Übung 4: Rosennachfrage, Teil II Lösungen

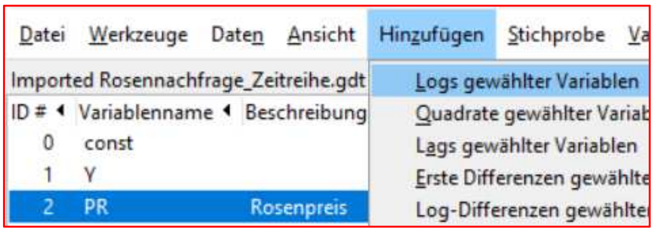
Durch Diskussionen mit anderen CAS-Teilnehmern haben Sie folgende Regressionsmodelle gesammelt:

1. Modell 1: $\ln y_t = \beta_1 + \beta_2 \ln PR_t + \beta_3 \ln PN_t + u_t$ $\quad t = 1,\ldots,16$
2. Modell 2: $\ln y_t = \beta_1 + \beta_2 \ln(PR_t / PN_t) + u_t$ $\quad t = 1,\ldots,16$
3. Modell 3: $\ln y_t = \beta_1 + \beta_2 \ln PR_t + \beta_3 \ln PN_t + \beta_4 \ln EINK_t + u_t$ $\quad t = 1,\ldots,16$
4. Modell 4: $\ln y_t = \beta_1 + \beta_2 \ln PR_t + \beta_3 \ln PN_t + \beta_4 \ln EINK_t + \beta_5 T + u_t$ $\quad t = 1,\ldots,16$
5. Modell 5: $\ln y_t = \beta_1 + \beta_2 \ln PR_t + \beta_3 \ln PN_t + \beta_4 \ln EINK_t + \beta_5 \ln T + u_t$ $\quad t = 1,\ldots,16$

Es gelte $u_t \sim$ iid $N(0;\sigma^2)$. iid: **i**ndependent and **i**dentically **d**istributed (unabhängig identisch verteilten Zufallsvariablen)

1. Definieren Sie folgende logarithmierte Variablen:
   l_y, l_PR, l_PN, l_EINK, l_RelP wobei l_ der natürliche Logarithmus symbolisiert.

   gretl: Hinzufügen / Logs gewählter Variablen→ wählen Sie jeweils Y,PR, PN, EINK,T, RelP



2. Welche Vorzeichen für die Regressionskoeffizienten erwarten Sie für das Modell 4?

   Rosenpreis: Je höher der Rosenpreis, desto kleiner die Rosennachfrage, ceteris paribus (c.p.)

   Nelkenpreis: Je höher der Nelkenpreis, desto höher die Rosennachfrage (Substitutionsgut)

   Einkommen: Je höher das Einkommen, desto höher die Nachfrage nach Rosen, c.p.

   Zeit: Beim Zeitablauf sollte der Rosenabsatz sinken, wie in der Graphik über die Absatzentwicklung bei Übung 3 ersichtlich war.

3. Schätzen Sie diese 5 Regressionsmodelle. Speichern Sie jeweils die geschätzten Werte →yhat1, …yhat5. Diese neuen Variablen erscheinen im Hauptfenster.

   gretl Output-Fenster: Speichern / geschätzte Werte

Modell 1

```
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  ---------------------------------------------------------------
  const        9,22776      0,568390      16,23      5,18e-010 ***
  l_PR        -1,76072      0,298206      -5,904     5,20e-05  ***
  l_PN         1,33978      0,527324       2,541     0,0246    **

Mittel d. abh. Var.     8,902209   Stdabw. d. abh. Var.    0,306877
Summe d. quad. Res.     0,382569   Stdfehler d. Regress.   0,171547
R-Quadrat               0,729174   Korrigiertes R-Quadrat  0,687509
F(2, 13)               17,50066    P-Wert(F)               0,000205
Log-Likelihood          7,164472   Akaike-Kriterium       -8,328944
Schwarz-Kriterium      -6,011178   Hannan-Quinn-Kriterium -8,210255
rho                    -0,052667   Durbin-Watson-Stat      2,058814
```

