Linköpings universitet
Linda Wänström
IDA/Statistik

Tentamen Statistik B, 732G71

2012-12-11

Skrivtid: 08.00-12.00
Tillåtna hjälpmedel: *Miniräknare.* *Formelsamling* (får innehålla markeringar och understrykningar men inte anteckningar). *Kursbok:* Bowerman, O'Connel, Koehler: Forecasting, Time series, and Regression (alla upplagor tillåtna - får innehålla markeringar, understrykningar och flärpar, men inte anteckningar).
Betyg: För godkänt betyg krävs 12 av 20 poäng. För väl godkänt betyg krävs 16 av 20 poäng.
Jourhavande lärare: Linda Wänström

Redovisa och motivera kort alla dina lösningar. Lycka till!

Uppgift 1.

Nedan följer delar av en MINITAB-utskrift från en enkel linjär regressionsanalys där modellen $y = \beta_0 + \beta_1 x + \epsilon$ har skattats, där y =pris (1000-tals kr) och x =storlek (i kvm) för lägenheter.

```
The regression equation is
Pris = - 397 + 45,9 Kvm

Predictor     Coef  SE Coef      T
Constant    -397,1    324,5  -1,22
Kvm         45,900    3,676  12,49

S = 448,545   R-Sq = 84,8%   R-Sq(adj) = 84,2%

Analysis of Variance

Source          DF        SS        MS       F
Regression       1  31368865  31368865  155,91
Residual Error  28   5633393    201193
Total           29  37002258
```

a) Beräkna ett 99%-igt konfidensintervall för lutningskoefficienten (β_1). Tolka intervallet. (2 p)
b) Utifrån ditt beräknade intervall i a), skulle du förkasta en nollhypotes att $\beta_1 = 0$? Motivera. (1 p)
c) Vad är korrelationskoefficienten mellan pris och storlek? (1 p)

Uppgift 2.

Nedan redovisas den totala vinsten per år (i milj. kr) för ett visst företag.

År	2007	2008	2009	2010	2011
Vinst	1.05	1.10	1.20	1.60	2.10

a) Antag att du vill skatta den genomsnittliga årliga procentuella vinstökningen med hjälp av en lämplig regressionsmodell. Ställ upp modellen. (1 p)
b) Skatta den genomsnittliga årliga procentuella vinstökningen för företaget. (3 p)
c) Gör en prognos för vinsten år 2012. (1 p)

Uppgift 3.

Antag att du har skattat följande modell på kvartalsdata för tidsperioden första kvartalet år 2001 till sista kvartalet år 2008:

$$y_t = \beta_0 + \beta_1 t + \beta_2 K_1 + \beta_3 K_2 + \beta_4 K_3 + \epsilon_t$$

där
y_t är avkastningen för en viss aktie (i 100 kr) vid tidpunkt t
t är tidpunkten och är kodad $t = 1, 2, ..., 32$
$K_1 = 1$ om kvartal 1; 0 annars
$K_2 = 1$ om kvartal 2; 0 annars
$K_3 = 1$ om kvartal 3; 0 annars

```
The regression equation is
y = 1,04 + 0,0368 t - 1,04 K1 - 0,726 K2 - 0,911 K3

Predictor       Coef   SE Coef       T      P
Constant     1,04190   0,08746   11,91  0,000
t           0,036804  0,003404   10,81  0,000
K1          -1,04459   0,08884  -11,76  0,000
K2          -0,72639   0,08851   -8,21  0,000
K3          -0,91132   0,08832  -10,32  0,000

S = 0,176504   R-Sq = 92,1%   R-Sq(adj) = 90,9%

Analysis of Variance

Source          DF       SS      MS      F      P
Regression       4   9,7685  2,4421  78,39  0,000
Residual Error  27   0,8412  0,0312
Total           31  10,6097

Source  DF  Seq SS
t        1  4,6145
K1       1  1,4536
K2       1  0,3834
K3       1  3,3171
```

a) Vad har du gjort för antagande om eventuell säsongvariation (enligt modellen du har valt)? Motivera. (1 p)

b) Gör ett lämpligt test för att testa om det finns säsongvariation i avkastningen. Dra slutsats från testet. (3 p)
c) Gör prognoser för alla fyra kvartal år 2009. (2 p)
d) En kollega till dig påstår att du har problem med multikolinjäritet. Hon säger att enligt din utskrift är $R^2 = 0.921$ och därför är $VIF = 1/(1-0.921) = 12.66$. Håller du med henne? Motivera. (1 p)

