqrt{2}}{2} \end{pmatrix}$$

$$M_1 = -\sqrt{2}$$

$$M_2 = \begin{vmatrix} -\sqrt{2} & 0 \\ 0 & -\frac{\sqrt{2}}{2} \end{vmatrix} = 1$$

Hinreichende Bedingung für Lokale Extrema

Für Funktionen in zwei Variablen erhalten wir folgende Bedingung:

Sei x_0 ein stationärer Punkt von f , und M_1 und M_2 die Hauptminoren von $H_f(x_0)$.

- (a) $M_2 < 0 \quad \Rightarrow \quad x_0$ ist Sattelpunkt
- (b) $M_2 > 0$ und $M_1 > 0 \quad \Rightarrow \quad x_0$ ist lokales Minimum
- (c) $M_2 > 0$ und $M_1 < 0 \quad \Rightarrow \quad x_0$ ist lokales Maximum
- (d) $M_2 = 0 \quad \Rightarrow \quad$ Keine Aussage möglich

Lokale Extrema / Beispiel

Berechne alle lokalen Extrema der Funktion

$$f(x, y) = \frac{1}{6}x^3 - x + \frac{1}{4}xy^2$$

Stationäre Punkte:

- $x_1 = (0, 2): \quad M_2 = -1 < 0 \quad \Rightarrow \quad$ Sattelpunkt
- $x_2 = (0, -2): \quad M_2 = -1 < 0 \quad \Rightarrow \quad$ Sattelpunkt
- $x_3 = (\sqrt{2}, 0): \quad M_2 = 1 > 0, M_1 = \sqrt{2} > 0 \quad \Rightarrow \quad$ lokales Minimum
- $x_4 = (-\sqrt{2}, 0): \quad M_2 = 1 > 0, M_1 = -\sqrt{2} < 0 \quad \Rightarrow \quad$ lokales Maximum

Hinreichende Bedingung für Lokale Extrema / Allgemein

Sei x_0 ein stationärer Punkt von f , und M_k der k -te Hauptminoren von $H_f(x_0)$.

- (a) Alle Hauptminoren $M_k > 0$
 $\Rightarrow \quad x_0$ ist ein **lokales Minimum** von f .
- (b) Für alle Hauptminoren gilt $(-1)^k M_k > 0$
 $\Rightarrow \quad x_0$ ist ein **lokales Maximum** von f .
- (c) $\det(H_f(x_0)) \neq 0$, aber weder (a) noch (b) sind erfüllt
 $\Rightarrow \quad x_0$ ist ein **Sattelpunkt** von f .
- (d) Andernfalls ist keine Aussage möglich, d.h. x_0 kann ein lokales Extremum sein, muß aber nicht.

Lokale Extrema / Beispiel

Wir suchen die lokalen Extrema der Funktion

$$f(x_1, x_2, x_3) = (x_1 - 1)^2 + (x_2 + 2)^2 + (x_3 + 1)^2$$

$$\nabla f = (2(x_1 - 1), 2(x_2 + 2), 2(x_3 + 1)) = (0, 0, 0)$$

Der einzige stationäre Punkt ist $x_0 = (1, -2, -1)$.

$$H_f(x) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Alle Hauptminoren sind positiv: $M_1 = 2, M_2 = 4, M_3 = 8$

$\Rightarrow x_0$ ist ein lokales Minimum.

Globale Extrema

Aufgabe:

Suche den höchsten Punkt der Welt.

Die Berechnung globaler Extrema von Funktionen in mehr als einer Variablen ist im allgemeinen sehr schwierig.

Wir wollen hier nur folgenden Spezialfall behandeln:

Sei x_0 ein stationärer Punkt von f , und M_k der k -te Hauptminoren von $H_f(x)$.

- Falls alle Hauptminoren $M_k > 0$ für **alle** x , dann ist x_0 ist ein **globales Minimum** von f .
- Falls für alle Hauptminor $(-1)^k M_k > 0$ für **alle** x , dann ist x_0 ist ein **globales Maximum** von f .

Globale Extrema / Beispiel

Wir suchen die globalen Extrema der Funktion

$$f(x_1, x_2, x_3) = (x_1 - 1)^2 + (x_2 + 2)^2 + (x_3 + 1)^2$$

Der einzige stationäre Punkt ist $x_0 = (1, -2, -1)$.

$$H_f(x) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Die Hesse-Matrix ist unabhängig von x und alle Hauptminoren sind (für alle x) positiv: $M_1 = 2, M_2 = 4, M_3 = 8$

$\Rightarrow x_0$ ist ein globales Minimum.

Optimierung unter Nebenbedingungen

Aufgabe:

Berechne

$$\max / \min f(x, y)$$

unter der Nebenbedingung (*constraint*)

$$g(x, y) = c$$

Beispiel:

Wir suchen die lokalen Extrema der Funktion

$$f(x, y) = x^2 + 2y^2$$

unter der Nebenbedingung

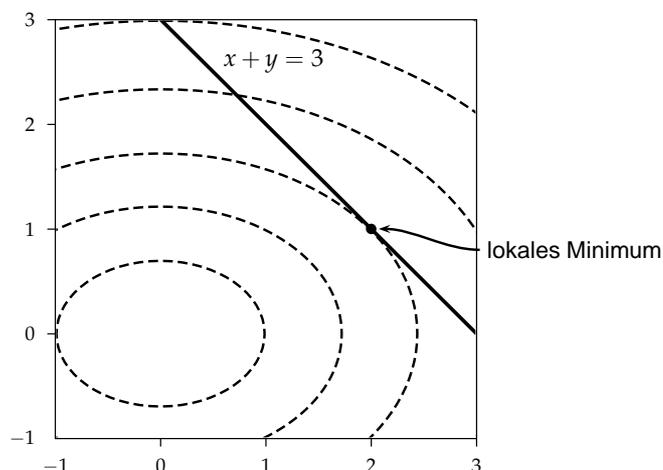
$$g(x, y) = x + y = 3$$

Graphische Lösung

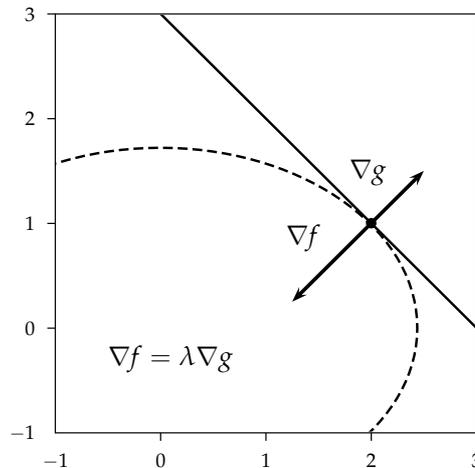
Im Falle von zwei Variablen können wir das Problem graphisch „lösen“.

- 1 Zeichne die Nebenbedingung $g(x, y) = c$ in die xy -Ebene ein. (Kurve in der Ebene)
- 2 Zeichne „geeignete“ Niveaulinien der zu optimierenden Funktion $f(x, y)$ ein.
- 3 Untersuche an hand der Zeichnung welche Niveaulinien den zulässigen Bereich schneiden und bestimme die ungefähre Lage der Extrema.

Graphische Lösung / Beispiel



Gradienten



Lagrangefunktion

Wir erzeugen uns aus f , g und einer Hilfsvariablen λ eine neue Funktion, die **Lagrange-Funktion**:

$$\mathcal{L}(x, y; \lambda) = f(x, y) - \lambda (g(x, y) - c)$$

Die Hilfsvariable λ heißt **Lagrange-Multiplikator**.

Wenn die Nebenbedingung erfüllt ist, so stimmen f und \mathcal{L} überein.

Lokale Extrema von f gegeben $g(x, y) = c$ sind stationäre Punkte der Lagrangefunktion \mathcal{L} :

$$\mathcal{L}_x = f_x - \lambda g_x = 0$$

$$\mathcal{L}_y = f_y - \lambda g_y = 0$$

$$\mathcal{L}_\lambda = -(g(x, y) - c) = 0$$

Lagrangefunktion / Beispiel

Wir suchen die lokalen Extrema der

$$f(x, y) = x^2 + 2y^2 \quad \text{gegeben} \quad g(x, y) = x + y = 3$$

Lagrangefunktion:

$$\mathcal{L}(x, y, \lambda) = (x^2 + 2y^2) - \lambda(x + y - 3)$$

Stationäre Punkte:

$$\mathcal{L}_x = 2x - \lambda = 0$$

$$\mathcal{L}_y = 4y - \lambda = 0$$

$$\mathcal{L}_\lambda = 3 - x - y = 0$$

⇒ einziger stationärer Punkt: $(x_0; \lambda_0) = (2, 1; 4)$

Geränderte Hesse-Matrix

Die Matrix

$$\bar{H}(x) = \begin{pmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{pmatrix}$$

heißt die **geränderte Hesse-Matrix**.

Hinreichende Bedingung für lokales Extremum:

Sei $(x_0; \lambda_0)$ ein stationärer Punkt von \mathcal{L} .

- $|\bar{H}(x_0)| > 0 \Rightarrow x_0$ ist lokales Maximum.
- $|\bar{H}(x_0)| < 0 \Rightarrow x_0$ ist lokales Minimum.

Geränderte Hesse-Matrix / Beispiel

Wir suchen die lokalen Extrema der

$$f(x, y) = x^2 + 2y^2 \quad \text{gegeben} \quad g(x, y) = x + y = 3$$

Lagrangefunktion: $(x_0; \lambda_0) = (2, 1; 4)$

$$\mathcal{L}(x, y, \lambda) = (x^2 + 2y^2) - \lambda(x + y - 3)$$

Stationärer Punkt: $(x_0; \lambda_0) = (2, 1; 4)$

Determinante der geränderte Hesse-Matrix:

$$|\bar{H}(x_0)| = \begin{vmatrix} 0 & g_x & g_y \\ g_x & \mathcal{L}_{xx} & \mathcal{L}_{xy} \\ g_y & \mathcal{L}_{yx} & \mathcal{L}_{yy} \end{vmatrix} = \begin{vmatrix} 0 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 4 \end{vmatrix} = -6 < 0$$

$\Rightarrow x_0 = (2, 1)$ ist ein lokales Minimum.

Lagrange-multiplikator / Eine Interpretation

Die Lage des Extremums x^* hängt von c der Nebenbedingung ab, $x^* = x^*(c)$, und somit auch der Extremalwert von f^* :

$$f^*(c) = f(x^*(c))$$

Wie ändert sich $f^*(c)$ mit c ?

Im Optimum stimmen \mathcal{L} und f überein. Daher ist

$$\frac{df^*(c)}{dc} = \frac{\partial \mathcal{L}}{\partial c} = \frac{\partial}{\partial c} (f(x) - \lambda g(x) + \lambda c) = \lambda$$

$$\frac{df^*(c)}{dc} = \lambda$$

Aufgabe:

Berechne

$$\max / \min f(\mathbf{x})$$

unter den Nebenbedingungen

$$g_1(\mathbf{x}) = c_1$$

$$\vdots$$

$$(m < n)$$

$$g_m(\mathbf{x}) = c_m$$

Lagrangefunktion

$$\mathcal{L}(\mathbf{x}; \lambda_1, \dots, \lambda_m) = f(\mathbf{x}) - \sum_{i=1}^m \lambda_i (g_i(\mathbf{x}) - c_i)$$

Kapitel 2

Kovarianz und Korrelation

Lernziele

- Mathematische und statistische Grundlagen der Portfoliotheorie
- Kovarianz und Korrelation
- Kovarianz- und Korrelationsmatrix
- Multivariate Normalverteilung
- Erzeugen von multinormalverteilten Zufallsvektoren

Erwartungswert einer Linearkombination

Der Erwartungswert einer **Linearkombination** von ZVen ist die Linearkombination der einzelnen Erwartungswerte:

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

Allgemein

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)$$

Voraussetzung: alle Erwartungswerte existieren.

Varianz einer Linearkombination

Die Varianz einer Linearkombination von ZVen ist **nicht** die Linearkombination der einzelnen Varianzen. Es gilt:

$$V(aX + bY + c) = a^2 V(X) + 2ab \operatorname{Cov}(X, Y) + b^2 V(Y)$$

Die Konstante c beeinflusst die Varianz nicht.

Bei der Varianz einer Summe tritt ein gemischter Term auf: die **Kovarianz** der beiden ZVen.

Nur wenn die Kovarianz der beiden ZVen Null ist, also beide unkorreliert sind, gilt:

„Die Varianz der Summe ist gleich die Summe der Varianzen“.

Voraussetzung: alle Varianzen existieren.

Kovarianz

Die Kovarianz zwischen zwei ZV X und Y ist definiert als

$$\begin{aligned} \operatorname{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\ &= \sum_{x,y} (x - \mu_x)(y - \mu_y) P(X = x, Y = y) \end{aligned}$$

Hier gilt auch ein Verschiebungssatz:

$$E[(X - E(X))(Y - E(Y))] = E[XY] - E(X)E(Y)$$

Bemerkung: Ist $\operatorname{Cov}(X, Y) = 0$, folgt

$$E[XY] = E(X)E(Y)$$

Normalverteilte Zufallsvariable

Reproduktionseigenschaft der Normalverteilung:

Seien die ZVen X und Y normalverteilt mit

$$X \sim N(\mu_x, \sigma_x^2), \quad Y \sim N(\mu_y, \sigma_y^2) \quad \text{und} \quad \text{Cov}(X, Y) = \sigma_{xy}$$

Dann ist

$$Z = aX + bY + c \sim N(\mu_z, \sigma_z^2)$$

$$\mu_z = a\mu_x + b\mu_y + c$$

$$\sigma_z^2 = a^2\sigma_x^2 + 2ab\sigma_{xy} + b^2\sigma_y^2$$

Korrelation

Die **Korrelation** zwischen zwei ZV X und Y ist definiert als

$$\text{Corr}(X, Y) = \rho(X, Y) = \rho = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}}$$

Es gilt immer

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

$$\text{Cov}(X, Y) = \text{Corr}(X, Y) \sqrt{V(X)V(Y)}$$

$$\text{Corr}(X, Y) = 0 \quad \Leftrightarrow \quad \text{Cov}(X, Y) = 0$$

Unkorrelierte Zufallsvariable

Sind X und Y unkorreliert, so gilt

$$V(X + Y) = V(X) + V(Y)$$

aber auch

$$V(X - Y) = V(X) + V(Y)$$

Allgemein gilt für unkorrelierte ZV X_i

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 V(X_i)$$

Unabhängige Zufallsvariable

Zwei Zufallsvariable X und Y heißen (stochastisch) **unabhängig** wenn

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$$

für all möglichen Merkmalsausprägungen x und y .

Unabhängige Zufallsvariable sind immer unkorreliert, i.e.

$$X, Y \text{ unabhängig} \Rightarrow \text{Corr}(X, Y) = \text{Cov}(X, Y) = 0$$

Die Umkehrung gilt jedoch **nicht!**

Beispiel 1 – 2-dimensionale Verteilung

Wir suchen für $Y = X_1 + X_2$ Erwartung und Varianz.

Die ZVen X_1 und X_2 besitzen die gemeinsame Verteilung

$P(X_1 = x_1, X_2 = x_2)$	$X_2 = -1$	$X_2 = 1$	$P(X_1 = x_1)$
$X_1 = 1$	0.12	0.48	0.60
$X_1 = 2$	0.08	0.32	0.40
$P(X_2 = x_2)$	0.20	0.80	1.00

Beispiel 1 / summe

Die Werte von $Y = X_1 + X_2$ erhält man über die Tabelle

$Y = X_1 + X_2$	$X_2 = -1$	$X_2 = 1$
$X_1 = 1$	0	2
$X_1 = 2$	1	3

mit der Verteilung

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	Σ
$P(Y = y)$	0.12	0.08	0.48	0.32	1.00

Beispiel 1 / Erwartungswert $E(Y)$

Für den Erwartungswert von Y erhalten wir

$$E(Y) = \sum y P(Y = y) = 0 \cdot 0.12 + \dots + 3 \cdot 0.32 = 2.0$$

Nach unseren Regeln ergibt sich

$$E(Y) = E(X_1 + X_2) = E(X_1) + E(X_2) = 1.4 + 0.6 = 2.0$$

$$E(X_1) = \sum x_1 P(X_1 = x_1) = 1 \cdot 0.60 + 2 \cdot 0.40 = 1.4$$

$$E(X_2) = \sum x_2 P(X_2 = x_2) = (-1) \cdot 0.20 + 1 \cdot 0.80 = 0.6$$

Beispiel 1 / Varianz $V(Y)$

Für die Varianz von Y erhalten wir

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] = E[Y^2] - [E(Y)]^2 = \\ &= 0^2 \cdot 0.12 + \dots + 3^2 \cdot 0.32 - 2.0^2 = 0.88 \end{aligned}$$

Nach unseren Regeln ergibt sich

$$\begin{aligned} V(Y) &= V(X_1 + X_2) = V(X_1) + 2\text{Cov}(X_1, X_2) + V(X_2) = \\ &= 0.24 + 2 \cdot 0.00 + 0.64 = 0.88 \end{aligned}$$

Beispiel 1 / Varianz $V(Y)$

$$\begin{aligned} V(X_1) &= \sum x_1^2 P(X_1 = x_1) - [E(X_1)]^2 = \\ &= 1^2 \cdot 0.60 + (2^2) \cdot 0.40 - 1.4^2 = 0.24 \end{aligned}$$

$$\begin{aligned} V(X_2) &= \sum x_2^2 P(X_2 = x_2) - [E(X_2)]^2 = \\ &= (-1)^2 \cdot 0.20 + 1^2 \cdot 0.80 - 0.6^2 = 0.64 \end{aligned}$$

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E[X_1 - E(X_1)][X_2 - E(X_2)] = \\ &= E[X_1 X_2] - E(X_1) E(X_2) = \\ &= \sum_{x_1, x_2} x_1 x_2 P(X_1 = x_1, X_2 = x_2) - E(X_1) E(X_2) = \\ &= 1 \cdot (-1) \cdot 0.12 + 1 \cdot 1 \cdot 0.48 + \dots + 2 \cdot 1 \cdot 0.32 - 1.4 \cdot 0.6 = \\ &= 0 \quad (\text{einfacher: } X_1 \text{ und } X_2 \text{ sind unabhängig}) \end{aligned}$$

Beispiel 2 – 2-dimensionale Verteilung

Die ZVen X_1 und X_2 besitzen die gleichen Randverteilungen wie in Beispiel 1. Die gemeinsame Verteilung sei hingegen

$P(X_1 = x_1, X_2 = x_2)$	$X_2 = -1$	$X_2 = 1$	$P(X_1 = x_1)$
$X_1 = 1$	0.00	0.60	0.60
$X_1 = 2$	0.20	0.20	0.40
$P(X_2 = x_2)$	0.20	0.80	1.00

Wir suchen für $Y = X_1 + X_2$ Erwartung und Varianz.

	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$	
$P(Y = y)$	0.00	0.20	0.60	0.20	1.00

Beispiel 2 / Erwartungswert $E(Y)$

Für den Erwartungswert von Y erhalten wir

$$E(Y) = \sum y P(Y = y) = 1 \cdot 0.20 + \dots + 3 \cdot 0.20 = 2.0$$

Erwartungswert von X_1 und X_2 sind dieselben wie in Beispiel 1.

Nach unseren Regeln ergibt sich daher wie in Beispiel 1

$$E(Y) = E(X_1 + X_2) = E(X_1) + E(X_2) = 1.4 + 0.6 = 2.0$$

$$E(X_1) = \sum x_1 P(X_1 = x_1) = 1 \cdot 0.60 + 2 \cdot 0.40 = 1.4$$

$$E(X_2) = \sum x_2 P(X_2 = x_2) = (-1) \cdot 0.20 + 1 \cdot 0.80 = 0.6$$

Beispiel 2 / Varianz $V(Y)$

Für die Varianz von Y erhalten wir

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] = E[Y^2] - [E(Y)]^2 = \\ &= 1^2 \cdot 0.20 + 2^2 \cdot 0.60 + 3^2 \cdot 0.20 - 2.0^2 = 0.48 \end{aligned}$$

Varianzen von X_1 und X_2 sind dieselben wie im Beispiel 1.

Die Kovarianz ist aber neu zu berechnen.

Nach unseren Regeln ergibt sich daher

$$\begin{aligned} V(Y) &= V(X_1 + X_2) = V(X_1) + 2 \operatorname{Cov}(X_1, X_2) + V(X_2) = \\ &= 0.24 + 2 \cdot (-0.24) + 0.64 = 0.48 \end{aligned}$$

Beispiel 2 / Varianz $V(Y)$

$$V(X_1) = 0.24 \quad (\text{Bsp. 1})$$

$$V(X_2) = 0.64 \quad (\text{Bsp. 1})$$

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E[X_1 - E(X_1)][X_2 - E(X_2)] = \\ &= E[X_1 X_2] - E(X_1) E(X_2) = \\ &= \sum_{x_1, x_2} x_1 x_2 P(X_1 = x_1, X_2 = x_2) - E(X_1) E(X_2) = \\ &= 1 \cdot (-1) \cdot 0.00 + 1 \cdot 1 \cdot 0.60 + \dots + 2 \cdot 1 \cdot 0.20 - 1.4 \cdot 0.6 = \\ &= -0.24 \end{aligned}$$

Beispiel 2 / Korrelation

Durch die Addition der Variablen X_1 zu X_2 wird die Varianz von $X_1 + X_2$ gegenüber der von X_2 alleine deutlich reduziert.

Die Kovarianz ist negativ, $\text{Cov}(X_1, X_2) = -0.24$, und daher auch die Korrelation:

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{V(X_1)V(X_2)}} = \frac{-0.24}{\sqrt{0.24 \cdot 0.64}} = -0.612$$

Beispiel 3 – Unkorrelierte Zufallsvariable

Gegeben seien n unkorreliert Zufallsvariable, X_1, \dots, X_n , die die gleichen Erwartungswerte, $E(X_1) = \dots = E(X_n) = \mu$ und die gleichen Varianzen, $V(X_1) = \dots = V(X_n) = \sigma^2$ besitzen.

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n \sigma^2 = n \sigma^2$$

Die Varianz steigt proportional mit der Anzahl der Summanden.

Die Standardabweichung steigt nur mit der Wurzel der Anzahl der Summanden:

$$\sqrt{V\left(\sum_{i=1}^n X_i\right)} = \sqrt{n \sigma^2} = \sqrt{n} \sigma$$

Beispiel 4 – Diversifikation

Die beiden ZVen R_1 und R_2 sind die Renditen von zwei verschiedenen Wertpapieren. Der aktuell Tageskurs sei bei beiden gleich.

Angenommen die ZVen R_1 und R_2 sind unkorreliert und haben gleichen Erwartungswert und gleiche Varianzen:

$$E(R_1) = E(R_2) = \mu, \quad V(R_1) = V(R_2) = \sigma^2$$

Wir stellen 3 Portfolios zusammen:

- A. Nur Papier 1,
- B. Nur Papier 2,
- C. Papier 1 und Papier 2 je mit einem Anteil von 1/2.

Beispiel 4 – Diversifikation

Wir berechnen von den 3 Portfolios Erwartungswert und Varianz.

$$E(A) = E(R_1) = \mu, \quad V(A) = V(R_1) = \sigma^2$$

$$E(B) = E(R_2) = \mu, \quad V(B) = V(R_2) = \sigma^2$$

$$E(C) = E\left(\frac{1}{2}R_1 + \frac{1}{2}R_2\right) = \frac{1}{2}E(R_1) + \frac{1}{2}E(R_2) = \mu,$$

$$V(C) = V\left(\frac{1}{2}R_1 + \frac{1}{2}R_2\right) = \left(\frac{1}{2}\right)^2V(R_1) + \left(\frac{1}{2}\right)^2V(R_2) = \frac{1}{2}\sigma^2$$

Wir sind indifferent zwischen den Papieren 1 und 2.

Portfolio C liefert hingegen mit derselben erwarteten Rendite nur die halbe Varianz.

Risikostreuung und Korrelation

Die Reduktion der Varianz durch Diversifikation (Risikostreuung) ist auch bei nicht zu stark positiven Korrelationen sinnvoll.

Besonders interessant wird sie bei negativen Korrelationen.

Risikostreuung und $\rho = +1$

Bei starker positiver Korrelation zwischen den Renditen ist die Varianz auch im Portfolio C hoch.

Aus $\text{Corr}(R_1, R_2) = 1$ erhalten wir

$$\text{Cov}(R_1, R_2) = \text{Corr}(R_1, R_2) \sqrt{V(R_1)V(R_2)} = \sigma^2$$

$$E(C) = E\left(\frac{1}{2}R_1 + \frac{1}{2}R_2\right) = \frac{1}{2}E(R_1) + \frac{1}{2}E(R_2) = \mu$$

$$\begin{aligned} V(C) &= V\left(\frac{1}{2}R_1 + \frac{1}{2}R_2\right) = \\ &= \left(\frac{1}{2}\right)^2 V(R_1) + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \text{Cov}(R_1, R_2) + \left(\frac{1}{2}\right)^2 V(R_2) \\ &= \sigma^2 \end{aligned}$$

Portfolio C ist nun gleich schlecht wie A oder B.

Risikostreuung und $\rho = -1$

Bei starker negativer Korrelation zwischen den Renditen verschwindet die Varianz im Portfolio C (fast).

Aus $\text{Corr}(R_1, R_2) = -1$ erhalten wir

$$\text{Cov}(R_1, R_2) = \text{Corr}(R_1, R_2) \sqrt{V(R_1)V(R_2)} = -\sigma^2$$

$$E(C) = E\left(\frac{1}{2}R_1 + \frac{1}{2}R_2\right) = \frac{1}{2}E(R_1) + \frac{1}{2}E(R_2) = \mu$$

$$\begin{aligned} V(C) &= V\left(\frac{1}{2}R_1 + \frac{1}{2}R_2\right) = \\ &= \left(\frac{1}{2}\right)^2 V(R_1) + 2 \cdot \frac{1}{2} \cdot \frac{1}{2} \text{Cov}(R_1, R_2) + \left(\frac{1}{2}\right)^2 V(R_2) \\ &= 0 \end{aligned}$$

Optimale Risikostreuung

Angenommen zwei Renditen R_1 und R_2 mit $E(R_1) = \mu_1$, $E(R_2) = \mu_2$ und $V(R_1) = \sigma_1^2$, $V(R_2) = \sigma_2^2$ und Kovarianz σ_{12} liegen vor.

