Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

CAS Datenanalyse

Zeitreihenanalyse
Teil 1: Trendbereinigung

Prof. Dr. Raúl Gimeno
FRM, CAIA, PRM

Inhaltsverzeichnis

1. Einführung

Zeitreihenanalyse:

- Zeitlich geordnete Folge von Beobachtungen (Zeitreihe) wird statistisch untersucht;
- Eigenart: Stochastische (nicht: deterministische) Abhängigkeit aufeinanderfolgender Beobachtungen → Basis für Prognostizierbarkeit

Regressionsanalyse	Zeitreihenanalyse
Entwicklung einer Zeitreihe wird durch bestimmte Variablen erklärt, die als kausale Einflussgrössen in Frage kommen → „äussere Methode“	Verhalten einer Zeitreihe wird „aus sich selbst heraus“ erklärt →„innere Methode“ Aufdeckung der Gesetzmässigkeiten, denen die Zeitreihe in Abhängigkeit von der Zeit unterliegt. Es wird damit unterstellt, dass sich die wesentlichen Einflussgrössen in dem Faktor Zeit niederschlagen.

Zwecke der Zeitreihenanalyse

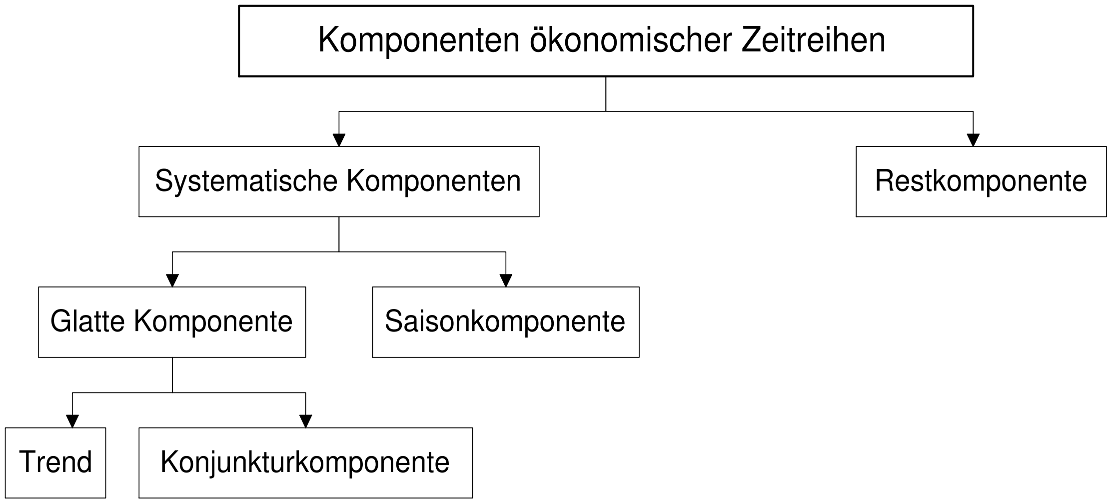
Deskription	Beschreibung des historischen Verlaufs einer Zeitreihe z.B. langfristige Preis- und Geldmengenentwicklung
Diagnose	Diagnose der aktuellen Tendenz einer Zeitreihe z.B. saisonbereinigte Arbeitslosen oder Konjunkturdiagnose
Prognose	Prognose der Entwicklung einer Zeitreihe in der Zukunft z.B. Absatzprognose, Branchenprognose oder Konjunkturprognose
Struktur- und Mustererkennung	Identifikation des datenerzeugenden Prozesses, Aufdeckung von Nichtlinearitäten (z.B. in Aktienrenditen), Aufdeckung von zeitlichen Clustern (z.B. Volatilitäten)
Kontrolle	Kontrolle der zeitlichen Entwicklung einer ökonomischen oder technischen Variablen z.B. Kontrolle eines Produktionsprozesses oder der Geldmengenentwicklung

2. Zeitreihenzerlegung und Komponentenmodell

2.1 Komponenten ökonomischer Zeitreihen

Ökonomische Zeitreihen lassen sich als Resultat eines Zusammenwirkens verschiedener **Bewegungskomponenten** auffassen.

Übersicht: Komponenten ökonomischer Zeitreihen

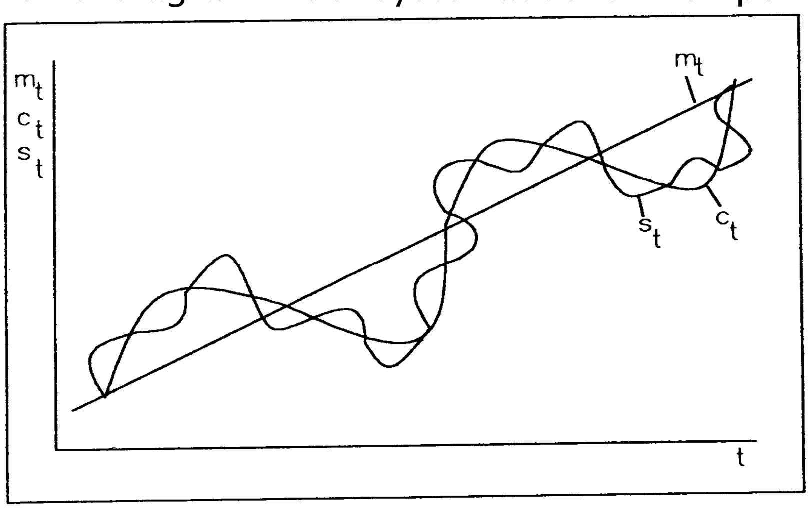


Zeitreihendiagramm

Zeitreihe: $x_1, x_2, \ldots, x_n$ oder $(x_t)_{t=1,2,\ldots,n}$

Zeitreihendiagramm: Lineare Verbindung der Wertepaare (t, x_t) in einem t,y_t-Koordinatensystem (Zeitreihenpolygon)

Zeitreihendiagramm der systematischen Komponenten



Das klassische Komponentenmodell

- Ziel der klassischen Zeitreihenanalyse: Zeitreihe in übersichtliche **Komponenten** zerlegen.
- Zusammensetzung der Zeitreihe x_t: Trend T_t, Konjunktur c_t, Saison S_t und eine Restkomponente u_t.
- Mögliche Zusammenfassung: Trend und Konjunktur zur glatten Komponente g_t oder Konjunktur und Saison zur zyklischen Komponente z_t.
- Der Trend erfasst die langfristigen Veränderungen des mittleren Niveaus, die Konjunktur die mehrjährigen Schwankungen und die Saisonkomponente die unterjährlichen, regelmässigen Schwankungen.
- Restgrösse u_t: **nicht** erklärte Einflüsse sowie Störungen

Trend

- Grundrichtung einer Zeitreihe → langfristige Entwicklungsrichtung der Reihe.
- Beispiele: langfristige Änderung der Betriebsgrösse, Wachstumstrend des Bruttoinlandprodukts eines Landes.

Auswahl der Methode

Kein Trend Kein Saisonmuster	Linearer Trend Kein Saisonmuster	Linearer Trend und Saisonmuster
↓	↓	
Exponentielle Glättungsmethode	Holt Methode	Holt-Winters Methode

Modelle für den Trend

Annahme: Zeitreihe besitzt keine Saisonschwankungen

Die wichtigsten Funktionen für den Trend:

Polynome

$T_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m$

(a) $T_t = 2 + 0.3t$

(b) $T_t = 4 - 0.2t + 0.05t^2$

(c) $T_t = 4 + 0.2t - 0.06t^2 + 0.004t^3$

(d) $T_t = 5 + 0.2t - 0.02t^2 + 0.0004t^3$

Exponentialpolynome

$T_t = \exp(a_0 + a_1 t + a_2 t^2 + \ldots + a_m t^m)$

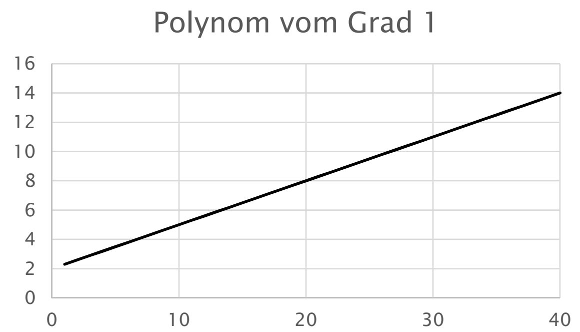
(a) $T_t = \exp(2.3 - 0.05t)$ und $T_t = \exp(1.3 - 0.05t)$

(b) $T_t = \exp(0.3 + 0.05t)$ und $T_t = \exp(0.3 + 0.03t)$

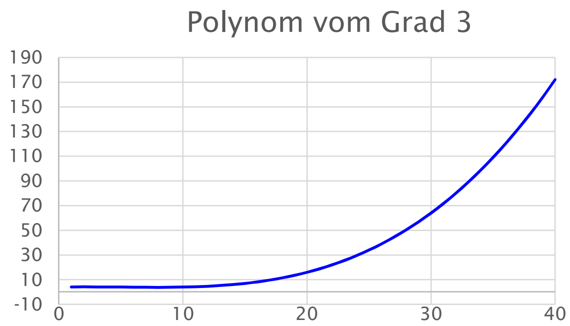
(c) $T_t = \exp(0.3 + 0.15t - 0.005t^2)$ $T_t = \exp(0.8 + 0.15t - 0.005t^2)$