Modell 2

```
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  ---------------------------------------------------------------
  const        8,71319     0,0533773      163,2       2,31e-024 ***
  l_RelP      -1,73605     0,295129       -5,882      3,99e-05  ***

Mittel d. abh. Var.      8,902209   Stdabw. d. abh. Var.     0,306877
Summe d. quad. Res.      0,406905   Stdfehler d. Regress.    0,170484
R-Quadrat                0,711946   Korrigiertes R-Quadrat   0,691370
F(1, 14)                 34,60194   P-Wert(F)                0,000040
Log-Likelihood           6,671089   Akaike-Kriterium        -9,342178
Schwarz-Kriterium       -7,797001   Hannan-Quinn-Kriterium  -9,263053
rho                     -0,158187   Durbin-Watson-Stat       2,276028
```

Modell 3

```
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient   p-Wert
  ------------------------------------------------------------
  const        6,28769      4,87459        1,290      0,2214
  l_PR        -1,85624      0,343780      -5,399      0,0002 ***
  l_PN         1,45408      0,572411       2,540      0,0259 **
  l_EINK       0,559553     0,921079       0,6075     0,5548

Mittel d. abh. Var.      8,902209   Stdabw. d. abh. Var.     0,306877
Summe d. quad. Res.      0,371154   Stdfehler d. Regress.    0,175868
R-Quadrat                0,737255   Korrigiertes R-Quadrat   0,671568
F(3, 12)                 11,22387   P-Wert(F)                0,000849
Log-Likelihood           7,406800   Akaike-Kriterium        -6,813600
Schwarz-Kriterium       -3,723245   Hannan-Quinn-Kriterium  -6,655348
rho                     -0,013701   Durbin-Watson-Stat       2,004954
```

Modell 4

```
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient   p-Wert
  ------------------------------------------------------------
  const        3,57216      4,69516        0,7608     0,4628
  l_PR        -1,17073      0,488324      -2,397      0,0354 **
  l_PN         0,737938     0,652863       1,130      0,2824
  l_EINK       1,15321      0,901989       1,279      0,2274
  T           -0,0301108    0,0164188     -1,834      0,0938 *

Mittel d. abh. Var.      8,902209   Stdabw. d. abh. Var.     0,306877
Summe d. quad. Res.      0,284245   Stdfehler d. Regress.    0,160750
R-Quadrat                0,798779   Korrigiertes R-Quadrat   0,725607
F(4, 11)                 10,91654   P-Wert(F)                0,000798
Log-Likelihood           9,541038   Akaike-Kriterium        -9,082076
Schwarz-Kriterium       -5,219132   Hannan-Quinn-Kriterium  -8,884262
rho                     -0,067449   Durbin-Watson-Stat       2,049078
```

Modell 5

```
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient   p-Wert
  ------------------------------------------------------------
  const        0,626824     6,14826        0,1020     0,9206
  l_PR        -1,27355      0,526649      -2,418      0,0341 **
  l_PN         0,937305     0,659191       1,422      0,1828
  l_EINK       1,71298      1,20084        1,426      0,1815
  l_T         -0,181597     0,127893      -1,420      0,1833

Mittel d. abh. Var.      8,902209   Stdabw. d. abh. Var.     0,306877
Summe d. quad. Res.      0,313664   Stdfehler d. Regress.    0,168864
R-Quadrat                0,777953   Korrigiertes R-Quadrat   0,697208
F(4, 11)                 9,634745   P-Wert(F)                0,001343
Log-Likelihood           8,753157   Akaike-Kriterium        -7,506314
Schwarz-Kriterium       -3,643370   Hannan-Quinn-Kriterium  -7,308499
rho                      0,091730   Durbin-Watson-Stat       1,782659
```

4. Interpretieren Sie die Regressionskoeffizienten des Regressionsmodells 4 und beurteilen Sie, ob die Parameterschätzungen plausibel sind.

b_2 = -1.17: Eine Erhöhung des Rosenpreises um 1% bewirkt einen Rückgang des Rosenabsatzes um durchschnittlich 1.17% erwarten, ceteris paribus.
Vorzeichen ist plausibel → bei steigendem Rosenpreis fällt der Rosenabsatz.

b_3 = 0.74: Eine Erhöhung des Nelkenpreises um 1% bewirkt einen Anstieg der Rosenabsatzes um durchschnittlich 0.74%, ceteris paribus.