Wir suchen die Kombination $\alpha_1 R_1 + \alpha_2 R_2$ mit minimaler Varianz unter der Einschränkung, dass die Summe der Gewichte 1 sei.

$$\min_{\alpha_1, \alpha_2} V[\alpha_1 R_1 + \alpha_2 R_2]$$

$$\text{NB: } \alpha_1 + \alpha_2 = 1$$

Lagrange-Ansatz:

$$L(\alpha_1, \alpha_2; \lambda) = V[\alpha_1 R_1 + \alpha_2 R_2] + \lambda (1 - \alpha_1 - \alpha_2)$$

Optimale Risikostreuung

$$L(\alpha_1, \alpha_2; \lambda) = \alpha_1^2 \sigma_1^2 + 2 \alpha_1 \alpha_2 \sigma_{12} + \alpha_2^2 \sigma_2^2 + \lambda (1 - \alpha_1 - \alpha_2)$$

Notwendige Bedingungen (stationäre Punkte):

$$\begin{aligned}L_{\alpha_1} : & 2 \alpha_1 \sigma_1^2 + 2 \alpha_2 \sigma_{12} - \lambda = 0 \\L_{\alpha_2} : & 2 \alpha_1 \sigma_{12} + 2 \alpha_2 \sigma_2^2 - \lambda = 0 \\L_{\lambda} : & 1 - \alpha_1 - \alpha_2 = 0\end{aligned}$$

Mit $\alpha_2 = 1 - \alpha_1$ ergibt die Lösung des Gleichungssystems

$$\alpha_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 - 2 \sigma_{12} + \sigma_2^2}$$

Varianz einer Linearkombination – Allgemein

Gegeben seien n Zufallsvariable, X_1, \dots, X_n , wobei die X_i die Erwartungswerte μ_i , die Standardabweichungen σ_i und die Kovarianzen σ_{ij} besitzen,

$$E(X_i) = \mu_i, \quad V(X_i) = \sigma_i^2 = \sigma_{ii}$$

$$\text{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ij}$$

Die Varianz einer Linearkombination dieser ZVen, $\sum_{i=1}^n a_i X_i$, ist

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij}$$

Varianz einer Linearkombination: $n = 2$

Setzen wir $n = 2$, so erhalten wir

$$\begin{aligned}V\left(\sum_{i=1}^2 a_i X_i\right) &= \sum_{i=1}^2 \sum_{j=1}^2 a_i a_j \sigma_{ij} = \\&= \sum_{j=1}^2 a_1 a_j \sigma_{1j} + \sum_{j=1}^2 a_2 a_j \sigma_{2j} = \\&= (a_1 a_1 \sigma_{11} + a_1 a_2 \sigma_{12}) + (a_2 a_1 \sigma_{21} + a_2 a_2 \sigma_{22}) = \\&= a_1^2 \sigma_1^2 + 2 a_1 a_2 \sigma_{12} + a_2^2 \sigma_2^2 = \\&= a_1^2 V(X_1) + 2 a_1 a_2 \text{Cov}(X_1, X_2) + a_2^2 V(X_2)\end{aligned}$$

Varianz einer Linearkombination: $n = 3$

Setzen wir $n = 3$, so erhalten wir

$$\begin{aligned}V\left(\sum_{i=1}^3 a_i X_i\right) &= \sum_{i=1}^3 \sum_{j=1}^3 a_i a_j \sigma_{ij} = \\&= \sum_{j=1}^3 a_1 a_j \sigma_{1j} + \sum_{j=1}^3 a_2 a_j \sigma_{2j} + \sum_{j=1}^3 a_3 a_j \sigma_{3j} = \\&= a_1^2 \sigma_1^2 + 2 a_1 a_2 \sigma_{12} + 2 a_1 a_3 \sigma_{13} + a_2^2 \sigma_2^2 + 2 a_2 a_3 \sigma_{23} + a_3^2 \sigma_3^2\end{aligned}$$

Kovarianzmatrix

Sei \mathbf{X} ein Spaltenvektor der ZVen X_i und $\boldsymbol{\mu}$ der zugehörige Vektor der Erwartungswerte.

$$\mathbf{X} = (X_1, \dots, X_n)', \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$$

Wir multiplizieren $(\mathbf{X} - \boldsymbol{\mu}) \cdot (\mathbf{X} - \boldsymbol{\mu})'$ und erhalten eine $(n \times n)$ -Matrix, in Elementschreibweise $[(X_i - \mu_i)(X_j - \mu_j)]_{n \times n}$.
Deren Erwartungswert heißt **Kovarianzmatrix** $\boldsymbol{\Sigma}$ von \mathbf{X} ,

$$\boldsymbol{\Sigma} = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'] = [E((X_i - \mu_i)(X_j - \mu_j))]_{n \times n}$$

bzw. mittels Kovarianzen $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$

$$\boldsymbol{\Sigma} = [\sigma_{ij}]_{n \times n}$$

Varianz in Matrixschreibweise

Die Varianz einer Linearkombination von ZVen in Matrixschreibweise ist

$$V\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij} = \mathbf{a}' \boldsymbol{\Sigma} \mathbf{a}$$

wobei $\mathbf{a} = (a_1, \dots, a_n)'$ der Spaltenvektor der Koeffizienten a_i ist.

Eigenschaften der Kovarianzmatrix

Die wichtigsten Eigenschaften der Kovarianzmatrix sind:

- Σ ist **symmetrisch**.

Da $\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E[(X_j - \mu_j)(X_i - \mu_i)] = \sigma_{ji}$ gilt, ist

$$\Sigma = [\sigma_{ij}] = [\sigma_{ji}] = \Sigma'$$

- Σ ist positiv **semidefinit**.

$$a' \Sigma a \geq 0 \quad \text{für alle Vektoren } a.$$

Das heißt, die Varianz einer Summe von ZVen ist immer größer oder gleich Null.

Korrelationsmatrix

Die Korrelation, ρ_{ij} , zwischen X_i und X_j erhält man aus

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii} \sigma_{jj}}}$$

wobei $\sigma_{ij} = \text{Cov}(X_i, X_j)$ und $\sigma_{ii} = \text{Cov}(X_i, X_i) = V(X_i)$.

Die Korrelationsmatrix R des Vektors X ist

$$R = [\rho_{ij}]_{n \times n} = \text{diag}\{\sigma_{11}, \dots, \sigma_{nn}\}^{-1/2} \cdot \Sigma \cdot \text{diag}\{\sigma_{11}, \dots, \sigma_{nn}\}^{-1/2}$$

Die Hauptdiagonale der Korrelationsmatrix besteht nur aus Einsen.

$$\text{diag}\{\sigma_{11}, \dots, \sigma_{nn}\}^{-1/2} = \text{diag}\{\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{nn}}\}^{-1}$$

ist Diagonalmatrix mit Standardabweichungen als Diagonalelemente.

Multivariate Normalverteilung

Sei X eine Spaltenvektor von Zufallsvariablen, $X = (X_1, \dots, X_n)'$, die gemeinsam normal verteilt sind.

$$X \sim N(\mu, \Sigma)$$

$\mu = (\mu_1, \dots, \mu_n)'$ ist der Vektor der Erwartungswerte der X_i .

$$\mu_i = E(X_i)$$

$\Sigma = [\sigma_{ij}]$ ist die Kovarianzmatrix des Vektors X .

$$\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)]$$

Man schreibt auch $X \sim MN(\mu, \Sigma)$. (multivariat normal)

Beispiel – Bivariate Normalverteilung

Sei X ein Spaltenvektor von Zufallsvariablen, $X = (X_1, X_2)'$, die gemeinsam normal verteilt sind ($n = 2$).

$$X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Die Erwartungswerte seien $\boldsymbol{\mu} = (3, 5)'$, die Kovarianzmatrix $\boldsymbol{\Sigma} = [\sigma_{ij}]$ ist gegeben durch $\sigma_{11} = V(X_1) = 4$, $\sigma_{22} = V(X_2) = 9$ und $\sigma_{12} = \text{Cov}(X_1, X_2) = 3$.

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 & 3 \\ 3 & 9 \end{pmatrix}\right)$$

Cholesky-Zerlegung der Kovarianzmatrix

Jede symmetrische, positiv definite ($n \times n$)-Matrix $\boldsymbol{\Sigma}$ kann als Produkt einer unteren Dreiecksmatrix L mit sich selbst zerlegt werden.

$$\boldsymbol{\Sigma} = L L'$$

Es gibt auch die Zerlegung

$$\boldsymbol{\Sigma} = \tilde{L} D \tilde{L}'$$

Hier ist \tilde{L} eine untere Dreiecksmatrix mit Einser in der Hauptdiagonale. D ist eine Diagonalmatrix. Es gilt: $\tilde{L} \sqrt{D} = L$.

(Untere Diagonalmatrix heißt, dass die Elemente oberhalb der Hauptdiagonale sind Null.)

Erzeugung von Normalverteilten Zufallsvektoren

Gesucht ist ein Vektor X mit $X \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ mit gegebenem $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$. Der Ansatz ist

$$X = L \boldsymbol{\epsilon} + \boldsymbol{\mu}$$

mit $\boldsymbol{\Sigma} = L L'$ und $\boldsymbol{\epsilon} \sim N(\mathbf{0}, I)$.

- 1 Berechne Cholesky-Faktor L der Kovarianzmatrix $\boldsymbol{\Sigma}$.
- 2 Generiere beliebig viele Vektoren $\boldsymbol{\epsilon}$ der Länge n mit den üblichen Zufallszahlengeneratoren für standard-normalverteilte univariate Zufallszahlen.
- 3 Einsetzen in $X = L \boldsymbol{\epsilon} + \boldsymbol{\mu}$ liefert multivariat normalverteilte Vektoren mit der gewünschten Kovarianz- bzw. Korrelationsstruktur.

Erzeugung von Normalverteilten Zufallsvektoren

Warum funktioniert die Methode?

- Der Erwartungswert von X ist der gewünschte:

$$E(X) = E(L\epsilon + \mu) = LE(\epsilon) + \mu = \mu$$

da $E(\epsilon) = 0$ gilt.

- Die Kovarianzmatrix von X ist Σ :

$$\begin{aligned} V(X) &= E[(X - \mu)(X - \mu)'] = E[(L\epsilon)(L\epsilon)'] = \\ &= E[(L\epsilon)(\epsilon'L')] = LE[\epsilon\epsilon']L' = LIL' = LL' = \Sigma \end{aligned}$$

- X ist normalverteilt, da jedes X_i aus einer Linearkombination der normalverteilten ϵ_j gebildet wird.
(Vgl. die Reproduktionseigenschaft)

Kapitel 3

Portfolio Management

Lernziele

- Konzept der modernen Portfolio-Theorie
- Capital Asset Pricing Model
- Optimieren eines Portfolios
- Analyse der Portfolioperformance

Einleitung

Einer der Bausteine der Optionspreistheorie ist Hedging des Basisobjekts mit der Option um ein risikofreies Portfolio zu erhalten.

Nicht jedes Portfolio ist bzw. soll risikofrei sein, nicht jeder Anleger hedgt.

In diesem Abschnitt sehen wir, wie ein riskantes Portfolio mit dem Ziel einen Ertrag zu garantieren und dabei das Risiko zu kontrollieren erstellt werden kann.

Annahmen

Wir treffen folgende Annahmen:

- Wir halten ein Portfolio eine *Periode* gegebener Länge, und untersuchen das Ergebnis am Ende dieser Periode.
- Während dieser Periode sind die Renditen *normalverteilt*.
- Daher genügt es die Renditen durch Erwartungswert und Standardabweichung (*Volatilität*) zu charakterisieren.

Notation

- Wir ziehen N Wertpapiere in die engere Wahl aus denen wir ein Portfolio zusammenstellen wollen. Im Folgenden beziehen wir uns nur auf Aktien.
- Der heutige Wert des Papiers i sei S_i und seine Rendite über unseren **Zeithorizont** T sei R_i .
- R_i ist die zukünftige Rendite und daher nicht bekannt.
- Die Renditen werden als **normalverteilt** angenommen mit Erwartungswert $(\mu_i T)$ und Standardabweichung $(\sigma_i \sqrt{T})$.
- Die Korrelationen zwischen den Renditen von Papieren i und Papieren j sind ρ_{ij} .
- Die Parameter μ_i , σ_i und ρ_{ij} beziehen sich auf *Drift*, *Volatilität* und Korrelation, jeweils auf die Zeiteinheit von 1 Jahr bezogen.

Ein Portfolio über den Zeithorizont T

- Ein nomineller Anlagebetrag Π wird auf N Papiere aufgeteilt.
- Wir halten von jedem Papier w_i Stück.
- Der Wert des gesamten Portfolios ist zu Beginn der Periode

$$\Pi = \sum_{i=1}^N w_i S_i$$

- Am Ende der Halteperiode ist der Wert des Portfolios

$$\Pi + \delta\Pi = \sum_{i=1}^N w_i S_i (1 + R_i)$$

Portfoliorendite und Gewichte

Für die relative Veränderung des Portfolios, die Portfoliorendite, $\frac{(\Pi + \delta\Pi) - \Pi}{\Pi} = \frac{\delta\Pi}{\Pi}$, ergibt sich

$$\frac{\delta\Pi}{\Pi} = \sum_{i=1}^N W_i R_i$$

wobei

$$W_i = \frac{w_i S_i}{\sum_{i=1}^N w_i S_i} = \frac{w_i S_i}{\Pi}$$

Die W_i sind Gewichte und geben den *wertmäßigen Anteil* für Papier i an der Gesamtinvestition Π an. Sie summieren sich zu Eins.

$$\sum_{i=1}^N W_i = 1$$

Drift und Volatilität der Portfoliorendite

Die erwartete Rendite unseres Portfolios bezogen auf die Halteperiode der Länge T , $(\mu_{\Pi} T)$, ist dann einfach

$$\mu_{\Pi} T = E\left[\frac{\delta\Pi}{\Pi}\right] = \sum_{i=1}^N W_i E[R_i] = \sum_{i=1}^N W_i (\mu_i T)$$

Der Drift des Portfolios ist

$$\mu_{\Pi} = \sum_{i=1}^N W_i \mu_i$$

Die Volatilität ist analog

$$\sigma_{\Pi} = \sqrt{\sum_{i=1}^N \sum_{j=1}^N W_i W_j \sigma_{ij}} = \sqrt{\sum_{i=1}^N \sum_{j=1}^N W_i W_j \rho_{ij} \sigma_i \sigma_j}$$

Stochastisches Modell der Portfoliorendite

Die Portfoliorendite hat die Form

$$\frac{\delta\Pi}{\Pi} = \mu_{\Pi} T + \sigma_{\Pi} \sqrt{T} Z$$

wobei Z eine standard-normalverteilte Zufallsvariable ist.

$\frac{\delta\Pi}{\Pi}$ ist normalverteilt.

(Reproduktionseigenschaft der Normalverteilung)

Mittel und Standardabweichung der Portfoliorendite hängen nur von den Mittel, Standardabweichungen und Korrelationen (Kovarianzen) der Renditen der Papiere ab.

Portfoliobildung

Einfaches Beispiel

- Angenommen wir haben nur Papiere in unserem Portfolio, die unkorreliert sind, $\rho_{ij} = 0, i \neq j$.
- Jedes Papier soll mit demselben Gewicht eingehen:
 $W_i = 1/N$.
- Der Drift der Portfoliorendite ist dann

$$\mu_{\Pi} = \frac{1}{N} \sum_{i=1}^N \mu_i$$

- Die Volatilität wird zu

$$\sigma_{\Pi} = \sqrt{\sum_{i=1}^N \frac{1}{N^2} \sigma_i^2}$$

Großes Portfolio ($N \rightarrow \infty$)

- Angenommen, alle Papiere haben dieselbe Varianz σ^2 .

$$\sigma_{\Pi} = \sqrt{\sum_{i=1}^N \frac{1}{N^2} \sigma^2} = \sqrt{N \frac{1}{N^2} \sigma^2} = \sqrt{\frac{1}{N} \sigma^2} = \frac{\sigma}{\sqrt{N}}$$

- σ/\sqrt{N} wird auch als $O(1/\sqrt{N})$, „groß O von $1/\sqrt{N}$ “, bezeichnet.
Wir sagen dazu: Die Volatilität ist $O(1/\sqrt{N})$.
Sie konvergiert mit $N \rightarrow \infty$ wie $1/\sqrt{N}$ gegen Null.
- Wenn wir die Anzahl der Papiere in unserem Portfolio erhöhen, geht die Standardabweichung unseres Portfolios gegen Null: Diversifikation reduziert die Volatilität ohne die erwartete Rendite zu reduzieren.

Ziel der Portfoliohaltung

Wir wünschen uns

- eine **hohe** Rendite, und
- eine **geringe** Volatilität.

Volatilität und Risiko:

- Volatilität wird hier als Risiko interpretiert. Große Volatilität bringt ein großes Risiko, dass kleine Renditen erzielt werden, mit. (Natürlich kommen umgekehrt auch hohe Renditen vor.)
- In Hinblick auf unsere Diskussion über die Nutzenfunktion, bedeute kleine Volatilität Risikoaversion. Wir unterstellen also risikoaverses Verhalten.
- Warum nur dieses hier sinnvoll ist, werden wir gleich sehen.

Moderne Portfolio-Theorie

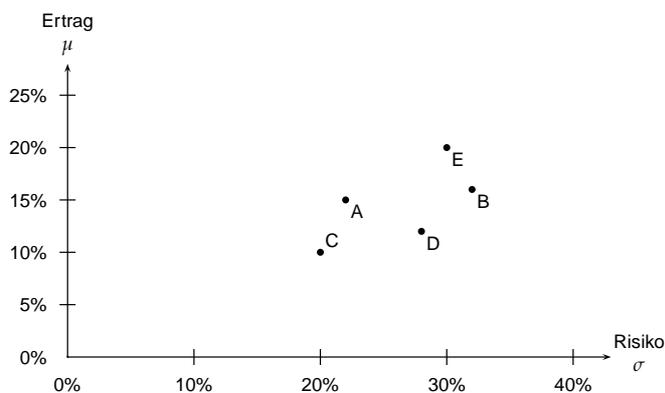
Wir suchen das „**beste**“ Portfolio, die geeignetste Zusammensetzung einer fix vorgegebenen Auswahl von Wertpapieren.

Markowitz definiert ein **effizientes** Portfolio als eines

- mit dem höchsten Ertrag für ein gegebenes Niveau von Unsicherheit, oder
- das Portfolio mit dem geringsten Risiko bei gegebenem Ertrag.

Zur Illustration legen wir ein $\sigma \times \mu$ -Diagramm an.

$\mu \times \sigma$ -Diagramm



Risiko und Ertrag von fünf Wertpapieren

Diversifikation

Setzen Portfolio aus den Papieren C und E zusammen.

Welchen Effekt haben unterschiedliche Kombinationen auf das Risiko und den Ertrag?

Aus der Formel für den Portfoliodrift und -volatilität erhalten wir

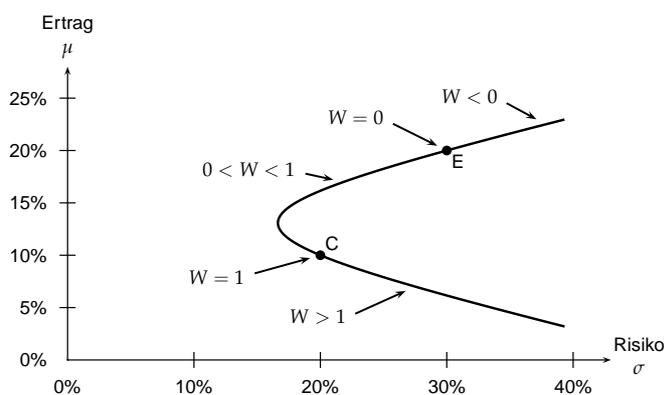
$$\mu_{\Pi} = W \mu_C + (1 - W) \mu_E$$

$$\sigma_{\Pi}^2 = W^2 \sigma_C^2 + 2W(1 - W) \rho_{CE} \sigma_C \sigma_E + (1 - W)^2 \sigma_E^2$$

W ist das Gewicht für Papier C, $(1 - W)$ das Gewicht von E.

Durch einsetzen verschiedener Werte für W erhalten wir eine Hyperbel. (Annahme: $\rho_{CE} = 0$)

Efficient Frontier / Hyperbel



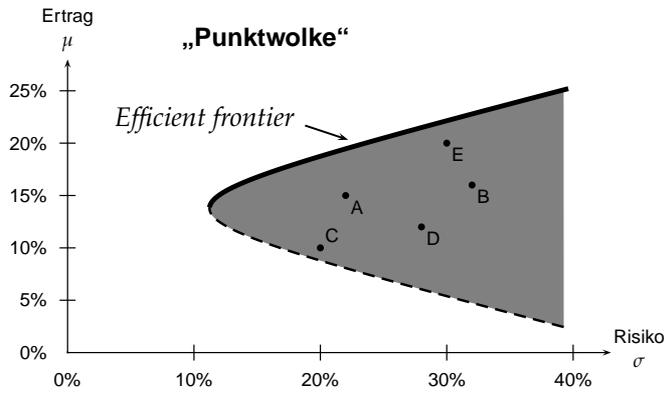
Kombination der Wertpapiere C und E

Efficient Frontier / Punkte

$$R_{\Pi} = W R_C + (1 - W) R_E$$

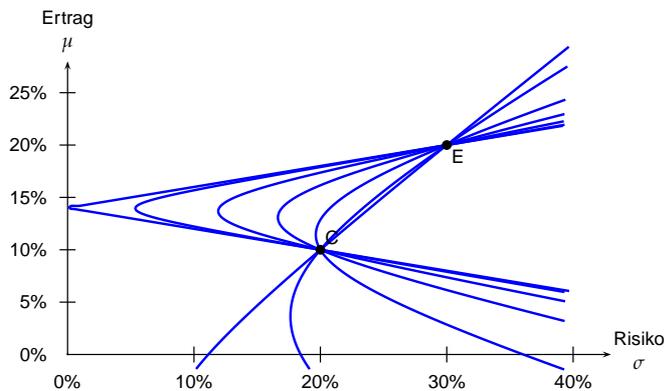
- $W = 1$: Nur Papier C.
- $W = 0$: Nur Papier E.
- $0 < W < 1$: Punkt auf der Hyperbel zwischen C und E.
- $W > 1$: Punkt unterhalb von C;
long in Papier C, *short* in Papier E.
- $W < 0$: Punkt oberhalb von E;
short in Papier C, *long* in Papier E.

Efficient Frontier



Kombination aller Wertpapiere

Efficient Frontier / Korrelation



Kombination der Wertpapiere C und E bei verschiedenen Korrelationen

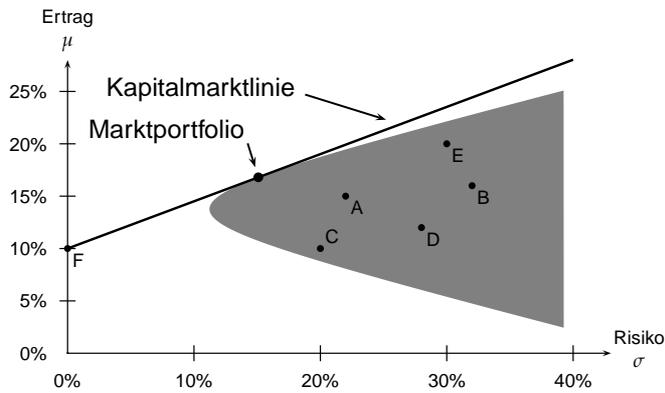
Kapitalmarktlinie und Marktportfolio

Wir betrachten Portfolios, die aus **allen** am Markt gehandelten riskanten Papieren zusammengesetzt werden, und ermitteln deren *Efficient Frontier*. Zusätzlich nehmen wir zusätzlich risikofreie Papiere auf.

Eine risikofreie Investition garantiert eine fixe Rendite von r . Die **Kapitalmarktlinie** beschreibt die *Efficient Frontier* von Portfolios, die risikofreie Papiere im Portfolio aufgenommen haben.

Das **Marktportfolio**, das Portfolio aller Wertpapiere, das sich automatisch am Markt einstellt, liegt am Berührungspunkt der Kapitalmarktlinie und der *Efficient Frontier* der riskanten Papiere.

Portfolio mit risikofreiem Wertpapier



Wahl des individuellen Portfolios

Wo wollen wir auf unserer *Efficient Frontier* sein?

Die „beste“ Position auf der *Efficient Frontier* eines individuellen Portfolios ist subjektiv.

Wir bieten dazu zwei Entscheidungsregeln an:

- Mittels der Steigung der Verbindungslinie zum risikofreien Papier. Das entspricht dem **maximalen Sharpe ratio**.
- Mittels **Nutzenfunktion** des Anlegers.

Maximaler Sharpe ratio

Die Steigung der Verbindungslinie des Portfolios II zum risikofreien Papier – Punkt $(0, r)$ zu Punkt (σ_{II}, μ_{II}) – ist

$$s = \frac{\mu_{II} - r}{\sigma_{II}}$$

heißt *Sharpe ratio* einer Anlage.

Maximierung des *Sharpe ratio* ist hier gleichbedeutend mit der Maximierung der Wahrscheinlichkeit, dass die Anlage II eine größere Rendite als die des risikofreien Zinssatzes r bringt.

Maximaler Sharpe ratio / Eigenschaften

Die Rendite R_{Π} des Portfolios Π ist normalverteilt,

$$R_{\Pi} \sim \mathcal{N}(\mu_{\Pi}, \sigma_{\Pi}^2)$$

Wahrscheinlichkeit, dass R_{Π} größer als der risikofreie Zinssatz r ist,

$$P(R_{\Pi} > r) = P\left(Z > \frac{r - \mu_{\Pi}}{\sigma_{\Pi}}\right) = P(Z > -s) = P(Z \leq s) = \Phi(s)$$

($\Phi(z)$ ist die Verteilungsfunktion der Standardnormalverteilung.)

Je größer s ist, desto größer ist die Wahrscheinlichkeit, dass die Rendite des gewählten Portfolios die einer risikofreien Anlage übertrifft.

Maximaler Sharpe ratio / Eigenschaften

Allgemein gilt

$$\Phi\left(\frac{\mu_{\Pi} - r^*}{\sigma_{\Pi}}\right)$$

ist die Wahrscheinlichkeit, dass die Rendite des Portfolios Π eine Rendite r^* übertrifft.

Wenn wir die Wahrscheinlichkeit maximieren wollen, dass wir besser als r^* abschließen, suchen wir auf der *Efficient Frontier* das Portfolio, das den steilsten Anstieg liefert.

$$\max_{\Pi} \frac{\mu_{\Pi} - r^*}{\sigma_{\Pi}}$$

Wahl des Portfolios durch Nutzenmaximierung

Wir zeichnen Indifferenzkurven der gegebenen Nutzenfunktion in das $\sigma \times \mu$ -Diagramm.

Die letzte nach oben verschobene Indifferenzkurve, die noch einen Punkt mit der *Efficient Frontier* gemeinsam hat, bestimmt das Portfolio.