(d) $T_t = \exp(0.3 - 0.15t + 0.005t^2)$ $T_t = \exp(0.8 + 0.15t - 0.005t^2)$

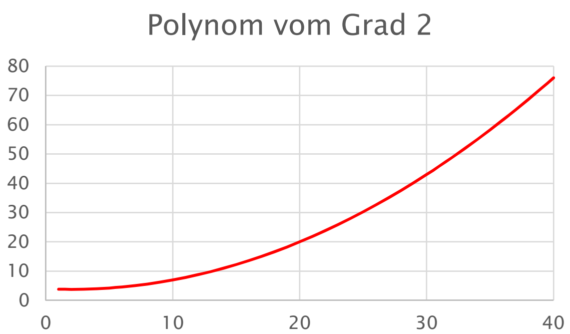
Abbildungen der Polynomen



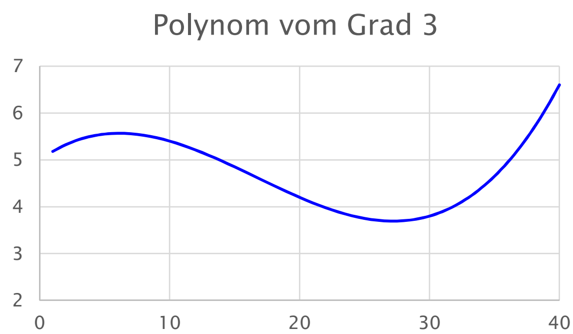
(a) $T_t = 2 + 0.3t$



(c) $T_t = 4 + 0.2t - 0.06t^2 + 0.004t^3$

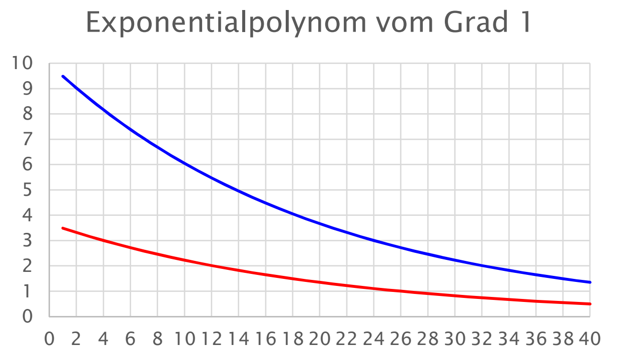


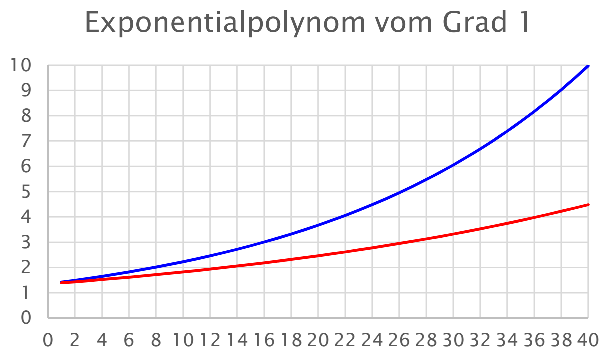
(b) $T_t = 4 - 0.2t + 0.05t^2$



(d) $T_t = 5 + 0.2t - 0.02t^2 + 0.0004t^3$

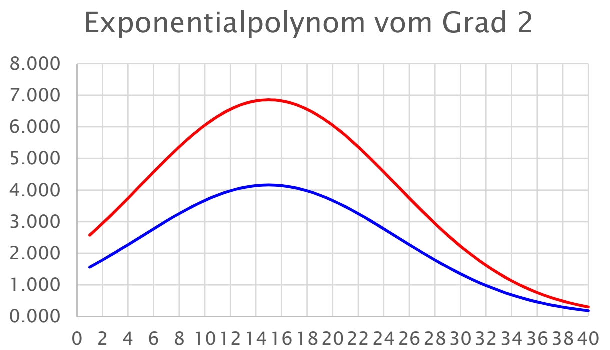
Abbildungen der Exponentialpolynome



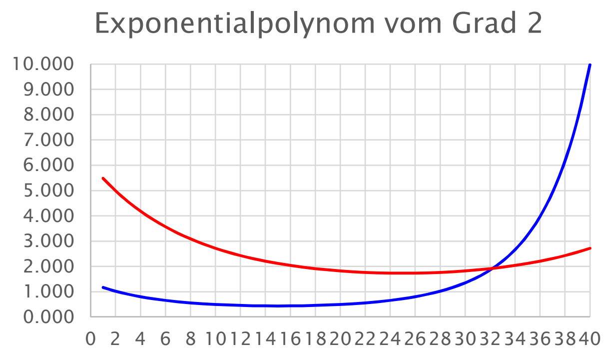


T_t = exp(2.3 - 0.05t) T_t = exp(1.3 - 0.05t)

T_t = exp(0.3 + 0.05t) T_t = exp(0.3 + 0.03t)



T_t = exp(0.3 + 0.15t - 0.005t^2)
T_t = exp(0.8 + 0.15t -0.005t^2)



T_t = exp(0.3 - 0.15t + 0.005t^2)
T_t = exp(0.8 + 0.15t -0.005t^2)

Das Komponentenmodell

Die zyklische Komponente

Die mittelfristigen Einflüsse, die auf eine Zeitreihe wirken und insbesondere durch konjunkturelle Schwankungen hervorgerufen werden, (Konjunkturzyklus).

Die Saisonkomponente

Zeitreihe mit unterjährigen Daten (Halbjahres-, Vierteljahres- oder Monatswerten)

Saisonkomponenten geben die durch jahreszeitliche Änderungen bedingten Einflüsse wieder.

Ursachen: Klima, Witterung, Volksgebräuchen, Festtagen, Sonderverkaufsbedingungen wie Winter- und Sommerschlussverkauf, Produktionsbedingungen usw.

Die Restkomponente

In der Restkomponente werden alle **einmaligen** Einflüsse zusammengefasst.

Das Komponentenmodell

Zeitreihenzerlegung: Separierung der Komponenten einer Zeitreihe

Additives Grundmodell:

(1a) $x_t = T_t + C_t + S_t + e_t$ oder (1b) $x_t = g_t + s_t + e_t$

Annahme: Zyklische Schwankungen mit konstanter Amplitude

Multiplikatives Grundmodell:

(2a) $x_t = T_t \cdot C_t \cdot S_t \cdot e_t$ oder (2b) $x_t = g_t \cdot S_t \cdot e_t$

Annahme: Zyklische Schwankungen mit proportional wachsender Amplitude

Überführung in ein additives Modell durch Logarithmierung → Transformation der Daten.

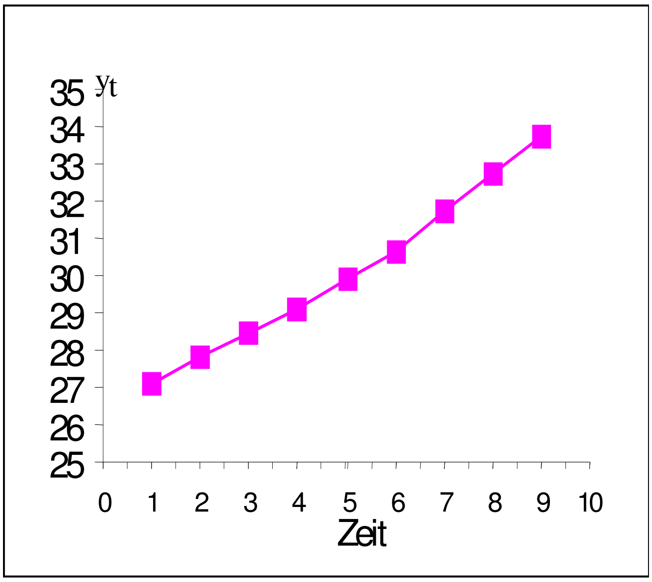
Aus (2b): $\ln y_t = \ln g_t + \ln S_t + \ln e_t$

Gemischtes Modell:

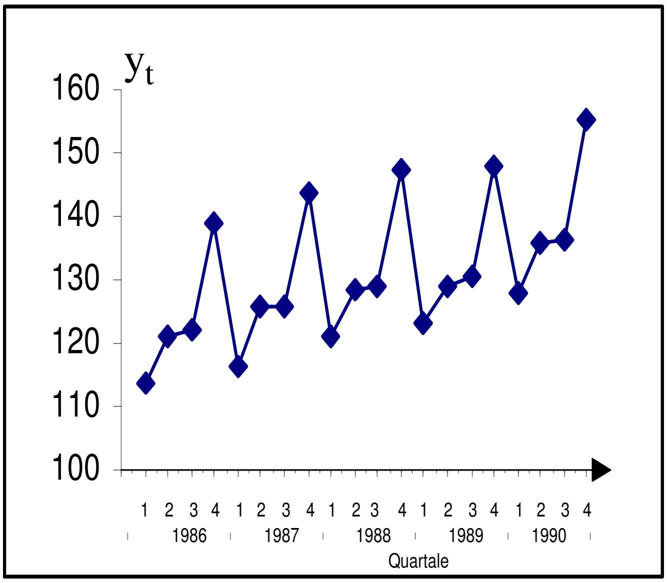
(3) $x_t = (T_t + c_t)S_t + e_t$

Zeitreihendiagramm

Zeitreihendiagramm des Kfz-Bestands



Zeitreihendiagramm der Löhne und Gehälter je Beschäftigten



2.2 Trend und glatte Komponente

Trendfunktionen

Wenn eine Zeitreihe in einem Zeitintervall keinen Strukturbruch aufweist, wird ihre Entwicklungstendenz durch eine Funktion der Zeit t modelliert.

Trendfunktion: $T_t = f(t)$

Regressionsfunktion mit der Zeit t als unabhängige Variable.

Trendbestimmung beim einfachen Grundmodell: $y_t = T_t + u_t$

Lineare Trendfunktion (konstante absolute Zuwächse):

$T_t = \beta_1 + \beta_2 \cdot t + u_t$

Beispiel: Kraftfahrzeugen

Wie aus der Abbildung hervorgegangen ist, wächst der Bestand an Kraftfahrzeugen relativ gleichmässig an, wobei die jährlichen Zuwächse nicht zu stark variieren → Trendkomponente der Zeitreihe durch eine lineare Trendfunktion nachbilden.