Som hjälp till uppgift 3 finns också följande MINITAB-utskrift:

```
The regression equation is
y = 0,300 + 0,0411 t


Predictor        Coef   SE Coef     T      P
Constant       0,3000    0,1618  1,85  0,074
t            0,041128  0,008559  4,81  0,000


S = 0,447034    R-Sq = 43,5%    R-Sq(adj) = 41,6%


Analysis of Variance

Source          DF       SS      MS      F      P
Regression       1   4,6145  4,6145  23,09  0,000
Residual Error  30   5,9952  0,1998
Total           31  10,6097
```

Uppgift 4.

En handlare säljer grönsaker i en saluhall. Nedan följer genomsnittliga priser och sålda kvantiteter under tre tertial.

		Morötter	Potatis	Palsternacka
Aug-Nov (tertial 1)	Pris (kr per kg)	10.2	6.7	17.4
	kg	150	210	70
Dec-Mars (tertial 2)	Pris (kr per kg)	14.1	7.9	25.8
	kg	110	220	30
April-Jul (tertial 3)	Pris (kr per kg)	15.7	8.1	27.3
	kg	120	230	20

a) Beräkna ett lämpligt sammansatt index som beskriver prisutvecklingen under året. Motivera valet av index. (3 p)
b) Hur stor har prisökningen varit mellan första och tredje tertialet? (1 p)

Formelsamling 2010-10-26

Enkel linjär regressionsanalys:

Modell:

$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i \qquad (= \alpha + \beta \cdot x_i + \varepsilon_i)$

där $\varepsilon \sim N(0, \sigma)$.

Anpassad regressionslinje:

$\hat{y} = b_0 + b_1 \cdot x \qquad (= a + b \cdot x_i)$

där

$$b_1 (= b) = \frac{\sum(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = \frac{\sum x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sum x_i^2 - n \cdot (\bar{x})^2} =$$

$$= \frac{\sum x_i \cdot y_i - \frac{(\sum x_i) \cdot (\sum y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{n \cdot \sum x_i \cdot y_i - (\sum x_i) \cdot (\sum y_i)}{n \cdot \sum x_i^2 - (\sum x_i)^2}$$

$b_0 (= a) = \bar{y} - b_1 \cdot \bar{x}$

Kvadratsummor:

Total: $SST = SS_{yy} = (n-1) \cdot s_y^2 = \sum(y_i - \bar{y})^2 = \sum y_i^2 - n \cdot (\bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

$SS_{xx} = (n-1) \cdot s_x^2 = \sum(x_i - \bar{x})^2 = \sum x_i^2 - n \cdot (\bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$

$SS_{xy} = \sum(x_i - \bar{x}) \cdot (y_i - \bar{y}) = \sum x_i \cdot y_i - n \cdot (\bar{x}) \cdot (\bar{y}) = \sum x_i \cdot y_i - \frac{(\sum x_i) \cdot (\sum y_i)}{n}$

Residual: $SSE = \sum(y_i - \hat{y}_i)^2 = SS_{yy} - b_1 \cdot SS_{xy} = \sum(y_i - \bar{y})^2 - b_1 \cdot \sum(x_i - \bar{x}) \cdot (y_i - \bar{y}) = \sum y_i^2 - b_0 \cdot \sum y_i - b_1 \cdot \sum x_i \cdot y_i$

