Problem sets II for ECON 4150, Spring 09

Problem set 6
Solve the exercises: 6.6, 6.9, 6.11, 7.14

Problem set 7
Solve exercise 6.15, review the log-normal distribution in combination with appendix 7A, and solve the following exercise:

Exercise 7.1
If we consider the regression equation

(1) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1,2,.....,n$

it will often happen that the disturbances $\{\varepsilon_i\}$ are heteroskedastic. In order to detect heteroskedasticity of the random disturbances textbooks will often recommend various kinds of residual plots. The idea is that the residuals mimic properties of the disturbances. But one has to be a bit careful. The residuals $\{\hat{\varepsilon}_i\}$ are defined by

(2) $\hat{\varepsilon}_i = Y_i - \hat{Y}_i = -(\hat{\beta}_1 - \beta_1)(X_i - \overline{X}) + (\varepsilon_i - \overline{\varepsilon})$

where

(3) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

(i) Deduce the last equality of (2).

Assume that the disturbances are homoskedastic, i.e. $Var(\varepsilon_i) = \sigma^2$ for all i.

Show that the following expressions hold:

(ii) $E(\hat{\varepsilon}_i) = 0$

(iii) $Var(\hat{\varepsilon}_i) = E(\hat{\varepsilon}_i^2) = \sigma^2 (1 - \frac{1}{n} - \frac{(X_i - \overline{X})^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2})$

(iv) Explain why the residuals $\{\hat{\varepsilon}_i\}$ must be correlated and then show that

(v) $\mathrm{Cov}(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = -\sigma^2 (\frac{1}{n} + \frac{(X_i - \overline{X})(X_j - \overline{X})}{\sum_{j=1}^{n} (X_j - \overline{X})^2})$

Hence, even if the disturbances are homoskedastic the variance of $\{\hat{\varepsilon}_i\}$ depends on the sample values of the explanatory variable X. This can make it difficult to interpret a graph in which one has plotted X_i against $\hat{\varepsilon}_i$. For this reason one will often normalise the residuals so that all residuals get the same variance. For example one could define

(4) $$\hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i}{s\sqrt{(1-v_{ii})}}$$

where $s^2 = \dfrac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{(n-2)}$ and $v_{ii} = \dfrac{1}{n} + \dfrac{(X_i - \overline{X})^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$

and then plot X_i against $\hat{\varepsilon}_i^*$ or X_i against $(\hat{\varepsilon}_i^*)^2$.
Finally, show that

(vi) $$\sum_{i=1}^{n} E(\hat{\varepsilon}_i^2) = \sigma^2 (n-2)$$

Problem set 8
Solve the exercises: 7.6, and in addition exercise 8.1.

Exercise 8.1
For å ta hensyn til kvalitative forklaringsvariable i en regresjonsligning benytter vi gjerne dummyvariable.

Spørsmål 1
Gi eksempler på økonometriske anvendelser av dummyvariable.

På et utvalg av norske lønnstakere skal vi estimere regresjonen

(1) $$X_i = \beta_0 + \beta_1 K_i + \beta_2 E_i + \varepsilon_i \qquad i = 1, 2, 3, \ldots\ldots\ldots\ldots, 200$$

der X_i betegner timelønnen til lønnstaker (i), K_i betegner dummyvariabelen for kjønn, som er lik 1 hvis person i er en kvinne og lik 0 hvis person i er en mann, E_i betegner antall år utdanning til lønnstaker (i). Det er like mange kvinner som menn i utvalget. Variabelen ε_i betegner det stokastiske restledd i regresjonen, som antaes å være uavhengige og identisk fordelte for alle i, med forventning 0 og varians σ^2.

Bruk det vedlagte datasett (lønn og kjønn des08) til å beregne regresjonen (1).

Spørsmål 2
Kommenter de empiriske resultatene. Hva vil du trekke frem som spesielt interessant ved disse resultatene.

Effekten av utdanning på timelønnen synes å være usikker. Estimer lønnsrelasjonen (1) når du har ekskludert utdanning, dvs.

(2) $X_i = \beta_0 + \beta_1 K_i + \varepsilon_i$

Minstekvadraters estimatorene på regresjonskoeffisientene β_0 og β_1 i regresjonen (2) er gitt ved formlene

(a) $\hat{\beta}_0 = \overline{X} - \hat{\beta}_1 \overline{K}$

(b) $$\hat{\beta}_1 = \frac{\sum_{i=1}^{200}(K_i - \overline{K})X_i}{\sum_{i=1}^{200}(K_i - \overline{K})^2}$$

der $\overline{X}$ betegner gjennomsnittlig timelønn i datasettet og $\overline{K}$ betegner gjennomsnittet av dummyvariabelen K_i.