Arbeitstabelle (mit originären Kfz-Bestandsdaten):

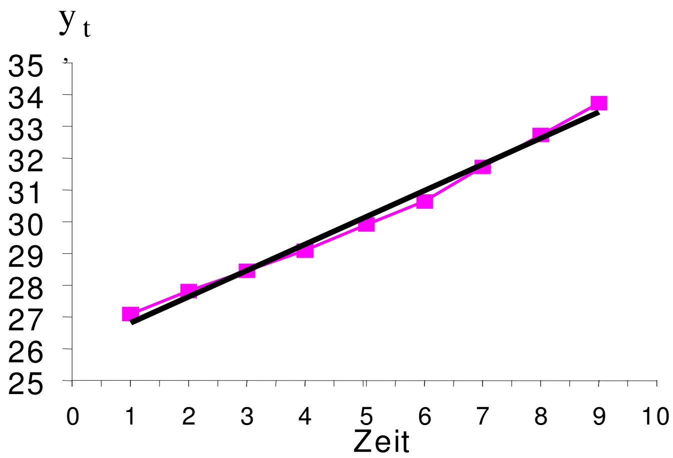
t	y_t	t^2	$y_t \cdot t$
1	27116	1	27116
2	27858	4	55716
3	28452	9	85356
4	29122	16	116488
5	29905	25	149525
6	30618	36	183708
7	31748	49	222236
8	32762	64	262096
9	33764	81	303876
Σ 45	271345	285	1406117

Trendextrapolation

Regressionsgerade: $\hat{y}_t = 26'033.4 + 823.2t$

Prognose des Kfz-Bestands durch Trendextrapolation

Für Jahr t= 10: $\hat{y}_{10} = 26'033.4 + 823.2(10) = 34'265.4$



Methode der gleitenden Durchschnitte

Flexible Methode zur Ermittlung der glatten Komponente, die ohne strenge Annahmen auskommt

Methodischer Ansatz und Grundidee

Zeitreihe wird geglättet, indem man sukzessive Mittelwerte aus einer feststehenden Anzahl benachbarter Zeitreihenwerte ermittelt (**gleiche** Länge der Stützbereiche), die jeweils der Mitte des betreffenden Zeitintervalls zugeordnet werden. Die Folge der Durchschnitte wird als gleitend bezeichnet, weil jeweils der älteste Zeitreihenwert durch den Zeitreihenwert ersetzt wird, der unmittelbar am rechten Rand ausserhalb des Stützbereichs liegt → die gebildeten Durchschnitte „gleiten" quasi entlang einer Zeitachse.

Glättungseffekt: extreme Zeitreihenwerte werden abgewichtet (Gewicht < 1).

Ordnung des gleitenden Durchschnitts = Anzahl der eingehenden Zeitreihenwerte = p

p-gliedriger gleitender Durchschnitt: $\bar{y}_t^p$

Zentrierter p-gliedriger gleitender Durchschnitt

Zentrierte p-gliedrige gleitende Durchschnitte (p ungerade)

Ein p-gliedriger gleitender Durchschnitt setzt sich aus $p=2q+1$ Beobachtungswerten zusammen, wobei jeweils q Werte vor und nach seiner zeitlichen Zentrierung liegen.

	$y_1=1$	$y_2=4$	$y_3=7$	$y_4=4$	$y_5=7$	$y_6=10$
t	1	2	3	4	5	6

3-gliedrige gleitende Durchschnitte: $\bar{y}_t^3 = \frac{1}{3}(y_{t-1} + y_t + y_{t+1})$ $\quad q = 1$ und $p = 2 \times 1 + 1 = 3$

t=2: $\bar{y}_2^3 = \frac{1}{3}(y_1 + y_2 + y_3)$ = (1 + 4 + 7)/3 = 12/3 = 4

t=3: $\bar{y}_3^3 = \frac{1}{3}(y_2 + y_3 + y_4)$ = (4 + 7 + 4)/3 = 15/3 = 5

t=4: $\bar{y}_4^3 = \frac{1}{3}(y_3 + y_4 + y_5)$ = (7 + 4 + 7)/3 = 18/3 = 6

t=5: $\bar{y}_5^3 = \frac{1}{3}(y_4 + y_5 + y_6)$ = (4 + 7 + 10)/3 = 21/3 = 7

Einfache p-gliedriger gleitender Durchschnitt

p-gliedriger gleitender Durchschnitt (p ungerade):

$$\bar{y}_t^p = \frac{1}{p}\left(\underbrace{y_{t-q} + \ldots + y_{t-1}}_{q} + y_t + \underbrace{y_{t+1} + \ldots + y_{t+q}}_{q}\right) = \frac{1}{p}\sum_{k=-q}^{q} y_{t+k} \qquad t = q+1, q+2, \ldots, n-q$$

In den Rändern lassen sich jeweils q gleitende Durchschnittswerte **nicht** berechnen.

p=3: $\bar{y}_t^3 = \frac{1}{3}(y_{t-1} + y_t + y_{t+1})$ $\qquad q = 1,\ p = 2q + 1 = 3$

p=5: $\bar{y}_t^5 = \frac{1}{5}(\underbrace{y_{t-2} + y_{t-1}}_{q\,=\,2} + y_t + \underbrace{y_{t+1} + y_{t+2}}_{q\,=\,2})$ $\qquad q = 2,\ p = 2q + 1 = 5$

Rekursionsformel:

$$\bar{y}_{t+1}^p = \bar{y}_t^p + \frac{1}{p}\left(y_{t+q+1} - y_{t-q}\right), \quad t = q+1, q+2, \ldots, n-q$$

Beispiel: $\bar{y}_{t+1}^3 = \frac{1}{3}(y_t + y_{t+1} + y_{t+2}) = \bar{y}_t^3 + \frac{1}{3}(y_{t+2} - y_{t-1})$

Beispiel: Auftragseingänge

Quartalsmässige Entwicklung des Index „Auftragseingänge im verarbeitenden Gewerbe“ über 3 Jahre:

Jahr	Q1	Q2	Q3	Q4
1		106.6	108.6	115.9
2	122.1	123.8	117.8	125.4
3	130.7	124.9	128.5	133.7
4	137.7			

Glättung mittels eines 3-gliedrigen gleitenden Durchschnitts→ erste und letzte Quartale bleiben unbesetzt. Der erste gleitende Durchschnittswert, der Q3 des 1. Jahres zugeordnet wird:

p = 3

$$\bar{y}^3_{1/III} = \frac{1}{3}\left(y_{1/II} + y_{1/III} + y_{1/IV}\right) = \frac{1}{3}(106.6 + 108.6 + 115.9) = 110.4$$

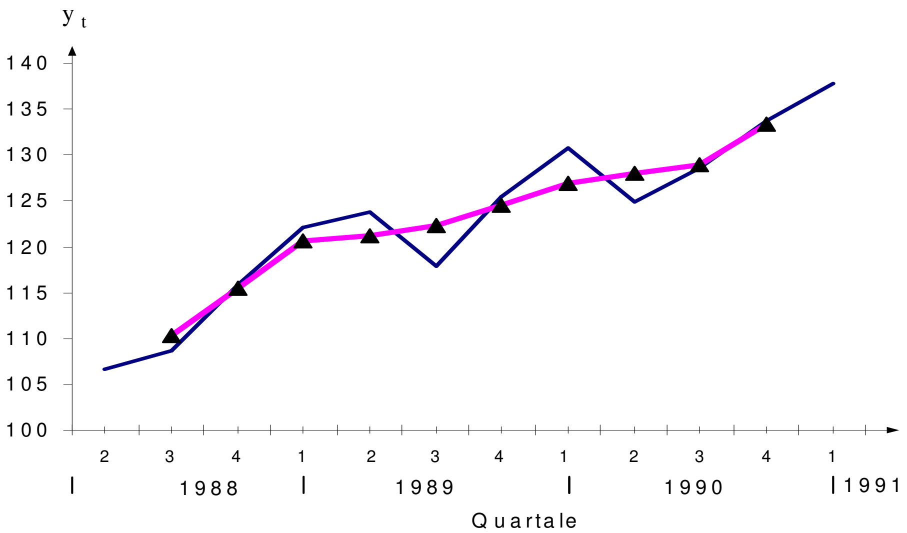
Der gleitende Durchschnittswert für 4Q des 1. Jahres:

$$\bar{y}^3_{1/IV} = \frac{1}{3}\left(y_{1/III} + y_{1/IV} + y_{2/I}\right) = \frac{1}{3}(108.6 + 115.9 + 122.1) = 115.5$$

Beispiel: Auftragseingänge

Zeit	y_t	$\bar{y}^3_t$
1/Q2	106.6	
1/Q3	108.6	110.4
1/Q4	115.9	115.5
2/Q1	122.1	120.6
2/Q2	123.8	121.2
2/Q3	117.8	122.3
2/Q4	125.4	124.6
3/Q1	130.7	127.0
3/Q2	124.9	128.0
3/Q3	128.5	129.0
3/Q4	133.7	133.3
4/Q1	137.7	

Index des Auftragseingangs mit 3-gliedrigem gleitenden Durchschnitt



Zentrierte p-gliedrige gleitende Durchschnitte (p gerade)

Wenn eine Zeitreihe saisonale Schwankungen aufweist, wird der Glättungseffekt durch zentrierte gleitende Durchschnitte ungerader Ordnung verzerrt → ein bestimmter Jahresabschnitt (z.B. Monat, Quartal) bleibt unberücksichtigt.

Ordnung des gleitenden Durchschnitts = Zykluslänge → saisonale Schwankungen ausschalten.

Bei Quartalsdaten: 4-gliedrige gleitende Durchschnitte

Bei Monatsdaten: 12-gliedrige gleitende Durchschnitte

Hierbei handelt sich um gleitende Durchschnitte gerader Ordnung.

Ein gleitender Durchschnittswert gerader Ordnung lässt sich jedoch keiner Zeiteinheit eindeutig zuordnen, da er auf der Zeitachse genau zwischen den beiden mittleren Perioden oder Zeitpunkten liegt.