Optimierungsproblem und *Efficient Frontier*

Die *Efficient Frontier* erhält man durch Minimierung der Standardabweichung bei gegebener erwarteter Portfoliorendite, μ_0 .

$$\min_{W_i} V(R_{\Pi})$$

$$\text{NB: } \begin{aligned} E(R_{\Pi}) &= \mu_0 \\ \sum_{i=1}^n W_i &= 1 \end{aligned}$$

Lösung sind die optimalen Gewichte, die wertmäßigen Anteile der einzelnen Aktien.

Die Minimierung wiederholt man für Werte aus einem Intervall von interessanten Renditen, μ_0 : $\mu_0 > r$.

(Minimierung mittels Lagrange-Ansatz)

Diskussion, Numerik

Es sind quadratische Optimierungsprobleme zu lösen für die es Lösungsalgorithmen gibt. Alternativ kann man iterative Minimierungsverfahren herabziehen, die allerdings numerisch suboptimal sind.

Diskussion um die Zuverlässigkeit des Ansatzes:

- Probleme treten i.A. bei der Verwendung der durchschnittlichen beobachteten Renditen auf. Dabei dient die Entwicklung in der Vergangenheit als Prognose für die zukünftige Rendite. Bekanntlich ist es schwierig zukünftige Kurse, wie Renditen vorherzusagen. („Vorhersagen ist schwierig, vor allem für die Zukunft.“)
- Die Varianzen der Renditen kennt man i.A. besser.

Diskussion, Numerik

- In Krisensituationen und Crashes brechen plötzlich die Korrelationsstrukturen zusammen, auf die das Verfahren aufbaut. Alle Papiere korrelieren plötzlich hoch positiv. (Z.B. die Kurse fallen alle gleichzeitig.)
Diversifikation hilft nicht, es sei denn sie streuen stärker. (. . . , Immobilien, Realitäten, Gold)
- Numerische Problem durch singuläre Korrelationsmatrizen können durch eine aufmerksam erstellte Auswahl von Papieren umgangen werden.
So sollten nicht zwei (oder mehrere) Papiere mit fast identischem Verhalten in das Portfolio genommen werden. (etwa Stamm- und Vorzugsaktie eines Unternehmens)
Für diese Probleme kann die numerische Mathematik Hilfestellungen anbieten.

Capital Asset Pricing Model (CAPM)

Das *Capital Asset Pricing Model*, **CAPM**, dient zur Bewertung einzelner Papiere bezüglich des Marktes, der durch einen Index repräsentiert wird.

Ein Index ist mit einem Portfolio vergleichbar: DAX30, S&P500, etc.

Dazu wird ein Beta-Koeffizient berechnet. Das Beta, β , eines Papiers **relativ zum Marktportfolio** M ist der Quotient aus der Kovarianz zwischen der Rendite des Papiers und der Rendite des Portfolios, und der Varianz der Rendite des Portfolios.

$$\beta_i = \frac{\text{Cov}(R_i, R_M)}{V(R_M)}$$

Die Interpretation erfolgt im Rahmen des *Single-Index Modells*.

Single-Index Modell

Wir setzen die Rendite jedes einzelnen Papiers zur Rendite **eines repräsentativen Index** M in Beziehung

$$R_i = \alpha_i + \beta_i R_M + \epsilon_i$$

Das lineare Modell für R_i besteht aus drei Teilen:

- Einem konstanten Drift, α_i ,
- Einer gemeinsamen Komponente mit dem Marktindex R_M , $\beta_i R_M$, und
- einem zufälligen Teil ϵ_i , der unkorreliert mit dem Index ist.

Die Koeffizienten werden durch Schätzen des linearen Regressionsmodells ermittelt.

Single-Index Modell / Interpretation

- α_i ist die Konstante in der Regressionsgleichung.
 $R_i = \alpha_i$ (im Durchschnitt), wenn $R_M = 0$.
Also, wenn $\alpha_i \neq 0$, besitzt Papier i eine autonome Rendite, auch wenn der Markt nur Null erwirtschaftet.
- β_i ist der Steigungskoeffizient.
Steigt die Marktrendite um 1, so steigt R_i um β_i .
 - Papiere, die sich mit dem Markt bewegen, haben positive Koeffizienten; Papiere, die sich gegen den Markt bewegen, negative (Gold ?).
 - Papiere mit einem $\beta_i > 1$ reagieren stärker als der Markt; Papiere mit einem $\beta_i < 1$ schwächer.
- Der Fehler der Regression, ϵ_i , ist mit R_M unkorreliert.

Single-Index Modell und Marktindex

Sei μ_M die erwartete Rendite des Index, und σ_M deren Standardabweichung.

- Erwartete Rendite des i -ten Papiers:

$$\mu_i = \alpha_i + \beta_i \mu_M$$

- Standardabweichung:

$$\sigma_i = \sqrt{\beta_i^2 \sigma_M^2 + e_i^2}$$

- e_i^2 ist die Varianz von ϵ_i .

Single-Index Modell und Portfoliorendite

Die Rendite eines Portfolio Π ist

$$\frac{\delta \Pi}{\Pi} = \sum_{i=1}^N W_i R_i = \left(\sum_{i=1}^N W_i \alpha_i \right) + R_M \left(\sum_{i=1}^N W_i \beta_i \right) + \sum_{i=1}^N W_i \epsilon_i$$

Die erwartete Rendite ist ($E(\epsilon_i) = 0$)

$$\mu_{\Pi} = \left(\sum_{i=1}^N W_i \alpha_i \right) + E(R_M) \left(\sum_{i=1}^N W_i \beta_i \right)$$

Wir erhalten daher für die erwartete Portfoliorendite

$$\mu_{\Pi} = \alpha_{\Pi} + \beta_{\Pi} \mu_M$$

mit

$$\alpha_{\Pi} = \sum_{i=1}^N W_i \alpha_i, \quad \beta_{\Pi} = \sum_{i=1}^N W_i \beta_i, \quad \mu_M = E(R_M)$$

Single-Index Modell und Portfoliorendite

Die Volatilität in Π ergibt sich als

$$\sigma_{\Pi} = \sqrt{\sum_{i=1}^N \sum_{j=1}^N W_i W_j \beta_i \beta_j \sigma_M^2 + \sum_{i=1}^N W_i^2 e_i^2}$$

Dies geht so einfach, weil alle ϵ_i unkorreliert mit R_M sind.

Unter der vereinfachten Annahme, dass alle Gewichte gleich ($1/N$) sind, alle $e_i^2 = e^2$ und alle $\beta_i = \beta$ ergibt sich für die Volatilität der Portfoliorendite

$$\begin{aligned} \sigma_{\Pi}^2 &= \sum_{i=1}^N \sum_{j=1}^N (1/N) (1/N) \beta \beta \sigma_M^2 + \sum_{i=1}^N (1/N)^2 e^2 \\ &= N^2 (1/N)^2 \beta^2 \sigma_M^2 + N (1/N)^2 e^2 \\ &= \beta^2 \sigma_M^2 + e^2 / N \end{aligned}$$

Single-Index Modell und Diversifikation

Die Volatilität der Portfoliorendite

$$\sigma_{\Pi}^2 = \beta^2 \sigma_M^2 + e^2/N$$

besteht aus 2 Teilen:

- Das **diversifizierbare Risiko** ist mit den ϵ_i verbunden: e^2/N .
Dessen Beitrag zur Varianz verschwindet mit steigendem N ,
($O(N^{-1/2})$).
- Das **systematische Risiko** korreliert mit dem Index: $\beta^2 \sigma_M^2$
Es kann durch Portfoliobildung nicht reduziert werden.

Single-Index Modell und optimales Portfolio

Die Optimierung unter Verwendung des *Single-Index* Modells sieht folgende Schritte vor:

1. Berechne α_i , β_i und e_i^2 zu allen Papieren.
2. Wähle einen Wert für die erwartete Portfoliorendite μ_{Π} .
3. Minimiere σ_{Π} unter dieser Nebenbedingung.
4. Wiederhole die Minimierung für verschiedene Portfoliorenditen um die *Efficient Frontier* zu erhalten.
5. Die Position auf der Kurve ist subjektiv zu entscheiden, oder nach dem maximalen *Sharpe ratio* Kriterium.

Multi-Index Modell

Die Idee des *Single-Index* Modells kann um andere repräsentative (i.A. korrelierte) Indizes R_{M_j} erweitert werden.

Zum Beispiel kann zusätzlich zu einem Aktienmarktindex

- ein repräsentativer Bondmarktindex,
- ein Index für Währungsmärkte, oder
- ein volkswirtschaftlicher Index

in die Regression miteinbezogen werden, wenn man glaubt, dass er für die Papiere von Bedeutung ist.

$$R_i = \alpha_i + \sum_{j=1}^k \beta_{ij} R_{M_j} + \epsilon_i$$

Performance-Messung

Wie misst man die Performance von Anlagestrategien?

Ein Kriterium basiert auf den Vergleich mit der risikofreien Rate. Ideal wäre die risikofreien Rate **konsistent** zu überbieten.

Ein anderes Kriterium hat zum Ziel nicht nur eine hohe Rendite zu erzielen, sondern dies auch mit so wenig Varianz wie möglich zu bewältigen.

Sharpe ratio und Treynor ratio

Zwei der häufigsten Maßzahlen von „Rendite pro Risikoeinheit“

- *Sharpe ratio* setzt „Ertrag zur Variabilität“ in Beziehung:

$$\text{Sharpe ratio} = \frac{\mu_{\Pi} - r}{\sigma_{\Pi}}$$

- *Treynor ratio* setzt „Ertrag zur Volatilität“ in Beziehung:

$$\text{Treynor ratio} = \frac{\mu_{\Pi} - r}{\beta_{\Pi}}$$

μ_{Π} und σ_{Π} bzw. β_{Π} sind die realisierten Werte in der Beobachtungsperiode.

σ_{Π} ist die beobachtete Standardabweichung,

β_{Π} ist ein Maß für die beobachtete „Volatilität“ des Portfolios.

Sharpe ratio und Treynor ratio

- Der *Sharpe ratio* wird üblicherweise verwendet, wenn man das Portfolio aller Investitionen betrachtet.
- Der *Treynor ratio* wird verwendet, wenn einzelne Komponenten eines gesamten Unternehmensportfolios bewertet werden.

Kapitel 4

Daten

Lernziele

- Datentypen
- Beobachtungsfrequenz
- Fluss- und Bestandsdaten
- Indizes: Preis-, Mengen- und Umsatzindex
- Preisbereinigung (Deflationieren)
- Verketteten von Indizes

Frequenz der Beobachtung

Die meisten Daten für wirtschaftliche Problemstellungen liegen in der folgenden **Periodizität** vor:

- Tick-Daten (*Intraday*-Kursdaten)
- Tagesdaten (Tagesschlußkurse, Verkaufszahlen)
- Wochendaten
- Monatsdaten (Aktienindizes, volkswirtschaftliche Aggregate)
- Quartalsdaten (Bilanzdaten von Unternehmen)
- Halbjahresdaten
- Jahresdaten (volkswirtschaftliche Aggregate, Lagerbestände nach Inventur)

Häufigkeit der Ereignisse

Speziell bei der Nachfrage nach einzelnen Gütern kann man unterscheiden:

- sehr selten nachgefragt: 0, 1 oder 2 Stück pro Woche;
- öfters nachgefragt: 0 bis 10 Stück;
- häufig nachgefragt.

Die später vorgestellten Methoden beruhen auf stetigen (meist bedingt normalverteilten) Variablen.

Problem:

Modellierung von 0 (**Null**) wenn die Wahrscheinlichkeit groß ist.

Lösungsmöglichkeit:

- Markoffmodell mit den Zuständen "0" und ">0";
- Schätzung der Übergangswahrscheinlichkeiten;
- konventionelle Prognosemethode für den Fall ">0".

Fluss- und Bestandsdaten

- **Flussdaten** (*flows*) messen Aktivitäten zwischen Wirtschaftssubjekten in einer Periode, deren Summe der Periode zugeordnet wird.
z.B.: Wöchentliche Verkaufszahlen, BIP, Investitionen, Stromproduktion.
- **Bestandsdaten** (*stocks*) messen den Bestand zu einem Zeitpunkt.
Die Veränderungen von Bestandsdaten sind Flussgrößen.
z.B.: Lagerbestand am 31. 12. jedes Jahres, Geldmenge, Kapitalstock, Aktienbestand, Anzahl an Beschäftigten oder Arbeitlosen.

Menge, Preis und Umsatz

- **Menge**, Volumen in einer Periode gemessen in realen Einheiten.
z.B.: Kilogramm innerhalb eines Monats, Stück pro Tag, BIP zu konstanten Preisen pro Quartal, kWh pro Jahr.
- **Preis**, in Währungseinheiten zu einem bestimmten Zeitpunkt gemessen.
z.B.: Preis eines Big Mac in London in BPD am 15. März, Aktienschlusskurs in USD, Preisindizes, Verbraucherpreisindex, Lohnsatz.
- **Umsatz** = Preis \times Menge.
Hängt von Preis- und Mengenentwicklung ab.
Beide werden meist getrennt modelliert.
z.B.: nominales BIP, Umsatz/Absatz in Euro, Bilanz.

Relativer Preis

Der **relative Preis** P_t^R ist der Quotient

$$P_t^R = \frac{P_t}{P_{0t}}$$

- P_t ist der Preis des betreffenden Gutes;
- P_{0t} ist der Preis eines **Referenzgutes**
(Die Preisentwicklung des Gutes wird in Relation zu einem Referenzgut beschrieben.)
oder eines geeigneten **Preisindizes**.
(Die Inflation der Preisreihe P_t wird elimiert.)

z.B.: Preis eines Mittelklassewagens relativ zum VPI;
Vergleich 1990 und 2005.

Preisbereinigung

Durch **Preisbereinigung** (*Deflationieren*) erhält man aus **nominalen** Größen mittels eines **Deflators** (ein Preisindex) „reale“ Werte, eigentlich Werte zu konstanten Preisen (zu Preisen des Basisjahrs des Deflators).

Für das BIP wird ein Preisindex $PBIP$ mit variablen Umsatzgewichten erhoben, sodass

$$BIP_t^{\text{real}} = \frac{BIP_t^{\text{nom}}}{PBIP_t}$$

$PBIP$ ist der Deflator des BIP.

Indices

Ein **Index** ist eine aggregierte Maßzahl, die eine Aussage über eine Gruppe verschiedener aber ähnlicher Kennzahlen macht.

(z.B.: Preise von Konsumgütern.)

Diese Maßzahl wird in Relation zu einer **Basisperiode** (oder Basisobjekt) angegeben und für diese Basisperiode (bzw. Basisobjekt) auf **100** normiert.

Indices / (2)

Wir unterscheiden

- Preis-,
- Mengen- und
- Umsatzindizes,

die sich auf ein einmal **fixiertes** Güterbündel beziehen.

Sie beschreiben die Preis-, Mengen-, und Umsatzentwicklung eines Güterbündels.

Im weiteren werden zwei Arten von Indizes unterschieden:

Indices nach **Laspeyres** und nach **Paasche**.

Bezeichnung:

0 bezeichnet die *Basis(Referenz)periode*.

1 bezeichnet die aktuelle *Berichtsperiode*.

Umsatzindex

Der **Umsatzindex**, U_{01} , beschreibt die Umsatzentwicklung eines Güterbündels.

$$U_{01} = \frac{\sum p_1 q_1}{\sum p_0 q_0} 100$$

Der Summationsindex ist hier weggelassen worden.

Eigentlich lautet die Formel

$$U_{01} = \frac{\sum_{i=1}^n p_{i1} q_{i1}}{\sum_{i=1}^n p_{i0} q_{i0}} 100$$

i ist der Laufindex und bezeichnet die verschiedenen Güter, die in den Index einbezogen werden.

Beispiel: Umsatzindex der Elektrohaushaltsgeräte von SIEMENS für die Jahre 1985 bis 2004 mit Basisjahr 1995 (1995=100).

Preisindex nach Laspeyres

Der **Preisindex nach Laspeyres**, P_{01}^L , ist ein Preisindex **fixen** Umsatzgewichten.

$$P_{01}^L = \sum w_{00} \frac{p_1}{p_0} 100 = \frac{\sum p_1 q_0}{\sum p_0 q_0} 100$$

Die Gewichte w_{i00} der einzelnen Güter werden in der Basisperiode (für das Basisobjekt) bestimmt und geben den Anteil von Gut i am Umsatz in der Basisperiode an.

$$w_{i00} = \frac{p_{i0} q_{i0}}{\sum p_{i0} q_{i0}}$$

Verbraucherpreisindex (VPI, engl *CPI*). Für einen „typischen“ Haushalt werden in Österreich – seit 1990 alle 5 Jahre – das Konsummuster erhoben und für die folgenden Jahre festgehalten.

Preisindex nach Paasche

Der **Preisindex nach Paasche** ist ein Preisindex mit **variablen** „Umsatz“gewichten.

Die Gewichte der einzelnen Güter werden in jeder Periode (für jedes Objekt) neu bestimmt.

$$P_{01}^p = \sum w_{01} \frac{p_1}{p_0} 100 = \frac{\sum p_1 q_1}{\sum p_0 q_1} 100$$

Die Gewichte w_{i01} geben den Anteil von Gut i am Umsatz gemessen in den aktuellen Mengen, aber zu den Preisen der Basisperiode an.

$$w_{i01} = \frac{p_{i0} q_{i1}}{\sum p_{i0} q_{i1}}$$

z.B.: Deflator des Bruttoinlandsprodukts.

Mengenindex nach Laspeyres

Der **Mengenindex nach Laspeyres** ist ein Mengenindex mit **fixen** Umsatzgewichten. (Analog zum Preisindex nach Laspeyres)

Die Gewichte der einzelnen Güter werden in der Basisperiode (für das Basisobjekt) bestimmt.

$$Q_{01}^L = \sum w_{00} \frac{q_1}{q_0} 100 = \frac{\sum p_0 q_1}{\sum p_0 q_0} 100$$

Mengenindex nach Paasche

Der **Mengenindex nach Paasche** ist ein Mengenindex mit **variablen** „Umsatz“gewichten.

Die Gewichte der einzelnen Güter werden in jeder Periode (für jedes Objekt) neu bestimmt.

$$Q_{01}^p = \sum w_{10} \frac{q_1}{q_0} 100 = \frac{\sum p_1 q_1}{\sum p_1 q_0} 100$$

Die Gewichte w_{i10} geben den Anteil von Gut i am Umsatz gemessen in den aktuellen Preisen, aber zu den Mengen der Basisperiode an.

$$w_{i10} = \frac{p_{i1} q_{i0}}{\sum p_{i1} q_{i0}}$$

Beispiel: BIP und Deflator

Das nominale BIP ist das Produkt aus BIP-Deflator und realem BIP:

$$BIP^{\text{nom}} = PBIP \cdot BIP^{\text{real}}$$

Das **nominelle BIP** der Berichtsperiode 1, ergibt sich aus allen Umsätzen, der in der jeweiligen Periode produzierten Güter minus Vorleistungen. Es hat die Form $\sum p_1 q_1$.

Das **reale BIP** ist die aktuelle Produktion bewertet zu Preisen des Basisjahres 0: $\sum p_0 q_1$.

Der **Deflator des BIP** ist daher ein Paasche'scher Preisindex:

$$\sum p_1 q_1 = \underbrace{\frac{\sum p_1 q_1}{\sum p_0 q_1}}_{P_{01}^P} \sum p_0 q_1$$

$$U_{01} = P_{01}^P \cdot Q_{01}^L$$

Dividiert man beide Seiten von $\sum p_1 q_1 = \frac{\sum p_1 q_1}{\sum p_0 q_1} \sum p_0 q_1$ durch $\sum p_0 q_0$ so erhalten wir

$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum p_0 q_1}{\sum p_0 q_0}$$
$$U_{01} = P_{01}^P \cdot Q_{01}^L$$

Das Produkt aus Paasche'schen Preisindex und dem Laspeyres'schen Mengenindex ergibt einen Umsatzindex.

Analog gilt für den Preisindex nach Laspeyres und dem Mengenindex nach Paasche

$$U_{01} = P_{01}^L \cdot Q_{01}^P$$

Verkettung von Indizes

Zur Berechnung des VPI wird alle 5 Jahre ein neues Konsummuster erhoben. Für eine längere Indexreihe müssen daher die 5-jährigen Indexreihen verkettet werden.

Um zwei Indexreihen zu **Verketteten** wird für eine Periode der Index sowohl nach der alten, wie nach der neuen Gewichtung berechnet, und anschließend die alten Werte auf das neue Niveau umgerechnet.

P_t^a ... alter Index für $t = 1, 2, \dots, t_0$

P_t^n ... neuer Index für $t = t_0, t_0 + 1, \dots$

P_t^K ... verketteter Index für $t = 1, 2, \dots, t_0, t_0 + 1, \dots$

$$P_t^K = \begin{cases} \kappa P_t^a & t = 1, \dots, t_0, \\ P_t^n & t = t_0, \dots \end{cases} \quad \text{mit } \kappa = \frac{P_{t_0}^n}{P_{t_0}^a}$$

Prognose

Lernziele

- Aufgabe der Prognose
- Problemtypen
- Ablauf einer Prognoseaufgabe
- Zeitreihe

Zeitreihenanalyse

Zeitreihenanalyse wird aus drei Gründen betrieben:

- Beschreibung des Verlaufs von Zeitreihen.
- **Prognose** von wirtschaftlichen Größen.
- Überprüfung von Verhaltensmodellen
 - um ökonomische Prozesse zu beschreiben, oder
 - um sie zu steuern.
z.B.: Wirtschaftspolitik, Lagerbestandskontrolle

Zeitreihenanalyse

Zeitreihenanalyse wird aus drei Gründen betrieben:

- Beschreibung des Verlaufs von Zeitreihen.
- **Prognose** von wirtschaftlichen Größen.
- Überprüfung von Verhaltensmodellen
 - um ökonomische Prozesse zu beschreiben, oder
 - um sie zu steuern.
z.B.: Wirtschaftspolitik, Lagerbestandskontrolle

Prognose

Prognostizieren ist schwierig, überhaupt wenn es sich um die Zukunft handelt.

Der Erfolg vieler betrieblicher Entscheidungen hängt (auch) von zukünftigen Entwicklungen ab und ist damit mit Unsicherheit behaftet.

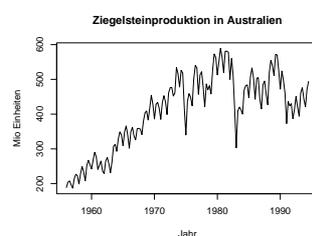
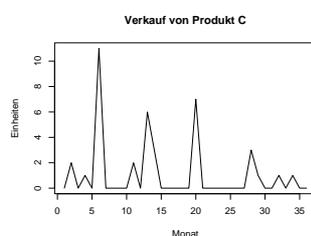
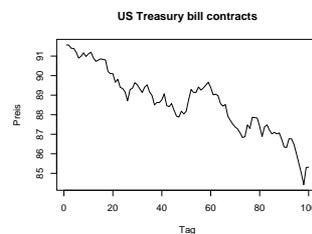
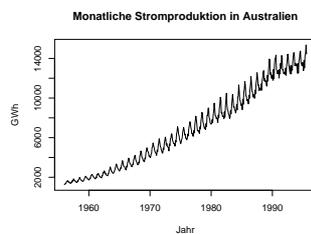
Die **Prognose** versucht die Unsicherheit zu reduzieren. Historisches Verhalten wird dabei in

- **systematische** und
- **nicht-systematische** („zufällige“) Effekte

zerlegt und in die Zukunft weitergeführt.

Systematische Effekte werden modelliert, nicht-systematische Effekte werden identifiziert und minimiert.

Historische Daten



Problemtypen

Wir unterscheiden drei Arten von Problemen:

- **Quantitative** Prognose:
Genügend quantifizierte Information verfügbar.
 - **Zeitreihen**
 - Strukturmodell (*Regressionsanalyse*)
- **Qualitative** Prognose:
Wenig quantifizierte Daten verfügbar, aber genügend (qualitative) Kenntnisse der Zusammenhänge.
- **Unvorhersagbar**:
Zu wenige Informationen verfügbar.

Ablauf einer Prognoseaufgabe

Eine Prognose läuft in 5 Schritten ab:

- ➊ Problemdefinition
- ➋ Sammeln relevanter Informationen
- ➌ Explorative Datenanalyse
- ➍ Auswahl and Anpassen des Modells
- ➎ Verwendung des Modells zur Prognose

1 · Problemdefinition

Schwierigster Teil der Aufgabe!

- **Zweck** der Prognose.
Wofür? Für wen? Wer ist betroffen? Kann der Prozess kontrolliert werden?
- Erfordert **Verständnis** des zugrundeliegenden Prozesses.
Beispiel: Lagerhaltung.
Wovon hängt der aktuelle Lagestand ab?
Beispiel: BIP.
Von welchen Faktoren hängt das (Wachstum des) BIP ab?

2 · Sammeln relevanter Informationen

Zwei Arten von Informationen verfügbar:

- statistische **Daten** (meist numerisch)
- gesammelte **Erfahrungen** („Expertenwissen“) und wissenschaftliche Erkenntnisse („Theorie“)

Beide werden zusammengefasst.

3 · Explorative Datenanalyse

Analysemethoden auf unterschiedlichem Niveau:

- Graphische Aufarbeitung und **visuelle** Beurteilung der Daten: Streudiagramme, Histogramme, Box-Plot, etc.
- **Statistische Kennzahlen** (Deskriptive Statistik): Min, Max, Mittelwerte, Standardabweichung, Korrelation, Autokorrelation, etc.
- **Glättungsmodelle:**
 - **Trends:** linear, nicht-linear, exponentielles Wachstum, Sättigungsverhalten.
 - **Saisonen:** regelmäßige Muster zu festen Zeitpunkten.
 - **Zyklen:** lange unregelmäßige Schwingungen.
 - **Ausreißer** und ungewöhnliche Datenpunkte.
 - **irreguläres**, unsystematisches Verhalten.

4 · Auswahl and Anpassen des Modells

- **Glättungsverfahren:** gleitender Durchschnitt, exponentielle Glättung, Verfahren nach Holts-Winter.
- **Lineare Regression** und linearisierbare Modelle.
- **ARMA** (*autoregressive moving average model*): univariates Modell für stationäre Zeitreihen (erweitert zu ARIMA für nicht-stationäre Reihe).
- **GARCH** (*generalized autoregressive conditional heteroscedastic model*): zur Modellierung von Varianzen von Renditen.
- Multivariate Modelle.

Diese Verfahren unterscheiden sich auch in der Eignung kurz-, mittel- oder langfristige Phänomene zu erfassen.

4 · Auswahl and Anpassen des Modells / (2)

Wenn das gewählte Modell mit den Daten vergleicht, spricht man von **Anpassen** des Modells.

