Übungsblatt 2: Zeitreihenanalyse

Lösungen

Aufgabe 1: Saisonkomponente

Was versteht man unter der Saisonkomponente im Komponentenmodell? Erläutern Sie das Konzept einer saisonbereinigten Zeitreihe. Unterscheiden Sie zwischen dem additiven und multiplikativen Modell.

Die Saisonkomponente spiegelt jährlich wiederkehrende zyklische Schwankungen in unterjährigen Zeitreihen (z.B. Quartals- oder Monatsdaten) wider. Sie ist auf natürliche oder institutionell bedingte jahrzeitliche Einflüsse zurückzuführen.
Während die Wellenlänge der Saisonzyklen konstant ist und ein Jahr umfasst, kann ihre Amplitude konstant oder variabel sein. In ersterem Fall liegt eine konstante Saisonfigur vor, die durch das additive Komponentenmodell erfasst werden kann. Die sich in letzterem Fall ergebende variable Saisonfigur wird in einem multiplikativen Komponentenmodell abgebildet, sofern die Saisonausschläge im Zeitverlauf zu- oder abnehmen.

Additives Modell: $x_t = T_t + C_t + S_t + e_t$

Multipikatives Modell: $x_t = T_t \times C_t \times S_t \times e_t$

wobei T = Trend, C = Zyklus, S = Saisonkomponente und e = Residualgrösse

Hat man die Saisonkomponente ermittelt, lässt sie sich aus der originären Zeitreihe „herausrechnen". Man erhält dadurch eine saisonbereinigte Zeitreihe, die angibt, wie sich eine Variable entwickelt hätte, wenn sie keinerlei saisonalen Einflüssen unterliegen würde. Wenn sich z.B. die Produktion in der Bauwirtschaft im Frühjahr erhöht, so kann dies durch eine veränderte Baukonjunktur und/oder witterungsbedingt verursacht sein. Die saisonbereinigte Produktion zeigt an, ob der Produktionsanstieg tatsächlich auf eine verbesserte ökonomische Entwicklung zurückgeführt werden kann.

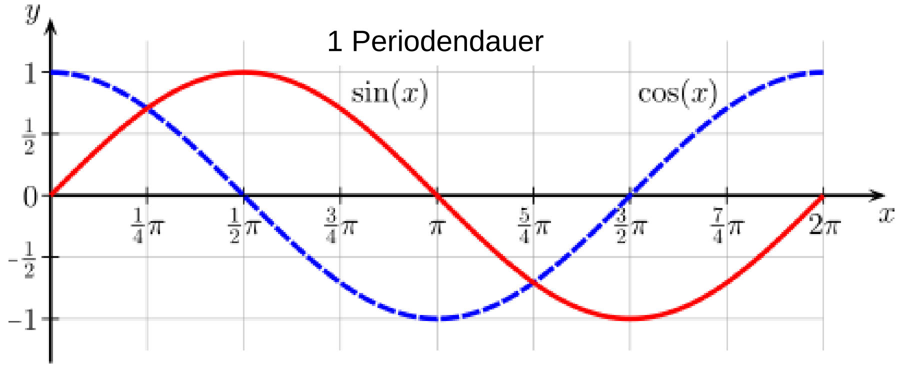
Aufgabe 2: Trigonometrisches Modell für Saisonkomponente

Als Modell für die Saisonschwankungen werden oft trigonometrische Funktionen verwendet.

Die Sinus- (sin) und Kosinusfunktionen (cos) stellen das Grundmodell einer zyklischen Funktion dar. Indem das Argument x mit einem Faktor λ multipliziert oder um einen additiven Term c ergänzt wird, lässt sich eine Vielzahl unterschiedlicher zyklischer Funktionen generieren. Die Multiplikation mit einem konstanten Faktor A erweitert die Palette zusätzlich.

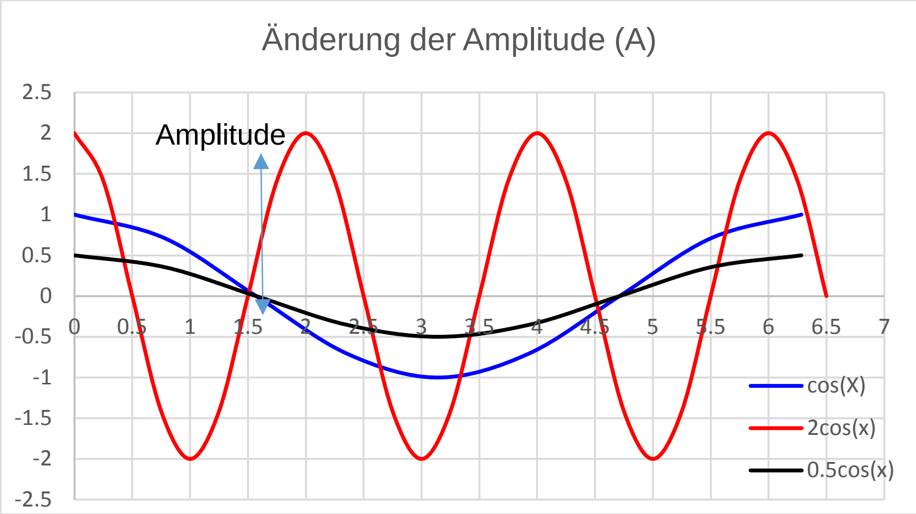
Allgemeine Sinusfunktion: $f(x) = A \sin(\lambda x + c)$