Spørsmål 3
Vis at formlene for minstekvadraters estimatene reduserer seg til
$\hat{\beta}_0 = \overline{X}_M$ og $\hat{\beta}_1 = \overline{X}_K - \overline{X}_M$
der $\overline{X}_M$ betegner gjennomsnittet av timelønnen for mennene i utvalget og $\overline{X}_K$ betegner gjennomsnittet av timelønnen for kvinnene.

Mange som forsker på arbeidsmarkedsrelaterte problemer og lønnsdannelse, vil hevde at lønnsrelasjoner er ikke-lineære, slik at de lineære regresjonene i (1) og (2) passer dårlig for dette formålet. I stedet for disse blir følgende spesifikasjoner foreslått

(3) $\ln X_i = \gamma_0 + \gamma_1 K_i + \gamma_i E_i + \varepsilon_i$

(4) $\ln X_i = \alpha_0 + \alpha_1 \ln K_i + \alpha_2 \ln E_i + \varepsilon_i$

der $\ln X_i$, $\ln K_i$ og $\ln E_i$ betegner den naturlige logaritmen til de angitte variable.

Spørsmål 4
Forklar hvorfor vi ikke kan benytte regresjon (4) med vårt valg av forklaringsvariable.

Estimer regresjonen (3) med bruk av det vedlagte datasett.

Spørsmål 5
Kommenter resultatene av denne regresjonen. Synes du det er grunn til å foretrekke regresjon (3) fremfor regresjon (1)?

Spørsmål 6
Hvordan vil du tolke parameteren γ_1 i regresjon (3)?

Spørsmål 7
Estimer timelønnen for en mannlig arbeider med 10 års utdanning når du: (a) benytter regresjon (1), og (b) benytter regresjon (3).

Problem set 9
Solve the exercises: 8.11, 8.16 and in addition exercise 9.1 below

Exercise 9.1
En interessant hypotese som diskuteres i makroteori er at offentlige utgifter bidrar til å fortrenge de samlede investeringer i et lands økonomi. Ideen er at store offentlige utgifter både direkte og indirekte via et økt rentenivå, bidrar til å redusere de samlede investeringer. Vi ønsker å undersøke denne hypotesen ved bruk av tverrsnittsdata for et utvalg på 30 viktige industriland i 1999. Dessverre innholder vårt datasett ikke observasjoner for rentenivået i landene, men gir oss data for variablene: offentlige utgifter (G), samlede investeringer (I), brutto nasjonalprodukt (X) og folketall (P).
Som bakgrunnsmodell for å undersøke denne hypotesen spesifiseres regresjonen

(1) $$I_j = \beta_0 + \beta_1 G_j + \beta_2 X_j + \varepsilon_j \qquad j = 1,2,3,..........,30$$
der ε_j betegner de stokastiske restleddene.

Utskrift 1 viser resultatene av denne regresjonsberegningen.

Spørsmål 1
(a) Gjør rede for innholdet i denne utskriften.
(b) Synes du hypotesen om at de offentlige utgifter fortrenger de samlede investeringer blir bekrefter eller avkreftet? Begrunn svaret.

Datasettet innholder både store og små land, for eksempel inngår både USA og Island i utvalget. Vi mistenker derfor sterkt de stokastiske restleddene $\{\varepsilon_j\}$ å være heteroskedastiske.

Spørsmål 2
(a) I Utskrift 2 har vi plottet residualene fra regresjon (1) mot brutto nasjonalprodukt X . Synes du de plottede punktene (de røde kryssene) tyder på at restleddene er heteroskedastiske?
For å undersøke mer presist om heteroskedastisitet er til sted i våre data, har vi også beregnet White's konsistente estimat på standardavvikene til estimatorene $\hat{\beta}_0, \hat{\beta}_1$ *og* $\hat{\beta}_2$. Resultatene er rapportert i Utskrift 3. White's estimerte standardavvik for estimatorene $\hat{\beta}_0, \hat{\beta}_1$ *og* $\hat{\beta}_2$ finner du i kolonnen betegnet HCSE.