Lösung: p+1 Zeitreihenwerte heranziehen und die beiden äusseren Zeitreihenwerte mit dem Faktor ½ gewichten.

Bei Zeitreihen mit saisonalen Schwankungen: Glättung mittels zentrierter gleitender Durchschnitte.

Zentrierte p-gliedrige gleitende Durchschnitte

Zentrierte gleitende Durchschnitte bei Quartalsdaten (p=4):

$$\bar{y}_t^4 = \frac{1}{4}\left(\underbrace{\frac{1}{2}y_{t-2} + y_{t-1}}_{2} + \underbrace{y_t + y_{t+1} + \frac{1}{2}}_{2}y_{t+2}\right)$$

Zentrierte gleitende Durchschnitte bei Monatsdaten (p=12):

$$\bar{y}_t^{12} = \frac{1}{12}\left(\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_{t-1} + y_t + y_{t+1} + \ldots + \frac{1}{2}y_{t+6}\right)$$

Formel für zentrierte p-gliedrige gleitende Durchschnitte ($q = p/2$):

$$\bar{y}_t^p = \frac{1}{p}\left(\frac{1}{2}y_{t-q} + \sum_{k=-q+1}^{q-1} y_{t-k} + \frac{1}{2}y_{t+q}\right), \qquad t = q+1, q+2, \ldots, n-q$$

An den Rändern des Beobachtungszeitraums lassen sich q=p/2 gleitende Durchschnittswerte nicht berechnen.

Die zentrierten gleitenden Durchschnitte entsprechen einer Mittelung jeweils zweier benachbarter unzentrierter gleitender Durchschnitte bei unveränderter Ordnung.

Zentrierte p-gliedrige gleitende Durchschnitte

Beispiel: Die Löhne und Gehälter je Beschäftigten weisen ein klares Saisonmuster auf. Im I. Quartal eines Jahres liegt der Tiefstand und nach den etwa gleichwertigen beiden mittleren Quartalen wird im IV. Quartal das saisonale Hoch erreicht.

Die langfristig steigende Tendenz dieser Zeitreihe kann daher am besten durch 4-gliedrige gleitende Durchschnitte beschrieben werden.

Jahr	Q1	Q2	Q3	Q4
1	113.6	121.3	122	138.8
2	116.3	125.7	125.7	143.5
3	121.1	128.6	129	147.3
4	123.2	129.2	130.3	147.9
5	128	135.7	136.2	155.5

Zentrierte p-gliedrige gleitende Durchschnitte

4-gliedrige gleitende Durchschnitte:

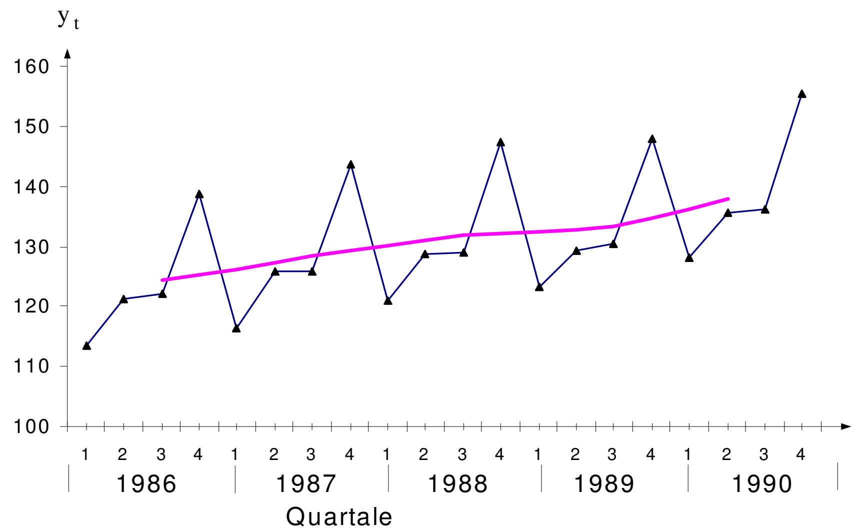
für 1/Q3: $$\bar{y}^4_{1/Q3} = \frac{1}{4}\left(\frac{1}{2}\cdot y_{1/Q1} + y_{1/Q2} + y_{1/Q3} + y_{1/Q4} + \frac{1}{2}\cdot y_{2/Q1}\right)$$
$$= \frac{1}{4}\left(\frac{1}{2}\cdot 113.6 + 121.3 + 122 + 138.8 + \frac{1}{2}\cdot 116.3\right) = 124.3$$

Durchschnitte für Q1 und Q2 lassen sich nicht berechnet q = p/2 = 2.

für 1/Q4: $$\bar{y}^4_{1/Q4} = \frac{1}{4}\left(\frac{1}{2}\cdot y_{1/Q2} + y_{1/Q3} + y_{1/Q4} + y_{2/Q1} + \frac{1}{2}\cdot y_{2/Q2}\right)$$
$$= \frac{1}{4}\left(\frac{1}{2}\cdot 121.3 + 122 + 138.8 + 116.3 + \frac{1}{2}\cdot 125.7\right) = 125.2$$

Beispiel: Löhne

Löhne und Gehälter je Beschäftigten mit zentriertem 4-gliedrigem gleitenden Durchschnitt



Exponentielle Glättung

Ziel: Glatte Komponente einer Zeitreihe herausfiltern.

Die Zeitreihenwerte werden nicht mehr gleich-, sondern exponentiell gewichtet. Das exponentielle Gewichtsschema weist den weiter zurückliegenden Werten **geometrisch abnehmende Gewichte** zu.

Vorteil: Eine Prognosegleichung → kurzfristig Vorhersage von Zeitreihen

Bei trendbehafteten Zeitreihen→ **exponentielle Glättung zweiter Ordnung**.

Exponentiellen Glättung erster Ordnung für das Grundmodell ohne Trend und ohne Saison.

Exponentielle Glättung erster Ordnung

Annahme: Zeitreihe (y_t) schwankt um einen konstanten Wert

Vorhersagewert für die Periode t+1 bei Ausnutzung aller verfügbaren Informationen bis zur Periode t:

$$\hat{y}_{t+1}(t) = \bar{y}_t = \frac{1}{t}\sum_{i=1}^{t} y_i$$

Prognosewert für Periode t+2: $\hat{y}_{t+2}(t+1) = \bar{y}_{t+1} = \frac{1}{t+1}\sum_{i=1}^{t+1} y_i$

Prognosewert für Periode t+h: $\hat{y}_{t+h}(t)$

t = aktuelle Periode, h = Prognosehorizont

Neuer Prognosewert in Abhängigkeit des vorhergehenden Prognosewerts:

$$\hat{y}_{t+2}(t+1) = \frac{1}{t+1}\sum_{i=1}^{t+1} y_i = \frac{t}{t+1}\bar{y}_t + \frac{1}{t+1}y_{t+1} = \frac{t}{t+1}\hat{y}_{t+1}(t) + \frac{1}{t+1}y_{t+1} .$$

Prognosewert der exponentiellen Glättung für Periode t+1:

$$\hat{y}_{t+1}(t) = (1-\alpha)\hat{y}_t(t-1) + \alpha y_t \qquad 0 < \alpha < 1 \qquad \text{(Rekursionsformel)}$$

Alter Prognosewert $\hat{y}_t(t-1)$: Gewichtung $=1-\alpha$,

Aktueller Beobachtungswert y_t: Gewichtung $= \alpha$

Exponentielle Glättung erster Ordnung: Beispiel

Die Unternehmensumsätze schwankten in einem 8-Jahres-Zeitraum bei keinem klar erkennbaren Trend.

Exponentielle Glättung erster Ordnung für die Vorhersage der Entwicklung.

Anfangswert y_0 = Prognosewert für die erste Periode des Beobachtungszeitraums → Zeitreihenwert unmittelbar vor Beginn des Stützbereichs.

Anfangswert = Umsatz im Jahr 0 → Prognosewert für das Jahr 1

Gewichtsfaktor $\alpha = 0.3$ → $\hat{y}_1(0) = y_0 = 12'752$

Prognosewert für das Jahr 2:

$$\hat{y}_2(1) = 0.7\hat{y}_1(0) + 0.3y_1 = 0.7 \cdot 12'752 + 0.3 \cdot 13'317 = 12'922$$

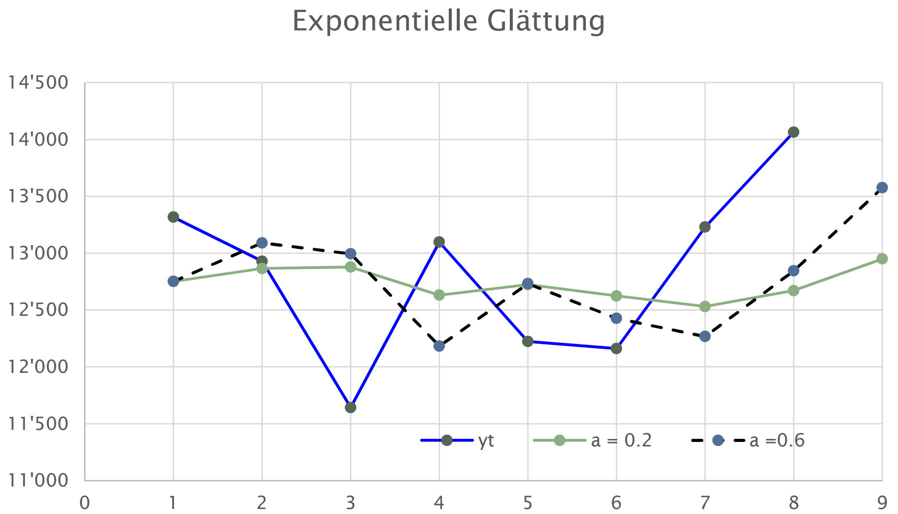
$$\hat{y}_3(2) = 0.7\hat{y}_2(1) + 0.3y_2 = 0.7 \cdot 12'922 + 0.3 \cdot 12'930 = 12'924$$