1. Skizzieren Sie die Funktionen sin(x) und cos(x). Nehmen Sie nur 4 Punkte 0, 0.5π, π, 2π

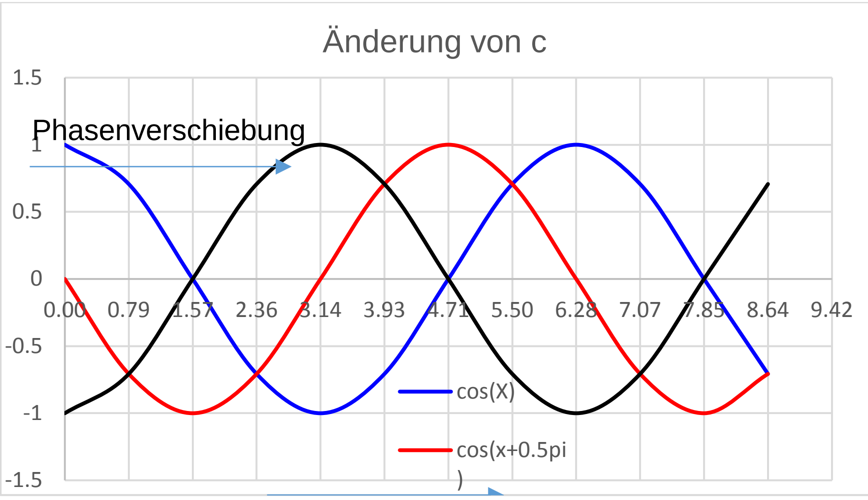


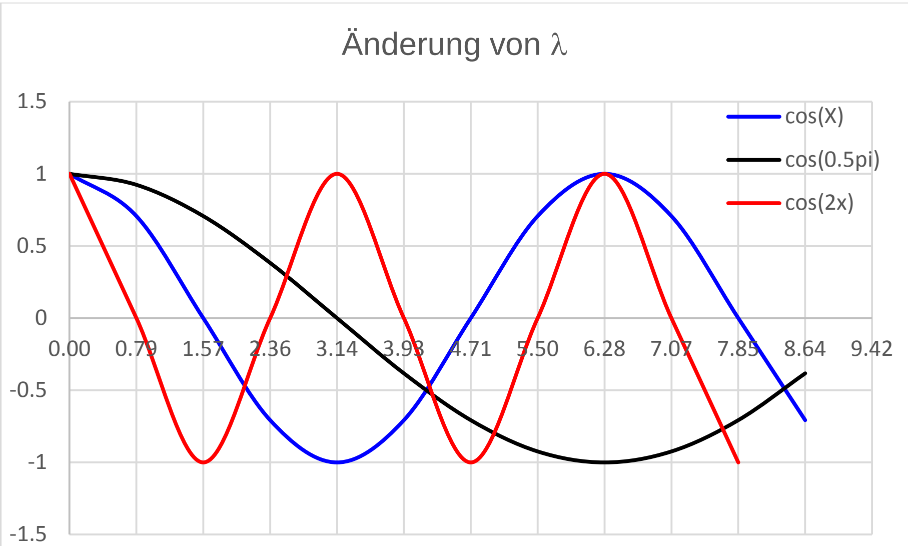
2. Erklären Sie die Auswirkung von A, λ und c auf die generierte Reihe

A = Amplitude → ändert den Ausschlag der Reihe



c = Phasenverschiebung → c positiv → Verschiebung der Reihe nach links





λ = Frequenz → Der multiplikative Faktor λ wirkt sich auf die Wiederholungsrate der Zyklen aus. Während sich für sin(x) ein Zyklus nach 2π wiederholt, hat sin(1.5x) einen Zyklus bereits nach $4/3\pi$ durchlaufen →(4/3)(3/2) = 2. Das gleiche gilt für die Kosinusfunktion

Ein hoher Wert von λ bedeutet, dass sich die Zyklen in sehr kurzer Folge wiederholen. Die Dauer eines Zyklus oder die Periodendauer P steht in umgekehrten Verhältnis zur Frequenz $P = 2\pi/\lambda$.

$y_t = T_t + S_t + e_t$

$S_t = \beta_1 \cos(Pxt) + \gamma_1 \sin(Pxt)$ wobei $P = 2\pi/\lambda$

3. Wie sind die Periodendauer und Frequenz bei Quartalszahlen zu definieren?

   Bei Quartalszahlen entspricht P = 4 der Dauer eines Jahres.

   Frequenz = $\lambda = 2\pi/P \rightarrow \lambda = \pi/2$

4. Definieren Sie die neuen Variablen

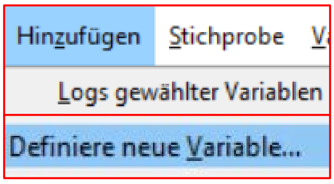
   Da monatliche Daten vorhanden sind entspricht P = 12 mit π = 3.1416

   cos1t = cos(time*3.1416/6) →Kosinus-Funktion

   sin1t = sin(time*3.1416/6) →Sinus-Funktion

   cos2t = cos(time*3.1416/3)

   sin2t = sin(time*3.1416/3)



   gretl Hauptfenster: Hinzufügen/ Zeittrend

   Hinzufügen/ Definiere neue Variable

Benutzen Sie die Datei USAutos.gdt für Ihre Schätzungen

5. Schätzen Sie folgende Modelle

   Modell 1: $y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 cos1t + \beta_5 sin1t + u$