(b) Tyder resultatene i Utskrift 3 på at restleddene er herteroskedastiske? Begrunn svaret. Fyll in tallene i kolonnen som er betegnet t-HCSE.

Som en tilleggstest for å avgjøre om restleddene er heteroskedastiske ønsker vi også å benytte Goldfeld-Quandt's test. Dataene er sortert etter størrelsen på bruttonasjonalproduktet, deretter er det kjørt to regresjoner. Utskrift 4 viser resultatene av regresjon (1) for de 15 landene med høyest bruttonasjonalprodukt, mens utskrift 5 viser de tilsvarende resultater for de 15 landene med lavest bruttonasjonalprodukt.

Spørsmål 3
Forklar kort innholdet i Goldfeld-Quandt's testen. Benytt resultater du finner i utskriftene 4 og 5 til å gjennomføre testprosedyren, og forklar hva dine resultater tyder på om restleddenes fordelinger.

Bruttonasjonalproduktet (X) i landene som inngår i vårt datasett varierer sterkt. For å redusere effekten av sterkt varierende (X) er det foreslått å spesifisere regresjonen mellom de tilsvarende pr. capita størrelsene.

Spørsmål 4
Regresjon (2) nedenfor viser denne spesifikasjon

(2) $$(I/P)_j = \beta_0 + \beta_1 (G/P)_j + \beta_2 (X/P)_j + u_j \qquad j = 1,2,...,30$$

Utskrift 6 viser resultatene av denne regresjonsberegningen. Bruk av Goldfeld-Quandt's test tyder på at restleddene (u_j) i (2) kan være homoskedastiske. Synes du at du kan finne støtte for en slik konklusjon i utskrift 6?

Forskere som er tvilende til hypotesen om at offentlige utgifter fortrenger de samlede investeringer (tvilerne), påstår at man bare kan avgjøre dette spørsmålet ved å benytte regresjonen

(3) $$(I/P)_j = \gamma_0 + \gamma_1 (G/P)_j + \gamma_2 (X/P)_j + \gamma_3 r_j + \delta_j \qquad j = 1,2,....,30$$

der r_j betegner rentenivået i de forskjellige landene.

La $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ og $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ betegne OLS estimatene i regresjonene (2) og (3).

Spørsmål 5
Tvilerne hevder at γ_1 kan være lik null selv om β_1 er forskjellig fra null. Diskuter denne påstanden. I diskusjonen kan du anta at regresjonen av r_j *på* $(G/P)_j$ *og* $(X/P)_j$ er lineær eller $r_j = \alpha_0 + \alpha_1 (G/P)_j + \alpha_2 (X/P)_j + u_j$.

Mot dette hevder tilhengerne av hypotesen at på grunn av liberaliseringen av de internasjonale kapitalmarkeder, vil rentenivået være tilnærmet likt i de forskjellige landene slik at $\hat{\beta}_1$ vil være tilnærmet lik $\hat{\gamma}_1$

Spørsmål 6

Drøft denne påstanden når vi antar at $r_j = \bar{r} + v_j$ der $\bar{r}$ betegner gjennomsnittet av rentenivåene i landene og v_j betegner svært små positive og negative avvik.