Vorzeichen ist plausibel: Werden die Nelken relativ zu Rosen teurer, werden die Konsumenten substituieren und mehr Rosen kaufen, ceteris paribus →Substitutionseffekt.

b_4 = 1.15: Eine Erhöhung des Einkommens um 1% bewirkt einen Anstieg des Rosenabsatzes um durchschnittlich 1.15% erwarten, ceteris paribus.

Vorzeichen ist plausibel: Steigt das verfügbare Einkommen, steigt der Rosenabsatz.

b_5 = -0.03: Mit einem zusätzlichen Quartal reduziert sich die Rosenabsatz im Durchschnitt um 3%, ceteris paribus.

5. Sind diese Koeffizienten 4 statistisch signifikant auf dem 5%-Signifikanzniveau?

Nur der Regressionskoeffizient b_2 ist auf dem 5% statistisch signifikant → zwei Sternchen.

Sie wissen, dass R^2 dem Quadrat des Korrelationskoeffizienten zwischen den tatsächlichen und geschätzten Werten entspricht: $R^2 = \rho(y_i, \hat{y}_i)$
Sie haben jeweils die geschätzten Werte yhat1,…,yhat5 gespeichert.
Berechnen Sie jetzt die entsprechenden Exponentialwerte in gretl: expy1 = exp(yhat1),…,expy5 = exp(yhat5).

6. Öffnen Sie die Korrelationsmatrix

gretl Hauptfenster: Ansicht Korrelationsmatrix, wählen Sie die Variablen exp(l_yhati) i = 1,..5, und y aus.

Hinweis: Wählen Sie die erste Zeile aus der Korrelationsmatrix

Korrelation von y mit den berechneten Exponentialwerten:

exp(l_yhat1)	exp(l_yhat1)	exp(l_yhat1)	exp(l_yhat1)	exp(l_yhat1)	
0.8558	0.8413	0.8503	0.8985	0.8764	y

7. Berechnen Sie die quadrierten Korrelationskoeffizienten

ρ_1^2	ρ_2^2	ρ_3^2	ρ_4^2	ρ_5^2	
0.7324	0.7078	0.7230	0.8073	0.7680	y

8. Welches Regressionsmodell würden Sie auswählen. Begründen Sie Ihre Auswahl.

Folgende Tabelle enthält eine Zusammenstellung der zur vergleichenden Kennzahlen mit den entsprechenden Kriterien.

	Modell 1	Modell 2	Modell 3	Modell 4	Modell 5
# Regressoren	K = 3	K = 2	K = 4	K = 5	K =5
Regressoren	l_PR, l_PN	l_PR/PN	l_PR, l_PN, l_EINK	l_PR, l_PN, l_EINK, T	l_PR, l_PN, l_EINK, lnT
$\overline{R}^2$	0.6875	0.6913	0.6715	0.7256	0.6972
Akaike	-8.328	**-9.34**	-6.81	-9.08	-7.506
SIC	-6	-7.79	-3.72	-5.21	-3.64
R^2	0.7324	0.7078	0.7230	0.8073	0.7680

Modell 4 weist das höchste R^2 und adjustierte Bestimmtheitsmass auf. Problematisch ist, dass nur der logarithmierte Rosenpreis (PR) statistisch signifikant ist. Um eine Modellfehlspezifikation zu vermeiden, könnte es dennoch Sinn machen, diese Regressoren zu berücksichtigen und Modell 4 auswählen. Anhand der Informationskriterien sollte das Modell 2 vorgezogen werden, ignoriert aber die Zeitkomponente und das Einkommen. Die Spezifikation in Logarithmen hat den Vorteil, dass die Koeffizienten als Elastizitäten interpretiert werden können.

9. Folgende Modelle wurden aus der Übung 3 und 4 ausgewählt:

Teil I, Modell 2: $y_t = \beta_1 + \beta_2 RelP + u_t$

Teil II, Modell 2: $\ln y_t = \beta_1 + \beta_2 \ln RelP + u_t$

Wie können jetzt diese Modelle miteinander verglichen werden? Welches Modell würden Sie vorziehen? Begründen Sie Ihre Antwort.