   Modell 2: $y_t = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 cos1t + \beta_5 sin1t + \beta_6 cos2t + \beta_7 sin2t + u$

```
Abhängige Variable: y

             Koeffizient   Std.-fehler   t-Quotient   p-Wert
  -----------------------------------------------------------
  const      96,5487       0,593874      162,6        5,51e-127 ***
  time        0,0622203    0,0249030       2,499      0,0140    **
  time2       0,00280448   0,000219292    12,79       4,54e-023 ***
  cos1t      -0,813794     0,274163       -2,968      0,0037    ***
  sin1t      -1,59936      0,275784       -5,799      7,22e-08  ***

Mittel d. abh. Var.     111,2171   Stdabw. d. abh. Var.    12,25795
Summe d. quad. Res.     426,6441   Stdfehler d. Regress.   2,025425
R-Quadrat               0,973709   Korrigiertes R-Quadrat  0,972698
F(4, 104)               962,9342   P-Wert(F)               3,49e-81
Log-Likelihood         -229,0351   Akaike-Kriterium        468,0702
```

Modell 1

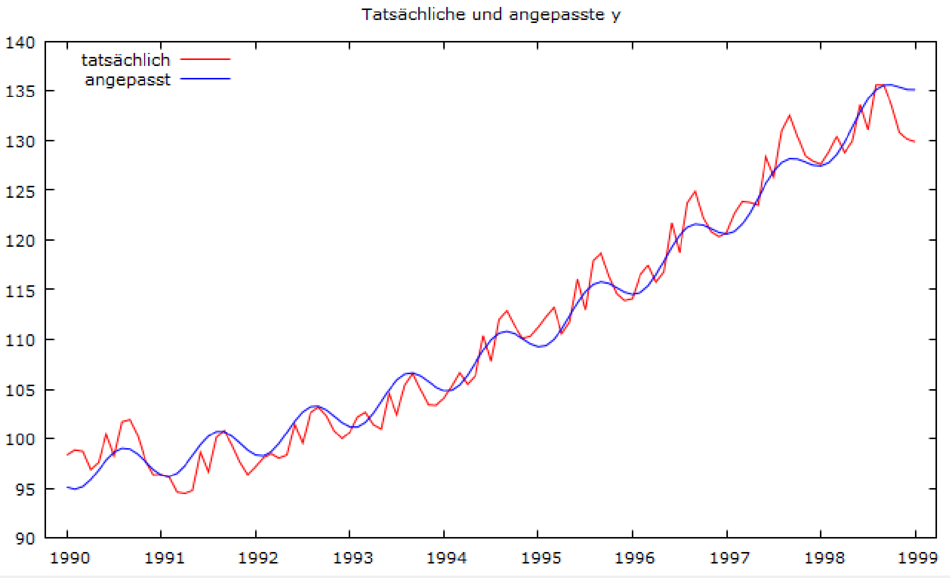
Abhängige Variable: y

Modell 2

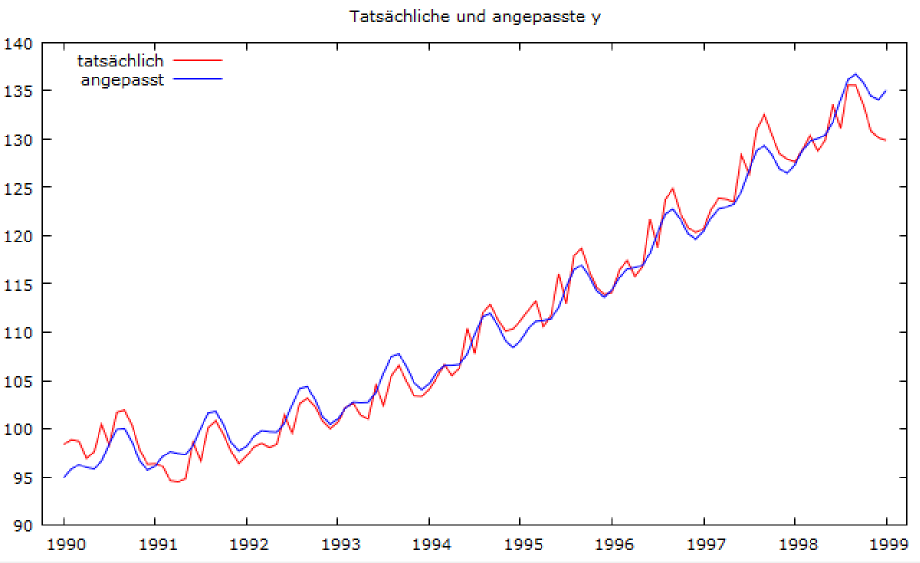
	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	96,5004	0,540015	178,7	3,54e-129	***
time	0,0628231	0,0226416	2,775	0,0066	***
time2	0,00280864	0,000199375	14,09	1,14e-025	***
cos1t	-0,813308	0,249172	-3,264	0,0015	***
sin1t	-1,59451	0,250628	-6,362	5,72e-09	***
cos2t	-1,13639	0,250091	-4,544	1,52e-05	***
sin2t	0,460389	0,248941	1,849	0,0673	*

Mittel d. abh. Var.	111,2171	Stdabw. d. abh. Var.	12,25795
Summe d. quad. Res.	345,5576	Stdfehler d. Regress.	1,840603
R-Quadrat	0,978706	Korrigiertes R-Quadrat	0,977453
F(6, 102)	781,3406	P-Wert(F)	7,28e-83
Log-Likelihood	-217,5470	Akaike-Kriterium	449,0941