Eine gute Anpassung liegt vor, wenn das beobachtete Verhalten (Trend, Zyklus, ...) gut vom Modell reproduziert werden.

Ein gutes Verständnis für drei Bereiche ist erforderlich:

- für Prozessablauf, betriebswirtschaftliche Zusammenhänge, etc.,
- für die mathematischen Eigenschaften des Modells,
- für die statistischen Werkzeuge um Modell und Daten zu Vergleichen.

5 · Verwendung des Modells zur Prognose

Letzter und einzig relevante Test:

Das Modell muss eine brauchbare Prognose liefern.

Vorgangsweise:

- In den Schritten 3 und 4 werden die letzten Werte der Zeitreihe oder einer Kontrollgruppe nicht miteinbezogen.
- Aus den verwendeten Daten wird eine Prognose für die Kontrolldaten errechnet und mit den tatsächlichen Werten verglichen.
- Die Differenz heißt **Prognosefehler**.
- Mittlere Quadratsumme (u.a. Kenngrößen) als Kriterium zur Bewertung des Prognosefehlers.

Definition: Zeitreihe

Eine **Zeitreihe** ist eine Realisation eines stochastischen Prozesses in diskreter Zeit

$$\{x_t, t \in \mathbb{Z}\}$$

- Der Prozess beginnt in $-\infty$ und geht bis $+\infty$.
- Die x_t sind Zufallsvariable.
- Folge von Zufallsvariablen, die stets dasselbe aber zu unterschiedlichen, äquidistanten Zeitpunkten messen.
- Eine Realisation hat einen Anfang, $t = 1$, und ein Ende, $t = T$.

$$\{x_t, t = 1, \dots, T\}$$

- Sowohl Zufallsvariablen als auch deren Realisationen werden in der Zeitreihenanalyse mit Kleinbuchstaben bezeichnet.

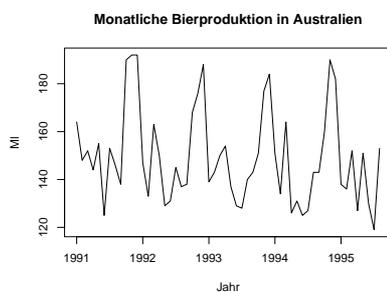
Basiswerkzeuge

Lernziele

- Beschreiben von Zeitreihen
- Graphische Darstellungen
- Univariate und bivariate Maßzahlen
- Transformationen von Zeitreihen
- Maßzahlen für die Prognosegüte

Zeitreihen-Plot

Zeitpfad wird visualisiert:



Muster und Charakteristika werden meist sichtbar:

- Trends
- Zyklen
- Saison
- Irreguläres Verhalten (keine Regelmäßigkeit)
- Ausreißer, atypische Beobachtungen

Muster

Wir können vier Muster unterscheiden:

- **Horizontales Muster.**
Daten fluktuieren um ein konstantes Mittel.
z.B.: Renditen, Bierproduktion
- **Saisonales Muster.**
Saisonbehafte Daten, fluktuieren mit fixer Periodizität.
z.B.: Stromverbrauch, Speiseeisproduktion
- **Zyklisches Muster.**
Daten fluktuieren in Zyklen unregelmäßiger Länge und Amplitude.
z.B.: Konjunkturzyklen
- **Trend.**
Daten steigen oder fallen über eine längere Periode.
z.B.: BIP-real, Stromproduktion

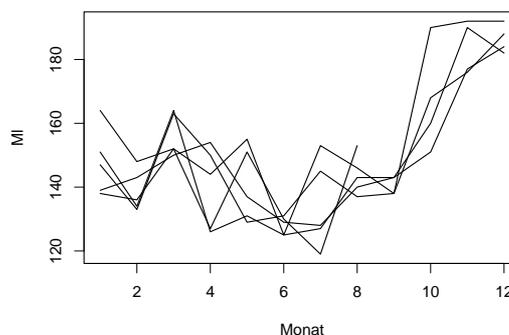
Überlagerung der Strukturen

- Ausgeprägte Strukturen machen die Wahl eines geeigneten Modells leicht.
- In kurzen Reihen sind einzelne Muster nicht zuverlässig erkennbar.
- Die Unterscheidung zwischen Trend und langen Zyklen ist allgemein ein schwieriges Problem.
- Zyklen, die kurz im Verhältnis zu Frequenz der Beobachtungen liegen sind schwer zu erkennen (oder erzeugen Scheinperiodizitäten).

Saisonale Plots

Saisonale Plots zeigen die Daten für einzelne „Saisonen“ (hier Monate).

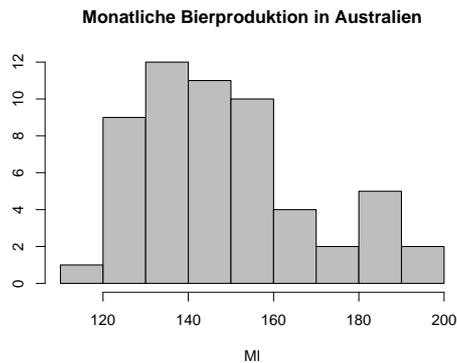
Monatliche Bierproduktion in Australien



Histogramm

Das **Histogramm** dient zur graphischen Beschreibung der Verteilung einer Variablen.

Nur sinnvoll bei horizontalem Muster (stationärer Struktur).

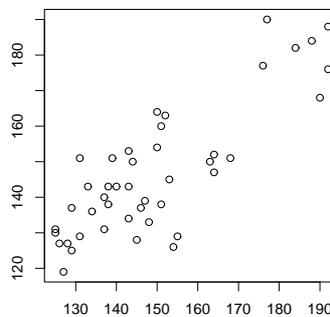


Streudiagramm (*scatter plot*)

Im **Streudiagramm** werden zwei Variable gegeneinander aufgetragen (Punktwolke).

Es erlaubt Korrelationen abzuschätzen.

(Korrelations-/Regressionsanalyse)



Univariate Maßzahlen

Maßzahlen für Zeitreihen $\{y_t\}, t = 1, \dots, T$.

- **Mittelwert**

$$\bar{y} = \frac{1}{T} \sum y_t$$

- **Median**

Me = „mittlerer Wert“

- **mittlerer absoluter Abstand**

$$\text{MAD} = \frac{1}{T} \sum |y_t - \bar{y}|$$

- **mittlerer quadratischer Abstand**

$$\text{MSD} = \frac{1}{T} \sum (y_t - \bar{y})^2$$

- **Varianz**

$$s_y^2 = \frac{1}{T-1} \sum (y_t - \bar{y})^2$$

- **Standardabweichung**

$$s_y = \sqrt{s^2} = \sqrt{\frac{1}{T-1} \sum (y_t - \bar{y})^2}$$

Summation über alle $t = 1, \dots, T$.

Bivariate Maßzahlen

Zwei Zeitreihen $\{x_t\}$ und $\{y_t\}$, $t = 1, \dots, T$.

- **Kovarianz** $\text{Cov}_{xy} = \frac{1}{T-1} \sum (x_t - \bar{x})(y_t - \bar{y})$

- **Korrelation** $r_{xy} = \frac{\text{Cov}_{xy}}{s_x s_y} = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}$

Der Korrelationskoeffizient liegt immer zwischen -1 und 1 :

$$-1 \leq r_{xy} \leq 1$$

Wir verwenden hier die Stichprobenvarianz bzw. -kovarianz.

Autokorrelation

Die **Autokorrelation k -ter Ordnung** misst die Korrelation zwischen Werten einer Reihe, die k Perioden voneinander entfernt sind:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Die Werte y_{t-k} und y_t sind k Perioden voneinander entfernt. Wir sagen y_{t-k} ist der Wert von y_t um k Perioden **verzögert**.

Durch Zusammenfassen der Autokorrelationen der Ordnung $0, 1, 2, 3, \dots$ erhält man die **Autokorrelationsfunktion** oder das **Korrelogramm**: $r_0 = 1, r_1, r_2, r_3, \dots$

Nur sinnvoll für stationäre Zeitreihen (horizontales Muster).

Autokorrelation erster Ordnung für die Australische Bierproduktion:

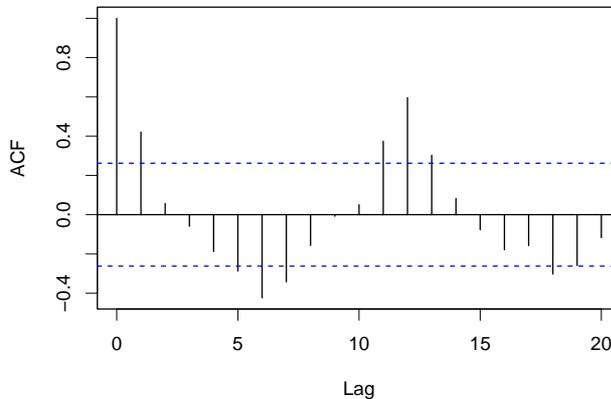
(1)	(2)	(3)	(4)	(5)	(6)	(7)
t	y_t	y_{t-1}	$(y_t - \bar{y})$	$(y_{t-1} - \bar{y})$	$(y_t - \bar{y})^2$	$(y_t - \bar{y}) \cdot (y_{t-1} - \bar{y})$
1	164	—	14.70	—	215.99	—
2	148	164	-1.30	14.70	1.70	-19.16
3	152	148	2.70	-1.30	7.27	-3.51
4	144	152	-5.30	2.70	28.13	-14.30
⋮						⋮
55	119	130	-30.30	-19.30	918.31	584.97
56	153	119	3.70	-30.30	13.66	-112.01
Σ	8361				21135.84	8893.51

$$T = 56, \bar{y} = 8361/56 = 149.30,$$

$$r_1 = \frac{\sum_{t=2}^T (y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} = \frac{8893.51}{21135.84} = 0.421$$

Korrelogramm

Korrelogramm der australischen Bierproduktion



Stationärer Prozess

Ein **stationärer Prozess** hat konstanten

- Erwartungswert $E(y_t) = \mu$ für alle t
- Varianz $V(y_t) = \sigma$ für alle t
- Autokovarianzen $\text{Cov}(y_t, y_{t+k}) = \gamma(k)$ für alle t
- Autokorrelationen $\rho_k = \gamma(k) / \gamma(0)$ für alle t

Der Zeitreihenplot eines stationären Prozesses hat:

- horizontaler Verlauf des Mittels,
- Streuung um Mittel gleichmäßig über die Zeit.

Die Autokorrelationsfunktion eines stationären Prozesses

- geht geometrisch gegen Null,
- macht saisonale Muster sichtbar.

Interpretation von r_k

In der Regressionsanalyse wird ein lineares Modell betrachtet:

$$y_t = a + b x_t + \epsilon_t$$

Die Methode der kleinsten Quadrate liefert die Schätzer

$$\hat{b} = \frac{\text{Cov}(x, y)}{s_x^2} = \text{Corr}(x, y) \frac{s_y}{s_x} \quad \text{und} \quad \hat{a} = \bar{y} - \hat{b} \bar{x}$$

Im Fall $x_t = y_{t-k}$ gilt $\sigma_x = \sigma_y$ und $\text{Corr}(x, y) = r_k$ und somit

$$y_t = a + \rho_k y_{t-k} + \epsilon_t$$

Diese Regression macht Sinn, und y_{t-k} kann zur Prognose y_t verwendet werden.

Ziel der Transformation

Reihen können mit Standardverfahren modelliert werden, wenn sie

- einen linearen Trend, oder
- konstantes Mittel und
- konstante Varianz

besitzen.

Exponentielles Wachstum ist nicht linear kann aber durch Logarithmierung **linearisiert** werden:

$$x_t = \log(y_t)$$

(log ist hier der natürliche Logarithmus.)

z.B.: Aktienindizes, reales BIP

Varianzstabilisierung

Allgemeiner als Logarithmierung ist die **Box-Cox Transformation**.

$$y_t^{BC} = \begin{cases} -y_t^p & \text{für } p < 0 \\ \log(y_t) & \text{für } p = 0 \\ y_t^p & \text{für } p > 0 \end{cases}$$

Steigt (selten fällt) die Varianz einer Reihe mit steigendem Mittel, so kann eine konstante Varianz durch die geeignete Wahl von p erreicht werden (**Varianzstabilisierung**). **Rücktransformation:**

$$y_t = \begin{cases} (-y_t^{BC})^{1/p} & \text{für } p < 0 \\ \exp(y_t^{BC}) & \text{für } p = 0 \\ (y_t^{BC})^{1/p} & \text{für } p > 0 \end{cases}$$

Varianzstabilisierung / (2)

Bedeutung des Parameters p :

- $p > 1$: Die Varianz für große Werte wird größer.
- $p = 1$: Reihe bleibt unverändert.
- $p < 1$: Die Varianz für große Werte wird kleiner.
- $p = 0$: Geeignete Transformation für exponentielles Wachstum.
- $p < 0$: Die Varianz für große Werte wird noch stärker reduziert.

Prognose

Das **Prognoseproblem** lautet:

Wir beobachten Daten $y_i, i = 1, \dots, t$, und wollen wissen, wie sich die Reihe in der Zukunft, \hat{y}_j , für die Zeitpunkte $j = (t + 1), (t + 2), \dots$ entwickelt.

Allgemein unterscheidet man

- **Punkt-** und
- **Intervallprognose.**

Ähnlich wie Punkt- und Intervallschätzer.

Naive Prognose

Der letzte Wert wird als Prognose verwendet:

$$\hat{y}_{t+1} = y_t$$

z.B.: Wenn der Bierabsatz in der letzten Jahr 80 MI betragen hat, dann prognostizieren wir für heuer ebenfalls 80 MI.

Saisonale Effekte können dabei berücksichtigt werden:

$$\hat{y}_{t+1} = y_{t-i+1}$$

wobei i die Länge der Periode ist.

z.B.: Wenn der Bierabsatz im August des letzten Jahres 60 MI betragen hat, dann prognostizieren wir für den heuerigen August ebenfalls 60 MI.

Einfache Prognosen

Angenommen unsere Reihe weist

- ein **horizontales Muster** auf:
Das Mittel der Reihe wird fortgeschrieben:

$$\hat{y}_{t+1} = \bar{y}$$

- einen **linearen Trend** auf:
Der Trend wird zur Prognose fortgeschrieben:

$$\hat{y}_{t+1} = \hat{a} + \hat{b}(t + 1)$$

\hat{a} und \hat{b} sind die Schätzer für das lineare Modell $y_t = a + bt + \epsilon_t$.

τ -Schritt Prognose und Prognosefehler

Die Prognose für τ Schritte in die Zukunft heißt **τ -Schritt Prognose**:

$$\hat{y}_{t+\tau}$$

Der **Prognosefehler** für τ ($\tau = 1, 2, \dots$) ist

$$e_{t+\tau} = y_{t+\tau} - \hat{y}_{t+\tau}$$

$y_{t+\tau}$ sind die zukünftigen, noch nicht beobachteten Werte.

Ziel ist es, den Prognosefehler so klein wie möglich zu machen.

in-sample / *out-of-sample* Prognose

Wir unterscheiden:

- ***in-sample* Prognose (ex post)**:
Ein Modell wird für $t = 1, \dots, T$ geschätzt. Für die Zeitpunkte der Schätzperiode (i.e. $1 \leq t \leq T$) werden Prognosen berechnet.
- ***out-of-sample* Prognose (ex ante)**:
Die Prognose wird für Zeiten außerhalb der Schätzperiode berechnet. z.B. für $T + \tau$, $\tau = 1, 2, \dots$

Out-of-sample Prognosen werden oft so durchgeführt, daß die Länge der Schätzperiode festgehalten wird, und diese Periode über die Zeit hinweg verschoben wird (*moving window*).

Der **Prognosehorizont** (z.B. 5-Schritt Prognose) wird festgehalten und mit der Prognoseperiode mitbewegt.

Prognoseintervall

Sind die Fehler ϵ_t bzw. e_t normalverteilt, so kann man auch **Prognoseintervalle** angeben.

95% Prognoseintervall:

$$[\hat{y}_{t+\tau} - 1.96 \hat{\sigma}_e(\tau), \hat{y}_{t+\tau} + 1.96 \hat{\sigma}_e(\tau)]$$

$\hat{\sigma}_e(\tau)$ ist die Standardabweichung des τ -Schritt Prognosefehlers.

Wie man diese bestimmt, werden wir später sehen.

Maße für die Prognosegüte

- **Mittlerer Fehler, Mean Error, ME**

misst den *Bias* (Verzerrtheit) einer k -Schritt Prognose:

$$ME = \frac{1}{k} \sum_{j=1}^k (y_{t+j} - \hat{y}_{t+j})$$

- **Mittlere Fehlerquadrate, Mean Square Error, MSE**

misst $Bias^2 + \text{Fehlervarianz}$:

$$MSE = \frac{1}{k} \sum_{j=1}^k (y_{t+j} - \hat{y}_{t+j})^2$$

Der MSE ist analog zur Fehlerquadratsumme konstruiert.

Maße für die Prognosegüte / (2)

- **Root Mean Square Error, RMSE**

analog zur Standardabweichung:

$$RMSE = \sqrt{MSE}$$

- **Mittlerer absoluter Fehler, Mean Absolute Error, MAE**

misst die absoluten Abweichungen (nicht die quadrierten).

$$MAE = \frac{1}{k} \sum_{j=1}^k |y_{t+j} - \hat{y}_{t+j}|$$

Maße für die Prognosegüte / (3)

Alle vier Kriterien können auch als **relative Fehler** (*Percent Error*) definiert werden.

- **Mean Absolute Percent Error, MAPE**

$$MAPE = \frac{100}{k} \sum_{j=1}^k \left| \frac{y_{t+j} - \hat{y}_{t+j}}{y_{t+j}} \right|$$

Hier wird der Fehler in Relation zum realisierten Wert gesetzt. „Große“ Fehler bei „großen“ y_t -Werten werden genauso bewertet wie „kleine“ Fehler bei „kleinen“ Realisationen.

Relative Fehler sind nur sinnvoll, wenn die verwendete Skala einen sinnvollen Nullpunkt hat.

Maße für die Prognosegüte / (4)

Wenn ein Modell für mehrere Reihen, $y_t^{(i)}$, $i = 1, \dots, n$ auf dessen Prognosegüte überprüft werden sollen, hält man die Schrittlänge k fest und mittelt über die Werte $\hat{y}_{t+k}^{(i)}$, $i = 1, \dots, n$, z.B.:

$$\text{MSE}_k = \frac{1}{n} \sum_{i=1}^n \left(y_{t+k}^{(i)} - \hat{y}_{t+k}^{(i)} \right)^2$$

Beispiele:

- Der datengenerierende Prozess ist bekannt oder vorgegeben. Mittels Simulation werden künstliche Datensätze generiert und die Prognosen aus den ausgewählten Modellen getestet.
- Die Modelle sollen für eine Vielzahl von Datensätzen eingesetzt werden.

Theil's U -Statistik

Theil's U -Statistik vergleicht die verwendete Methode mit der Güte der naiven Methode.

$$U = \sqrt{\frac{\sum_{j=1}^{k-1} \left(\frac{y_{t+j+1} - \hat{y}_{t+j+1}}{y_t} \right)^2}{\sum_{j=1}^{k-1} \left(\frac{y_{t+j+1} - y_{t+j}}{y_t} \right)^2}}$$

Interpretation:

- $U = 1$: die naive Methode ist genauso gut.
- $U < 1$: die gewählte Prognosetechnik ist besser als die naive Methode.
- $U > 1$: die naive Methode produziert bessere Prognosen.

Kapitel 7

Zerlegung von Zeitreihen

Lernziele

- Klassische Zerlegung von Zeitreihen
- Saisonbereinigungsverfahren: Gleitende Durchschnitte
- Glättungsverfahren zur Prognose: Exponentielle Glättung
- Verfahren von Holt und Holt-Winters
- Modellwahl

Zerlegungsmodelle

Die Zeitreihe Y_t wird in drei Komponenten zerlegt:

T_t ... **Trend-Zyklus**

S_t ... **Saisonale Komponente**

E_t ... **irreguläre Komponente**

$$Y_t = f(T_t, S_t, E_t)$$

Die Funktion $f(\cdot)$ kann eine Summe oder Produkt sein:

additive Zerlegung

multiplikative Zerlegung

$$Y_t = T_t + S_t + E_t$$

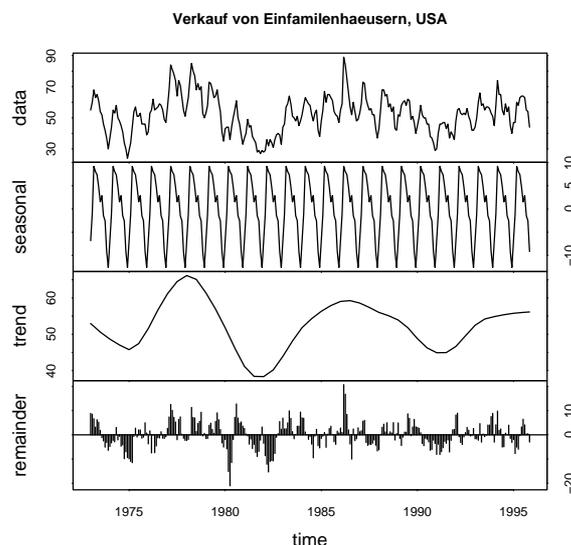
$$Y_t = T_t \cdot S_t \cdot E_t$$

Der saisonale Effekt ist keine Funktion des Niveaus T_t .

Der saisonale Effekt wird durch T_t verstärkt.

Die multiplikative Zerlegung wird durch Logarithmieren additiv:

$$Y_t = T_t \cdot S_t \cdot E_t \quad \Rightarrow \quad \log Y_t = \log T_t + \log S_t + \log E_t$$



Saisonbereinigung

Die **saisonbereinigte Reihe** zu Y_t erhalten wir indem wir die Saisonkomponente von der beobachteten Reihe abziehen,

$$(Y_t - S_t) = T_t + E_T$$

sodass Trend-Zyklus und irreguläre Komponente übrigbleiben.

Vorgangsweise:

- 1 Berechnung der Trendkomponente T_t .
(z.B. linearer, exponentieller, kein Trend oder Glättung)
- 2 Berechnung der trendbereinigten (*detrended*) Reihe $(Y_t - T_t)$.
- 3 Ermittlung der Saisonkomponente, S_t , aus $(Y_t - T_t)$.
- 4 Berechnung von $(Y_t - S_t)$.

Gleitender Durchschnitt

Der **gleitende Durchschnitt** (*moving average*), **k MA**, ist ein Verfahren zur Berechnung der Trend-Zyklus-Komponente.

k ist ein **Zeitfenster** mit einer **ungeraden** Anzahl von Beobachtungen: $k = 2m + 1$.

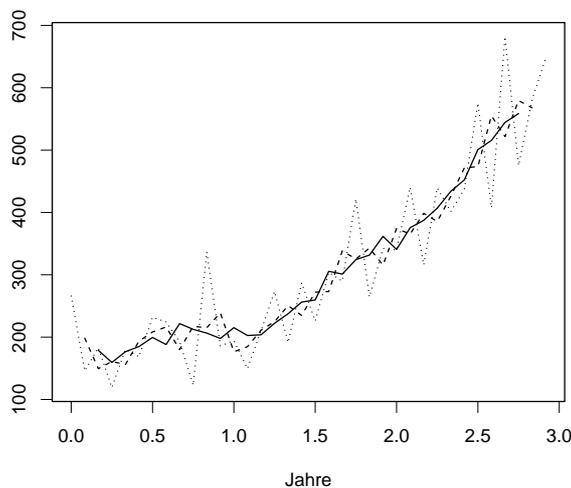
$$T_t = \frac{1}{2m+1} \sum_{j=-m}^m Y_{t+j} = \frac{1}{2m+1} (Y_{t-m} + \dots + Y_t + \dots + Y_{t+m})$$

Die neue Reihe T_t ist um die ersten m und letzten m Beobachtungen kürzer als Y_t .

Die Stärke der Glättung hängt von der Wahl von k ab.

- Für große k wird Y_t stark,
- für kleine k wenig geglättet.

Haarshampooverkauf – 5 MA



Zentrierter gleitender Durchschnitt

Der gleitende Durchschnitt kann nur für ungerade Fensterweiten ausgerechnet werden, Periodenlängen sind aber oft gerade. (z.B.: 4 Quartale, 12 Monate)

In diesem Fall kann der **zentrierte gleitende Durchschnitt** (*centered moving average*) verwendet werden.

Zum Beispiel:

Der zentrierte 4 MA ist definiert als ein 2 MA eines 4 MA, bezeichnet als **2 × 4 MA**.

$$T_3 = \frac{1}{2}(T'_1 + T'_2) = \frac{1}{8}(Y_1 + 2Y_2 + 2Y_3 + 2Y_4 + Y_5)$$

wobei die (T'_i) einen 4 MA bezeichnet.

Zentrierter gleitender Durchschnitt

2 × 4 MA und 12 × 4 MA

Allgemein gilt für einen **2 × 4 MA**:

$$T_t = (0.5Y_{t-2} + Y_{t-1} + Y_t + Y_{t+1} + 0.5Y_{t+2})/4$$

Für Monatsdaten verwendet man den **2 × 12 MA**:

$$T_t = (0.5Y_{t-6} + Y_{t-5} + \dots + Y_t + \dots + Y_{t+5} + 0.5Y_{t+6})/12$$

Der 2 × 4 MA ist ein spezieller 5 MA mit nicht-konstanten Gewichten.

Gewichteter gleitender Durchschnitt

Der zentrierte MA ist ein Spezialfall eines **gewichteten gleitenden Durchschnitts** (*weighted k moving average*):

$$T_t = \sum_{j=-m}^m a_j Y_{t+j} \quad \text{mit} \quad \sum_{j=-m}^m a_j = 1, \quad a_j \geq 0$$

Die $a_j \geq 0$ heißen Gewichte. $k = 2m + 1$ ist ungerade.

Beispiel:

Für den 2x4 MA lauten die Gewichte:

$$\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}$$

Additive Zerlegung für Monatsdaten

Wir schreiben den Zeitindex t in der Form: $t = J.M = \text{Jahr.Monat}$

- 1 Trend-Zyklus Komponente über einen 2×12 MA ermitteln: T_t .