Jahr	1	2	3	4	5	6	7	8	9
y_t	13‘117	12‘930	11‘643	13‘098	12‘223	12‘161	13‘230	14‘065	
$\hat{y}_t(t-1)$	12‘752	12‘922	12‘924	12‘540	12‘707	12‘562	12‘442	12‘678	13‘094

Exponentielle Glättung erster Ordnung: Beispiel

Die Ein-Schritt-Prognosen für den Umsatz können auf diese Weise sukzessive für die Folgejahre bestimmt werden:

Umsätze und exponentielle Glättung



Prognose im Stützbereich: Ex-post-Prognose
Prognose ausserhalb des Stützbereichs: Ex-ante-Prognose

Gewichtungsschema der exponentiellen Glättung

Aktueller Prognosewert (4): $\hat{y}_{t+1}(t) = (1-\alpha)\hat{y}_t(t-1) + \alpha y_t \qquad 0 < \alpha < 1$

Vorheriger Prognosewert (5): $\hat{y}_t(t\text{-}1) = (1-\alpha)\hat{y}_{t\text{-}1}(t-2) + \alpha y_{t\text{-}1} \qquad 0 < \alpha < 1$

Nach Einsetzen von (5) in (4) erhält man

$$\hat{y}_{t+1}(t) = (1-\alpha)^2 \hat{y}_{t-1}(t-2) + \alpha(1\text{-}\alpha)y_{t-1} + \alpha y_t$$

Nach fortlaufender Substitution der alten Prognosewerte in (4):

$$(6) \quad \hat{y}_{t+1}(t) = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \alpha(1-\alpha)^3 y_{t-3} + \ldots \qquad = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i y_{t-i}$$

wenn man den Regress unendlich oft durchführt.

Prognosewert $\hat{y}_{t+1}(t)$ (6): Gewogenes arithmetisches Mittel aller zurückliegender Zeitreihenwerte mit geometrisch abnehmenden Gewichten $\alpha(1-\alpha)^i$ und $\sum_{i=0}^{\infty} \alpha(1-\alpha)^i = 1$

(allmähliche Niveauverschiebung wird hierdurch berücksichtigt)

Exponentielle Glättung erster Ordnung

Prognosewert zum Zeitpunkt t für die Periode t+1 mit einem Beobachtungszeitraum der Länge n:

$$\hat{y}_{t+1}(t) = \sum_{i=0}^{n-1} \alpha(1-\alpha)^i y_{t-i} + \alpha(1-\alpha)^n y_0$$

n=3: y_0, y_1, y_2, y_3 — $\hat{y}_4(3) = \alpha y_3 + \alpha(1-\alpha)y_2 + \alpha(1-\alpha)^2 y_1 + \alpha(1-\alpha)^3 y_0$

t: 0, 1, 2, 3, 4

y_{t-3}, y_{t-2}, y_{t-1}, y_t, Prognose

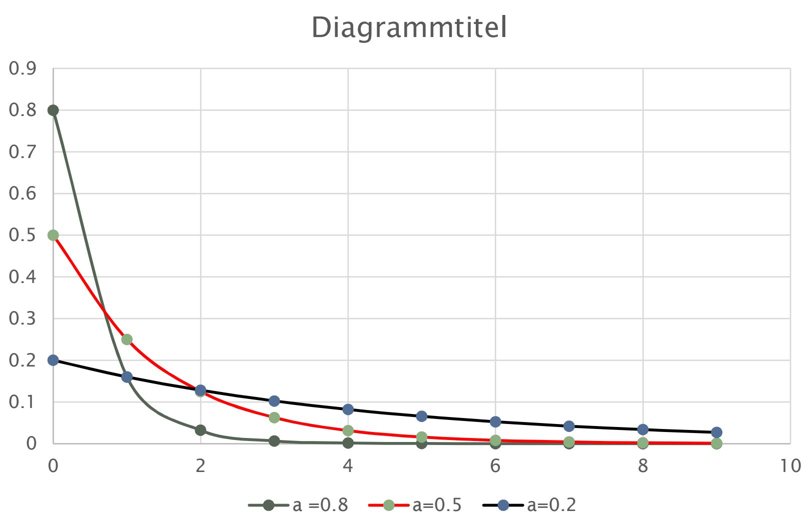
Der Anfangswert y_0 wird mit wachsendem n vernachlässigbar, da der Faktor $(1-\alpha)^n$ gegen null tendiert.

Der aktuellste Wert (y_t) geht mit dem höchsten Gewicht (α) in die Prognose ein, die Gewichte nehmen exponentiell für die zurückliegenden Werte ab.

Festlegung von y_0 (für Initialisierung des Verfahrens):
Zeitreihenwert oder Mittelwert der Zeitreihenwerte vor Beginn des Stützzeitraums

Gewichtungsschema der exponentiellen Glättung

Verhalten der Gewichtsfunktion $(1-\alpha)^i \cdot \alpha$ bei alternativen Werten von α



Fehlerkorrekturformel

Darstellung als Fehlerkorrekturformel:

(7) $\hat{y}_{t+1}(t) = (1-\alpha)\,\hat{y}_t(t-1) + \alpha y_t = \hat{y}_t(t-1) + \alpha(y_t - \hat{y}_t(t-1)) = \hat{y}_t(t-1) + \alpha\, e_t$

mit $e_t = y_t - \hat{y}_t(t-1)$

Prognose korrigiert sich quasi selbständig:

- bei Unterschätzung ($e_t > 0$) erfolgt automatisch ein Aufschlag,
- bei Überschätzung ($e_t < 0$) erfolgt ein Abschlag.

Der Prognosefehler e_t wird mit dem Gewicht α berücksichtigt.
Wenn α klein ist, wird der Prognosefehler kaum berücksichtigt.

Beispiel mit Fehlerkorrekturformel

Fehlerkorrekturformel (7) für die Umsatzdaten
Startwert: $\hat{y}_1(0) = y_0 = 12'752$

Prognosefehler im Jahr 1: $e_1 = y_1 - \hat{y}_1(0) = 13'317 - 12'752 = 565$ $\rightarrow$ Überschätzung

Ein-Schritt-Prognose für das Jahr 2 mit $\alpha = 0.3$:

$\hat{y}_2(1) = \hat{y}_1(0) + 0.3e_1 = 12'752 + \underbrace{0.3(565)}_{\text{Aufschlag}} = 12'922$

Analog ergeben sich die Prognosewerte für die Folgejahre unter Verwendung der Fehlerkorrekturformel:

Jahr	1	2	3	4	5	6	7	8	9
y_t	13'317	12'993	11'643	13'098	12'223	12'161	13'230	14'065	
$\hat{y}_t(t-1)$	12752	12922	12924	12540	12707	12562	12442	12678	13094
e_t	565	8	-1281	558	-484	-401	788	1387	
$0.3e_t$	170	2	-384	167	-145	-120	236	416	

Bedeutung des Gewichtsfaktors α für die Glättung

Reagibilität und Einfluss der Zeitreihenwerte

	α klein	α gross
Glättungseffekt der Vorhersage	gross	klein
Reagibilität auf irreguläre Schwankungen	klein	gross
Berücksichtigung neuer Zeitreihenwerte	schwach	stark
Berücksichtigung älterer Zeitreihenwerte	stark	schwach

Wahl des Glättungsparameters α

Optimaler Wert für α durch Vergleich der Anpassung alternativer Werte zwischen 0 und 1 in einem Stützbereich.

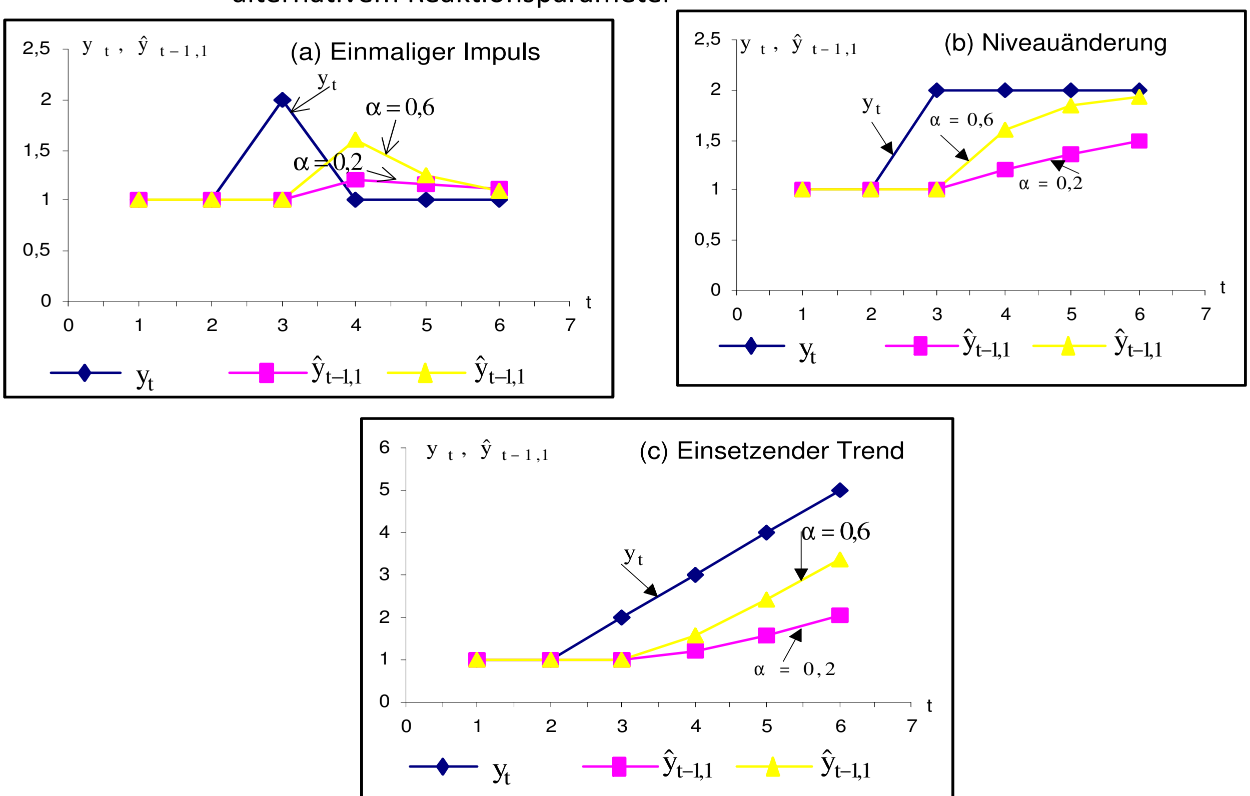
Kriterien: Mean Square Error (MSE) oder Root Mean Square Error (RMSE)