6. Zeigen Sie die originäre Zeitreihe mit den angepassten Daten



Modell 1



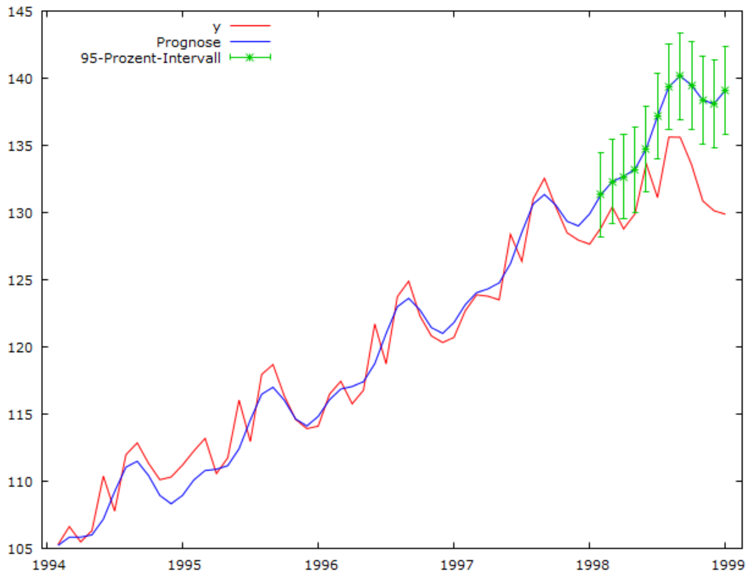
Modell 2

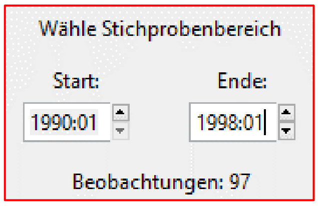
7. Welches Modell würden Sie vorziehen?

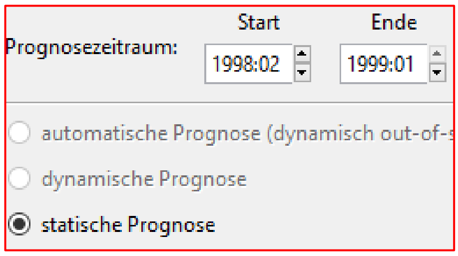
   Modell 2 hat eine bessere Anpassungsgüte. Das adjustierte R^2 ist grösser und das Akaike-Informationskriterium ist kleiner als beim Modell 1.

   Die Komplexität des Modells lässt sich erhöhen, indem zusätzliche Kosinus- und Sinus-Terme in die Regression aufgenommen werden.

8. Erstellen Sie mittels Modell 2 Prognosen für den Prognosezeitraum 1998:2 – 1999:01







Ergebnisse der Prognosefehleranalyse

	y_t	Prognose	Std.-Fehler	95%-Intervall
1998:08	135,62	139,36	1,620	136,15 – 142,58
1998:09	135,61	140,17	1,628	136,93 – 143,40
1998:10	133,58	139,48	1,635	136,23 – 142,72
1998:11	130,87	138,36	1,640	135,10 – 141,62
1998:12	130,14	138,11	1,646	134,84 – 141,38
1999:01	129,88	139,10	1,657	135,81 – 142,39

Prognose-Evaluationsstatistiken

Mittlerer Fehler	-4,8017
Mittlerer quadratischer Fehler	28,989
Wurzel d. mittl. quad. Fehlers	5,3842
Mittlerer absoluter Fehler	4,8017
Mittlerer prozentualer Fehler	-3,6565
Mittlerer absoluter prozentualer Fehler	3,6565

Aufgabe 3: Census-Verfahren X-12-Arima

Das Census-Verfahren wird unter anderem vom U.S. Bureau of the Census, der OECD, von der Europäischen Zentralbank und vielen nationalen statistischen Behörden verwendet.

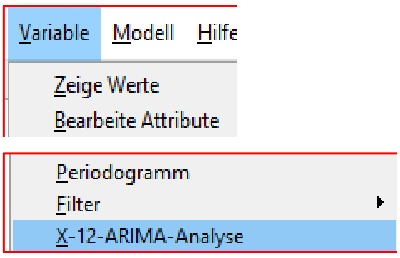
Neben eigenständigen Modulen zur Erkennung und Berücksichtigung extremer Werte (Ausreisser) und zur Bereinigung um Kalendereinflüsse besteht das Verfahren im Kern aus einer iterativen Prozedur zur Bestimmung der Trendkomponente und der daraus resultierenden Saisonfaktoren.

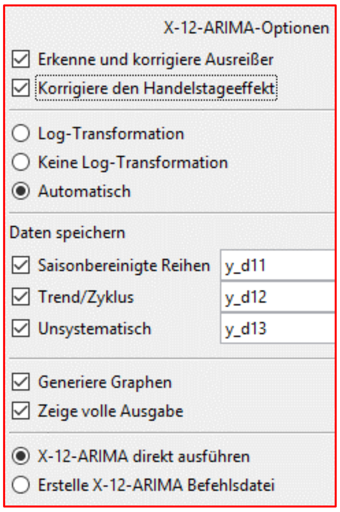
Die Bestimmung der Trendkomponente basiert im Wesentlichen auf gleitenden Durchschnitten.

Aus den Originalwerten y_t und den gewichteten gleitenden Durchschnitten $y_t(\varnothing 13)$ werden für jeden Monat (jedes Quartal) Saisonfaktoren s_t berechnet, für die wiederum gleitende Durchschnitte $s_t(\varnothing)$ gebildet werden. Nachdem die Ursprungsreihe um den Einfluss dieser Saisonfaktoren bereinigt wurde, wird das Verfahren erneut auf die verbleibende Restgrösse angewandt.