Regression: $SSR = \sum(\hat{y}_i - \bar{y})^2 = SST - SSE$

Variansskattning

$$\widehat{\sigma^2} = s^2 = s_e^2 = MSE = \frac{SSE}{n-2} = \frac{\sum(y_i - \hat{y}_i)^2}{n-2}$$

$$s = s_e = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}}$$

Förklaringsgrad:

$$r^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{SSE}{SST}$$

Korrelationskoefficient:

$$r = \frac{\sum(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}} = \frac{\sum x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{\sqrt{(\sum x_i^2 - n \cdot (\bar{x})^2) \cdot (\sum y_i^2 - n \cdot (\bar{y})^2)}} =$$

$$= \frac{\sum x_i \cdot y_i - \frac{(\sum x_i) \cdot (\sum y_i)}{n}}{\sqrt{(\sum x_i^2 - \frac{(\sum x_i)^2}{n}) \cdot (\sum y_i^2 - \frac{(\sum y_i)^2}{n})}} = \frac{n \cdot \sum x_i \cdot y_i - (\sum x_i) \cdot (\sum y_i)}{\sqrt{(n \cdot \sum x_i^2 - (\sum x_i)^2) \cdot (n \cdot \sum y_i^2 - (\sum y_i)^2)}}$$

Konfidensintervall, prognosintervall och hypotesprövning

Stickprovsfördelningar:

$$b_1 \sim N\left(\beta_1, \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2}}\right)$$

$$b_0 \sim N\left(\beta_0, \sigma \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\sum(x_i - \bar{x})^2}}\right)$$

$$b_0 + b_1 \cdot x_0 \sim N\left(\beta_0 + \beta_1 \cdot x_0, \sigma \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_i - \bar{x})^2}}\right)$$

Konfidensintervall för β_1*:*

$$b_1 \pm t_{[\alpha/2]}(n-2) \cdot \frac{s}{\sqrt{\sum(x_i-\overline{x})^2}} \qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

Konfidensintervall för β_0*:*

$$b_0 \pm t_{[\alpha/2]}(n-2) \cdot s \cdot \sqrt{\left(\frac{1}{n} + \frac{(\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)} \qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

Konfidensintervall för $\mu_{y|x_0} = \beta_0 + \beta_1 \cdot x_0$*:*

$$b_0 + b_1 \cdot x_0 \pm t_{[\alpha/2]}(n-2) \cdot s \cdot \sqrt{\left(\frac{1}{n} + \frac{(x_0-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)} \qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

Prognosintervall för $y_0 = \beta_0 + \beta_1 \cdot x_0 + \varepsilon_0$*:*

$$b_0 + b_1 \cdot x_0 \pm t_{[\alpha/2]}(n-2) \cdot s \cdot \sqrt{\left(1 + \frac{1}{n} + \frac{(x_0-\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)} \qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$$

Formellt t-test av $H_0 : \beta_0 = 0$*:*

Testfunktion: $t = \frac{b_0}{s_{b_0}} = \frac{b_0}{s \cdot \sqrt{\left(\frac{1}{n} + \frac{(\overline{x})^2}{\sum(x_i-\overline{x})^2}\right)}}$ $\qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$

Jämför med $\pm t_{[\alpha/2]}(n-2)$

Formellt t-test av $H_0 : \beta_1 = 0$ *dvs inget samband mellan y och x:*

Testfunktion: $t = \frac{b_1}{s_{b_1}} = \frac{b_1}{\frac{s}{\sqrt{\sum(x_i-\overline{x})^2}}}$ $\qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$

Jämför med $\pm t_{[\alpha/2]}(n-2)$

Formellt t-test av $H_0 : \beta_1 = B$ *(där B är något annat än 0):*

Testfunktion: $t = \frac{b_1 - B}{s_{b_1}} = \frac{b_1 - B}{\frac{s}{\sqrt{\sum(x_i-\overline{x})^2}}}$ $\qquad \left(\sum(x_i-\overline{x})^2 = \sum x_i^2 - n\cdot(\overline{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)$

Jämför med $\pm t_{[\alpha/2]}(n-2)$

Vid enkelsidiga mothypotseser jämförs t med $t_{[\alpha]}(n-2)$ (eller med -$t_{[\alpha]}(n-2)$ beroende på mothypotesens riktning).