Die Informationskriterien können zum Vergleich der Modelle 2 nicht herangezogen werden, da eine Spezifikation als log-log und die andere ohne Transformation dargestellt wird.

In diesem Fall kann R^2 herangezogen werden:

Ausgewählte Modelle	R^2
$y_t = \beta_1 + \beta_2 RelP + u_t$	0.7823
$\ln y_t = \beta_1 + \beta_2 \ln RelP + u_t$	0.7078

Die log-log Spezifikation ist nicht besser als das Modell 2 aus Übung 3.

10. Testen Sie das Regressionsmodell 1 mit dem F-Test!

Modell 1: $\ln y_t = \beta_1 + \beta_2 \ln PR_t + \beta_3 \ln PN_t + u_t$

Frage: Hat wenigstens einer der Preise (von Rosen oder Nelken) Einfluss auf die Rosenabsatzmenge?

i. Stellen Sie die Nullhypothese und alternative Hypothese auf.

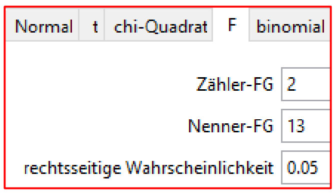
Nullhypothese H_0: $\beta_2 = \beta_3 = 0$

H_1: mindestens ein Koeffizient hat einen Einfluss $\Leftrightarrow$ H_1: H_0 ist falsch

ii. Bestimmen Sie den kritischen F-Wert (F_c) auf dem 5%-Signifikanzniveau.

Kritischer Wert $F_c(0.95,2,13) = 3.81$

Zähler-Freiheitsgrade	K–1 = 3 – 1 = 2
Nenner-Freiheitsgrade	N – k = 16 -3 = 13



gretl Hauptfenster: Werkzeuge/Statistische Tabellen/F/

rechtsseitige Wahrsch. = 0.05

iii. Berechnen Sie den F-test mittels Bestimmtheitsmass $F = \frac{R^2}{1-R^2}\frac{N-k}{L}$

$$F = \frac{R^2}{1-R^2}\frac{N-k}{L} = \frac{0.72917}{0.27082}\cdot\frac{13}{2} = 17.5$$

Hinweis: R^2 ist hier aus der log-log Spezifikation zu entnehmen.

Diese Berechnung dient nur dem Umgang mit dieser Formel und ist nicht notwendig, da der F-Wert von gretl automatisch ermittelt wird.

iv. Was ist Ihre Schlussfolgerung?

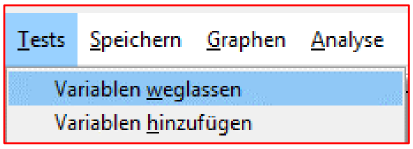
$F_e = 17.5 > F_c = 3.81 \rightarrow H_0$ kann verworfen werden

p-Wert $\cong 0 < 5\% \rightarrow H_0$ kann verworfen werden

→ Mindestens ein Koeffizient ist statistisch signifikant verschieden von 0!

11. Führen Sie in gretl den Test auf „Weglassen der Variablen" durch. Nehmen Sie die Variablen l_PR und l_PN vom Modell 1 weg. Was ist Ihre Schlussfolgerung?

gretl Output-Fenster: Test / variablen weglassen / l_PR und l_PN weglassen




```
  Nullhypothese: Die Regressionskoeffizienten sind Null für die Variablen
    l_PR, l_PN
  Teststatistik: F(2, 13) = 17,5007, p-Wert 0,000205346
  Das Weglassen von Variablen verbesserte 0 von 3 Informationskriterien.

Modell 19: KQ, benutze die Beobachtungen 1971:3-1975:2 (T = 16)
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  -------------------------------------------------------------
  const        8,90221      0,0767192      116,0      1,43e-023 ***

Mittel d. abh. Var.     8,902209   Stdabw. d. abh. Var.    0,306877
Summe d. quad. Res.     1,412599   Stdfehler d. Regress.   0,306877
R-Quadrat               0,000000   Korrigiertes R-Quadrat  0,000000
```

p-Wert $\cong 0 \rightarrow H_0$ kann verworfen werden

→ Mindestens ein Koeffizient ist statistisch signifikant von 0 verschieden!