- 2 *Detrending*, Trend aus Y_t eliminieren:

$$(Y_t - T_t) = S_t + E_t$$

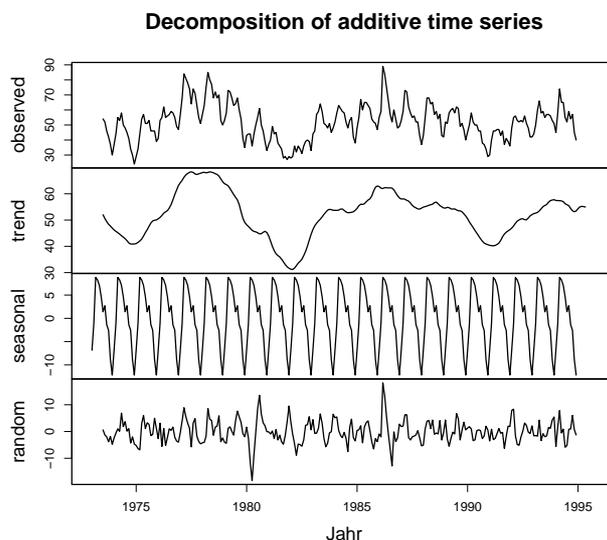
- 3 Saisonale Indizes (für all 12 Monate) berechnen:

$$S_{J.M} = s_M = \frac{1}{N_M} \sum (Y - T)_{J.M} \quad M = 1, \dots, 12$$

N_M bezeichnet die Anzahl der jeweiligen Monate.

- 4 Die irreguläre Komponente E_t ergibt sich als Residuum.

$$E_t = Y_t - T_t - S_t$$



Multiplikative Zerlegung für Monatsdaten

Wir gehen analog zum additiven Modell vor.

- 1 Trend-Zyklus Komponente über einen 2×12 MA ermitteln: T_t .

- 2 *Detrending*, Trend aus Y_t eliminieren:

$$R_t = (Y_t / T_t) = S_t \cdot E_t$$

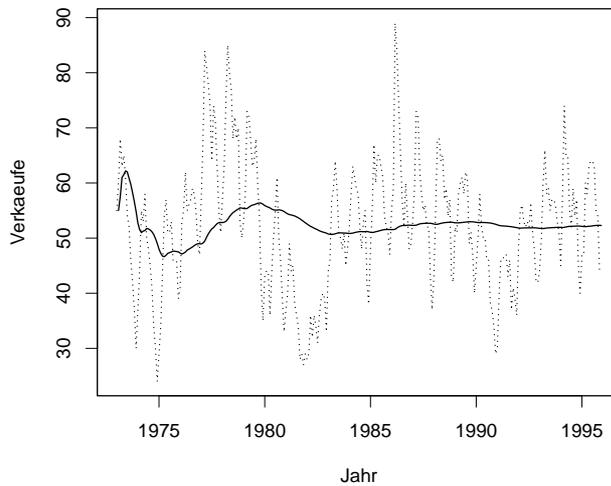
- 3 Saisonale Indizes (für all 12 Monate) berechnen:

$$S_{J.M} = s_M = \frac{1}{N_M} \sum R_{J.M} \quad M = 1, \dots, 12$$

- 4 Die irreguläre Komponente E_t ergibt sich als Residuum.

$$E_t = Y_t / (T_t \cdot S_t)$$

Hausverkäufe: Prognose – Mittelwert



Gleitender Durchschnitt

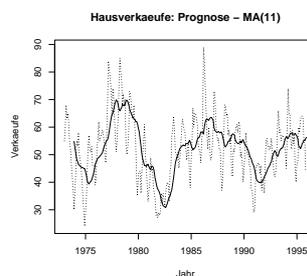
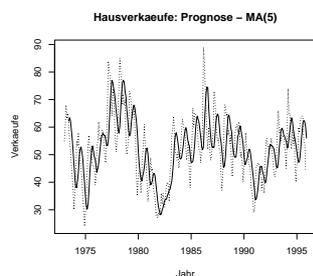
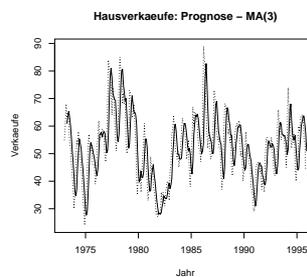
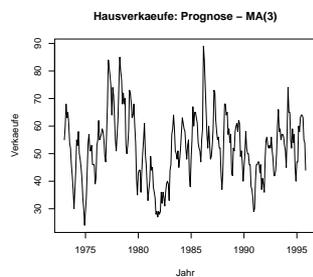
Für Reihen mit Trend ist ein **gleitender Durchschnitt** (*moving average*) **MA(k)** besser geeignet:

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^t Y_i$$

Man beachte den Unterschied von k MA und $MA(k)$.

Die Prognose kann auch rekursiv aus der letzten Prognose berechnet werden:

$$F_{t+1} = F_t + \frac{1}{k}(Y_t - Y_{t-k})$$



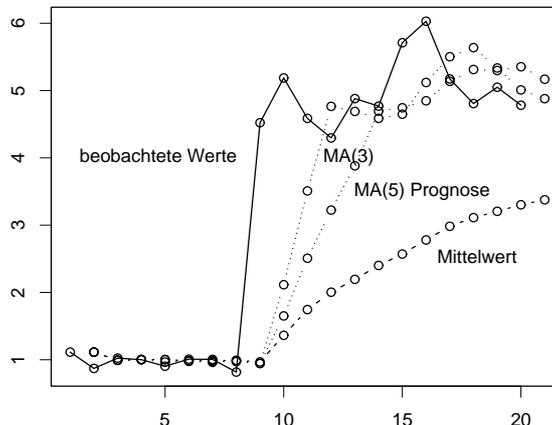
MA(k), Mittelwert und naive Prognose

Spezialfälle der MA(k) Prognose:

- Für $k = t$ fallen Mittelwert und MA zusammen.
- Für $k = 1$ erhalten wir $F_{t+1} = Y_t$, i.e. die naive Prognose. Der zuletzt beobachtete Wert wird als Prognose verwendet. (Das ist optimal für einen Random Walk ohne Drift.)

Generell gilt:

Sprünge oder Verschiebungen werden bei großem k nur langsam berücksichtigt.



Einfaches exponentielles Glätten

Das **einfache exponentielle Glätten**, (*simple exponential smoothing*), **SES**, ist ein gewichteter gleitender Durchschnitt mit geometrisch fallenden Gewichten und nur einem Parameter: $0 < \alpha < 1$.

$$F_{t+1} = F_t + \alpha(Y_t - F_t)$$

Die Prognose für morgen ist die letzte Prognose plus einer Fehlerkorrektur für die letzte Prognose.

- Ist α groß ($\alpha \approx 1$), wird stark auf den letzten Fehler reagiert, die Anpassung ist schnell.
- Ist α klein ($\alpha \approx 0$), wird kaum auf den Fehler reagiert, die Anpassung ist langsam.

Geeignet für **Reihen ohne Trend** und **ohne Saison**.

Einfaches exponentielles Glätten / (2)

Andere Darstellung der Rekursion:

$$F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$$

F_{t+1} ist ein gewichteter Durchschnitt aus Y_t und F_t .

Explizite Darstellung:

$$F_{t+1} = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i Y_{t-i} + (1 - \alpha)^t F_1$$

Prognose für die weitere Zukunft:

$$F_{t+m} = F_{t+1} \quad \text{für } m > 0$$

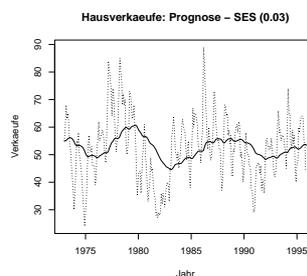
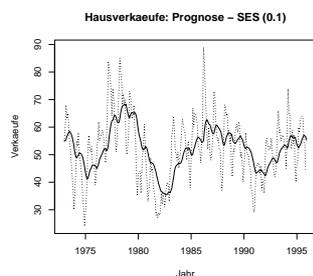
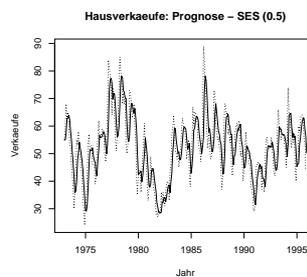
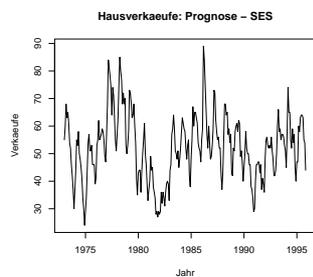
SES – Anfangswerte

Für den Anfangswert kann

- die erste Beobachtung, oder
- ein Mittel aus den ersten Beobachtungen

verwenden.

Probleme entstehen bei kurzen Datensätzen und einem kleinen α .



SES – optimales α

Das „optimale“ α wird z.B. so gewählt, dass die Summe der quadrierten (1-Schritt-) Prognosefehler minimiert wird.

Berechnung:

- numerisches Verfahren zur *nicht-linearen Optimierung*, oder
- Gitter-Suche, *grid search*.

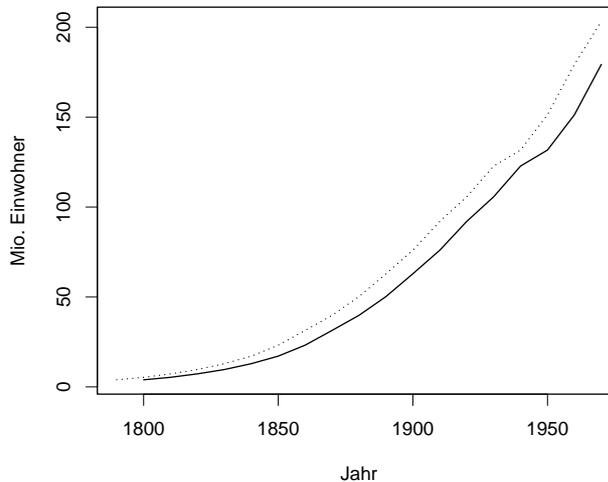
Diese ist sehr einfach.

Man wählt ein Gitter für das Intervall $(0, 1)$, z.B.

$\alpha = 0.1, 0.2, \dots, 0.9$, und berechnet für jedes gewählte α die mittlere Fehlerquadratsumme, MSE.

Das α , das den kleinsten MSE liefert, ist das beste.

USA: Prognose – SES



Holt's Methode

Das Verfahren von **Holt** erweitert das SES auf **Reihen mit linearem Trend**.

Diese Verfahren benötigt dafür zwei Parameter: $0 < \alpha, \beta < 1$.

Das Modell lautet

$$\text{Niveau: } L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1})$$

$$\text{Trend: } b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$$

$$\text{Prognose: } F_{t+m} = L_t + b_t m$$

(*Trend* bedeutet hier die Steigung des Trends.)

Holt's Methode / (2)

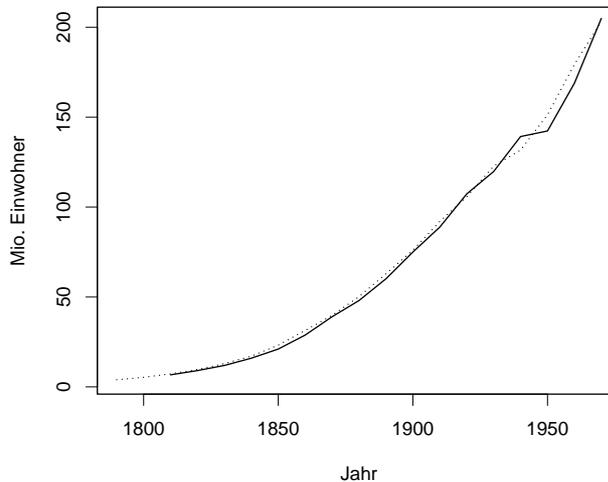
- Die Trendgleichung ist ein SES der (1-Perioden-) Steigung der Reihe mit dem Parameter β .
- Die Niveaugleichung ist ähnlich einem SES von Y mit dem Unterschied, dass der letzte Niveauwert um die Steigung verschoben wird.
- Die Prognosegleichung gibt die m -Schritt-Prognose an. Daher wird b_t mit m multipliziert.

Wir benötigen hier zwei Anfangswerte: L_1 und b_1 .

z.B.: $L_1 = Y_1$ und $b_1 = (Y_3 - Y_1)/2$

Die optimalen Parameter können wiederum über ein Gitter (α, β) (z.B.: mit $\alpha, \beta = 0.1, \dots, 0.9$) gewählt werden.

USA: Prognose – Holt



Die Methode von Holt-Winters

Die **Methode von Holt-Winters** berücksichtigt auch saisonale Effekte und ist damit zur Prognose von Reihen mit **linearem Trend und Saison** geeignet.

Es treten hier 3 Glättungsparameter auf:

$$\alpha, \beta \text{ und } \gamma \text{ mit } 0 < \alpha, \beta, \gamma < 1$$

Die optimalen Parameterwerte werden wie oben, entweder über nicht-lineare Optimierung oder über die Gittersuche ermittelt.

Wir unterscheiden hier zwischen einem Modell mit additiver und mit multiplikativer Saisonkomponente.

Holt-Winters mit additiver Saison

Im additiven Modell ist die Saison vom Niveau der Reihe unabhängig.

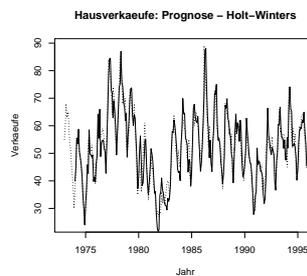
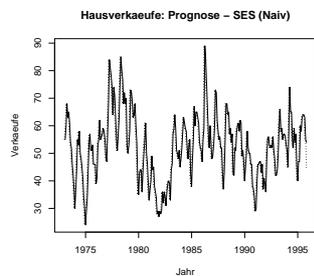
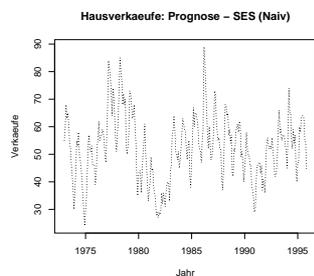
Niveau: $L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1})$

Trend: $b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$

Saison: $S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s}$

Prognose: $F_{t+m} = L_t + b_t m + S_{t-s+m}$

- Neu ist die Saisongleichung. Hier wird ein gleitender Durchschnitt der $(Y_t - L_t)$ Komponente (Beobachtung minus geglättetem Niveau) beschrieben.
- s ist 12 für Monatsdaten, und 4 für Quartalsdaten, etc.
- Das Niveau L ergibt sich aus Y minus dem Saisoneffekt.



Holt-Winters mit multiplikativer Saison

Im multiplikativen Modell ist der Saisoneffekt proportional zum Niveau der Reihe. Der Trend selbst ist linear.

Niveau: $L_t = \alpha(Y_t / S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1})$

Trend: $b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1}$

Saison: $S_t = \gamma(Y_t / L_t) + (1 - \gamma)S_{t-s}$

Prognose: $F_{t+m} = (L_t + b_t m) \cdot S_{t-s+m}$

- Die Saisongleichung beschreibt einen gleitenden Durchschnitt des Quotienten (Y_t / L_t) .

Holt-Winters – Anfangswerte

Die benötigten Anfangswerte sind:

$$L_s, b_s \text{ und } S_1, \dots, S_s$$

Für das multiplikative Modell gibt es folgende Möglichkeit:

- $L_s = (Y_1 + \dots + Y_s)/s$,
- $b_s = [(Y_{s+1} - Y_1)/s + \dots + (Y_{s+s} - Y_s)/s]/s$

Bemerkung: $(Y_{s+1} - Y_1)/s$ ist die durchschnittliche 1-Perioden-Steigung als Durchschnitt über s Perioden berechnet.

- $S_1 = Y_1/L_s, \dots, S_s = Y_s/L_s$.

Diskussion

- Die vorgestellten Methoden sind einfach und benötigen höchstens drei Parameter.
- Sie lassen sich leicht automatisieren, so dass sie auf eine große Zahl von Reihen (z.B. in großen Lagerhaltungsproblemen) anwendbar sind.
- Die Zielfunktion (minimale Fehlerquadratsumme) kann bei Vorliegen von Ausreißern z.B. durch die Minimierung des MAPE, *mean absolute percentage error*, ersetzt werden.
- Die vorgestellten Modelle sind Spezialfälle des stochastischen ARIMA Modells, das später behandelt wird.
- Reihen mit exponentiellem Wachstum und multiplikativer Saison lassen sich durch Logarithmieren in Reihen mit linearem Trend und additiver Saison überführen.

Modellwahl

Die einzelnen Schritte zur Wahl eines geeigneten Modells sind:

1. Teilen der Daten in zwei Teile:
 - Schätzdatensatz, *initialization set, estimation sample*:
Für diesen Datensatz wird das Modell angepasst.
 - Testdatensatz, *test set*:
Hier wird die Prognosegüte des geschätzten Modells gemessen.
2. Wahl der Glättungsmethode.
3. Schätzen des Modells für die Initialisierungsmenge.
4. Prognosegüte des Modells mit den geschätzten Parametern aus Schritt 3 an Hand des Testsatzes messen.
5. Bei Unzufriedenheit mit dem Ergebnis nochmals zu Schritt 2.

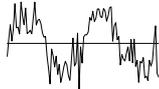
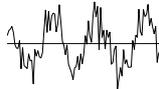
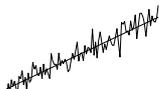
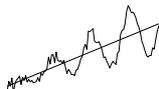
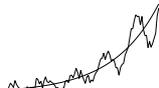
Pegel's Klassifikation

Pegel (1969) klassifiziert Zeitreihen nach additiven oder multiplikativen Trend und Saisonkomponente.

Daraus ergeben sich eine Reihe von Modellen, die die Methode von Holt-Winters erweitern.

Die Modelle von Holt-Winters werden zu Spezialfällen.

In der Praxis sucht man sich ein geeignetes Modell aus.

	saisonale Komponente		
	keine	additiv	multiplikativ
kein T.			
additiver T.			
multipl. T.			

Kapitel 8

Einfache Regression

Lernziele

- Lineares Regressionsmodell
- Anpassen des linearen Regressionsmodells, OLS
- Eigenschaften der Schätzer für das Modell
- Testen und Konfidenzintervalle
- Überprüfung der Modellannahmen
- Prognose mittels linearer Regression
- Multiples Regressionsmodell

Das lineare Regressionsmodell

Das einfache lineare Regressionsmodell lautet

$$y_t = \alpha + \beta x_t + \varepsilon_t \quad t = 1, \dots, n$$

y_t ... abhängige Variable α ... Interzept, Konstante
 x_t ... unabhängige Variable β ... Steigung
 ε_t ... Fehler, Residuum

Modellvoraussetzung

Annahmen für das einfache lineare Regressionsmodell:

- Die Beziehung zwischen y und x ist linear.
- Die x_t sind deterministische (nicht-stochastische) Variablen. Ihre Werte sind genau bekannt.
- Die Residuen ε_t sind

- unkorreliert,

$$\text{Cov}(\varepsilon_{t_i}, \varepsilon_{t_j}) = 0 \quad \text{für } i \neq j$$

- haben Erwartungswert 0 und konstante Varianz,

$$E(\varepsilon_t) = 0, \quad V(\varepsilon_t) = \sigma_\varepsilon^2, \quad \text{für alle } t$$

- und sind normalverteilt,

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

Der bedingte Erwartungswert

Der bedingte Erwartungswert $E(y_t|x_t)$ von y_t gegen x_t ist

$$E(y_t|x_t) = \alpha + \beta x_t$$

da $E(\varepsilon_t) = 0$ und x_t bekannt ist.

Das Mittel von y_t wird als Funktion von x_t angesehen.

- $E(y_t)$ bezeichnet das univariate (**unbedingte**) Mittel von y_t , i.e., der Mittelwert aller y_t über die Zeit wenn die Information x_t nicht verwendet/bekannt ist.
- $E(y_t|x_t)$ ist eine **bedingte Erwartung**. Die Information x_t wird zur Erklärung/Modellierung von y_t verwendet.

Die bedingte Varianz

- Die **unbedingte Varianz** ergibt sich als

$$V(y_t) = E([y_t - E(y_t)]^2)$$

- Die **bedingte Varianz** $V(y_t|x_t)$ ist analog

$$V(y_t|x_t) = E\left((y_t - E(y_t|x_t))^2|x_t\right) = \sigma_\varepsilon^2$$

In unserem Modell ist die bedingte Varianz von y_t konstant:

$$y_t - E(y_t|x_t) = (\alpha + \beta x_t + \varepsilon_t) - (\alpha + \beta x_t) = \sigma_\varepsilon^2$$

Und somit

$$V(y_t|x_t) = E(\varepsilon_t^2|x_t) = E(\varepsilon_t^2) = \sigma_\varepsilon^2$$

Varianzreduktion

Zur Berechnung von $V(y_t|x_t)$ wird die Information x_t verwendet. Daher gilt immer (sofern das Modell einen Erklärungswert hat)

$$V(y_t) > V(y_t|x_t) = \sigma_\varepsilon^2$$

Die Regression reduziert die Varianz in den Daten. Sie erklärt somit einen Teil der Varianz in y_t .

Anpassen eines Regressionsmodells

Die Parameter des Regressionsmodells werden so gewählt, daß die **Fehlerquadratsumme der Residuen minimal** wird.

$$\min_{\alpha, \beta} \sum_{t=1}^n \varepsilon_t^2 = \min_{\alpha, \beta} \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \min_{\alpha, \beta} \sum_{t=1}^n (y_t - (\alpha + \beta x_t))^2$$

Das Minimierungsproblem wird mit der Methode der kleinsten Quadrate (*ordinary least squares*, **OLS**) gelöst.

Man berechnet dazu die ersten partiellen Ableitungen und setzt diese gleich Null.

$$\hat{\alpha} = \frac{1}{n} \sum y_t - \hat{\beta} \frac{1}{n} \sum x_t \quad \hat{\beta} = \frac{n \sum x_t y_t - \sum x_t \sum y_t}{n \sum x_t^2 - (\sum x_t)^2}$$

OLS-Schätzer

Man kann diese Schätzer auch anders darstellen:

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad \hat{\beta} = \frac{\text{Cov}(x, y)}{V(x)} = \text{Corr}(x, y) \frac{s_y}{s_x}$$

Der Schätzer für β wird Null, falls x_t und y_t unkorreliert sind. Dann wird die Variable x_t nicht in die Regressionsgleichung aufgenommen.

Nach der Schätzung lautet unser Modell:

$$y_t = \hat{\alpha} + \hat{\beta} x_t + \hat{\varepsilon}_t$$

$\hat{\varepsilon}_t$ sind die geschätzten Fehler.

Verteilung der Schätzer

Unter Annahme (Nullhypothese), daß unser Modell das *wahre Modell* ist, sind die OLS-Schätzer $\hat{\alpha}$ und $\hat{\beta}$ normalverteilt mit Erwartungswert α bzw. β :

$$\hat{\alpha} \sim N(\alpha, \sigma_{\hat{\alpha}}^2) \quad \hat{\beta} \sim N(\beta, \sigma_{\hat{\beta}}^2)$$

mit

$$\sigma_{\hat{\alpha}}^2 = \sigma_{\varepsilon}^2 \frac{\sum x_t^2}{n \sum (x_t - \bar{x})^2} = \sigma_{\varepsilon}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_t - \bar{x})^2} \right)$$
$$\sigma_{\hat{\beta}}^2 = \sigma_{\varepsilon}^2 \frac{1}{\sum (x_t - \bar{x})^2}$$

Verteilung der Schätzer / (2)

σ_ε^2 wird durch die **Stichprobenvarianz des Fehlers** s^2 ersetzt:

$$\hat{\sigma}_\varepsilon^2 = s^2 = \frac{1}{n-2} \sum \hat{\varepsilon}_t^2, \quad \text{mit } \hat{\varepsilon}_t = y_t - (\hat{\alpha} + \hat{\beta} x_t)$$

Damit wird erhalten wir die Schätzer $\hat{\sigma}_\alpha^2$ und $\hat{\sigma}_\beta^2$ für σ_α^2 bzw. σ_β^2 .

Eigenschaften der OLS-Schätzer

Die OLS-Lösung liefert den **BLUE**, den besten linearen unverzerrten Schätzer (*best linear unbiased estimator*):

- **Best**: der Schätzer hat die kleinste Varianz.
- **Linear**: das Modell ist linear.
- **Unbiased** oder **unverzerrt**: Erwartungswert des Schätzers ist der Wert des Parameters ist.
(Ansonsten wäre er verzerrt.)
- **Estimator**: Schätzer

t-Test der geschätzten Parameter

Wir können obige Verteilung benutzen um die Hypothese

$$H_0: \beta = b \quad H_1: \beta \neq b$$

zu testen. Die Prüfgröße (*Teststatistik*)

$$\frac{\hat{\beta} - b}{\hat{\sigma}_\beta} \sim t_{n-2}$$

ist dabei t -verteilt mit $n - 2$ Freiheitsgraden.

(Analog für α .)

Konfidenzintervalle

Das $(1 - \alpha)$ -Konfidenzintervall für β ergibt sich aus

$$[\hat{\beta} - t_{n-2, 1-\alpha/2} \hat{\sigma}_{\beta}, \hat{\beta} + t_{n-2, 1-\alpha/2} \hat{\sigma}_{\beta}]$$

Das Bestimmtheitsmaß

Wir wollen messen, wie gut sich das Modell an die Daten anpasst und ob unsere Schätzung die Annahmen des Regressionsmodells erfüllt.

Das **Bestimmtheitsmaß** R^2 ist definiert durch als der Quotient aus der Summe der Abweichungsquadrate im geschätzten Modell und der Summe der Abweichungsquadrate der beobachteten Daten.

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2}$$

R^2 gibt den Anteil der durch die Regression erklärten Varianz an.

Das Bestimmtheitsmaß / (2)

$$\begin{array}{ccccc} (y_t - \bar{y}) & = & (y_t - \hat{y}_t) & + & (\hat{y}_t - \bar{y}) \\ \uparrow & & \uparrow & & \uparrow \\ \text{totale} & & \text{nicht-erklärte} & & \text{erklärte} \\ \text{Abweichung} & & \text{Abweichung} & & \text{Abweichung} \end{array}$$

Quadrieren und Summieren ergibt

$$\begin{array}{ccccc} \sum(y_t - \bar{y})^2 & = & \sum(y_t - \hat{y}_t)^2 & + & \sum(\hat{y}_t - \bar{y})^2 \\ \uparrow & & \uparrow & & \uparrow \\ \text{SST} & & \text{SSE} & & \text{SSR} \end{array}$$

SST ... *total sum of squares*

SSE ... *sum of squared errors*

SSR ... *sum of squares from regression*

(Der Term $2 \sum(y_t - \hat{y}_t)(\hat{y}_t - \bar{y})$ ist Null.)

F-test

Das Bestimmtheitsmaß R^2 kann ebenfalls getestet werden.
Der Test $H_0: R^2 = 0$ ist dabei äquivalent dem Test $H_0: \beta = 0$.
Die Prüfgröße lautet

$$\frac{R^2}{1 - R^2}(n - 2) \sim F_{1, n-2}$$

und ist F -verteilt mit 1 und $n - 2$ Freiheitsgraden.
Der Test heißt kurz **F-Test**.