(10) $$\mathrm{MSE}(e) = \frac{1}{n-1}\sum_{t=1}^{n}\left(y_t - \hat{y}_t(t-1)\right)^2 = \frac{1}{n-1}\sum_{t=1}^{n} e_t^2$$

(11) $$\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$$

Praxis: α-Wert zwischen 0.1 und 0.3 → weiter zurückliegende Zeitreihenwerte für die Prognose bedeutsam. Bei einer sich allmählich verändernden zentralen Tendenz einer Zeitreihe, empfiehlt sich die Wahl eines grösseren α-Wertes oder der Übergang zu einer exponentiellen Glättung zweiter Ordnung.

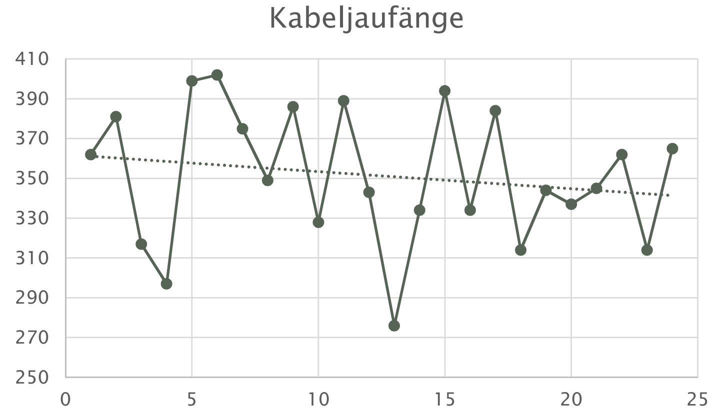
Abb.: Reaktion der Vorhersage auf verschiedene Ereignisse bei alternativem Reaktionsparameter



Exponentielle Glättung: Beispiel Kabeljaufänge

Die Fischerei Bay Company verzeichnete für die letzten zwei Jahre folgende Kabeljaufänge.

	Jahr 1	Jahr 2
Januar	362	276
Februar	381	334
März	317	394
April	297	334
Mai	399	384
Juni	402	314
Juli	375	344
August	349	337
September	386	345
Oktober	328	362
November	389	314
Dezember	343	365



Ein Trend und Saisonkomponente sind nicht ersichtlich. Wir benutzen ein Modell ohne Trend und Saisonkomponente, obwohl sich der Mittelwert über die Zeit geringfügig ändern kann: $y_t = \beta + e_t$

Exponentielle Glättung: Beispiel Kabeljaufänge

	alpha0.1		Prognose
Zeit	y_t	$\hat{y}_t(t-1)$	fehler e_t
0		360.667	
1	362	360.667	1.333
2	381	360.800	20.200
3	317	362.820	-45.820
4	297	358.238	-61.238
5	399	352.114	46.886
6	402	356.803	45.197
7	375	361.323	13.677
8	349	362.690	-13.690
9	386	361.321	24.679
10	328	363.789	-35.789
11	389	360.210	28.790
12	343	363.089	-20.089
13	276	361.080	-85.080
14	334	352.572	-18.572
15	394	350.715	43.285
16	334	355.044	-21.044
17	384	352.939	31.061
18	314	356.045	-42.045
19	344	351.841	-7.841
20	337	351.057	-14.057
21	345	349.651	-4.651
22	362	349.186	12.814
23	314	350.467	-36.467
24	365	346.821	18.179

1: Anfangswert als Durchschnitt der ersten 12 Werte

$$\hat{y}_1(0) = \frac{1}{12}\sum_{t=1}^{12} y_t = \frac{1}{12}(362 + 381 + \ldots + 343) = 360.67$$

2: Berechnung Prognosewerte $\hat{y}_t(t-1)$ mit $\alpha = 0.1$

$$\hat{y}_2(1) = (1-\alpha)\hat{y}_1(0) + \alpha y_1 = 0.9 \cdot 360.67 + 0.1 \cdot 362 = 360.80$$

$$\hat{y}_3(2) = (1-\alpha)\hat{y}_2(1) + \alpha y_2 = 0.9 \cdot 360.80 + 0.1 \cdot 381 = 362.82$$

$$ESS = \sum_{t=1}^{24} (y_t - \hat{y}_t(t-1))^2 = 28'735.11$$

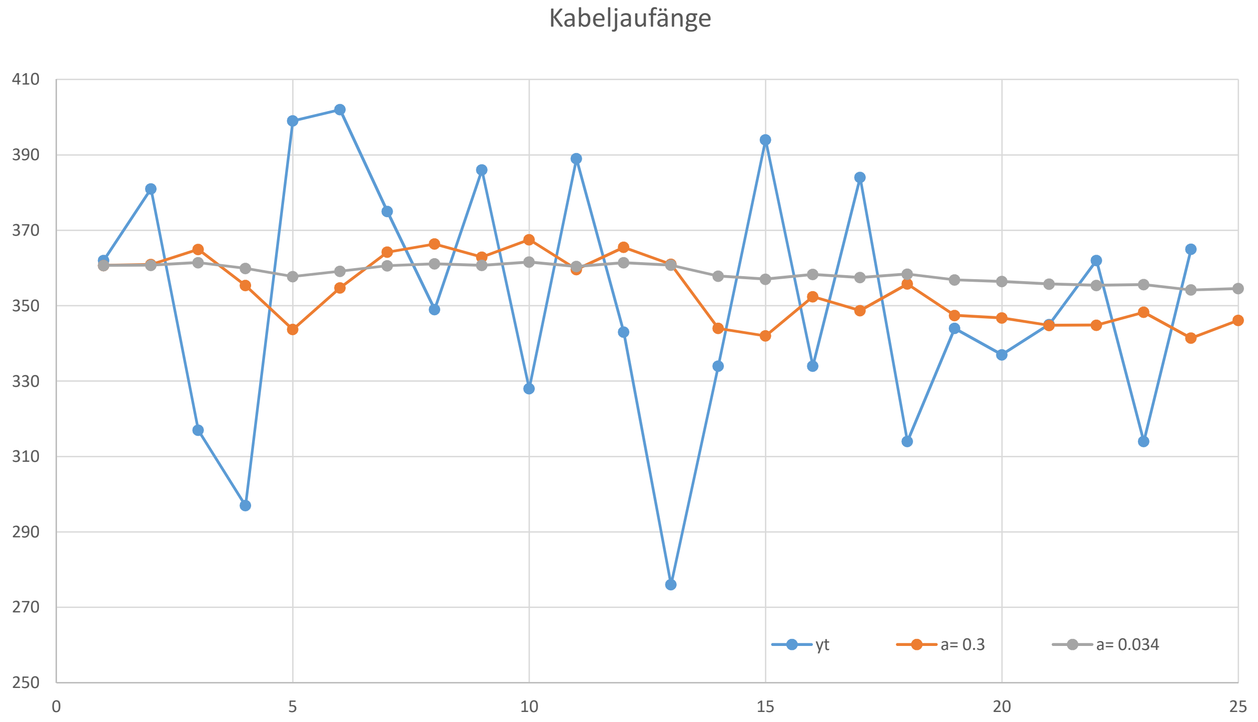
$$MSE = \frac{ESS}{T-1} = \frac{28'735.11}{23} = 1'249.4$$

$$\sigma = \sqrt{MSE} = \sqrt{1'249.4} = 35.34$$

3: Bestimmung von optimalem α

Mittels Excel Solver Funktion

Exponentielle Glättung: Beispiel Kabeljaufänge



Holt Verfahren

Voraussetzung: Die Zeitreihe weist einen langfristigen, **linearen Trend** auf $\rightarrow y_t = \mu + \beta t + u_t$

Niveau-komponente (μ)   Lineare Trendkomponente (βt)

Glättungsverfahren anhand von zwei Parametern α und γ mit $\alpha, \gamma \in [0,1]$

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

$\rightarrow$ exponentielles Glätten 1. Ordnung der (diskretisierten) 1-Perioden-Steigungen der Zeitreihe mit dem Parameter γ.