Gehen Sie auf die Webseite http://gretl.sourceforge.net/win32/ und installieren Sie das Paket X-13-ARIMA

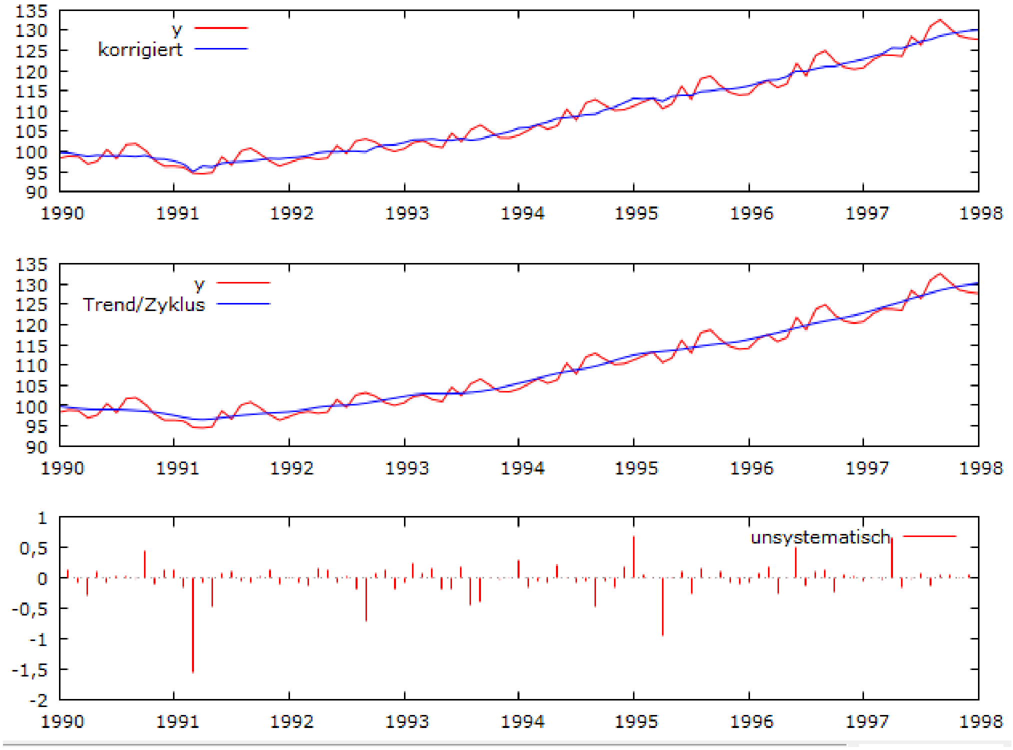
Für Mac: http://gretl.sourceforge.net/osx.html
Für Linux: http://gretl.sourceforge.net/#dl





1. Benutzen Sie die gretl Funktion X-12-ARIMA für die Bereinigung der Zeitreihe der registrierten Autos.

gretl Hauptfenster: Variable / X-12-ARIMA-Analyse



Das X-12-Verfahren ist viel mächtiger als das trigonometrische Komponentenmodell, da die Residuen keine Systematik mehr enthalten.

Aufgabe 4: Hodrick-Prescott Filter (HP-Filter)

1. Erklären Sie die Idee dieser Glättungsmethode

   Die Idee der Methode besteht darin, die Abwägung zwischen einer möglichst guten Anpassung der vorhandenen Daten einerseits und einer möglichst glatten Trendkomponente andererseits explizit vorzugeben.

2. Erklären Sie beide Komponente der Zielfunktion für den HP-Filter.

$$\text{Min} \sum_{t=1}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=1}^{T}[(\tau_{t+1}-\tau_t)-(\tau_t-\tau_{t-1})]^2$$

Die Zeitreihe wird in eine Trend- und zyklischen Komponente zerlegt: $y_t = t_t + c_t$

$\sum_{t=1}^{T}(y_t-\tau_t)^2$ →Summe der quadratischen Abweichungen zwischen Trendkomponente und Wert der Zeitreihe y_t, die möglichst klein sein soll.

$\sum_{t=1}^{T}[(\tau_{t+1}-\tau_t)-(\tau_t-\tau_{t-1})]^2$ → misst die Glattheit der Trendfunktion. Im Idealfall eines linearen Trends wird die darin gemessene Veränderung der Steigung des Trends von Periode zu Periode gleich null sein. Je stärker die Steigung der Trendkomponente schwankt, desto grösser wird diese zweite Komponente.

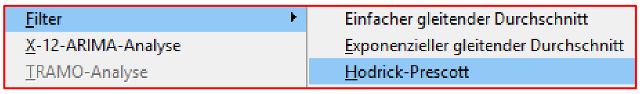
3. Erklären Sie die Auswirkung auf die Glättung für kleine und grosse Gewichtungsparameter λ.

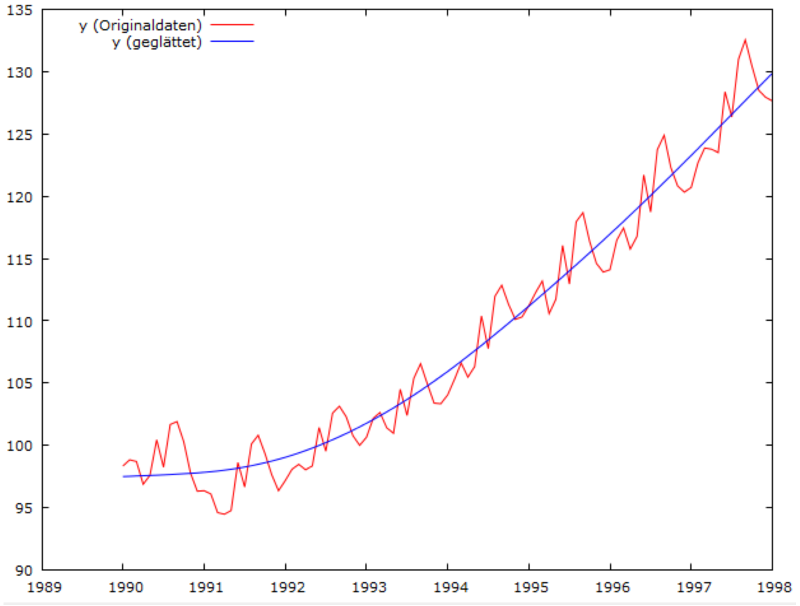
   Der Gewichtungsparameter λ gibt an, wie die beiden Aspekte relativ zueinander gewichtet werden sollen. Für $\lambda = 0$ spielt die Glattheit keine Rolle und der Trend wird mit der Reihe selbst zusammenfallen.
   Für einen grossen λ -Wert wird besonderes Gewicht auf die Glattheit gelegt, so dass sich die Trendkomponente τ mit wachsenden Werten für λ einem linearen Trend annähern wird. Für die praktische Arbeit werden häufig folgende Werte eingesetzt:

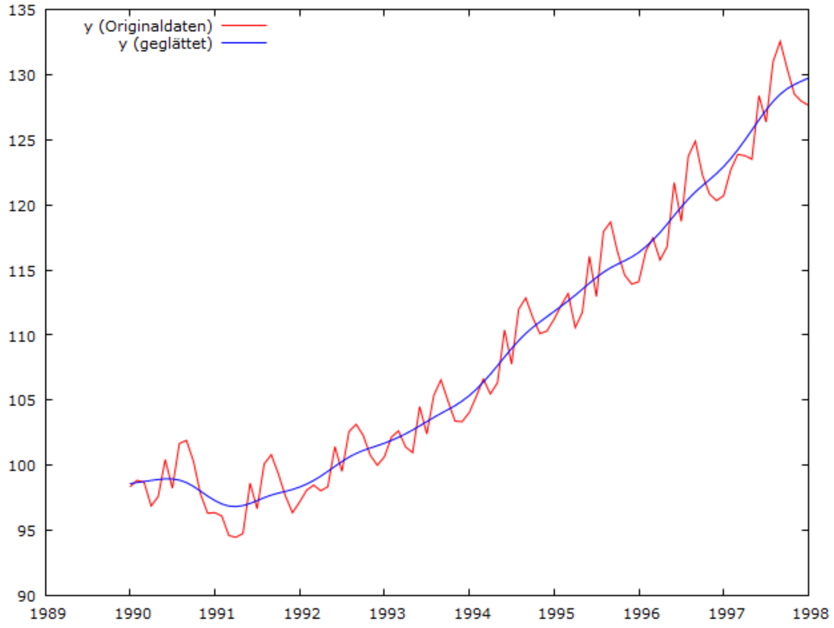
   - Jahresdaten $\lambda = 100$
   - Quartalsdaten $\lambda = 1600$
   - Monatsdaten $\lambda = 14‘400$

   Hinweis: gretl erkennt die Frequenz der Daten und setzt den entsprechenden Gewichtungsparameter automatisch ein.

4. Glätten Sie die Zeitreihe der registrierten Autos anhand des HP-Filters. Anschliessend wiederholen Sie die Glättung mit $\lambda = 100$. Was stellen Sie fest?







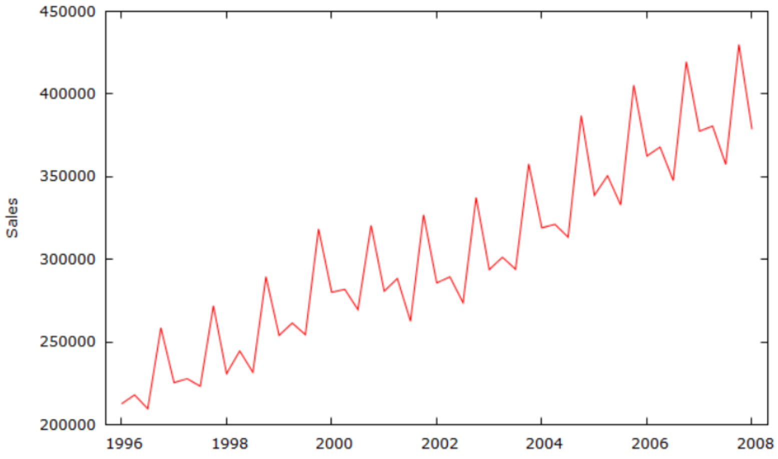
Glättung mit

$\lambda = 100$

Aufgabe 5: Saisondummies

Die Daten stellen die US retail & food services sales dar, entsprechen den Einzelhandelsumsätzen für die Periode 1996:Q1 bis 2008:Q1.
Das additive Komponentenmodell wird angewendet.

1. Erstellen Sie das Zeitreihendiagram. Was stellen Sie fest?



Diese Zeitreihe weist einen Trend und ein bestimmtes Saisonmuster.

2. Schätzen Sie das Trendmodell 1: $y_t = \beta_1 + \beta_2 \mathbf{t} + u_t$

```
Abhängige Variable: Sales

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  -------------------------------------------------------------
  const        211461        6876,48        30,75      8,78e-033 ***
  time           3678,10      239,408       15,36      5,99e-020 ***

Mittel d. abh. Var.      303413,3   Stdabw. d. abh. Var.    57550,40
Summe d. quad. Res.      2,64e+10   Stdfehler d. Regress.   23700,20
R-Quadrat                0,833940   Korrigiertes R-Quadrat  0,830407
```

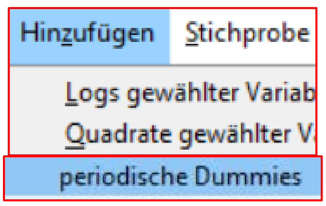
3. Erklären Sie was Saisondummies sind.