Formellt F-test av $H_0 : \beta_1 = 0$*:*

Testfunktion: $F = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}$

Jämför med $F_{[\alpha]}(1, n-2)$

Multipel linjär regressionsanalys:

Modell:

$y_i = \beta_0 + \beta_1 \cdot x_{i1} + \beta_2 \cdot x_{i2} + \ldots + \beta_k x_{ik} + \varepsilon_i$

där $\varepsilon_i \sim N(0, \sigma)$.

Anpassad modell:

$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_k \cdot x_k$

Kvadratsummor:

$SST = SSE + SSR$

Total: $SST = (n-1) \cdot s_y^2 = \sum(y_i - \bar{y})^2 = \sum y_i^2 - n \cdot (\bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$

Residual: $SSE = \sum(y_i - \hat{y}_i)^2$

Regression: $SSR = \sum(\hat{y}_i - \bar{y})^2 = SST - SSE$

SSE har $n-k-1$ frihetsgrader, SSR har k frihetsgrader.

Variansskattning:

$\widehat{\sigma^2} = s^2 = s_e^2 = MSE = \frac{SSE}{n-k-1}$

Förklaringsgrad:

$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

Justerad förklaringsgrad:

$R_{adj}^2 = \bar{R}^2 = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)} = 1 - \frac{\sum(y_i - \hat{y}_i)^2/(n-k-1)}{\sum(y_i - \bar{y}_i)^2/(n-1)} = 1 - \frac{s_e^2}{s_y^2}$

Konfidensintervall och hypotesprövning

Stickprovsfördelningar:

$b_j \sim N(\beta_j, \sigma_{b_j})$

Formellt F-test av $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$:

Testfunktion: $F = \frac{MSR}{MSE} = \frac{SSR/k}{SSE/(n-k-1)}$

Jämför med $F_{[\alpha]}(k, n-k-1)$

Konfidensintervall för β_j:

$b_j \pm t_{[\alpha/2]}(n-k-1) \cdot s_{b_j}$

där s_{b_j} hämtas från datorutskrift.

Formellt t-test av $H_0 : \beta_j = 0$:

Testfuktion: $t = \frac{b_j}{s_{b_j}}$

Jämför med $t_{[\alpha/2]}(n-k-1)$

Konfidensintervall för $\mu_{y|x_{01},\ldots,x_{0k}}$:

$\hat{y}_0 \pm t_{[\alpha/2]}(n-k-1) \cdot s\sqrt{\text{Distance value}}$

där $s = \sqrt{MSE}$ och "Distance value" (eller $s \cdot \sqrt{\text{Distance value}}$) bestäms från datorutskrift.

Prognosintervall för y_0:

$\hat{y}_0 \pm t_{[\alpha/2]}(n-k-1) \cdot s\sqrt{1 + \text{Distance value}}$

där $s = \sqrt{MSE}$ och "Distance value" (eller $s \cdot \sqrt{1 + \text{Distance value}}$) bestäms från datorutskrift.

Partiellt F-test av $H_0 : \beta_{g+1} = \ldots = \beta_k = 0$:

Testfunktion: $F = \dfrac{(SSE_R - SSE_C)/(k-g)}{SSE_C/(n-k-1)} = \dfrac{(SSR_C - SSR_R)/(k-g)}{SSE_C/(n-k-1)}$

där

SSE_R =Residualkvadratsumman i den mindre (reducerade) modellen,
SSE_C =Residualkvadratsumman in den större (kompletta) modellen,
SSR_R =Regressionskvadratsumman i den mindre (reducerade) modellen,
SSR_C =Regressionskvadratsumman in den större (kompletta) modellen, och
$k-g$=skillnaden i antal förklaringsvariabler mellan modellerna.

Jämför med $F_{[\alpha]}(k-g, n-k-1)$.

Alternativ formel: $F = \dfrac{(R^2_{UR} - R^2_R)/r}{(1-R^2_{UR})/(n-k-1)}$

där R^2_{UR} =Förklaringsgraden i den större (kompletta, "unrestricted") modellen och R^2_R =Förklaringsgraden i den mindre (reducerade, "restricted") modellen och $r = k-g$

Jämför med $F_{[\alpha]}(r, n-k-1) = F_{[\alpha]}(k-g, n-k-1)$.