Überprüfung der Modellannahmen

- **Unkorreliertheit der Residuen.** Aus der Autokorrelation könnte man die Residuen prognostizieren. Damit wäre unser Modell nicht vollständig.
- **Konstanz der Varianz der Residuen.** Wenn die Varianzen in verschiedenen Teilperioden unterschiedlich sind, sind die geschätzten Varianzen von $\hat{\alpha}$ und $\hat{\beta}$ nicht korrekt.
- **Abweichungen von der Normalverteilung** kann mittels Histogrammen und Test auf Normalverteilung entdeckt werden.

Prognose mittels linearer Regression

Zur Prognose müssen die Werte der unabhängigen Variable $x_{t+\tau}$, $\tau = 1, \dots, k$ bekannt sein.

- **Prognosefunktion**

$$\hat{y}_{t+\tau} = \hat{\alpha} + \hat{\beta} x_{t+\tau}$$

- **Standardfehler der Prognose**

$$\text{se}(\hat{y}_{t+\tau}) = \hat{\sigma}_\varepsilon \sqrt{1 + \frac{1}{t} + \frac{(x_{t+\tau} - \bar{x})^2}{\sum_{i=1}^t (x_i - \bar{x})^2}}$$

Der Standardfehler der Prognose ist minimal, wenn $x_{t+\tau} = \bar{x}$ ist.

- **Prognoseintervall**

$$[\hat{y}_{t+\tau} - t_{t-2} \cdot \text{se}(\hat{y}_{t+\tau}), \hat{y}_{t+\tau} + t_{t-2} \cdot \text{se}(\hat{y}_{t+\tau})]$$

Multiple Regressionsmodell

Wir erweitern das Modell auf mehrere erklärende Variable

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_m x_{mt} + \varepsilon_t$$

Die Koeffizienten $\alpha, \beta_1, \beta_2, \dots, \beta_m$ werden wieder mittels OLS geschätzt.

Es gibt eine Reihe von Verfahren um die Anzahl m der notwendigen Variablen zu bestimmen (so wenig wie möglich!).

Das multiple Bestimmtheitsmaß

Das **multiple Bestimmtheitsmaß** ist analog zum einfachen Bestimmtheitsmaß definiert.

Liegt eine Regression mit m unabhängigen Variablen, x_{jt} , mit $j = 1, \dots, m$ vor, so ist der Test auf $R^2 = 0$ äquivalent dem Test, dass alle $\beta_j = 0$ sind (bzw. dass kein x_j einen Beitrag zur Varianzreduktion in y zu leisten im Stande ist).

F-Test:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0 \quad (R^2 = 0)$$

$$H_1: \text{zumindest ein } \beta_j \neq 0 \quad (R^2 \neq 0)$$

Die Prüfgröße

$$\frac{R^2}{1 - R^2} \frac{n - (m + 1)}{m} \sim F_{m, n - (m + 1)}$$

ist F -verteilt mit m und $n - (m + 1)$ Freiheitsgraden.

Kapitel 9

ARMA Modelle

Lernziele

- Stationäre und nicht-stationäre Prozesse:
White noise und random walk
- ARMA: Autoregressive moving average Modelle
- Modellbildung
- Schätzung von ARMA Modellen
- Modellwahl und Modellüberprüfung
- Prognose
- Integrierte ARMA Modelle: ARIMA

Schwach stationäre Prozesse

Kovarianz oder **schwach stationäre** Prozesse

$$\{y_t\}, \quad t = \dots, -2, -1, 0, 1, 2, \dots$$

haben die Eigenschaft:

- **mittelwertstationär:**

$$E(y_t) = E(y_{t-s}) = \mu$$

- **kovarianzstationär:**

$$V(y_t) = E(y_t - \mu)^2 = V(y_{t-s}) = \sigma_y^2$$

$$\text{Cov}(y_t, y_{t-s}) = E(y_t - \mu)(y_{t-s} - \mu) = \text{Cov}(y_{t-j}, y_{t-s-j}) = \gamma_s$$

μ , σ_y^2 und γ_s sind **konstant** und **unabhängig** von t .

Autokorrelationsfunktion (ACF)

Die **Autokorrelation** zwischen y_t und y_{t-s} ist definiert als

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \text{Corr}(y_t, y_{t-s})$$

$$\gamma_s = \text{Cov}(y_t, y_{t-s}), \quad \gamma_0 = \sigma_y^2.$$

Im Speziellen gilt:

$$\rho_0 = 1 \quad \text{und} \quad -1 \leq \rho_s \leq 1.$$

Fasst man die ρ_s , $s \geq 0$, zusammen, erhält man die **Autokorrelationsfunktion, ACF**:

$$1, \rho_1, \rho_2, \rho_3, \dots$$

Beispiel

White noise, WN

Ein Prozess $\{y_t\}$, $y_t = \epsilon_t$, mit

$$E(\epsilon_t) = 0, \quad V(\epsilon_t) = \sigma_\epsilon^2, \quad \rho_0 = 1, \quad \rho_1 = \rho_2 = \rho_3 = \dots = 0$$

heißt *white noise*, WN, oder **Weißes Rauschen**.

Ein *white noise* Prozess ist (kovarianz-) stationär.

Wold Darstellung

Jeder schwach stationäre Prozess $\{y_t\}$ lässt sich als unendliche gewichtete Summe eines vergangenen white noise Prozesses darstellen:

$$y_t - \mu = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

Die ψ_j heißen **Impulse-response-Koeffizienten**; die Funktion $\{\psi_j, j \geq 0\}$, **Transferfunktion**, Impuls-Antwort-Funktion.

Sie erfüllt die Bedingung

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty$$

D.h., die ψ_j sind **quadratisch summierbar**.

Wold Darstellung / (2)

- Aus $E(\epsilon_t) = 0$ folgt

$$E(y_t) = \mu$$

- Aus $V(\epsilon_t) = \sigma_\epsilon^2$ und der quadratischen Summierbarkeit der ψ_j folgt

$$V(y_t) = \sum \psi_j^2 \sigma_\epsilon^2 = \sigma_\epsilon^2 \sum \psi_j^2$$

da die ϵ_t unkorreliert sind.

Lagoperator

Der **Lagoperator** L ist definiert als

$$Ly_t = y_{t-1}$$

Durch Mehrfaches anwenden des Lagoperators erhält man

$$\begin{aligned}L^2 y_t &= Ly_{t-1} = y_{t-2} \\L^3 y_t &= L^2 y_{t-1} = Ly_{t-2} = y_{t-3}, \\&\dots \\L^s y_t &= y_{t-s}\end{aligned}$$

Beispiel:

$$\begin{aligned}(1 - L)y_t &= y_t - y_{t-1}, \\(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3)y_t &= y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} - \alpha_3 y_{t-3}. \\(1 - \alpha_1 L - \alpha_2 L^2 - \alpha_3 L^3) &\text{ heißt auch Lagpolynom der Ordnung 3.}\end{aligned}$$

Wold Darstellung mittels Lagoperator

Die Wold Darstellung kann auch mit Hilfe eines unendlichen Lagpolynoms $\Psi(L)$ angegeben werden:

$$y_t - \mu = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \sum_{j=0}^{\infty} \psi_j (L^j \epsilon_t) = \sum_{j=0}^{\infty} (\psi_j L^j) \epsilon_t = \Psi(L) \epsilon_t$$

wobei $\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j$ ist.

Random Walk (RW)

Ein Prozess $\{y_t\}$ mit $y_t = y_{t-1} + \epsilon_t$ heißt *random walk*, RW, (ohne Drift).

Ein Prozess $\{y_t\}$ mit

$$y_t = c + y_{t-1} + \epsilon_t$$

heißt *random walk mit Drift*. c ist der Driftparameter.

Der Prozess ist in rekursiver Darstellung gegeben.

Explizite Darstellung des RWs:

$$y_t = y_0 + \sum_{j=1}^t \epsilon_j \quad \text{bzw.} \quad y_t = y_0 + ct + \sum_{j=1}^t \epsilon_j$$

Ein *random walk* ist **nicht stationär**.

Bedingter und unbedingter Erwartungswert

Die Informationsmenge I_t ist

$$I_t = \{y_t, \epsilon_t, y_{t-1}, \epsilon_{t-1}, \dots, y_1, \epsilon_1, y_0\}$$

Der **bedingte Erwartungswert** eines *random walks* y_t bezüglich der Informationsmengen I_{t-1} , I_{t-s} und I_0 ist

$$\begin{aligned} E(y_t | I_{t-1}) &= c + y_{t-1} \\ E(y_t | I_{t-s}) &= s c + y_{t-s} \\ E(y_t | I_0) &= t c + y_0 \end{aligned}$$

Die Abhängigkeit des bedingten Erwartungswertes vom Anfangswert verschwindet nicht mit $s \rightarrow \infty$.

Der unbedingte Erwartungswert $E(y_t)$ existiert nicht.

Bedingte und unbedingte Varianz

Die **bedingte Varianz** eines *random walks* y_t ist

$$\begin{aligned} V(y_t | I_{t-1}) &= \sigma_\epsilon^2 \\ V(y_t | I_{t-s}) &= s \sigma_\epsilon^2 \\ V(y_t | I_0) &= t \sigma_\epsilon^2 \end{aligned}$$

Die bedingte Varianz ist nicht konstant und nimmt ausgehend von $t = 0$ mit t zu.

Die unbedingte Varianz existiert nicht.

Die Kovarianz $\text{Cov}(y_t, y_{t+s})$ ist $t \sigma_\epsilon^2$.

Random Walk und dynamischer Multiplikator

Der *random walk* Prozess hat die Eigenschaft, dass vergangene Schocks nicht vergessen werden. Jeder (vergangene) Schock, ϵ_{t-s} , geht zur Gänze in das aktuelle Niveau, y_t , ein. Kein Schock wird vergessen.

$$y_t = y_0 + c t + \sum_{j=1}^t \epsilon_j \quad \Rightarrow \quad \frac{\partial y_t}{\partial \epsilon_{t-s}} = 1$$

Man sagt auch die **Persistenz** eines Schocks ist Eins.

Mit diesem Modell können *irreversible ökonomische Entscheidungen* beschrieben werden.

ARMA

Ein *autoregressiver moving average Prozess* der Ordnung (p, q) , **ARMA** (p, q) , ist ein schwach stationärer Prozess mit dem Bildungsgesetz

$$\alpha_p(L)(y_t - \mu) = \beta_q(L)\epsilon_t$$

wobei

$$\begin{aligned}\alpha_p(L) &= 1 - \alpha_1 L - \dots - \alpha_p L^p \\ \beta_q(L) &= 1 - \beta_1 L - \dots - \beta_q L^q\end{aligned}$$

$\alpha_p(L)$... **AR-Polynom** der Ordnung p

$\beta_q(L)$... **MA-Polynom** der Ordnung q

ϵ_t ... *white noise*.

ARMA / (2)

Beispiele:

ARMA(0,0), $\mu = 0$: $y_t = \epsilon_t$ white noise

ARMA(0,0), $\mu \neq 0$: $y_t = \mu + \epsilon_t$

AR(1): $(1 - \alpha_1 L)(y_t - \mu) = \epsilon_t$

MA(1): $(y_t - \mu) = (1 - \beta_1 L)\epsilon_t$

ARMA(1,1): $(1 - \alpha_1 L)(y_t - \mu) = (1 - \beta_1 L)\epsilon_t$

ARMA(1,2): $(1 - \alpha_1 L)(y_t - \mu) = (1 - \beta_1 L - \beta_2 L^2)\epsilon_t$

AR(12): $(1 - \alpha_1 L - \dots - \alpha_{12} L^{12})(y_t - \mu) = \epsilon_t$

Vergleich ARMA und Wold Darstellung

ARMA Modelle liefern eine Approximation der $\Psi(L)$ -Polynoms aus der Wold Darstellung mittels einer rationalen Funktion.

Durch Division (sofern zulässig) erhält man aus

$$\alpha_p(L)(y_t - \mu) = \beta_q(L)\epsilon_t$$

$$y_t - \mu = \frac{\beta(L)}{\alpha(L)}\epsilon_t = \Psi(L)\epsilon_t$$

Beispiel:

$$\text{ARMA}(1,0): y_t - \mu = \frac{1}{1 - \alpha_1 L}\epsilon_t = \sum_{j=0}^{\infty} \alpha_1^j \epsilon_{t-j} = \Psi(L)\epsilon_t$$

$$\text{ARMA}(1,1): y_t - \mu = \frac{1 - \beta_1 L}{1 - \alpha_1 L}\epsilon_t = \Psi(L)\epsilon_t$$

$$\text{MA}(\infty): y_t - \mu = (1 - \beta_1 L - \beta_2 L^2 - \dots)\epsilon_t = \beta_{\infty}(L)\epsilon_t = \Psi(L)\epsilon_t$$

Principle of Parsimony

Mittels ARMA-Modellen können alle (schwach-) stationären Prozesse dargestellt werden, sofern die Ordnung der Polynome groß genug gewählt wird: $p \rightarrow \infty$ oder $q \rightarrow \infty$.

In der Regel muss man annehmen, dass der zugrundeliegende Prozess sehr kompliziert ist, das eigentlich ein $MA(\infty)$ zur Modellierung notwendig wäre.

Bei der **Modellbildung** wird der zugrundeliegenden Prozess durch ein **sparsam parametrisiertes** ARMA-Modell (ARMA-Modell mit niedriger Ordnung) approximiert: *principle of parsimony*.

Das Problem besteht darin ein „gutes“ und zugleich sparsam parametrisiertes Modell zu finden.

AR(1) Prozess

Das Modell für einen AR(1) Prozess lautet:

$$(1 - \alpha L)(y_t - \mu) = \epsilon_t \quad \text{mit } |\alpha| < 1$$

oder

$$y_t - \alpha y_{t-1} = c + \epsilon_t \quad \text{mit } c = (1 - \alpha)\mu$$

Für $\mu = 0$ erhalten wir $y_t - \alpha y_{t-1} = \epsilon_t$.

Explizite Darstellung:

$$y_t = \alpha^t y_0 + \sum_{j=1}^t \alpha^{t-j} \epsilon_j \quad \text{bzw.} \quad y_t = \alpha^t y_0 + c \sum_{j=1}^{t-1} \alpha^j + \sum_{j=1}^t \alpha^{t-j} \epsilon_j$$

Erwartungswert und Varianz eines AR(1)

Aus der expliziten Darstellung erhalten wir direkt den **bedingten Erwartungswert** und die **bedingte Varianz** für $\mu = 0$:

$$E(y_t | y_0) = \alpha^t y_0 \quad \text{und} \quad V(y_t | y_0) = \sigma_\epsilon^2 \sum_{j=1}^t \alpha^{2(t-j)}$$

Die Abhängigkeit vom Anfangswert verschwindet mit $|\alpha| < 1$, wenn wir den Prozess im Zeitpunkt $-\infty$ starten lassen:

$$E(y_t | y_0) = \alpha^t y_0 \quad \rightarrow \quad E(y_t | y_{-\infty}) = E(y_t) = 0$$

$$V(y_t | y_0) = \sigma_\epsilon^2 \sum_{j=1}^t \alpha^{2(t-j)} \quad \rightarrow \quad V(y_t | y_{-\infty}) = V(y_t) = \frac{\sigma_\epsilon^2}{1 - \alpha^2}$$

Der **unbedingte Erwartungswert** ist konstant und gleich Null:

$$E(y_t) = 0.$$

Stationarität eines AR(1)

- Der AR(1) Prozess ist für

$$|\alpha| < 1$$

ein stationärer Prozess.

- Ein Prozess, gegeben durch die Differenzgleichung

$$y_t - \alpha y_{t-1} = y_t - \alpha L y_t = 0$$

ist stationär, wenn er einen **stabilen Fixpunkt** hat.

Das ist genau dann der Fall, wenn die Wurzeln des charakteristische Polynoms

$$1 - \alpha z = 0$$

außerhalb des Einheitskreises liegen. (Hier: $|z| = |1/\alpha| > 1$)

Autokorrelationsfunktion eines AR(1)

Die ACF fällt geometrisch (exponentiell):

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\text{Cov}(y_t, y_{t-s})}{\text{V}(y_t)} = \alpha^s$$

Ein AR(1) Prozess beschreibt ein **Vergessen vergangener Schocks**. Ein Schock, der s Perioden zurückliegt, wird mit $\psi_s = \alpha^s$ gewichtet, die ACF fällt daher mit α^s .

Allgemein gilt für AR(p)-Prozesse, dass die ACF (betragsmäßig) geometrisch fällt. Sie muss aber nicht monoton fallen wie beim AR(1).

Varianz und ACF sind von μ unabhängig.

Prognose eines AR(1)

Die τ -**Schritt Prognose** ist der bedingte Erwartungswert, $E(y_{t+\tau}|y_t)$.

Die Varianz des Prognosefehlers der τ -Schritt Prognose ist

$$\text{V}(y_{t+\tau}|y_t) = E\left([y_{t+\tau} - E(y_{t+\tau}|y_t)]^2|y_t\right)$$

Aus der Definition einer AR(1) erhalten wir $y_{t+1} = c + \alpha y_t + \epsilon_{t+1}$ mit $c = \mu(1 - \alpha)$.

$\tau = 1$ (1-Schritt Prognose):

$$E(y_{t+1}|y_t) = c + \alpha y_t, \quad \text{V}(y_{t+1}|y_t) = \sigma_\epsilon^2$$

$\tau = 2$ (2-Schritt Prognose):

$$E(y_{t+2}|y_t) = c(1 + \alpha) + \alpha^2 y_t, \quad \text{V}(y_{t+2}|y_t) = \sigma_\epsilon^2(1 + \alpha^2)$$

Prognose eines AR(1)

Allgemein (τ -Schritt Prognose):

$$E(y_{t+\tau}|y_t) = c \sum_{j=1}^{\tau-1} \alpha^j + \alpha^\tau y_t$$

$$V(y_{t+\tau}|y_t) = \sigma_\epsilon^2 \sum_{j=1}^{\tau} \alpha^{2(\tau-j)}$$

- Die Prognose konvergiert mit $\tau \rightarrow \infty$ gegen das arithmetische Mittel.
- Die Prognosevarianz konvergiert mit $\tau \rightarrow \infty$ gegen die unbedingte Varianz.
- Je größer $|\alpha|$ ist, desto langsamer erfolgt die Konvergenz.

MA(1) Prozess

Das **Modell** für einen MA(1) lautet

$$y_t - \mu = \epsilon_t - \beta \epsilon_{t-1} = (1 - \beta L)\epsilon_t$$

bzw.

$$y_t = \mu + \epsilon_t - \beta \epsilon_{t-1} = \mu + (1 - \beta L)\epsilon_t$$

Erwartungswert eines MA(1)

Der **bedingte Erwartungswert** $E(y_t|y_{t-1})$ ergibt sich als:

$$\begin{aligned} E(y_t|y_{t-1}) &= E(\mu + \epsilon_t - \beta \epsilon_{t-1} | \mu + \epsilon_{t-1} - \beta \epsilon_{t-2}) \\ &= E(\epsilon_t - \beta \epsilon_{t-1} | \epsilon_{t-1}, \epsilon_{t-2}) \\ &= \mu + \beta \epsilon_{t-1} \end{aligned}$$

$$\begin{aligned} E(y_t|y_{t-2}) &= E(\mu + \epsilon_t - \beta \epsilon_{t-1} | \mu + \epsilon_{t-2} - \beta \epsilon_{t-3}) \\ &= E(\mu + \epsilon_t - \beta \epsilon_{t-1} | \epsilon_{t-2}, \epsilon_{t-3}) \\ &= \mu \end{aligned}$$

$$E(y_t|y_{t-s}) = \mu \quad \text{für } s > 1$$

Die bedingten Erwartungswerte $E(y_t|y_{t-s})$ für $s > 1$ sind gleich dem **unbedingten Erwartungswert**:

$$E(y_t|y_{t-s}) = \mu = E(y_t) \quad (\text{für } s > 1)$$

Der Prozess hat ein Gedächtnis von genau einer Periode.

Varianz eines MA(1)

Für die **bedingten Varianzen** erhalten wir

$$V(y_t|y_{t-1}) = \sigma_\epsilon^2$$

$$V(y_t|y_{t-2}) = \sigma_\epsilon^2(1 + \beta^2)$$

$$V(y_t|y_{t-s}) = \sigma_\epsilon^2(1 + \beta^2) \quad \text{für } s > 1$$

da die Kovarianzen der ϵ_t Null sind.

Die bedingten Varianzen sind für $s > 1$ gleich der **unbedingten Varianz**:

$$V(y_t|y_{t-s}) = \sigma_\epsilon^2(1 + \beta^2) = V(y_t) \quad (\text{für } s > 1)$$

Die Varianz von y_t existiert immer, unabhängig davon, welchen Wert β annimmt.

Autokovarianz eines MA(1)

Die **Autokovarianzen** haben ebenfalls eine einfache Struktur.

$$\begin{aligned} \text{Cov}(y_t, y_{t-1}) &= \text{Cov}(\epsilon_t - \beta \epsilon_{t-1}, \epsilon_{t-1} - \beta \epsilon_{t-2}) \\ &= -\beta \text{Cov}(\epsilon_{t-1}, \epsilon_{t-1}) \\ &= -\beta V(\epsilon_{t-1}) \\ &= -\beta \sigma_\epsilon^2 \end{aligned}$$

$$\begin{aligned} \text{Cov}(y_t, y_{t-2}) &= \text{Cov}(\epsilon_t - \beta \epsilon_{t-1}, \epsilon_{t-2} - \beta \epsilon_{t-3}) \\ &= 0 \end{aligned}$$

$$\text{Cov}(y_t, y_{t-s}) = 0 \quad \text{für } s > 1$$

Autokorrelationsfunktion eines MA(1)

Für die **Autokorrelationsfunktion** erhalten wir

$$\text{Corr}(y_t, y_{t-1}) = \frac{\text{Cov}(y_t, y_{t-1})}{V(y_t)} = \frac{-\beta \sigma_\epsilon^2}{(1 + \beta^2) \sigma_\epsilon^2} = -\frac{\beta}{1 + \beta^2}$$

$$\text{Corr}(y_t, y_{t-s}) = 0 \quad \text{für } s > 1$$

- Die ACF bricht nach dem Lag 1 ab.
- Für $s \geq 2$ zeigt die ACF das Muster der ACF eines white noise.
- Allgemein bricht die ACF eines MA(q) Prozesses nach dem Lag q ab.

Invertierbarkeitsbedingung

MA(1) Prozesse mit Parameter β und Parameter $(1/\beta)$ besitzen die selbe ACF, und haben somit die selben stochastischen Eigenschaften. Daher beschränkt man sich der Eindeutigkeit wegen auf den Bereich $|\beta| < 1$.

Diese Bedingung heißt **Invertierbarkeitsbedingung**.

Beispiel:

Der MA(1) Prozess $\{y_t\}$ mit $E(y_t) = 3$, $V(y_t) = 2.50$ und $\rho_1 = 0.40$ lässt sich auf 2 Arten darstellen:

$$y_t - 3 = u_t - 0.5 u_{t-1} \quad \text{mit} \quad u_t \sim N(0, 2) \quad (\text{iid})$$

und

$$y_t - 3 = v_t - \frac{1}{0.5} v_{t-1} \quad \text{mit} \quad v_t \sim N(0, 0.5) \quad (\text{iid})$$

Prognose eines MA(1)

Die τ -**Schritt Prognose** ist der bedingte Erwartungswert $E(y_{t+\tau}|y_t)$.

$\tau = 1$ (1-Schritt Prognose) für MA(1):

$$E(y_{t+1}|y_t) = \mu - \beta \epsilon_t, \quad V(y_{t+1}|y_t) = \sigma_\epsilon^2$$

τ -Schritt Prognose ($\tau > 1$):

$$E(y_{t+\tau}|y_t) = \mu, \quad V(y_{t+\tau}|y_t) = (1 + \beta^2) \sigma_\epsilon^2$$

- Ein MA(1) liefert nur für die 1-Schritt-Prognose eine kleinere Prognosevarianz als das arithmetische Mittel.
- Ein MA(q) liefert nur bis zur q -Schritt-Prognose eine kleinere Prognosevarianz als das arithmetische Mittel.

ARMA(p, q) Prozess

Zur Modellierung sollen ARMA Prozesse folgende Bedingungen erfüllen:

- **Stationarität:** die Wurzeln (Nullstellen) des AR-Polynoms liegen außerhalb des Einheitskreises.
- **Invertierbarkeit:** die Wurzeln des MA-Polynoms liegen ausserhalb des Einheitskreises.

Die Parameter des Prozesses müssen geschätzt, die Modellvoraussetzungen überprüft werden.

Schätzer für Mittel und Varianz

Der **Mittelwert** (arithmetisches Mittel) einer Reihe y_1, \dots, y_T der Länge T ist

$$\bar{y} = \hat{\mu} = \frac{1}{T} \sum_{t=1}^T y_t$$

μ wird in der Regel durch \bar{y} ersetzt.

Die **Varianz** ist gegeben als

$$\hat{\sigma}_y^2 = \frac{1}{T} \sum_{t=1}^T (y_t - \bar{y})^2$$

Sie ist nur dann brauchbar, wenn nur wenig Autokorrelation in y_t vorliegt oder die Reihe sehr lang ist.

t-Test für das Mittel

Das Mittel ist für große T ungefähr normalverteilt

$$\bar{y} \sim N(\mu, \sigma_y^2/T)$$

sofern die zugrundeliegende Verteilung eine Varianz besitzt und die Autokorrelationen in der Reihe nicht zu stark sind.

Die Verteilung des Stichprobenmittels kann zum Testen auf das Mittel verwendet werden.

$$H_0 : \mu = 0 \quad H_1 : \mu > 0 \quad Z = \frac{\bar{y}}{\sqrt{\hat{\sigma}_y^2/T}} \sim N(0,1)$$

Die Prüfgröße Z ist für große T (näherungsweise) standardnormalverteilt.

Test auf die Differenz zweier Mittel

Test auf Gleichheit der Mittelwerte in A und B :

$$H_0 : \mu_A = \mu_B \quad H_1 : \mu_A \neq \mu_B$$

Die Prüfgröße

$$Z = \frac{(\bar{y}_A - \bar{y}_B)}{\sqrt{\hat{\sigma}_A^2/T_A + \hat{\sigma}_B^2/T_B}} \sim N(0,1)$$

ist (näherungsweise) standardnormalverteilt.

Voraussetzung:

- Stichproben sind voneinander unabhängig.
- Beobachtungen sind nicht oder nur schwach autokorreliert.
- Die Stichprobenumfänge T_A und T_B sind hinreichend groß.