Hinweis: Dieser Schritt ist nicht notwendig, da gretl den F-Test mit dem p-Wert automatisch berechnet.

12. Interpretieren Sie beim Modell 1 konkret folgende Restriktion: $\beta_2 = -\beta_3$

Das Steigen des Rosenpreises um 1% hat die gleiche Wirkung auf den Rosenabsatz, wie das Sinken des Nelkenpreises um 1%, ceteris paribus.

13. Stellen Sie das restringierte Modell auf und schätzen Sie es.

Regressionsmodell: $\ln y = \beta_1 + \beta_2 \ln PR_t + \beta_3 \ln PN_t + u$

Restringiertes Modell: $\ln y = \beta_1 + \beta_2 \ln PR_t - \beta_2 \ln PN_t + u$

$\leftrightarrow \ln y = \beta_1 + \beta_2 [\ln PR_t - \ln PN_t] + u$

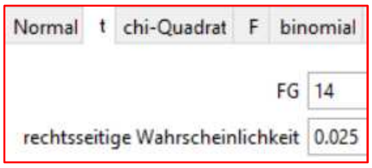
Definieren Sie die neue Variable: l_diff = lnPR – lnPN.

```
Abhängige Variable: l_Y

              Koeffizient   Std.-fehler   t-Quotient     p-Wert
  --------------------------------------------------------------
  const         8,71319      0,0533773     163,2       2,31e-024 ***
  l_diff       -1,73605      0,295129       -5,882     3,99e-05  ***

Mittel d. abh. Var.      8,902209   Stdabw. d. abh. Var.     0,306877
Summe d. quad. Res.      0,406905   Stdfehler d. Regress.    0,170484
R-Quadrat                0,711946   Korrigiertes R-Quadrat   0,691370
F(1, 14)                 34,60194   P-Wert(F)                0,000040
```

14. Testen Sie die Signifikanz von b_2 mittels t-Tests.



Zweiseitiger Test: H_0: $\beta_2 = 0$ H_1: $\beta_2 \neq 0$

K = 2 (Interzept + 1 erklärende Variablen)

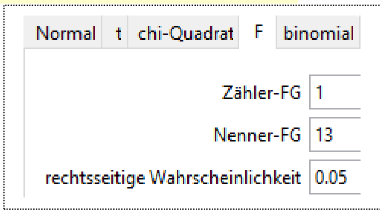
Freiheitsgrade für t-Test: N-K = 16 – 2 = 14

$t_e = -5.88 \rightarrow |t_e| > t_c = 2.178 \rightarrow H_0$ kann verworfen werden

p-Wert $\cong 0 < 0.05$ → H_0 kann verworfen werden → β_2 ist von null verschieden

→ die Daten aus der Stichprobe sind mit der Restriktion $\beta_2 = -\beta_3$ kompatibel

15. Testen Sie anhand des F-Tests auf dem 5%-Signifikanzniveau, ob die Restriktion falsch ist.



Berechnen Sie den F-Wert mittels $F = \frac{(RSS_r - RSS)/L}{RSS/(N-K)}$

Nur eine lineare Restriktion → L = 1

Zähler-FG = L = 1

Nenner- FG = N – K = 16 – 3 = 13

$$F = \frac{(RSS_r - RSS)}{RSS}\frac{(N-K)}{L} = \frac{(0.406905 - 0.382569)}{0.382569} \cdot \frac{13}{1} = 0.827$$

$F_e = 0.827 < F_c(0.95, 1, 13) = 4.66$ → H_0 kann nicht verworfen werden!

Die Daten aus der Stichprobe sind mit der Restriktion $\beta_2 = -\beta_3$ kompatibel.

Hinweis: Diese Berechnung dient nur dem Umgang mit dieser Formel und ist in der Praxis nicht erforderlich, da gretl einen eingebauten Test besitzt.