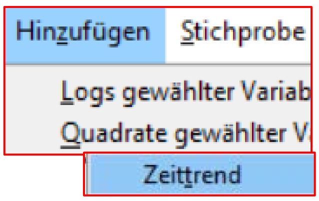
   Saisondummies sind qualitative Variablen, die jeweils für eine Teilperiode des Jahres den Wert 1 und sonst den Wert 0 annehmen. Für Quartalsdaten gibt es vier Saisondummies, für Monatsdaten 12. Durch den Einsatz von Saisondummies werden die Saisonkomponenten mittels Regressionsanalyse aufgedeckt.

4. Schätzen Sie das Modell 2 mit den entsprechenden Saisondummies.

   Modell 2: $y_t = \beta_1 + \beta_2 t + \beta_3 D_2 + \beta_4 D_3 + \beta_5 D_4 + u$

Fügen Sie zuerst die Dummyvariablen für die Saisons sowie eine Trendvariable mittels gretl Menu:
gretl Hauptfenster: Hinzufügen / periodische Dummies *Hinzufügen / Zeittrend*





Abhängige Variable: Sales

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	205459	3416,85	60,13	7,62e-041	***
time	3608,74	100,337	35,97	4,42e-032	***
dq2	1754,41	3644,95	0,4813	0,6329	
dq3	-14503,9	3643,57	-3,981	0,0003	***
dq4	43639,5	3644,95	11,97	8,41e-015	***

Mittel d. abh. Var.	296011,2	Stdabw. d. abh. Var.	53501,94
Summe d. quad. Res.	3,05e+09	Stdfehler d. Regress.	8728,700
R-Quadrat	0,975803	Korrigiertes R-Quadrat	0,973383
F(4, 40)	403,2687	P-Wert(F)	9,71e-32
Log-Likelihood	-469,5488	Akaike-Kriterium	949,0977

5. Welche implizite Annahme legt dieser Spezifikation mit Saisondummies zugrunde? Welches Quartal ist das Referenzquartal?

   Die Saisoneffekte ändern sich über die Zeit nicht und wiederholen sich zu regelmässigen Zeitintervallen.
   Q1 stellt das Referenzquartal dar.

6. Welches Modell weist die beste Anpassungsgüte auf?

   Modell 2 weist die grösste adjustierte Bestimmtheitsmass auf.

7. Berechnen Sie die normierten Saisonfaktoren anhand der Regressionsergebnisse.H57:H60

   D = (0+1754.4 -14503.9+43'639.5)/4 = 7'722.5

   Normierte Saisonkomponenten

   S_1 = -7'722.5

   S_2 = 1754.7 -7'722.5 = -5968.1

   S_3 = 14'503.9 -7'722.5 = -22'226.4

   S_4 = 43'639.5 -7'722.5 = 35'917

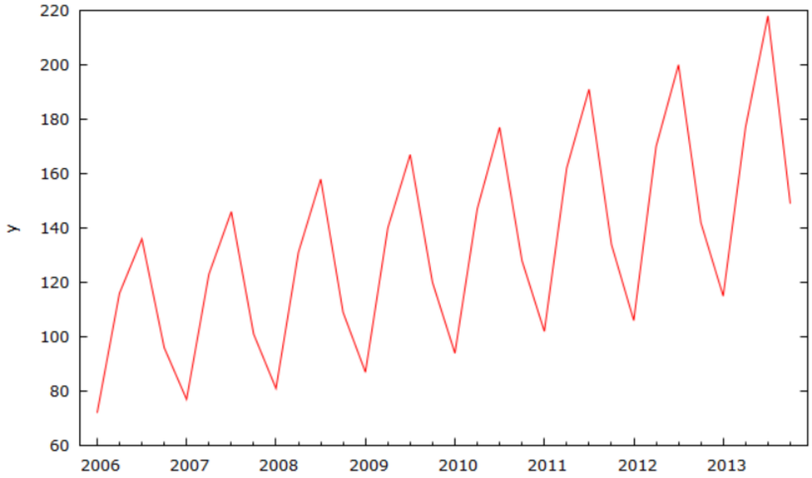
8. Interpretieren Sie den Saisonfaktor S_3.

   Die dritten Quartalszahlen liegen im Durchschnitt um 22'226 registrierte Autos unter der glatten Komponente.

Aufgabe 6: Holt-Winters Modell

Benutzen Sie die Datei Sportgetränke.gdt

1. Erstellen Sie das Zeitreihendiagram. Was stellen Sie fest?

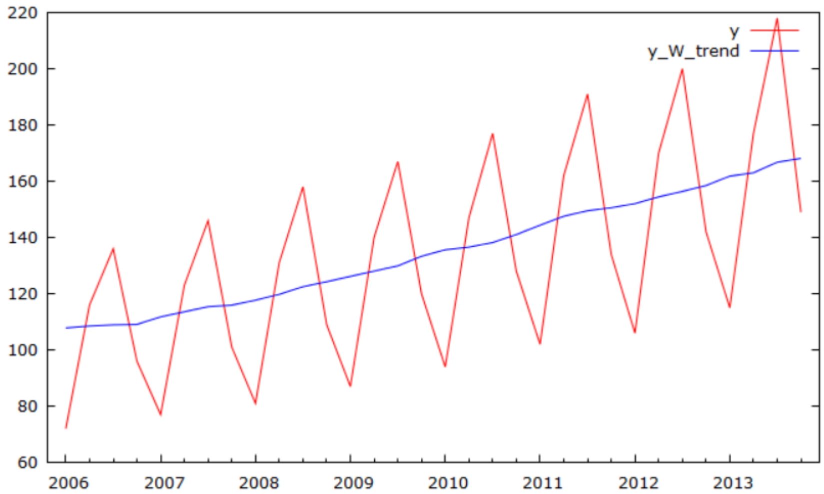


Ein positiver Trend ist sichtbar. Die saisonalen Schwankungen nehmen mit der Zeit zu.