Schiefe einer Verteilung

Eine Maßzahl für die **Schiefe** ist

$$\hat{S} = \frac{1}{T\hat{\sigma}_y^3} \sum_{t=1}^T (y_t - \bar{y})^3$$

Die Verteilung ist

$$\left. \begin{array}{l} \text{linksschief} \\ \text{symmetrisch} \\ \text{rechtsschief} \end{array} \right\} \dots S \left\{ \begin{array}{l} < \\ = \\ > \end{array} \right\} 0 \dots \mu \left\{ \begin{array}{l} < \\ = \\ > \end{array} \right\} \tilde{\mu}$$

$\tilde{\mu}$ bezeichnet den Median.

Die Normalverteilung ist symmetrisch. Ihre Schiefe ist Null.

Kurtosis einer Verteilung

Eine Maßzahl für die **Kurtosis** (Wölbung)

$$\hat{K} = \frac{1}{T\hat{\sigma}_y^4} \sum_{t=1}^T (y_t - \bar{y})^4$$

$(\hat{K} - 3)$ heißt **Exzess-Kurtosis**.

- Die Normalverteilung besitzt eine Kurtosis von 3.
- Ist die Kurtosis größer als 3, so besitzt die Verteilung höhere Schwänze als die Normalverteilung (**leptokurtische** Verteilung).

Beispiel:

Die t -Verteilung mit ν Freiheitsgraden ist symmetrisch.

Ihre Kurtosis beträgt $3(\nu - 2)/(\nu - 4)$.

Test auf Normalverteilung

Die Schätzer \hat{S} und \hat{K} sind bei Zugrundeliegen einer Normalverteilung näherungsweise normalverteilt:

$$\hat{S} \sim N(0, 6/T) \quad \hat{K} \sim N(3, 24/T)$$

Der **Jarque-Bera Test** testet simultan auf $S = 0$ und $K = 3$ einer Verteilung.

Unter der Nullhypothese einer normalverteilten Zufallsvariable ist die Teststatistik χ^2 verteilt mit 2 Freiheitsgraden:

$$\frac{T-k}{6} [\hat{S}^2 + \frac{1}{4}(\hat{K} - 3)^2] \sim \chi^2$$

- Beim Test der Verteilung einer Variablen ist $k = 0$.
- Testet man Residuen einer Regression auf Normalverteilung, ist k die Anzahl der Regressoren.

Autokorrelation eines ARMA(p, q)

Die Autokorrelationsfunktion eines ARMA(p, q)-Prozesses ergibt sich durch Überlagerung der ACF des AR- und MA-Teils.

Die ersten q Werte werden durch die AR- und MA-Komponenten gemeinsam, das Ausklingen durch das AR-Polynom bestimmt.

Der Schätzer für die ACF ist die

Stichprobenautokorrelationsfunktion (SACF):

$$\hat{\rho}_k = r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Die Autokorrelationen der Ordnungen $0, 1, 2, 3, \dots$ werden zum Korrelogramm zusammengefasst.

Verteilung von $\hat{\rho}_k (= r_k)$

Die SACF eines (normalverteilten) **stationären Prozesses**, $\{y_t\}$, mit $\rho_k \neq 0$ für $k \leq m$ und $\rho_s = 0$ für $s > m$ hat für $s > m$

$$E(\hat{\rho}_s) = -\frac{1}{T} \quad \text{und} \quad V(\hat{\rho}_s) = \frac{1}{T} \left(1 + 2 \sum_{k=1}^{s-1} \rho_k^2 \right)$$

Für einen *white noise* $y_t = \epsilon_t$ gilt: $m = 0$, $\rho_s = 0$ für $s > m = 0$.

$$E(\hat{\rho}_s) = -\frac{1}{T} (\approx 0) \quad \text{und} \quad V(\hat{\rho}_s) = \frac{1}{T}$$

Die $\hat{\rho}_k$ sind normalverteilt: $\hat{\rho}_k \sim N(-\frac{1}{T}, \frac{1}{T})$. (\Rightarrow Test auf *white noise*)

Test auf white noise: Omnibustest

Die **Box-Pierce-Statistik**, Q , testet, ob bestimmte Autokorrelationskoeffizienten eines stationären Prozesses gleich Null sind.

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0 \quad H_1 : \rho_k \neq 0 \quad \text{für ein } k, 1 \leq k \leq s$$

Die Box-Pierce-Statistik ist χ^2 -verteilt mit s Freiheitsgraden.

$$Q = T \sum_{k=1}^s \hat{\rho}_k^2 \sim \chi_s^2$$

Die **Ljung-Box-Statistik**, \tilde{Q} , ist für kleine T besser geeignet.

$$\tilde{Q} = T(T+2) \sum_{k=1}^s \frac{\hat{\rho}_k^2}{T-k} \sim \chi_s^2$$

Schätzen eines ARMA Modells

Schätzen eines Modells bedeutet, die Parameter des Modells so zu wählen, dass die Eigenschaften des geschätzten Modells möglichst gut mit den Eigenschaften der beobachteten Reihe übereinstimmen.

Man sagt auch: **Das Modell wird an die Daten angepasst.**

Es gibt verschiedene Methoden ARMA Modelle zu schätzen, u.a.

- eine Kleinst-Quadrat-Methode, die *Unconditional Least Squares* Methode, ULS.
- die *Maximum Likelihood* Methode, ML, basierend auf der Normalverteilung

Unconditional Least Squares Methode, ULS

Wir schreiben das ARMA(p, q) Modell wie folgt an

$$y_t = c + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - \dots - \beta_q \epsilon_{t-q}$$

Die Parameter α_i und β_j werden analog zur linearen Regression so gewählt, dass

$$\sum_{t=1}^T \epsilon_t^2 \rightarrow \min$$

Für die ersten Beobachtungen fehlen aber die notwendigen Werte fuer y_{t-i} und ϵ_{t-i} wenn $t - i \leq 0$ ist.

Diese werden durch die Mittel \bar{y} bzw. 0 geschätzt:

$$y_{-j} = \bar{y} \quad \text{und} \quad \epsilon_{-k} = 0 \quad \text{für } j, k \geq 0$$

Eigenschaften der ULS

Die Schätzmethode heißt **unconditional least squares** methode (unbedingte Methode der Kleinsten Quadrate).

Die Schätzung verwendet das unbedingte Mittel der Zufallsvariablen für die nicht beobachteten Werte.

Die geschätzten Parameter können (in großen Stichproben) analog zum bekannten Regressionsmodell getestet werden.

Diese sind

- asymptotisch unverzerrt,
- konsistent und
- asymptotisch normal verteilt.

Maximum Likelihood Schätzer

Beim *Maximum Likelihood* Schätzer (ML) werden die Parameter so gewählt, dass sie am besten zu den Beobachtungen passen; also am wahrscheinlichsten (*likely*) sind.

Dazu wird die sogenannte *Likelihood*-Funktion maximiert.

Beispiel:

Wir suchen den ML Schätzer für den Mittelwert μ einer normalverteilten ZV mit Varianz $\sigma^2 = 1$:

$$L(\mu; x_1, \dots, x_n) = 1/(\sqrt{2\pi})^n \exp\left(-\sum_{i=1}^n (x_i - \mu)^2/2\right)$$

Differenzieren nach μ und Null setzen ergibt

$$\sum_{i=1}^n (x_i - \mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Maximum Likelihood Methode für ARMA

Wir gehen einem multinormalverteilten Zufallsvektor $\mathbf{y} = (y_1, \dots, y_T)'$ aus mit

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \Sigma)$$

$\boldsymbol{\mu}$ ist der Vektor der (konstanten) Mittel für die einzelnen Zeitpunkte.

Σ ist die zugehörige Kovarianzmatrix.

$$\Sigma = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & & \gamma_{T-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & & \gamma_{T-3} \\ \vdots & & & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \dots & \gamma_0 \end{pmatrix}$$

Likelihood für AR(1)

Die Kovarianzen eines AR(1) Prozesses sind

$$\text{Cov}(y_t, y_{t+s}) = \gamma_s = \alpha^s \sigma_y^2 \quad \text{wobei} \quad \sigma_y^2 = \gamma_0 = \sigma_\epsilon^2 / (1 - \alpha^2)$$

Daher ist Σ eine Funktion von α und σ_ϵ^2 .

$$\Sigma = \Sigma(\alpha, \sigma_\epsilon^2) = \frac{\sigma_\epsilon^2}{1 - \alpha^2} \begin{pmatrix} 1 & \alpha & \dots & \alpha^{T-1} \\ \alpha & 1 & & \alpha^{T-2} \\ \dots & & \ddots & \dots \\ \alpha^{T-1} & \alpha^{T-2} & \dots & 1 \end{pmatrix}$$

Die *Likelihood*-Funktion ist bei gegebenen Daten $\mathbf{y} = (y_1, \dots, y_T)'$

$$L(\mu, \alpha, \sigma_\epsilon^2 | \mathbf{y}) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})'\right)$$

Likelihood für MA(1)

Die Kovarianzen eines MA(1) Prozesses sind

$$\text{Cov}(y_t, y_t) = \gamma_0 = (1 + \beta^2) \sigma_\epsilon^2,$$

$$\text{Cov}(y_t, y_{t+1}) = \gamma_1 = -\beta \sigma_\epsilon^2, \text{ und}$$

$$\text{Cov}(y_t, y_{t+s}) = \gamma_s = 0, \text{ für } s > 1.$$

Daher ist Σ eine Funktion von β und σ_ϵ^2 .

$$\Sigma = \Sigma(\beta, \sigma_\epsilon^2) = \sigma_\epsilon^2 \begin{pmatrix} 1 + \beta^2 & -\beta & 0 & \dots & 0 \\ -\beta & 1 + \beta^2 & -\beta & & 0 \\ \dots & & & \ddots & 0 \\ 0 & 0 & 0 & \dots & 1 + \beta^2 \end{pmatrix}$$

Die *Likelihood*-Funktion ist bei gegebenen Daten $\mathbf{y} = (y_1, \dots, y_T)'$

$$L(\mu, \beta, \sigma_\epsilon^2 | \mathbf{y}) = (2\pi)^{-T/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)\Sigma^{-1}(\mathbf{y} - \mu)'\right)$$

Maximierung der Log-Likelihood

Allgemein ist für ein beliebiges ARMA(p, q) Modell Σ eine Funktion von allen Parametern und der Fehlervarianz

$$\Sigma = \Sigma(\alpha_1, \dots, \alpha_p; \beta_1, \dots, \beta_q; \sigma_\epsilon^2)$$

μ wird durch \bar{y} ersetzt, und die Log-Likelihood Funktion numerisch maximiert (bzw. minus Log-Likelihood minimiert).

Numerische Minimierungsverfahren sind u.a.

- Grid-Search
- Simplex-Verfahren
- Methode des steilsten Abstiegs
(*method of steepest decent*)

Numerische Verfahren

- **Grid-Search:** Dieses einfachste Verfahren dient nur zur Illustration. Man gibt ein grobes Raster (Gitter) z.B. im Suchraum $(\alpha, \sigma_\epsilon^2)$ vor. Dann probiert man alle gewählten Paare durch. In der Umgebung des gefunden Minimums verfeinert man das Gitter, usw.
- **Methode des steilsten Abstiegs:** Der Gradient nach den zu minimierenden Parametern gibt die Richtung des steilsten Abstiegs. In diese Richtung wird ausgehend von einem Startwert der nächste Wert gesucht, usw., bis sich der Funktionswert kaum mehr ändert.
- **Simplex-Verfahren:** Dieses ist nicht mit der Linearen Optimierung zu verwechseln. Es umgeht die Berechnung der Ableitungen in der Methode des steilsten Abstiegs.

Eigenschaften der ML

Die geschätzten Parameter können (in großen Stichproben) analog zum bekannten Regressionsmodell getestet werden.

Diese sind

- asymptotisch unverzerrt,
- konsistent,
- asymptotisch normal verteilt, und
- asymptotisch effizient.

Signifikanz der Parameter

Wir wollen nun ein geschätztes ARMA Modell auf seine Eignung überprüfen.

Angenommen wir haben eine Ordnung (p, q) gewählt und die Parameter des ARMA(p, q) geschätzt.

Kriterien für ein brauchbares Modell:

- ➊ **Parameter α_p und β_q signifikant (von Null verschieden).**
Man führt dazu den t -Test für die Parameter durch, die zum größten Lag im AR- bzw MA-Polynom gehören, $\hat{\alpha}_p$ und $\hat{\beta}_q$.
Ist z.B. $\hat{\alpha}_p$ nicht signifikant von Null verschieden, so sollte die AR-Ordnung um 1 reduziert oder erhöht werden. „Mittlere“ Parameter „dürfen“ Null sein.

Residuen

- ➋ **Residuen des geschätzten Modells sind *white noise*.**
Ist die gesamte Autokorrelationsstruktur der Reihe durch unser Modell erfasst?
Dazu plottet man die Residuen und kontrolliert visuell die SACF der Residuen. Dann führt man den Ljung-Box Test auf Autokorrelation in den Residuen durch. Sind die Residuen nicht *white noise*, so hat das Modell nicht die gesamte Autokorrelationsstruktur der Reihe erfasst: p, q sind zu klein.
- ➌ **Modellierte Reihe erfasst weitgehend die wichtigen Charakteristika.**
Dies sollte stets die erste Frage sein. Dazu vergleicht man visuell den Pfad der Reihe, den das Modell erzeugt, mit dem beobachteten. Interessant dabei ist z.B. ob Wendepunkte (Richtungswechsel), und wieviele davon, erfasst werden.

Prognose

🕒 **Modell liefert vernünftige Prognosen.**

Dazu führt man eine Prognose, sowohl *ex post*, als auch *ex ante* durch, und misst die Prognosegüte mit z.B. dem RMSE.

In der Praxis stellt sich oft eine anderes als das „beste“ geschätzte Modell als bestes Prognosemodell heraus.

Nun stellt sich die Frage, ob die Modellwahl, die Wahl der Ordnungen p und q , richtig war, selbst wenn das geschätzte Modell alle obigen Kriterien zu erfüllen scheint.

Modellwahl

Ziel der Anpassung eines Modells ist es,

- die Fehlerquadratsumme zu minimieren, oder
- die Likelihood zu maximieren.

Nimmt man eine weitere Variable in die Gleichung auf, so wird sie i.A. die Fehlerquadratsumme zumindest ein wenig verringern.

Die (vermeintliche?) Anpassung eines Modells an die Daten wird immer besser, je mehr Parameter in das Modell aufgenommen werden.

Frage: Wie viele Variable (hier Lags) sollen aufgenommen werden?

Die oben besprochenen Kriterien sind nicht immer einfach zu handhaben.

Informationskriterium und *Parsimony*

Die am leichtesten handhabbaren Instrumente zur Modellwahl sind **Informationskriterien**. Wir geben

- **Akaikes Informationskriterium**, AIC, und das
- **Schwarz' Kriterium**, auch Schwarz'-Bayes'sche Informationskriterium, SIC (SBC) an.

Sie helfen auch (non nested) Modelle wie ARMA(1,2) und ARMA(2,1) zu vergleichen.

Informationskriterien können durch das **Prinzip der Sparsamkeit** (*parsimony*) motiviert werden.

Um die vermeintliche Anpassung des Modells an die Daten bei Hinzunahme weiterer Parameter in das Modell zu vermeiden, wird eine eine Bestrafung, ein **Pönale**, für jede eingeführte Variable verwendet.

Informationskriterien: AIC und SIC

- **Akaike's Informationskriterium, AIC**, ist für ARMA(p, q) Prozesse gegeben durch

$$\text{AIC}(p, q) = \log(\hat{\sigma}_{\epsilon, p+q}^2) + \frac{2}{T}(p + q)$$

- **Schwarz'-Informationskriterium, SIC**, bzw. **Schwarz-Bayes'sche Informationskriterium, SBC**, ist gegeben durch

$$\text{SIC}(p, q) = \log(\hat{\sigma}_{\epsilon, p+q}^2) + \frac{\log(T)}{T}(p + q)$$

$\hat{\sigma}_{\epsilon, p+q}^2$ steht für die minimierte Fehlervarianz.
 $\log(\cdot)$ steht stets für den natürlichen Logarithmus.

Informationskriterien: Anzahl an Lags

Frage: Wie viele Variable (hier Lags) sollen aufgenommen werden?

- Laut AIC wird nur dann eine Variable (Lag) aufgenommen, wenn sie zumindest die (logarithmierte) Fehlerquadratsumme um $(2/T)$ reduziert.
- Ähnlich, aber mit einem größeren Pönale bewertet das SIC: $\log(T)/T$.

Es werden somit solange Variable (Lags) aufgenommen, bis die letzte Variable das geforderte Mindestmaß an Reduktion der Fehlerquadratsumme unterschreitet.

Modellsuche

Praktisch geht man so vor, dass man *a priori* eine maximale Ordnung für p und q annimmt, und alle Modelle bis zu diesen Ordnungen schätzt.

Die maximale Ordnung hängt davon ab, ob man saisonale Effekte berücksichtigen muss oder nicht, und, welche Struktur die ACF erahnen lässt.

Beispiel: $p_{max} = 4$ und $q_{max} = 4$.

Es sind $(4 + 1) \times (4 + 1) = 25$ Modelle zu schätzen, aus denen wir eines auswählen werden.

- Für jedes in Betracht kommende Modell werden AIC oder SIC berechnet.
- Es wird nun das Modell gewählt, das das kleinste AIC bzw. kleinste SIC aufweist.

AIC, SIC in Log-Likelihood-Funktion

Eine andere Darstellung für die Informationskriterien erhält man, wenn folgende Näherung verwendet wird:

$$\log(\hat{\sigma}_{\epsilon, p+q}^2) \approx -\frac{2}{T} l(\hat{\theta}_{p+q})$$

- $\hat{\theta}_{p+q}$ ist der Vektor der geschätzten Parameter des Modells,

$$\hat{\theta}_{p+q} = (\hat{\alpha}_1, \dots, \hat{\alpha}_p, \hat{\beta}_1, \dots, \hat{\beta}_q, \hat{\sigma}_\epsilon^2)$$

- $l(\hat{\theta}_{p+q})$ steht für den Logarithmus der maximierten Likelihood-Funktion des Modells, $l(\theta_{p+q}) = \log(L(\theta_{p+q}))$.

Die Likelihood-Funktion wird **maximiert**, während die Fehlerquadratsumme **minimiert** wird.

Trendstationarität

Ein Prozess heißt **trendstationär** wenn er durch Berücksichtigung eines linearen Trends in einen stationären Prozess transformiert werden kann.

Man modelliert üblicherweise Trend und stationären Teil simultan.

Beispiel: x_t trendstationär

$$x_t = a + b t + u_t \quad \text{bzw.} \quad x_t - (a + b t) = u_t,$$

wobei u_t ein stationärer Prozess ist.

Man regrestriert x_t auf die Zeit t , und modelliert den stationären Teil u_t (z.B. als AR(3)).

Differenzenstationarität

Ein Prozess heißt **differenzstationär** wenn er durch Differenzbildung in einen stationären Prozess transformiert werden kann.

Hier wird die Veränderung zwischen den Beobachtungen x_t und x_{t-1} berechnet und als neue Variable modelliert:

$$y_t = (1 - L)x_t$$

Beispiel: x_t differenzstationär

$$x_t = c + x_{t-1} + v_t \quad \text{bzw.} \quad x_t - x_{t-1} = c + v_t$$

wobei v_t ein stationärer Prozess ist.

Beispiel: *random walk*, $x_t = x_{t-1} + \epsilon_t$.

Differenzenstationär: ARIMA(p, d, q)

Man bezeichnet differenzenstationäre Prozesse als **ARIMA(p, d, q)**, **integrierte ARMA Modelle** mit Integrationsordnung d .

$d = 1$, ARIMA($p, 1, q$):

Hier wird eine einmalige Differenzenbildung durchgeführt.

Durch **Integration** (Summierung) der y_t , $y_t = (1 - L)x_t$, kann wieder der ursprüngliche Prozess x_t erzeugt werden: $x_t = x_{t-1} + y_t$

$$x_t = x_0 + \sum_{j=1}^t y_j$$

Integrierte Prozesse zeigen in der SACF einen **sehr langsamen Abfall**, vergleichbar mit dem eines random walks.

ARIMA(p, d, q) – Beispiele

- Ein ARIMA($p, 0, q$) Prozess ist ein ARMA Prozess. $d = 0$.
- Ein Random Walk ist ein ARIMA($0, 1, 0$), $d = 1$.
- Ist x_t ein ARIMA($p, 2, q$), dann ist die 2-malige Differenz

$$y_t = (1 - L)^2 x_t$$

stationär.

Saisonaler ARMA (stationär)

Saison ist ein mit fixer Periodizität, S , wiederkehrendes (stochastisches) Verhalten.

Beispiele:

- Wochentagsmuster in Tagesdaten von 5-Tages-Wochen, $S = 5$,
- Monatsmuster in Monatsdaten, $S = 12$,
- Quartalsmuster in Quartalsdaten, $S = 4$.

Die ACF zeigt hier Ausschläge bei Lag S und Vielfachen davon.

Im ARMA Modell führt man dann einen Lag von S (evt. Vielfache davon) ein.

Additive und multiplikative SARMA (stationär)

Man unterscheidet

- **additive saisonale ARMA**, z.B.,

$$(1 + \alpha_S L^S + \alpha_{(2S)} L^{2S})(y_t - \mu) = \epsilon_t$$

mit dem saisonalen AR-Polynom $\alpha_P(L^S)$, und

- **multiplikative saisonale ARMA Modelle**

$$\alpha_P(L) A_P(L^S)(y_t - \mu) = \beta_Q(L) B_Q(L^S) \epsilon_t$$

mit den saisonalen AR-Polynom $A_P(L^S)$ und dem saisonalen MA-Polynom $B_Q(L^S)$. Man schreibt auch

$$\text{ARMA}(p, q)(P, Q)_S$$

Kapitel 10

Stochastisches Modell

Lernziele

- Zeitreihendaten und Renditen
- Deterministische und stochastische Renditen
- Stochastischer Prozess
- Wiener Prozess, ein stetiger *Random Walk*
- Vergleich deterministisches und stochastisches Wachstum
- Eigenschaften stochastischer Prozesse

Einleitung

Diskrete Prozesse (z.B. das Binomialmodell) modellieren einen Aktienkurs durch eine Zufallsfolge S_0, S_1, \dots, S_N zu diskreten Zeitschritten $0, \delta t, 2\delta t, \dots, N\delta t$, mit $\delta = (T - t)/N$.

Reale Aktienkurse lassen sich mit **stetigen** Modellen besser beschreiben. Der Aktienkurs ist dann ein stetiger „Zufallspfad“ $S(t)$, d.h., der Kurs kann zum jedem Zeitpunkt $t \in \mathbb{R}$ einen anderen Wert annehmen.

Wir werden das Konzept eines stetigen **Zufallspfads** als Limes einer diskreten Zufallsfolge mit immer kürzeren Zeitschritten plausibel machen.

Dazu betrachten wir zunächst nochmals das deterministische Wachstumsmodell und erweitern es um eine Zufallskomponente. Das Ergebnis ist ein **stochastischer Prozess** in stetiger Zeit der mit Hilfe des sogenannten **Wiener Prozesses** beschrieben wird.

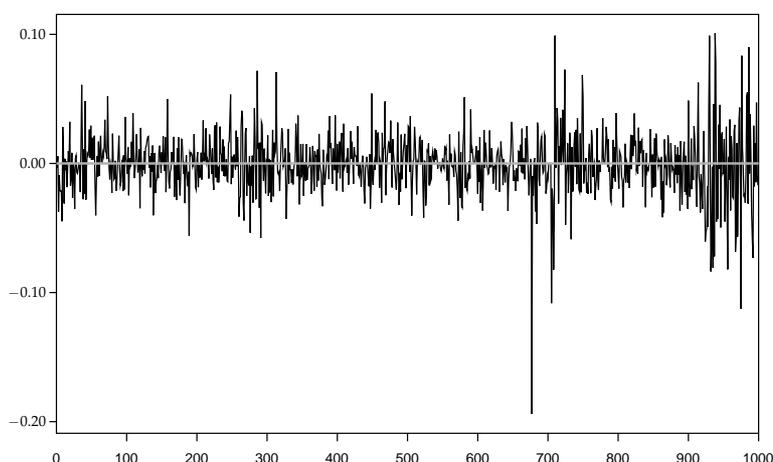
Kurs und Rendite

Aktienpreisverlauf (Bayer AG 7. 1. 1999 – 4. 1. 2003)



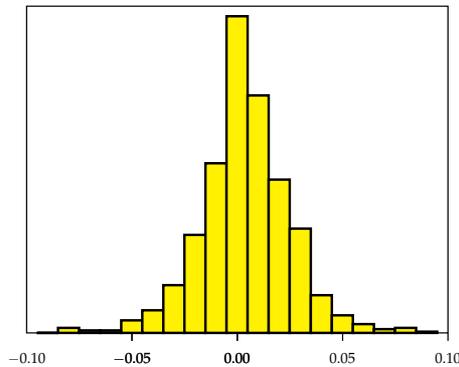
Kurs und Rendite

Renditenverlauf (Bayer AG 7. 1. 1999 – 4. 1. 2003)



Kurs und Rendite

Verteilung der Renditen (Bayer AG 7. 1. 1999 – 4. 1. 2003)



$$\bar{x} = -0.0007$$

$$s = 0.0233$$

Deterministische Rendite

Deterministische Renditen sind für eine fixe Zeitspanne δt konstant.

$$R_i = \frac{S_{i+1} - S_i}{S_i} = \mu \delta t$$

μ ist die Wachstumsrate von S und wird auf Periode von einem Jahr bezogen. Sie heißt im stochastischen Modell **Drift**.

δt misst die Zeitspanne von i zu $(i + 1)$ und skaliert μ geeignet. Z.B. $\mu \delta t = \mu (1/250)$ ist die tagliche Wachstumsrate.

Für die Änderung von S erhalten wir

$$\delta S_i = S_i \mu \delta t$$

Dieses Modell wurde schon im Kapitel QF-Einführung verwendet.

Deterministischer Prozess S

Für *infinitesimal* („unendlich“) kleine Zeitintervalle schreiben wir

$$dS = S(t) \mu dt$$

Wir können das interpretieren als

- **Differentialgleichung**

$$\frac{dS}{dt} = S \mu \quad [\text{mit Lösung } S(t) = S_0 e^{\mu t}]$$

- Anleitung zum Erzeugen eines deterministischen Prozesses:

1. Starte bei S_0 .

2. Für $i = 1, \dots, N$: $S_{i+1} = S_i + \delta S_i = S_i + S_i \cdot \mu \delta t \doteq S_i e^{\mu \delta t}$

Der stetige Prozess kann durch sehr kleine $\delta t = T/N$ beliebig genau diskret **approximiert** werden.