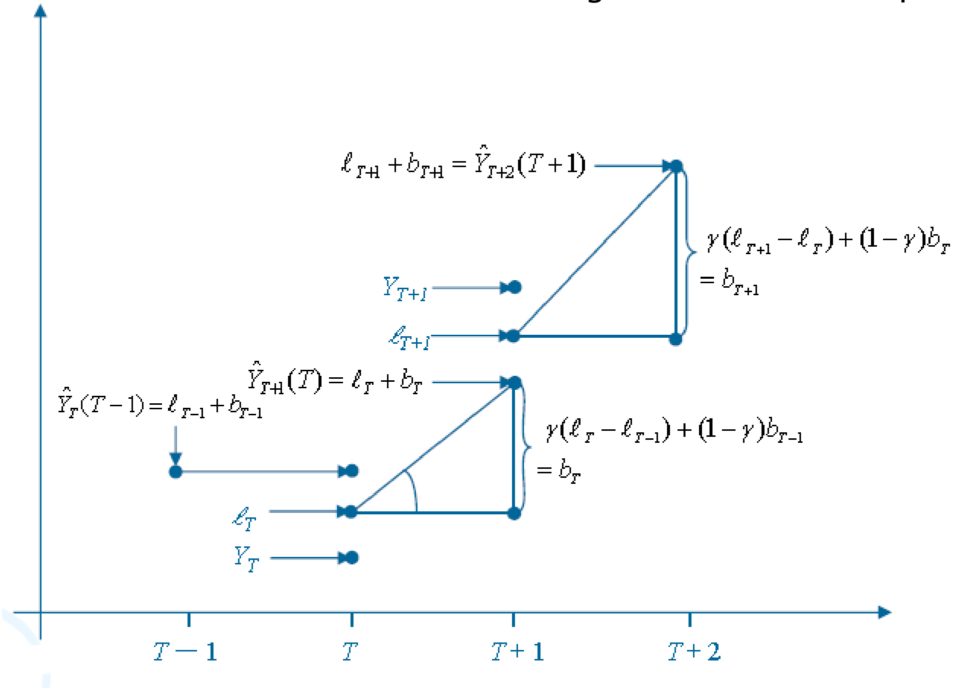
Niveauschätzung: $L_t = \alpha y_t + (1-\alpha)\underbrace{(L_{t-1} + b_{t-1})}_{\hat{y}_t(t-1)} = \alpha y_t + (1-\alpha)\hat{y}_t(t-1)$

Die Niveaugleichung ist ähnlich einem exponentiellem Glätten 1.Ordnung von y mit dem Unterschied, dass der letzte Niveauwert (L_{t-1}) um die Steigung (b_{t-1}) verschoben wird.

Grafische Darstellung

$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$

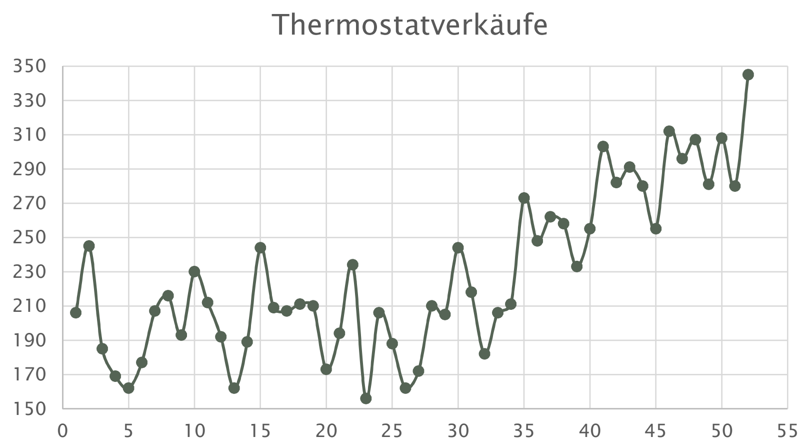
$\hat{y}_T(T-1)$: Prognosewert zum Zeitpunkt T-1 für Periode T



Holt Verfahren: Beispiel

Wöchentliche Verkäufe von Thermostaten

1	206	162	188	262	281
2	245	189	162	258	308
3	185	244	172	233	280
4	169	209	210	255	345
5	162	207	205	303	
6	177	211	244	282	
7	207	210	218	291	
8	216	173	182	280	
9	193	194	206	255	
10	230	234	211	312	
11	212	156	273	296	
12	192	206	248	307	



Beobachtungen:
- ✓ Steigender Trend
- ✓ Wachstumsrate hat sich im Laufe der Zeit geändert
- ✓ Keine Saisonkomponente ersichtlich

Holt Verfahren: Beispiel

Schritt 1: Die Parameter α und β werden mittels OLS-Schätzung für die erste Hälfte der Daten geschätzt.

gretl Output-Fenster

	coefficient	std. error	t-ratio	p-value	
const	202.625	10.3199	19.63	2.72e-016	***
time	-0.368205	0.668238	-0.5510	0.5867	

Mean dependent var	197.6538	S.D. dependent var	25.19673
Sum squared resid	15673.61	S.E. of regression	25.55517
R-squared	0.012492	Adjusted R-squared	-0.028654

Geschätzte Regressionsgerade: $\hat{y}_t$ = 202.625 – 0.3682**t**

Startwerte: L_0 = 202.625

b_0 = -0.3682

Holt Verfahren: Beispiel

Schritt 2: Ein-Schritt-Prognose berechnen

Startwerte: L_0 = 202.625 ; b_0 = -0.3682 und α = 0.2 und γ = 0.1

Ein-Schritt-Prognose in Periode 0:

$\hat{y}_1(0) = L_0 + b_0 = 202.625 - 0.3682 = 202.2568$

Ein-Schritt-Prognose in Periode 1:

$L_1 = \alpha y_1 + (1-\alpha)(L_0 + b_0) = 0.2(206) + 0.8(202.625 - 0.3682) = 203.005$

$b_1 = \gamma(L_1 - L_0) + (1-\gamma)b_0 = 0.1(203.005 - 202.625) + 0.9(-0.3682) = -0.293$

$\hat{y}_2(1) = L_1 + b_1 = 203.005 - 0.2933 = 202.712$

Prognosefehler $e_2 = y_2 - \hat{y}_2(1) = 245 - 202.712 = 42.287$ → Überschätzung

Ein-Schritt-Prognose in Periode 2:

$L_2 = \alpha y_2 + (1-\alpha)\,\hat{y}_2(1) = 0.2(245) + 0.8(202.712) = 211.16$

$b_2 = \gamma(L_2 - L_1) + (1-\gamma)b_1 = 0.1(211.16 - 203.005) + 0.9(-0.293) = 0.552$

$\hat{y}_3(2) = L_2 + b_2 = 211.16 + 0.552 = 211.722$

Prognosefehler $e_3 = y_3 - \hat{y}_3(2) = 185 - 211.722 = -26.72$ → Unterschätzung

Holt Verfahren: Beispiel

A	B	C	D	E		
	n	α	γ	RSS	MSE	σ
	52	0.2	0.1	39182.479	783.65	27.994
	y_t	L_t	b_t	$\hat{y}_t(t-1)$	e_t	e_t^2
0		202.625	-0.37			
1	206	203.0054	-0.29	202.2568	3.7432	14.012
2	245	211.1697	0.552	202.7121	42.2879	1788.266
3	185	206.3777	0.018	211.7221	-26.7221	714.071
4	169	198.9165	-0.73	206.3957	-37.3957	1398.436
5	162	190.9493	-1.45	198.1866	-36.1866	1309.470
6	177	186.9965	-1.7	189.4956	-12.4956	156.140
7	207	189.6343	-1.27	185.2929	21.7071	471.198
8	216	193.8919	-0.72	188.3649	27.6351	763.699
9	193	193.1401	-0.72	193.1752	-0.1752	0.031
10	230	199.9359	0.031	192.4199	37.5801	1412.263
11	212	202.3738	0.272	199.9673	12.0327	144.786
12	192	200.5167	0.059	202.6459	-10.6459	113.334
13	162	192.8606	-0.71	200.5758	-38.5758	1488.091
14	189	191.5186	-0.78	192.1482	-3.1482	9.911
15	244	201.3946	0.29	190.7432	53.2568	2836.288
16	209	203.1474	0.436	201.6843	7.3157	53.519
17	207	204.2668	0.504	203.5835	3.4165	11.672
18	211	206.0170	0.629	204.7712	6.2288	38.798
19	210	207.3168	0.696	206.646	3.3540	11.250
20	173	201.0103	-0	208.0128	-35.0128	1225.897
21	194	199.6048	-0.14	201.0061	-7.0061	49.085
22	234	206.3684	0.546	199.4605	34.5395	1192.975
23	156	196.7319	-0.47	206.9149	-50.9149	2592.326
24	206	198.2081	-0.28	196.2601	9.7399	94.866
25	188	195.9448	-0.48	197.931	-9.9310	98.626

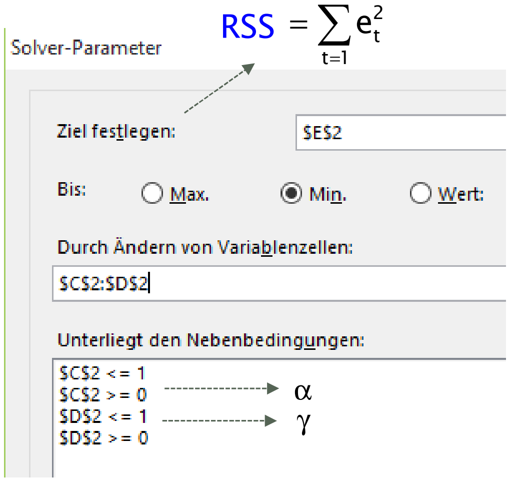
Schritt 3: Finde die optimale Kombination für α und γ

Minimiere die Fehlerquadratsumme

→ Min RSS!

→ Excel-Solver

$$RSS = \sum_{t=1}^{n} e_t^2$$



Holt Verfahren: Glättungsparameter

Für α nahe bei 0 gehen auch länger zurückliegende Beobachtungen noch stärker in die Niveauschätzung ein.

$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \qquad \alpha \cong 0 \rightarrow L_t \cong L_{t-1} + b_{t-1}$

Die letzte Beobachtung erhält ein umso grösseres Gewicht bei der Niveauschätzung, je näher der Glättungsparameter α an 1 liegt.

$L_t = \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1}) \qquad \alpha \cong 1 \rightarrow L_t \cong y_t$

Für γ nahe bei 0 geht die länger zurückliegende Trendentwicklung stärker in die aktuelle Trendschätzung ein.

$b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1} \qquad \gamma \cong 0 \rightarrow b_t \cong b_{t-1}$

Die letzte Trendentwicklung erhält ein umso grösseres Gewicht bei der Trendschätzung, je näher der Glättungsparameter γ an 1 liegt.

