Viðskipta- og Hagfræðideild **Tölfræði II, 04.51.09-960**
Háskóli Íslands **Skriflegt próf**

6. desember 2006, 9.00-12.00. Leyfileg hjálpargögn: Skriffæri, strokleður

Rökstyðjið öll svör og gerið grein fyrir forsendum. Þar sem til teljið vanta forsendur gefið ykkur það sem þið teljið nauðsynlegt.

1. (10%) Gert er ráð fyrir að sambandi Y við breyturnar X_1 og X_2 sé á forminu:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

Áhugi er á að meta þýðingu X_1 breytunnar fyrir Y. Gögn er fengin með random úrtaki, n mælingar á (Y, X), þ.e. $(x_1, y_1), \ldots, (x_n, y_n)$. Stikinn (parameterinn) β_1 er metinn með:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_{i1} - \bar{x}_1)}{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}$$

Hvað segir $\hat{\beta}_1$ þetta um þýðingu X_1 breytunnar fyrir Y? Getur $\hat{\beta}_1$ verið „unbiased" metill (estimator) fyrir β_1?

2. (10%) Tvívíða hendingin (X, Y) hefur eftirfarandi líkindamassafall (pmf):

$$f(x, y) = P(X = x, Y = y) = \binom{2}{x}\binom{2-x}{y} p_x^x p_y^y (1 - p_x - p_y)^{2-x-y}$$
$$x = 0, 1, 2 \quad y = 0, 1, 2, \quad 0 \leq x + y \leq 2, \quad \binom{n}{k} = n!/(k!(n-k)!)$$

a) Setjið upp tvívíðu dreifinguna. b) Finnið jaðardreifingu (marginal distribution) X, $f(x) = P(X = x)$, c) Finnið skilyrtu dreifinguna $f(y|x)$ d) Hefur jaðardreifingin nafn? e) Hefur skilyrta dreifingin nafn? f) Finnið regression-fallið $E(Y|X = x)$ g) Er hægt að giska á $V(Y|X = x)$?

3. (10%) Hending X hefur þéttifall:

$$f(x) = cx^{\alpha-1}(1 - x) \quad \alpha > 0, 0 < x < 1$$

Finnið fastann c. Finnið dreifingu $Y = 1/(1 - X)$.

4. (10%) Gefið að $X_1, \ldots, X_n$ eru óháðar einsdreifðar með þéttifall:

$$f(x) = cx^{\alpha-1}(1 - x) \quad \alpha > 0, 0 < x < 1$$

Leiðið út MM (method-of-moments) metil fyrir α.

5. (10%) Hvert er innihald central-limit-theorem. Segið frá skrefum í sönnun á central-limit-theorem.

6. (10%) Hugleiðið eftirfarandi kenningaprófun.

$$H_0 : \theta = 0$$
$$X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1}$$

Stingið upp á og rökstyðjið aðferð hér.

7. (10%) Gerið ráð fyrir að hermuð séu eftirfarandi ferli:

$$X_t = X_{t-1} + \varepsilon_t$$
$$Y_t = Y_{t-1} + \epsilon_t \quad \varepsilon_t \text{ og } \epsilon_t \text{ óháðir white-noise } t = 1, T$$

Síðan er reiknaður fylgnistuðull $r = r_{XY}$ og t-gildi hans $t = \frac{r\sqrt{T-2}}{\sqrt{1-r^2}}$. Þetta er framkvæmt 100 sinnum fyrir T=50, 200 og 1000. Lýsið, t.d. með myndum og orðum, hvers konar dreifingu má vænta af t-gildunum? Hvað segir þetta um svona fylgnireikning?

8. (10%) Gefið random úrtak $(X_1, \ldots, X_n)$ úr eftirfarandi dreifingu:

$$f(x|\theta) = \theta e^{-\theta x}, \quad \theta > 0, \quad x > 0$$

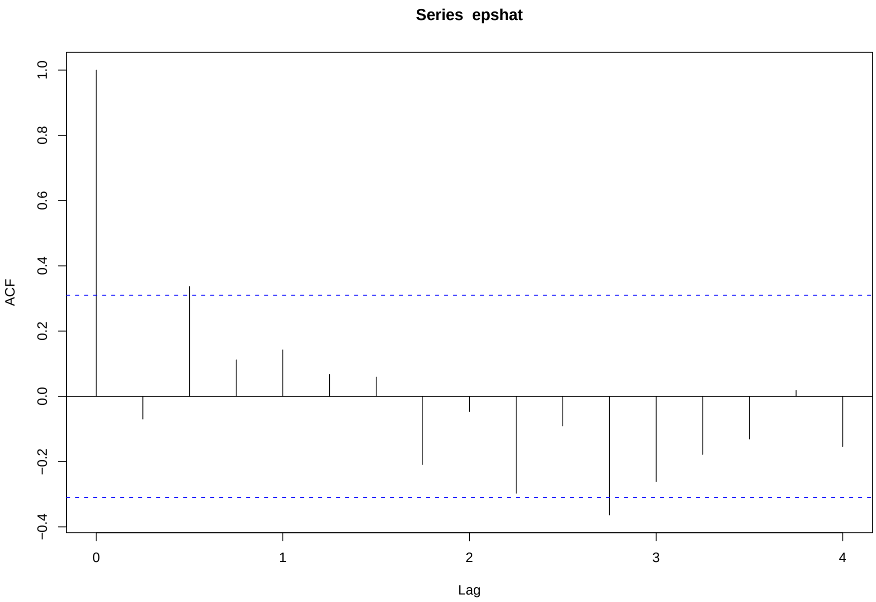
Finnið ML-metil fyrir θ. Finnið mörk á hversu nákvæmlega er hægt að álykta um θ. Út á hvað gengur ML? Hverjir eru helstu eiginleikar ML og hvers vegna er ML eftirsóknarverð aðferð? (ML=maximum-likelihood=mesti sennileiki)

9. (10%) Skýrið hugtakið ECM líkan. Af hverju er ECM líkan sett fram?

10. (10%) Meta á langtímasamband tekna (Y) og neyslu (C). Meðal þess sem er skoðað eru model 5 og model 3 í töflunum hér að neðan. Metnir afgangsliðir úr líkaninu í model 5 er teiknaðir í mynd 1 og metið sjálffylgnifall er sýnt í mynd 2. Túlkið útkomurnar. Hvers vegna var model 3 metið? Hvers konar ferli gæti verið að baki afgangsliðunum úr model 5? Prófið kenninguna stikarnir (parametrar) d_1_C_1 og d_1_Y_1 séu samtímis 0.

    Gögnin eru ársfjórðungsgögn. Hér var tekinn logaritmi af breytunum Y og C (af hverju?). Breyturnar dq2,dq3 og dq4 eru 0/1 dummy breytur, dqj=1 ef ársfjórðungur=j og 0 annars. l_C stendur fyrir logaritma neyslu, l_C_1 er l_C tafin um eitt tímabil og d_l_C er breyting á logaritma neyslu milli tímabila og d_l_C_1 breyting tafin um eitt tímabil. Samsvarandi er gert við Y-breytuna.

Mynd 1: Metnir afgangsliðir úr model 5.



Mynd 2: Correlogram fyrir metna afgangsliði úr model 5.

Model 5: OLS estimates using the 40 observations 1975:1–1984:4
Dependent variable: d_l_C

Variable	Coefficient	Std. Error	t-statistic	p-value
const	0.241409	0.498542	0.4842	0.6314
d_l_Y	0.351530	0.146346	2.4021	0.0221
dq2	0.0607805	0.0139721	4.3501	0.0001
dq3	0.0762504	0.0128031	5.9556	0.0000
dq4	0.0983064	0.0108668	9.0465	0.0000
l_C_1	−0.427595	0.152467	−2.8045	0.0084
l_Y_1	0.395603	0.140032	2.8251	0.0080

Mean of dependent variable	0.00408217
S.D. of dependent variable	0.0518630
Sum of squared residuals	0.00940448
Standard error of residuals ($\hat{\sigma}$)	0.0168815
Unadjusted R^2	0.910349
Adjusted $\bar{R}^2$	0.894049
$F(6, 33)$	55.8491
Durbin–Watson statistic	2.00929
First-order autocorrelation coeff.	−0.0701989
Log-likelihood	110.351

Model 3: OLS estimates using the 40 observations 1975:1–1984:4
Dependent variable: d_l_C

Variable	Coefficient	Std. Error	t-statistic	p-value
const	0.0149497	0.448848	0.0333	0.9736
d_l_Y	0.271974	0.134327	2.0247	0.0516
d_l_Y_1	0.446794	0.134573	3.3201	0.0023
l_C_1	−0.0854443	0.167921	−0.5088	0.6145
l_Y_1	0.0779438	0.153051	0.5093	0.6142
dq2	0.0453691	0.0176474	2.5709	0.0152
dq3	0.0864898	0.0114277	7.5685	0.0000
dq4	0.105960	0.00966811	10.9597	0.0000
d_l_C_1	−0.416507	0.165384	−2.5184	0.0172

Mean of dependent variable	0.00408217
S.D. of dependent variable	0.0518630
Sum of squared residuals	0.00660522
Standard error of residuals ($\hat{\sigma}$)	0.0145970
Unadjusted R^2	0.937034
Adjusted $\bar{R}^2$	0.920785
$F(8, 31)$	57.6660
Durbin–Watson statistic	1.86471
First-order autocorrelation coeff.	0.00253799
Log-likelihood	117.418