Prozess S und Wachstumsrate μ

Sei S_0 der Anfangswert. Dann erhalten wir für den Zeitpunkt T aus der Lösung der Differentialgleichung

$$S(T) = S_0 e^{\mu T}$$

Ohne Zufallskomponente wächst S exponentiell mit Wachstumsrate μ .

Im diskreten Zinsmodell erhält man (bei Approximation durch einen stetigen Prozess) die Rendite als

$$R = \log(S(t + \delta t)) - \log(S(t))$$

Der **logarithmierte** Prozess steigt linear mit μ :

$$\log(S(T)) = \log(S_0) + \mu T$$

Prognostizierbarkeit von Kursen

Ein typischer Verlauf eines Aktienkurses lässt die Schwierigkeiten bei dessen Prognose erahnen. Da hier viele Zufallskomponenten zusammenspielen, benötigt jedes Kursmodell eine wahrscheinlichkeitstheoretische Fundierung.

Langfristig mag das deterministisch exponentielle Modell den Eindruck erwecken eine gute Approximation zu sein, kurz- und mittelfristig ist es sicher keines.

Annahme normalverteilter Renditen

Die Verteilung der Renditen kann mittels Normalverteilung **modelliert** werden.

Jede normalverteilte Zufallsvariable R kann dargestellt werden als

$$R = \mu + \sigma Z$$

wobei $Z \sim N(0, 1)$, $\mu = E(R)$ und $\sigma = \sqrt{V(R)}$ ist.

Die obige Beziehung folgt aus der Umkehrung der Standardisierung einer normalverteilten ZVen R :

$$Z = \frac{R - \mu}{\sigma}$$

Die Normalverteilung liefert allerdings zu kleine Wahrscheinlichkeiten für sehr große Renditen (vgl. Kurssprünge).

Rendite, Drift und Volatilität

Wir fügen zum deterministischen Teil (bestimmt durch **Drift** μ) eine stochastische Komponente (bestimmt durch **Volatilität** σ) hinzu:

$$R_i = \frac{S_{i+1} - S_i}{S_i} = \mu \delta t + \sigma Z_i \sqrt{\delta t}$$

wobei $Z_i \sim N(0, 1)$ und zwei ZVen Z_i und Z_j für $i \neq j$ unkorreliert sind.

R_i ist daher normalverteilt mit

$$R_i \sim N(\mu \delta t, \sigma^2 \delta t)$$

Die Renditen sind **unkorreliert**.

Andernfalls könnte man den nächsten Aktienkurs prognostizieren und Arbitragegewinne erzielen.

Skalierung mit δt und $\sqrt{\delta t}$

Für sehr kleine Zeitschritte δt sind diskrete und stetige Rendite in erster Näherung gleich

$$R_i = \frac{S_{i+1} - S_i}{S_i} \doteq \log(S_{i+1}) - \log(S_i)$$

Die Gesamrendite über den Zeitraum $T = 1$ Jahr ist daher die Summe der Renditen der einzelnen (sehr kurzen) Zeitintervalle der Länge $\delta t = T/N = 1/N$:

$$R_T \doteq \sum_{i=1}^N [\log(S_i) - \log(S_{i-1})] \doteq \sum_{i=1}^N R_i$$

Erwartungswert und Varianz der Gesamrendite muss daher gleichmäßige über die (gleichlangen) Zeitintervalle aufgeteilt werden.

(Wir nehmen an, dass Drift μ und Volatilität σ konstant sind.)

Skalierung mit δt und $\sqrt{\delta t}$

Angenommen $R_T \sim N(\mu, \sigma^2)$ und $R_i \sim N(a\mu, b\sigma^2)$.

Aus den Rechenregeln für Erwartungswert und Varianz für die Summe *unkorrelierter* ZV folgt:

$$E(R_T) = E\left(\sum_{i=1}^N R_i\right) = \sum_{i=1}^N E(R_i) = N a \mu = \mu$$

$$V(R_T) = V\left(\sum_{i=1}^N R_i\right) = \sum_{i=1}^N V(R_i) = N b \sigma^2 = \sigma^2$$

Daher gilt

$$a = 1/N = \delta t \quad \text{und} \quad b = 1/N = \delta t$$

Stochastischer Prozess S

Für die Änderung von S erhalten wir

$$\delta S_i = S_i \mu \delta t + S_i \sigma Z_i \sqrt{\delta t}$$

Für *infinitesimal* („unendlich“) kleine Zeitintervalle könnten wir (mathematisch nicht sehr korrekt) schreiben

$$dS = S(t) \mu dt + S(t) \sigma Z(t) \sqrt{dt}$$

Wie lässt sich das interpretieren ?

- Als Differentialgleichung?
- Anleitung zum Erzeugen eines stochastischen Prozesses?

Generieren von Kurspfaden

Mit unserer Formel können wir einen *möglichen* Kurspfad S generieren:

$$S_{i+1} = S_i (1 + \mu \delta t + \sigma \sqrt{\delta t} Z_i)$$

Die folgenden Parameter sind zu wählen:

- den Drift μ , die Volatilität σ ,
- einen Startwert, S_0 ,
- die Schrittweite δt , und
- die Gesamtzahl der Schritte, bzw. die Länge der Periode.

Die standardnormalverteilten Z_i erhält per **Zufallszahlengenerator**.

(Praktisch wird aber immer der logarithmierte Pfad generiert.)

Der Wiener Prozess

Was soll $(Z(t)\sqrt{dt})$ bedeuten ?

Wir ersetzen $Z\sqrt{dt}$ durch eine **normalverteilte** Zufallsvariable dW .

$W(t)$ steht für den sogenannten **Wiener Prozess**.

dW hat die Eigenschaften:

$$dW = Z\sqrt{dt}, \quad E(dW) = 0 \quad \text{und} \quad V(dW) = E(dW^2) = dt$$

Wir schreiben daher für dS

$$dS = \mu S dt + \sigma S dW \quad \text{bzw.} \quad \frac{dS}{S} \sim N(\mu dt, \sigma^2 dt)$$

Stochastische Differentialgleichung für dS

$$dS = \mu S dt + \sigma S dW$$

ist eine **stochastische Differentialgleichung**, auf die ein weites Gebiet der Theorie der Finanzmärkte aufbaut.

Die Lösung der Differentialgleichung ist eine geometrische **Brown'sche Bewegung**.

Bemerkung:

Auch hier sind $S(t)$ und $dW(t)$ Funktionen des Zeitpunkts t .

Lösung für S im Intervall $(0, T)$

$$dS = \mu S dt + \sigma S dW \quad \text{bzw.} \quad \frac{dS}{S} \sim N(\mu dt, \sigma^2 dt)$$

Der logarithmierte Prozess $\log(S)$ gehorcht

$$\log(S(T)) \sim N(\log(S_0) + (\mu - \frac{\sigma^2}{2})T, \sigma^2 T)$$

Der Prozess S selbst ist daher **lognormal** verteilt mit

$$\begin{aligned} E[S(T)] &= S_0 \exp(\mu T) \\ V[S(T)] &= S_0^2 \exp(2\mu T) [\exp(\sigma^2 T) - 1] \end{aligned}$$

Unterschied zwischen Wachstumsrate und Drift

- Deterministisches Modell:

$$\log(S(T)) = \log(S(0)) + \mu T$$

μ heißt Wachstumsrate: $S(T) = S(0) \exp(\mu T)$.

- Stochastisches Modell:

$$\log(S(T)) \sim N(\log(S(0)) + (\mu - \frac{\sigma^2}{2})T, \sigma^2 T)$$

μ heißt Drift.

Markov-Eigenschaft

Unser stochastischer Prozess hat folgende wichtige Eigenschaft:

$$P(S_i \in A | S_j, j < i) = P(S_i \in A | S_{i-1})$$

D.h., die bedingte Wahrscheinlichkeit, dass S_i in A gegeben alle früheren Werte S_0, \dots, S_{i-1} hängt nur vom letzten Wert S_{i-1} ab.

Insbesondere gilt für den **bedingten** Erwartungswert

$$E(S_i | S_j, j < i) = E(S_i | S_{i-1})$$

Diese Eigenschaft wird **Markov-Eigenschaft** genannt.

Wenn S_{i-1} eine Zufallsvariable ist, dann hängt $E(S_i | S_{i-1})$ von deren Realisation s_{i-1} ab. $E(S_i | S_{i-1})$ ist also eine Funktion von S_{i-1} und selbst wieder eine Zufallsvariable.

Martingal-Eigenschaft

Wenn der Drift $\mu = 0$ ist, so gilt für unseren Prozess:

$$\begin{aligned} E(S_{i+1} | S_i) &= E(S_i + S_i \sigma Z_{i+1} \sqrt{\delta t}) \\ &= E(S_i) + \sigma \sqrt{\delta t} E(S_i) E(Z_{i+1}) \\ &= E(S_i) \end{aligned}$$

D.h., der bedingte Erwartungswert von S_{i+1} , gegeben der letzte Wert S_i , ist gleich dem Erwartungswert des letzten Wertes.

Diese Eigenschaft wird die **Martingal-Eigenschaft** eines Prozesses genannt.

Sie spielt in der Theorie zur Bewertung von Derivaten in arbitragefreien Märkten eine wichtige Rolle.

Kapitel 11

Stochastisches Integral

Lernziele

- Wiener Prozess und Brownsche Bewegung
- Stochastisches Integral
- Stochastische Differentialgleichung
- Itô's Lemma
- $\pi \times$ Daumen-Regeln

Ein Binomialprozess

Wir werfen innerhalb einer Periode der Länge T eine Münze N mal:
Bei Zahl erhalten wir $\sqrt{T/N}$ €, bei Kopf müssen wir $\sqrt{T/N}$ € zahlen.

Unser Gewinn/Verlust R_i ist eine Zufallsvariable mit Erwartungswert $E(R_i) = 0$ und Varianz $V(R_i) = T/N$.

Unser Guthaben zum Zeitpunkt t (nach n Zeitschritten) beträgt

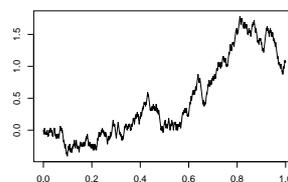
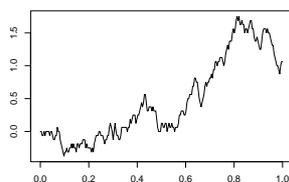
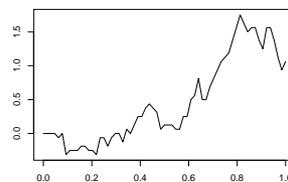
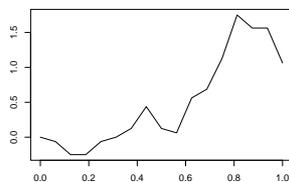
$$W(t) = W_n = \sum_{i=1}^n R_i \quad (t = nT/N)$$

und ist ebenfalls eine Zufallsvariable mit Erwartungswert $E[W(t)] = 0$ und Varianz $V[W(t)] = nT/N = t$.

Es gilt für $t_0 < t_1$:

$$E[W(t_1) - W(t_0)] = 0 \quad \text{und} \quad V[W(t_1) - W(t_0)] = t_1 - t_0$$

Ein Binomialprozess



Wiener Prozess und Brownsche Bewegung

Der Grenzprozess $W(t)$ für $N \rightarrow \infty$ hat folgende Eigenschaften:

- $W(t)$ ist stetig. (Aber nirgends differenzierbar!)
- Die Änderung $W(t_1) - W(t_0)$ ist unabhängig von $W(t_0)$.
- $W(t)$ erfüllt die Markov-Eigenschaft.
- $W(t_1) - W(t_0)$ (für $t_0 < t_1$) ist normalverteilt mit
 - Erwartungswert $E[W(t_1) - W(t_0)] = 0$ und
 - Varianz $V[W(t_1) - W(t_0)] = t_1 - t_0$.

Dieser Prozess heißt **Wiener Prozess**.

Der Prozess $X(t) = \mu t + \sigma W(t)$ heißt **Brownsche Bewegung**.

Das Riemann-Integral

Integrale können mit Hilfe von **Riemann-Summen** berechnet werden:

$$\int_0^T g(t) dt = \lim_{n \rightarrow \infty} \sum_{i=1}^n g(\tau_i) (t_i - t_{i-1})$$

mit $t_i = \frac{i}{n}T$ und $\tau_i \in (t_{i-1}, t_i)$.

Das Riemann-Stieltjes-Integral

Der Erwartungswert $E[g(X)]$ einer Funktion g einer Zufallsvariable X mit Dichte f und Verteilungsfunktion F ist

$$E[g(X)] = \int_a^b g(x)f(x) dx = \int_a^b g(x) dF(x)$$

Wir fassen hier $dF(x) = f(x)dx$ als Differential von F auf.

Das Integral $\int g(x)dF(x)$ wird als **Riemann-Stieltjes-Integral** bezeichnet.

$$\int_a^b g(x) dF(x) = \lim_{n \rightarrow \infty} \sum_{i=1}^n g(\xi_i) (F(x_i) - F(x_{i-1}))$$

mit $x_i = a + \frac{i}{n}(b - a)$ und $\xi_i \in (x_{i-1}, x_i)$.

Die Länge des Intervalls (x_{i-1}, x_i) wird dabei mit F „gewichtet“.

Riemann-Integral und Differential

Für den deterministischen Prozess gilt

$$dX(t) = \mu dt$$

Man kann das als „Kurzform“ des Integrals interpretieren

$$X(T) = X_0 + \int_0^T dX(t) = X_0 + \int_0^T \mu dt$$

Wir erweitern die deterministische Differentialgleichung um einen stochastischen Term:

$$dX(t) = \mu dt + \sigma dW(t)$$

Was bedeutet das ?

Stochastisches Integral

$$X(T) = X_0 + \int_0^T f(t) dW(t) = X_0 + \lim_{n \rightarrow \infty} \sum_{i=1}^n f(t_{i-1}) (W(t_i) - W(t_{i-1}))$$

mit $t_i = \frac{i}{n}T$ heißt das **stochastische Integral (Itô Integral)** von f .

Achtung: Die Funktion f wird immer am Anfang des Intervalls (t_{i-1}, t_i) ausgewertet (d.h., links).

Als „Kurzform“ verwenden wir

$$dX(t) = f(t) dW(t)$$

Diese Darstellung wird in Anlehnung an gewöhnliche Differentialgleichungen auch als **stochastische Differentialgleichung** bezeichnet.

Stochastisches Integral

Der Wiener Prozess $W(t)$ ist für alle t eine Zufallsvariable. Daher ist auch

$$\sum_{i=1}^n f(t_{i-1}) (W(t_i) - W(t_{i-1}))$$

und somit auch der Grenzwert

$$X(T) = \int_0^T f(t) dW(t)$$

eine Zufallsvariable.

Das Integral ist eindeutig definiert im Sinne der quadratischen Konvergenz bei $n \rightarrow \infty$: Der Erwartungswert des Abweichungsquadrats strebt gegen 0.

$$E \left[\left(X(T) - \sum_{i=1}^n f(t_{i-1}) (W(t_i) - W(t_{i-1})) \right)^2 \right] \rightarrow 0$$

Itô's Lemma

Wie berechnet man $\int f(W) dW(t)$? z.B. $\int W dW = ?$

Sei δt sehr klein, und sei $h = \frac{\delta t}{n}$. Taylor-Formel:

$$\begin{aligned} f(W(t+h)) - f(W(t)) &= f'(W(t))(W(t+h) - W(t)) + \frac{1}{2}f''(W(t))(W(t+h) - W(t))^2 + \dots \end{aligned}$$

Das gilt aber für jedes Intervall $[t + (j-1)h, t + jh]$:

$$\begin{aligned} f(W(t+jh)) - f(W(t+(j-1)h)) &= f'(W(t+(j-1)h))(W(t+jh) - W(t+(j-1)h)) \\ &+ \frac{1}{2}f''(W(t+(j-1)h))(W(t+jh) - W(t+(j-1)h))^2 + \dots \end{aligned}$$

Itô's Lemma / (2)

Daraus folgt (unter Verwendung von $f''(W(t)) \approx f''(W(t+(j-1)h))$)

$$\begin{aligned} \sum_{j=1}^n f(W(t+jh)) - f(W(t+(j-1)h)) &= \sum_{j=1}^n f'(W(t+(j-1)h))(W(t+jh) - W(t+(j-1)h)) \\ &+ \frac{1}{2} \sum_{j=1}^n f''(W(t+(j-1)h))(W(t+jh) - W(t+(j-1)h))^2 + \dots \end{aligned}$$

Linke Seite der Gleichung:

$$\begin{aligned} \sum_{j=1}^n f(W(t+jh)) - f(W(t+(j-1)h)) &= f(W(t+nh)) - f(W(t)) = f(W(t+\delta t)) - f(W(t)) \end{aligned}$$

Rechte Seite, 1. Term (per Definition des stochastischen Integrals):

$$\begin{aligned} \sum_{j=1}^n f'(W(t+(j-1)h))(W(t+jh) - W(t+(j-1)h)) &\rightarrow \int_t^{t+\delta t} f'(W) dW(\tau) \quad \text{für } h \rightarrow 0 \end{aligned}$$

Rechte Seite, 2. Term (quadratischen Konvergenz):

$(W(t+jh) - W(t+(j-1)h))^2$ durch

$$E[(W(t+jh) - W(t+(j-1)h))^2] = h = \frac{\delta t}{n}$$

$$\begin{aligned} \frac{1}{2} \sum_{j=1}^n f''(W(t+(j-1)h))(W(t+jh) - W(t+(j-1)h))^2 &\approx \frac{1}{2} \sum_{j=1}^n f''(W(t+(j-1)h))((t+jh) - (t+(j-1)h)) \\ &\rightarrow \frac{1}{2} \int_t^{t+\delta t} f''(W(\tau)) d\tau \quad \text{für } h \rightarrow 0 \end{aligned}$$

Itô's Lemma

Insgesamt (alle anderen Terme verschwinden im Grenzübergang)

$$f(W(t + \delta t)) - f(W(t)) = \int_t^{t+\delta t} f'(W(\tau)) dW(\tau) + \frac{1}{2} \int_t^{t+\delta t} f''(W(\tau)) d\tau$$

Durch Zusammenfügen vieler kleiner Intervalle $(t, t + \delta t)$

$$f(W(T)) = f(W(0)) + \int_0^T f'(W(t)) dW(t) + \frac{1}{2} \int_0^T f''(W(t)) dt$$

Itô's Lemma in „Kurzform“:

$$df = f'(W) dW + \frac{1}{2} f''(W) dt$$

Itô's Lemma und Taylor-Formel / Kurzversion

Sei f eine Funktion der Brownschen Bewegung W . Nehmen wir an, dW wäre eine „unendlich“ kleine Änderung von $W(t)$ und wir dürften ganz naiv die Taylor-Formel verwenden:

$$df = f(W + dW) - f(W) = f'(W) dW + \frac{1}{2} f''(W) (dW)^2$$

Man kann beweisen, dass $\mathbb{E}(\int_0^t (dW)^2) = t$. Daher könnten wir schreiben

$$(dW)^2 = dt$$

Wir erhalten dadurch Itô's Formel

$$df = f'(W) dW + \frac{1}{2} f''(W) dt$$

Leider ist diese Vorgangsweise **falsch**, da W nicht differenzierbar ist! (Funktioniert aber als $\pi \times$ Daumen-Regel überraschend oft.)

Beispiel

Sei $f(W) = W^2$ und $W(0) = 0$. Dann ist

$$f' = 2W \quad \text{und} \quad f'' = 2$$

Aus Itô's Lemma folgt

$$df = 2W dW + dt$$

bzw.

$$W^2 = f(W) = f(0) + \int_0^T 2W dW + \frac{1}{2} \int_0^T 2 dt = \int_0^T 2W dW + T$$

und daher gilt

$$\int_0^T W dW = \frac{1}{2} (W^2(T) - T)$$

Itô's Lemma für Brownsche Bewegungen

Wir definieren eine allgemeine Brownsche Bewegung X durch $dX = a(X) dt + b(X) dW$. Sei $f(X)$ eine Funktion von X . Dann ist

$$df = f'(X) dX + \frac{1}{2} b^2 f''(X) dt$$

Die Terme dieser Gleichung hängen oft noch direkt von der Zeit t ab:

$$dX = a(X, t) dt + b(X, t) dW$$

Dann gilt für eine Funktion $f(X, t)$:

$$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial X} dX + \frac{1}{2} b^2 \frac{\partial^2 f}{\partial X^2} dt$$

Beispiel / Brownsche Bewegung mit Drift

$$dS = \mu dt + \sigma dW$$

ist die „Kurzform“ für

$$\begin{aligned} S(T) &= S(0) + \int_0^T \mu dt + \int_0^T \sigma dW \\ &= S(0) + \mu T + \sigma(W(T) - W(0)) \end{aligned}$$

Es gilt:

$$\begin{aligned} \int_0^T dW &= \lim_{n \rightarrow \infty} \sum_{i=1}^n [W(\frac{i}{n}T) - W(\frac{i-1}{n}T)] \\ &= \lim_{n \rightarrow \infty} [W(\frac{n}{n}T) - W(0)] \\ &= W(T) - W(0) \end{aligned}$$

Beispiel / Modell für Aktienkurs – geometrische Brownsche Bewegung

Unser Modell für Aktienkurse lautet

$$dS = \mu S dt + \sigma S dW$$

Itô's Lemma mit $f(S) = \log S$ ergibt ($a(S) = \mu S$, $b(S) = \sigma S$)

$$\begin{aligned} df &= f'(S) dS + \frac{1}{2} \sigma^2 S^2 f''(S) dt \\ &= \frac{1}{S} (\mu S dt + \sigma S dW) - \frac{1}{2} \sigma^2 dt \\ &= (\mu - \frac{1}{2} \sigma^2) dt + \sigma dW \end{aligned}$$

bzw.

$$\log S(T) = \log S(0) + (\mu - \frac{1}{2} \sigma^2) T + \sigma(W(T) - W(0))$$

D.h., der logarithmierte Aktienkurs ist eine Brownsche Bewegung mit Drift $(\mu - \frac{1}{2} \sigma^2)$ und Volatilität σ .

Wilmott's $\pi \times$ Daumen-Regeln für *Dummies*

- Stochastische Differentialgleichungen sind eine Art Vorschrift zum Erzeugen eines Zufallspfades.
- Jede Funktion eines Zufallspfades ist selbst wieder ein Zufallspfad.
- Zum Berechnen verwenden wir die Taylor-Formel.
- Dabei behalten wir nur Terme mit dt und dW („ $= \sqrt{dt}$ “).
- $(dW)^2$ wird durch dt ($= E(dW^2)$) ersetzt.

Achtung!

Die Ergebnisse dieser $\pi \times$ Daumen-Regeln müssen selbstverständlich immer mit „harter“ Mathematik (von einem Mathematiker) verifiziert werden!

Kapitel 12

Das Black-Scholes Modell

Lernziele

- Delta-Hedging
- Black-Scholes Gleichung
- Black-Scholes Formel

Delta-Hedging

Wir legen ein Portfolio an, bestehend aus einer *long*-Position in einer Option und einer *short*-Position von Δ Einheiten im zugrundeliegenden Papier. Wert des Portfolios:

$$\Pi = V(S, t) - \Delta S$$

Der Wert $V(S, t)$ der Option (bei vorgegebenem E , r und T) hängt vom Wert des Papiers $S(t)$ und der Zeitspanne $(T - t)$ ab.

Wir nehmen an, dass der Kurs des zugrunde liegenden Papiers einem **lognormalen** *Random Walk* gehorcht: (geometrische Brownsche Bewegung)

$$dS = \mu S dt + \sigma S dW$$

Wie ändert sich der Wert des Portfolios ?

$$d\Pi = dV - \Delta dS$$

Delta-Hedging / (2)

Nach dem Itô-Lemma gilt

$$dV = \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} dS + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt$$

und somit

$$\begin{aligned} d\Pi &= \frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} dS + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt - \Delta dS \\ &= \underbrace{\left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right)}_{\text{deterministisch}} dt + \underbrace{\left(\frac{\partial V}{\partial S} - \Delta \right)}_{\text{stochastisch}} dS \end{aligned}$$

Im risikolosem Portfolio ist die stochastische Komponente gleich Null:

$$\frac{\partial V}{\partial S} - \Delta = 0$$

Delta-Hedging / (3)

Im **risikolosen** Portfolio muss gelten:

$$\Delta = \frac{\partial V}{\partial S}$$

$V(S(t), t)$ und damit $\frac{\partial V}{\partial S}$ ändert sich mit t . Im risikolosem Portfolio muss daher der Anteil Δ ständig angepasst (*rebalance*) werden (**dynamisches Hedgen**).

Für die Änderung des Wertes des risikolosen Portfolios gilt

$$d\Pi = \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt$$

Wegen der **Arbitragefreiheit** muss diese Änderung gleich der risikolosen Rate r sein:

$$d\Pi = r\Pi dt$$

Die Black-Scholes Gleichung

In einem risikofreiem Portfolio gilt

$$\left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt = r \Pi dt = r \left(V - r S \frac{\partial V}{\partial S} \right) dt$$

Dies gilt für alle (infinitesimal kleinen) Zeitschritte. Wir erhalten daher die sogenannte **Black-Scholes (Differential-) Gleichung**:

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r S \frac{\partial V}{\partial S} - r V = 0$$

Das ist eine sogenannte *lineare parabolische partielle Differentialgleichung* (2. Ordnung) – *Diffusionsgleichung*.

Zur Lösung benötigen wir ausserdem den Endwert, i.e., den Wert der Option am Auszahlungstag.

Voraussetzungen für das Black-Scholes Modell

- Der Kurs des zugrundeliegenden Papiers ist **lognormal** verteilt.
- Die **risikolose** Rate ist eine bekannte Funktion der Zeit.
- Es werden **keine Dividenden** ausbezahlt.
- **Delta-Hedging** erfolgt ständig.
- Es gibt **keine Transaktionskosten**.
- **Keine Arbitragemöglichkeiten**.

Erwartungswert im Black-Scholes Modell

Der **faire** Preis einer Option ist der Zeitwert der **erwarteten** Auszahlung zur Fälligkeit unter einem **risikoneutralen geometrischen Random Walk**:

$$dS = rS dt + \sigma S dW$$

Für den Preis einer (europäische) Call-Option gilt daher

$$V(t) = e^{-r(T-t)} \mathbb{E}[\max(S(T) - E, 0)]$$

Die Black-Scholes Formel

Für der Wert der **Call-Option** erhält man:
(nach einer langen Rechnung)

$$V(t) = S \Phi(d_1) - E e^{-r(T-t)} \Phi(d_2)$$

$$d_1 = \frac{\log(S/E) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = \frac{\log(S/E) + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$$

$\Phi(x)$ ist die Verteilungsfunktion der Standard-Normalverteilung.

$S = S(t)$ ist der Wert des zugrundeliegenden Papiers *heute*.