16. Testen Sie diese Restriktion mittels gretl Restriktionen Funktion.

gretl output-Fenster: Test / lineare Restriktionen / b[2] + b[3] = 0

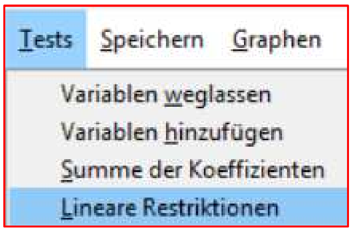
```
Restriktion:
 b[l_PR] + b[l_PN] = 0

Teststatistik: F(1, 13) = 0,826986, mit p-Wert = 0,379697

Restringierte Schätzungen:

              Koeffizient   Std.-fehler   t-Quotient     p-Wert
  --------------------------------------------------------------
  const         8,71319      0,0533773     163,2       2,31e-024 ***
  l_PR         -1,73605      0,295129       -5,882     3,99e-05  ***
  l_PN          1,73605      0,295129        5,882     3,99e-05  ***

  Standardfehler der Regression = 0,170484
```



p-Wert = 0.379 > α = 5% → H_0 kann nicht verworfen werden

Die Daten aus der Stichprobe sind mit der Restriktion $\beta_2 = -\beta_3$ kompatibel.

Hinweis: Das ist die effektivste Methode, um diese Restriktion zu testen.

17. Testen Sie im Regressionsmodel 4, ob die Variablen l_PN, l_EINK und T gemeinsam statistisch signifikant sind.

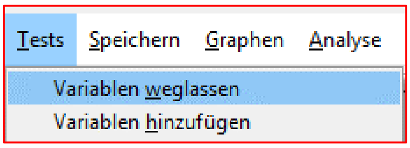
gretl: Tests / Variablen weglassen →l_PN, l_EINK und T auswählen

```
  Nullhypothese: Die Regressionskoeffizienten sind Null für die Variablen
    l_PN, l_EINK, T
  Teststatistik: F(3, 11) = 3,71884, p-Wert 0,0456033
  Das Weglassen von Variablen verbesserte 0 von 3 Informationskriterien.

Modell 26: KQ, benutze die Beobachtungen 1971:3-1975:2 (T = 16)
Abhängige Variable: l_Y

             Koeffizient   Std.-fehler   t-Quotient    p-Wert
  --------------------------------------------------------------
  const        10,4620       0,347849       30,08     4,03e-014 ***
  l_PR         -1,39276      0,307297       -4,532    0,0005    ***

Mittel d. abh. Var.      8,902209   Stdabw. d. abh. Var.    0,306877
Summe d. quad. Res.      0,572535   Stdfehler d. Regress.   0,202226
R-Quadrat                0,594694   Korrigiertes R-Quadrat  0,565743
F(1, 14)                 20,54179   P-Wert(F)               0,000469
```



p-Wert = 4.5% < 5% → H_0: $b_3 = b_4 = b_5 = 0$ kann auf dem 5%-Niveau verworfen werden

Konklusion: Diese Variablen sind gemeinsam statistisch signifikant → mindestens eine Variable ist von null verschieden.

18. Testen Sie im Modell 4, ob die Preiselastizität **-1** entspricht.

Die Preiselastizität ist das Verhältnis der prozentualen Änderung des Rosenabsatzes zur prozentualen Veränderung des Rosenpreises.

H_0: $b_2 = -1$ $\qquad$ H_1: $b_2 \neq -1$

$t_e = (-1.1707 + 1)/0.4883 = -0.349$

$|t_e| < 2$ → H_0 kann nicht verworfen werden

Die Daten sind mit einer Preiselastizität von 1 kompatibel → Wenn der Rosenpreis um 1% steigt, bildet sich der Rosenabsatz im Durchschnitt um 1% zurück, ceteris paribus.

19. Testen Sie im Modell 4, ob die Kreuzpreiselastizität **1** entspricht.

Die Kreuzpreiselastizität ist das Verhältnis der prozentualen Änderung des Rosenabsatzes zur prozentualen Veränderung des Nelkenpreises.

H_0: $b_3 = 1$ $\qquad$ H_1: $b_3 \neq 1$

$t_e = (0.7379 - 1)/0.6528 = -0.4015$

$|t_e| < 2$ → H_0 kann nicht verworfen werden

Die Daten sind mit einer Kreuzpreiselastizität von 1 kompatibel → Wenn der Nelkenpreis um 1% steigt, erhöht sich der Rosenabsatz im Durchschnitt um 1%, ceteris paribus.