$b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1} \qquad \gamma \cong 1 \rightarrow b_t \cong L_t - L_{t-1}$

Die Glättungsparameter α, γ können durch Minimierung der Fehlerquadratsumme (RSS) zwischen Einschritt-Vorhersage und wahrem Wert der Zeitreihen geschätzt werden.

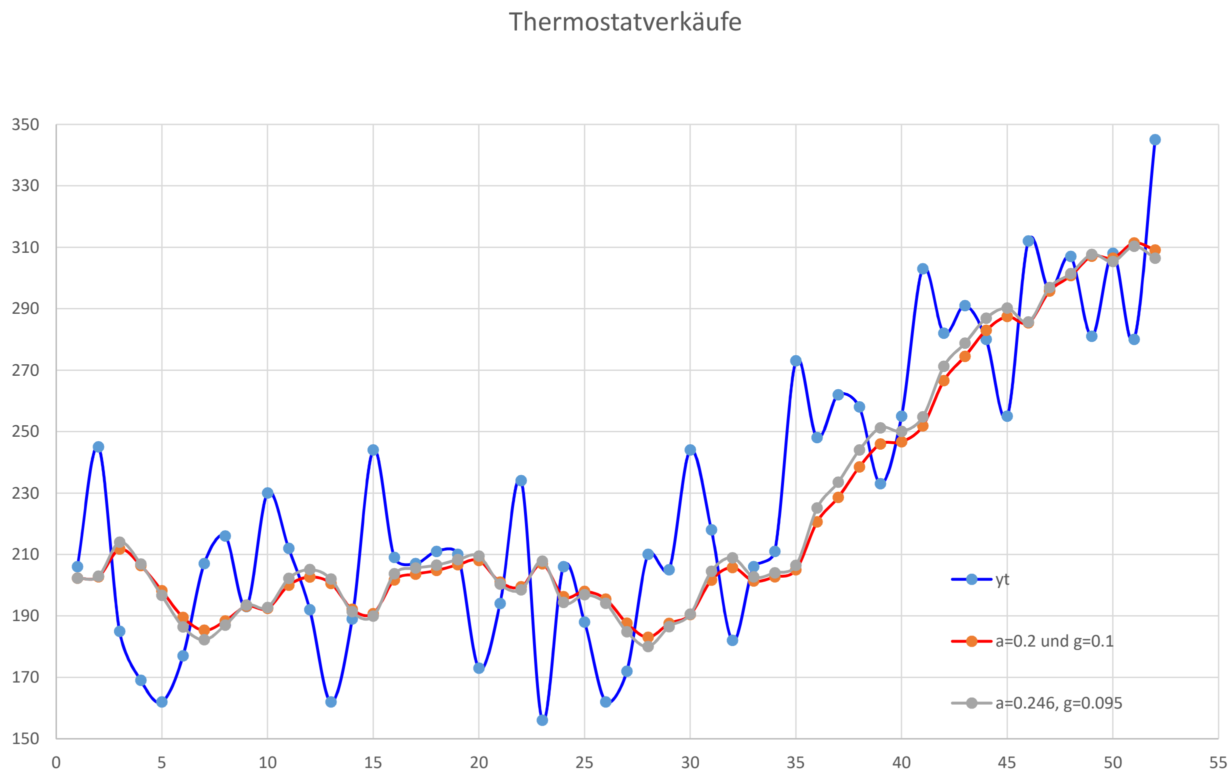
Holt Verfahren

n	α	γ	SSE	MSE	σ
52	0.2468	0.095	38884.250	777.685	27.887

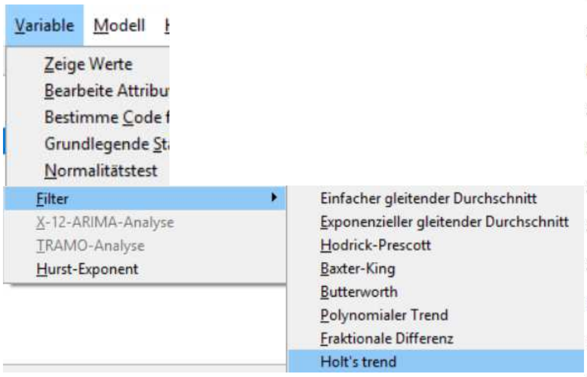
	y_t	L_t	b_t	$\hat{y}_t(t-1)$	e_t	e_t^2
0		202.625	-0.37			
1	206	203.1808	-0.28	202.2568	3.7432	14.012
2	245	213.2924	0.707	202.9004	42.0996	1772.376
3	185	206.8414	0.027	213.9998	-28.9998	840.988
4	169	197.5209	-0.86	206.8684	-37.8684	1434.017
5	162	188.1040	-1.67	196.6594	-34.6594	1201.272
6	177	184.1017	-1.9	186.4292	-9.4292	88.910
7	207	188.3260	-1.31	182.2057	24.7943	614.758
8	216	194.1673	-0.63	187.0117	28.9883	840.320
9	193	193.4016	-0.65	193.5332	-0.5332	0.284
10	230	201.9486	0.227	192.755	37.2450	1387.191
11	212	204.6009	0.458	202.1759	9.8241	96.512
12	192	201.8353	0.151	205.0587	-13.0587	170.531
13	162	192.1163	-0.79	201.9867	-39.9867	1598.937
14	189	190.7545	-0.84	191.3295	-2.3295	5.427
15	244	203.2640	0.428	189.913	54.0870	2925.402
45	255	281.4910	4.145	290.1732	-35.1732	1237.156
46	312	292.1440	4.764	285.6364	26.3636	695.040
47	296	296.6839	4.743	296.908	-0.9080	0.824
48	307	302.8023	4.873	301.4265	5.5735	31.064
49	281	301.0910	4.248	307.6757	-26.6757	711.594
50	308	305.9955	4.31	305.3386	2.6614	7.083
51	280	302.8248	3.599	310.3055	-30.3055	918.422
52	345	315.9460	4.504	306.4237	38.5763	1488.130

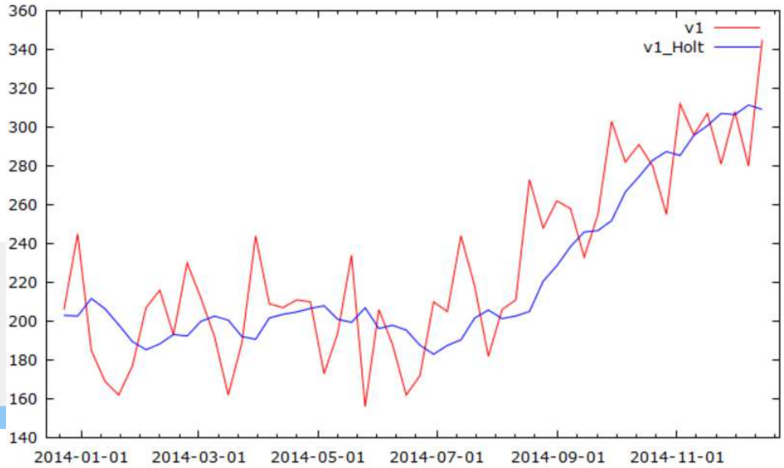
Ergebgnis der Optimierung
$\alpha = 0.247$
$\gamma = 0.095$

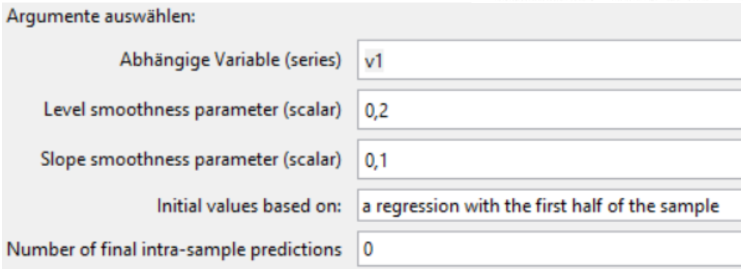
Holt Verfahren: Beispiel Thermostate



gretl







Holt Verfahren: Beispiel Thermostate

h-Schritte-Prognose: $\hat{y}_{t+h}(t) = L_t + hb_t$

Optimale Kombination: (α,γ) = (0.2468, 0.0951)

In Periode 52 (letzte Beobachtung), die Ein-Schritt-Prognose für Periode 53:

$\hat{y}_{53}(52)$ = L_{52} + b_{52} = 315.946 + 4.504 = 320.45

In Periode 52, die Zwei-Schritte-Prognose (h=2) für Periode 54:

$\hat{y}_{54}(52)$= L_{52} + $2b_{52}$ = 315. 946 + 2(4.504) = 324.95

In Periode 52, die Drei-Schritte-Prognose (h=3) für Periode 55:

$\hat{y}_{55}(52)$= L_{52} + $3b_{52}$ = 315. 946 + 3(4.504) = 329.45

Holt Verfahren: Beispiel Thermostate

h-Schritte-Prognose: $\hat{y}_{t+h}(t) = L_t + hb_t$

Wenn $y_{53} = 330$ beobachtet wird,

- Glättungsparameter α und γ aktualisieren, indem RSS minimiert wird oder
- nächste Trend- und Niveauschätzung berechnen

$L_{53} = \alpha y_{53} + (1-\alpha)(L_{52} + b_{52}) = \alpha y_{53} + (1-\alpha)\,\hat{y}_{53}(52)$

$= 0.247(330) + 0.753(315.946 + 4.504) = 322.81$

$b_{53} = \gamma(L_{53} - L_{52}) + (1-\gamma)b_{52}$

$= 0.095(322.77 - 315.9073) + 0.905(4.4946) = 4.728$

In Periode 53 (letzte Beobachtung), die Ein-Schritt-Prognose für Periode 53:

$\hat{y}_{54}(53) = L_{53} + b_{53} = 322.81 + 4.728 = 327.53$

In Periode 53 (letzte Beobachtung), die Zwei-Schritt-Prognose für Periode 53:

$\hat{y}_{55}(53) = L_{53} + 2\beta_{53} = 322.53 + 2(4.728) = 332.2104$