Variance Inflation Factor (VIF):

$VIF = \dfrac{1}{1-R_j^2}$

där R_j^2 =Förklaringsgraden i modell där x_j är y-variabel och övriga x-variabler är förklaringsvariabler.

Sekventiella kvadratsummor:

$SSR = SSR(x_1) + SSR(x_2|x_1) + \ldots + SSR(x_k|x_1, \ldots, x_{k-1})$

där $SSR(x_j|x_1, \ldots, x_{j-1})$ är tillskottet till SSR då variabel x_j läggs till en modell med variablerna $x_1, x_2, \ldots, x_{j-1}$.

Ett partiellt F-test av $H_0 : \beta_{g+1} = \ldots = \beta_k = 0$ kan då göras med testfunktionen

$F = \dfrac{(SSR(x_{g+1}|x_1, \ldots, x_g) + SSR(x_{g+2}|x_1, \ldots, x_{g+1}) + \ldots + SSR(x_k|x_1, \ldots, x_{k-1}))/(k-g)}{MSE}$, Jämför med $F_{[\alpha]}(k-g, n-k-1)$

förutsatt att variablerna matas in i ordningen $x_1, x_2, \ldots, x_k$ i modellen.

Exponentiella samband och elasticitetsmodeller:

Logaritmbeteckningar: $\lg x$ betyder 10-logaritmen av x, $\log x$ står för logaritm och man kan välja om man vill använda $\lg x$ eller $\ln x$ (den naturliga logaritmen). Samma sorts logaritm måste användas genomgående i en och samma analys.

Exponentiell modell: $y = \beta_0 \cdot (\beta_1)^x \cdot \delta$

där $\log \delta \sim N(0, \sigma)$

$\log y = \log \beta_0 + (\log \beta_1) \cdot x + \log \delta$

Anpassad modell: $\hat{y} = b_0 \cdot (b_1)^x$

där

$$\log b_1 = \frac{\sum(x_i - \bar{x}) \cdot (\log y_i - \overline{\log y})}{\sum(x_i - \bar{x})^2} = \frac{\sum x_i \cdot \log y_i - n \cdot \bar{x} \cdot \overline{\log y}}{\sum x_i^2 - n \cdot (\bar{x})^2} =$$

$$= \frac{\sum x_i \cdot \log y_i - \frac{(\sum x_i) \cdot (\sum \log y_i)}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}$$

och $\log b_0 = \overline{\log y} - (\log b_1) \cdot \bar{x}$ $\quad [\overline{\log y} = \frac{1}{n} \sum \log y_i]$

Kvadratsummor, variansskattning och test:

$SST = \sum(\log y_i - \overline{\log y})^2 = \sum(\log y_i)^2 - n \cdot (\overline{\log y})^2$

$SSE = SST - (\log b_1) \cdot \sum(x_i - \overline{x}) \cdot (\log y_i - \overline{\log y}) = SST - (\log b_1) \cdot (\sum x_i \cdot \log y_i - n \cdot \overline{x} \cdot \overline{\log y}) =$

$\sum(\log y_i)^2 - (\log b_0) \cdot \sum \log y_i - (\log b_1) \cdot \sum x_i \cdot \log y_i$

$\widehat{\sigma^2} = \dfrac{SSE}{n-2}$

Test av $H_0 : \beta_1 = 1$ dvs inget samband mellan y och $x \Longleftrightarrow \log \beta_1 = 0$:

Testfunktion $t = \dfrac{\log b_1}{\sqrt{\frac{SSE/(n-2)}{\sum(x_i - \overline{x})^2}}}$, jämför med $t_{[\alpha/2]}(n-2)$

Elasticitetsmodeller:

Formler enligt AJÅ:

x_1=Pris, x_2=Inkomst

Modeller:

$\hat{y} = a \cdot x_1^e, \quad \hat{y} = a \cdot x_2^E, \quad \hat{y} = a \cdot x_1^e \cdot x_2^E$

e =priselasticitet, E =inkomstelasticitet

Anpassning av t.ex. $\hat{y} = a \cdot x_1^e$:

$\lg \hat{y} = a' + e \cdot \lg x_1, \quad a' = \lg a$

$e = \dfrac{n \cdot \sum(\lg y) \cdot (\lg x_1) - (\sum \lg y) \cdot (\sum \lg x_1)}{n \cdot \sum(\lg x_1)^2 - (\sum \lg x_1)^2}$

$SST = \sum(\lg y - \overline{\lg y})^2 = \sum(\lg y)^2 - \frac{\cdot(\sum \lg y)^2}{n}$

$SSE = SST - e \cdot \sum(\lg x_1 - \overline{\lg x}) \cdot (\lg y - \overline{\lg y}) = \sum(\lg y)^2 - a' \cdot \sum \lg y - e \cdot \sum(\lg x_1) \cdot (\lg y)$

$\widehat{\sigma^2} = \dfrac{SSE}{n-2} \qquad [\overline{\lg x} = \frac{1}{n} \sum \lg x_i$ och $\overline{\lg y} = \frac{1}{n} \sum \lg y_i]$

Test av H_0 : priselasticiteten $= B$ där B är ett ifrågasatt värde på priselasticiteten:

Testfunktion $t = \dfrac{e - B}{\sqrt{\frac{SSE/(n-2)}{\sum(\lg x_1 - \overline{\lg x_1})^2}}}$, jämför med $t_{[\alpha/2]}(n-2)$ och vid enkelsidig mothypotes med $t_{[\alpha]}^{(n-2)}$ eller $-t_{[\alpha]}^{(n-2)}$.

Formler enligt Mikroekonomin, Fö-anteckningar och datorövningar:

$Q = C \cdot (P)^{E_P} \cdot \delta, \; Q = C \cdot (I)^{E_I} \cdot \delta$

$Q = C \cdot (P)^{E_P} \cdot (I)^{E_I} \cdot \delta$

$\log Q = \log C + E_P \cdot \log P + \log \delta$

$\log Q = \log C + E_I \cdot \log I + \log \delta$

$\log Q = \log C + E_P \cdot \log P + E_I \cdot \log I + \log \delta$

där $\log \delta \sim N(0, \sigma)$

Exempel på anpassad modell: $\widehat{Q} = c \cdot (P)^{\widehat{E_P}}$, där $\widehat{E_P} = \dfrac{\sum(\log P_i - \overline{\log P}) \cdot (\log Q_i - \overline{\log Q})}{\sum(\log P_i - \overline{\log P})^2} =$

$= \dfrac{\sum(\log P_i) \cdot (\log Q_i) - n \cdot \overline{\log P} \cdot \overline{\log Q}}{\sum(\log P_i)^2 - n \cdot (\overline{\log P})^2}$ och

$\log c = \overline{\log Q} - \widehat{E_P} \cdot \overline{\log P}$ $\quad\quad [\overline{\log P} = \frac{1}{n}\sum \log P_i$ och $\overline{\log Q} = \frac{1}{n}\sum \log Q_i]$

Kvadratsummor, variansskattning och test:

$SST = \sum(\log Q_i - \overline{\log Q})^2 = \sum(\log Q_i)^2 - n \cdot (\overline{\log Q})^2$

$SSE = SST - \widehat{E_P} \cdot \sum(\log P_i - \overline{\log P}) \cdot (\log Q_i - \overline{\log Q}) = SST - \widehat{E_P} \cdot [\sum(\log P_i) \cdot (\log Q_i) - n \cdot \overline{\log P} \cdot \overline{\log Q}] =$
$= \sum(\log Q_i)^2 - (\log c) \cdot \sum \log Q_i - \widehat{E_P} \cdot \sum(\log P_i) \cdot (\log Q_i)$

$\widehat{\sigma^2} = \frac{SSE}{n-2}$

Test av $H_0 : E_P = B$ där B är ett ifrågasatt värde på E_P:

Testfunktion $t = \frac{\widehat{E_P} - B}{\sqrt{\frac{SSE/(n-2)}{\sum(\log P_i - \overline{\log P})^2}}}$, jämför med $t_{[\alpha/2]}(n-2)$ och vid enkelsidig mothypotes med $t_{[\alpha]}^{(n-2)}$ eller $-t_{[\alpha]}^{(n-2)}$.