Utskrift 1

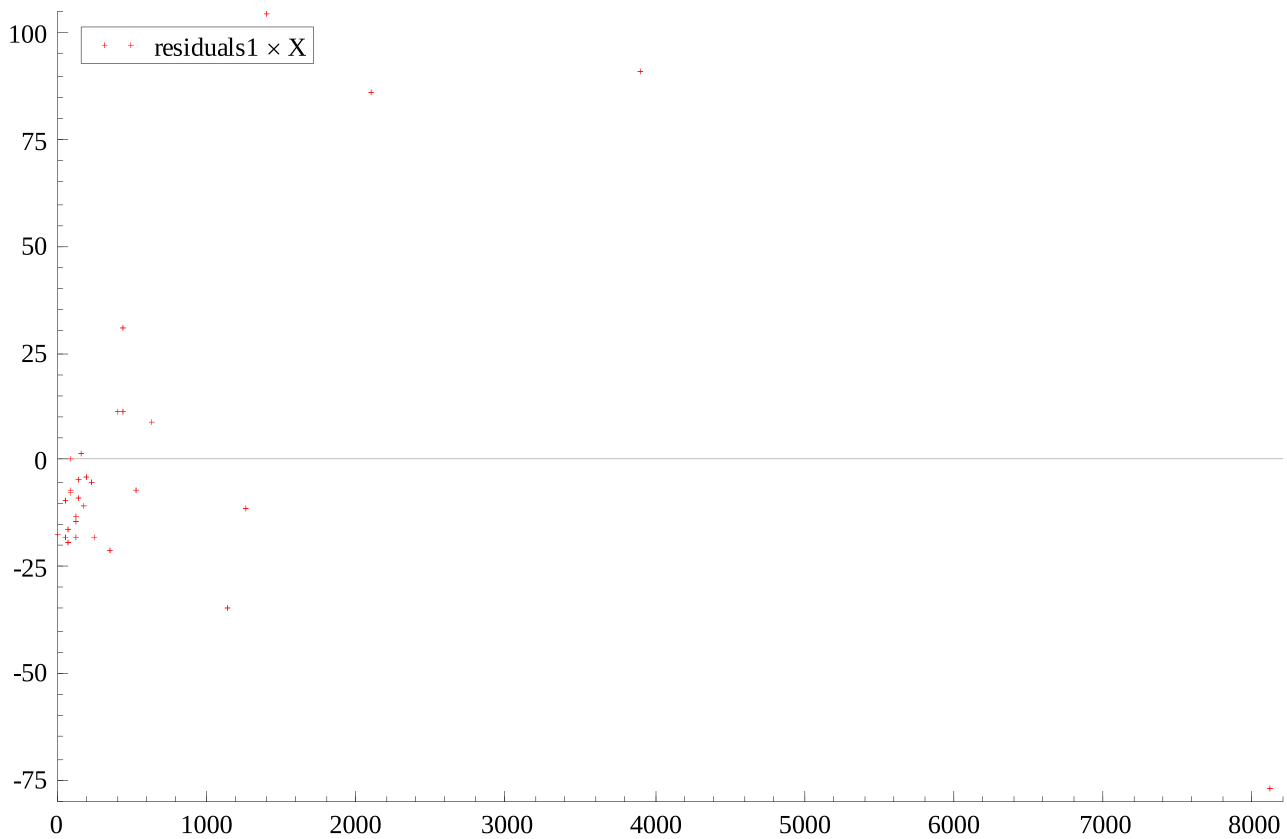
EQ(1) Modelling (I) by OLS-CS (using makrodata.xls)

The estimation sample is: 1 to 30

	Coefficient	Std.Error	t-value	t-prob
Constant	18.0926	7.786	2.32	0.028
G	-1.07496	0.1383	-7.77	0.000
X	0.359251	0.02074	17.3	0.000

sigma	37.7334	RSS	38442.8687
R^2	0.987796	F(2,27) =	1093 [0.000]**
		DW	1.38
no. of observations	30	no. of parameters	3
mean (I)	162.47	var (I)	105000

Utskrift 2



Utskrift 3

	Coefficients	SE	t-SE	HCSE	t-HCSE
Constant	18.093	7.7857	2.3238	5.8031	
G	-1.0750	0.13834	-7.7705	0.34912	
X	0.35925	0.020736	17.325	0.052798	

Utskrift 4

EQ(1) Modelling *I* by OLS-CS (using makrodata22.xls)
The estimation sample is: 1 to 15

	Coefficient	Std.Error	t-value	t-prob
Constant	35.2641	17.21	2.05	0.063
G	-1.12410	0.1972	-5.70	0.000
X	0.362737	0.02915	12.4	0.000

sigma	52.6624	RSS	33279.9365
R^2	0.987192	F(2,12) =	462.5 [0.000]**
		DW	1.21
no. of observations	15	no. of parameters	3
mean(*I*)	297.747	var(*I*)	173229

Utskrift 5

EQ(2) Modelling *I* by OLS-CS (using makrodata22.xls)
The estimation sample is: 1 to 15

	Coefficient	Std.Error	t-value	t-prob
Constant	0.630508	3.923	0.161	0.875
G	-0.820569	0.2567	-3.20	0.008
X	0.378605	0.05563	6.81	0.000

sigma	6.02232	RSS	435.220297
R^2	0.834817	F(2,12) =	30.32 [0.000]**
		DW	1.28
no. of observations	15	no. of parameters	3
mean(*I*)	27.2133	var(*I*)	175.652

Utskrift 6

EQ(3) Modelling (I/P) by OLS-CS (using makrodata.xls)
The estimation sample is: 1 to 30

	Coefficient	Std.Error	t-value	t-prob	HCSE
Constant	0.110689	0.3254	0.340	0.736	0.20291
G/P	-0.684616	0.1613	-4.24	0.000	0.2567
X/P	0.329780	0.03347	9.85	0.000	0.05294

sigma	0.877237	RSS	20.7777011
R^2	0.87778	F(2,27) =	96.96 [0.000]**
		DW	1.79
no. of observations	30	no. of parameters	3
mean(I/P)	3.9529	var(I/P)	5.66673

Problem set 10
Solve the exercises: 9.4, 9.10, 9.13 and in addition the exercise 10.1 below.