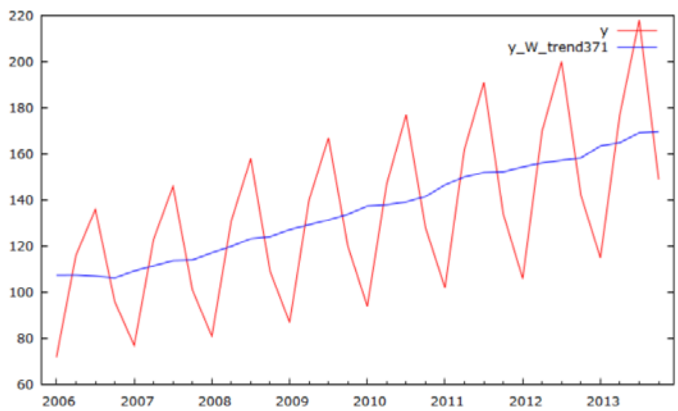
2. Wann wird das Holt-Winters Verfahren angewendet?

   Wenn die saisonalen Schwankungen mit dem Trendniveau zunehmen.

3. Glätten Sie mittels Winters-Methode die Zeitreihe mit den Parametern α = 0.3, γ = 0.1 und δ = 0.7



Glättungsparameter α = 0.3, γ = 0.7 und δ = 0.1



Aufgabe 7: Multiplikatives Modell

Sie erhalten folgende Tabelle mit den Quartalsumsätzen von Traktoren eines Unternehmens. Alle Zahlen sind in Millionen Euros angegeben. Ein multiplikatives Modell für die Saisonbereinigung wurde angewandt.

1	2	3	4	5	6	7	8	9
		t	y_t	GD	S_{ij}	S*	y*	Trend
2005	Q1	1	362			0.960519635	376.879334	342.8
	Q2	2	385			1.022149385	376.657273	360.6
	Q3	3	432	382.5	1.12941176	1.140020527	378.940545	378.4
	Q4	4	341	388	0.87886598	0.877310453	388.687948	396.2
2006	Q1	5	382	399.25	0.95679399	0.960519635	397.701396	414
	Q2	6	409	413.25	0.98971567	1.022149385	400.137207	431.8
	Q3	7	498	430.375	1.15713041	1.140020527	436.834239	449.6
	Q4	8	387	A	B	C	D	E
2007	Q1	9	473	478.25	0.98902248	0.960519635	492.441782	485.2
	Q2	10	513	499.625	1.02677008	1.022149385	501.883587	503
	Q3	11	582	519.375	1.12057762	1.140020527	510.517123	520.8
	Q4	12	474	536.875	0.88288708	0.877310453	540.287647	538.6
2008	Q1	13	544	557.875	0.97512884	0.960519635	566.360104	556.4
	Q2	14	582	580.625	1.00236814	1.022149385	569.388397	574.2
	Q3	15	681	601.5	1.13216958	1.140020527	597.357665	592
	Q4	16	557	627.625	0.88747262	0.877310453	634.894977	609.8
2009	Q1	17	628	654.75	0.95914471	0.960519635	653.812767	627.6
	Q2	18	707	670.625	1.05424045	1.022149385	691.67972	645.4
	Q3	19	773	674.875	1.1453973	1.140020527	678.057966	663.2
	Q4	20	592	677	0.87444609	0.877310453	674.789634	681
2010	Q1	21	627	689.375	0.90951949	0.960519635	652.771664	698.8
	Q2	22	725	708.125	1.02383054	1.022149385	709.28967	716.6
	Q3	23	854			1.140020527	749.109318	734.4
	Q4	24	661			0.877310453	753.439102	752.2

Spalte 5: gleitende Durchschnitte (GD)

Spalte 6: Saisonkomponente (S_{ij})

Spalte 7: Normierte Saisonkomponente S*

Spalte 8: Saisonbereinigte Werte y*

Spalte 9: Trendkomponente

Leider sind die Werte für **2006:Q4** verloren gegangen

Die unnormierten Saisonfaktoren wurden wie folgt berechnet:

Q1	0.95792
Q2	1.01938
Q3	1.13693
Q4	0.87493

Summe = 3.98918

1. Schreiben Sie das multiplikative Komponentenmodell auf und erklären Sie wann es angewendet werden sollte.

   $y_t = g_t \times S_t \times e_t$ wobei g_t = glatte Komponente und S_t = Saisonkomponente
   Das multiplikative Komponentenmodell ist besonders geeignet, wenn die saisonalen Schwankungen abhängig von der trendmässigen Entwicklung sind. (bleiben nicht konstant).

2. Bestimmen Sie den Wert A

   Quartalsdaten → 5-gliedriger gleitender Durchschnitt
   A = (0.5*409 + 498 + 387 + 473+ 0.5*513)/ 4 = 454.75

3. Bestimmen Sie den Wert B

   Trendbereinigter Wert B = 387 / 454.75 = 0.8510

4. Berechnen Sie den Normalisierungsfaktor

   NF = 4 / 3.98918 = 1.00271

5. Berechnen Sie die normierte Saisonkomponente für das 4. Quartal, C.

   S_4 = 0.87493 * 1.00271 = 0.87731

6. Berechnen Sie den saisonbereinigten Wert D

   y* = 387 / 0.87731 = 441.12

7. Berechnen Sie den geschätzten Trendwert E mittels des geschätzten Trendmodells

   $\hat{y}_t = 325 + 17.8\mathbf{t} + e$

   Trendwert = (325 + 17.8 * 8) = 467.4

8. Berechnen Sie den entsprechenden Prognosefehler

   Prognostizierter Wert $\hat{y}_t$ = 467.4 x 0.8773 = 410.055

   e = y – $\hat{y}$ = 387 – 410.055 = -23.0549