Index

Sammansatta fastbasindex:

$I_t = i_{1,t} \cdot w_1 + i_{2,t} \cdot w_2 + \ldots + i_{n,t} \cdot w_n$

där n är antalet ingående varor/tjänster, $i_{1,t}, \ldots, i_{n,t}$ är enkla prisindex för ingående varor, alla med basår t_0 och $w_1, \ldots, w_n$ väljs enligt ett viktsystem:

Laspeyre: $w_i = \frac{p_{i,t_0} \cdot q_{i,t_0}}{\sum_j p_{j,t_0} \cdot q_{j,t_0}}$

Paasche: $w_i = \frac{p_{i,t_0} \cdot q_{i,t}}{\sum_j p_{j,t_0} \cdot q_{j,t}}$

Kedjeprisindex:

$I_t = L_{0,1} \cdot L_{1,2} \cdot \ldots \cdot L_{t-1,t} \cdot 100$

där

$$L_{t-1,t} = \sum_{i=1}^{n} \frac{p_{i,t}}{p_{i,t-1}} \cdot w_{i,t-1,t}$$

är årslänken från år $t-1$ till t för n ingående varor/tjänster. $w_{i,t-1,t}$ väljs enligt ett viktsystem:

Laspeyre: $w_{i,t-1,t}^{\mathcal{L}} = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t-1}{\text{Totala försäljningsvärdet år } t-1}$

Paasche: $w_{i,t-1,t}^{\mathcal{P}} = \frac{\text{Försäljningsvärdet för vara } i \text{ år } t \text{ i priser för år } t-1}{\text{Totala försäljningsvärdet år } t \text{ i priser för år } t-1}$

Med representantvaror byts "Försäljningsvärdet för vara i" mot "Försäljningsvärdet för varugrupp i" i vikterna.

Implicitprisindex:

$I_t = \frac{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i löpande priser}}{\text{Försäljningsvärdet av varan/tjänsten/gruppen år } t \text{ i basårets priser}} \cdot 100$

Relativprisindex:

$I_t^R = \frac{I_t^v}{I_t^0} \cdot 100$

där I_t^v =Prisindex för aktuell vara/tjänst/grupp och I_t^0 =Prisindex för den större jämförelsegruppen, t ex KPI.

Tidsserieanalys

Tidsserieregression:

Modell:

$y_t = TR_t + SN_t + \varepsilon_t$

där

$TR_t = \beta_0 + \beta_1 \cdot t$ eller $TR_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2$

och

$SN_t = \sum_{i=1}^{L-1} \beta_{si} \cdot x_{si,t}$

med

$L =$Antal säsonger och $x_{si,t} = 1$ om t tillhör säsong i och $= 0$ annars.

Durbin-Watson's test:

Test av H_0 : Residualerna är okorrelerade.

Testfunktion $d = \dfrac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$

där $e_t = y_t - \hat{y}_t$.

Jämförelser:

Om $d < 1 \Rightarrow$ Förkasta H_0, positiv seriell korrelation
Om $d > 3 \Rightarrow$ Förkasta H_0, positiv seriell korrelation
Om $1 \leq d \leq 3 \Rightarrow H_0$ kan ej förkastas.

Komponentuppdelning:

Modeller:

Multiplikativ modell: $y_t = TR_t \cdot SN_t \cdot CL_t \cdot IR_t$

Additiv modell: $y_t = TR_t + SN_t + CL_t + IR_t$

Enkel exponentiell utjämning:

Modell: $y_t = \mu + \varepsilon_t$

Uppdateringsschema för skattning av $\mu : S_t = \alpha \cdot y_t + (1 - \alpha) \cdot S_{t-1}$ $\quad 0 < \alpha < 1$

Prognos: $\hat{y}_{t+\tau} = S_t$

Prognosintervall: $S_t \pm z \cdot s \cdot \sqrt{1 + \alpha^2}$

där $z =1.96$ för 95% intervall, 2.576 för 99% intervall och

$s = \sqrt{\frac{1}{n-1} \cdot \sum_{t=1}^{n} (y_t - \bar{y})^2}$