Exercise 10.1
Vi er interessert i å studere sammenhengen mellom verdien på bedrifter uttrykt ved bedriftenes aksjekapital og størrelsen på utbytte (dividende) bedriftene deler ut til sine aksjeeiere. Vårt datasett er aggregerte tall i faste priser og dekker årene 1923- 1996. P_t betegner den samlede verdi på bedriftenes aksjekapital i år t, D_t betegner den samlede verdi på utbytte og endelig betegner C_t det samlede konsum. Variabelen C_t er inkludert for å vise utviklingen av realinntekten i perioden.

For å analysere denne sammenhengen er følgende spesifikasjon foreslått:

(1) $P_t = \beta_0 + \beta_1 D_t + \beta_2 C_t + \varepsilon_t \qquad t = 1, 2, 3, ..., 74$

der ε_t betegner stokastiske restledd.

Spørsmål 1
(i) Bruk det vedlagte datasett (anders2.data) til å beregne regresjonen (1), og gjør rede for om du synes resultatene er rimelige.
(ii) Gjør rede for hvordan tallene i kolonnene betegnet t-value og t-prob er fremkommet.

Suppler den beregnede regresjon ovenfor med regresjonene:

(2) $P_t = \beta_0 + \beta_1 D_t + u_t \qquad t = 1, 2, 3, ..., 74$

(3) $P_t = \beta_1 D_t + \beta_2 C_t + \varepsilon_t \qquad t = 1, 2, 3, ..., 74$

Spørsmål 2

Ved å sammenligne regresjonsberegningene 1 og 2 ser vi at estimatet på β_0 har forskjellige egenskaper alt etter som C_t er inkludert i eller ekskludert fra regresjonen. Likeså ser vi ved å sammenligne beregningene for 1 og 3 at estimatet på β_2 har forskjellige egenskaper alt etter som et konstantledd er inkludert i eller ekskludert fra regresjonen. Hvordan vil du forklare disse forskjellene?

I det følgende skal vi basere oss på regresjonen (2)

Ved bruk av tidsrekkedata i regresjonsberegninger vil autokorrelerte restledd ofte være et problem man møter. Siden vi benytter årsdata i vår undersøkelse, er det rimelig å gå ut fra at restleddene er autokorrelerte av orden 1. Vi skal derfor anta at restleddene u_t i regresjon (2) tilfredsstiller ligningen

(3) $\quad u_t = \rho u_{t-1} + e_t \qquad |\rho| < 1$

der e_t er stokastisk uavhengige med forventning $0 \;\; og \;\; \text{var}\, ians\; \sigma^2$.

For å undersøke om autokorrelasjon er til stede vil de fleste programmene beregne den så kalte Durbin-Watson observatoren, betegnet DW i utskriftene.

Spørsmål 3

(i) I spesifikasjonen (3) forutsetter vi at $-1 < \rho < 1$. Forklar hvorfor dette er en viktig forutsetning.

(ii) Gi en kort begrunnelse for Durbin_Watson testen.

(iii) Bruk regresjonsberegningen for regresjon (2) til å test null hypotesen H_0 : $\rho = 0$ mot den alternative hypotesen $H_1: \quad \rho > 0.$ Velg signifikansnivå $\alpha = 0.05$

I modeller for å forklare aksjeverdien P_t oppfattes P_t ofte som nåverdien av fremtidige aksjeutbytter. Lar vi $D_{t+1}^t, D_{t+2}^t, D_{t+3}^t, \ldots, D_{t+k}^t, \ldots$, betegne prognosene vi stiller opp på tidspunkt t for de fremtidige utbytter, vil bakgrunnsmodellen være gitt ved

(4) $$P_t = \beta_0 + \sum_{i=1}^{\infty} \gamma^i D_{t+i}^t$$

der diskonteringsfaktoren $0 < \gamma < 1$. Det kan vises ved å ta utgangspunkt i og å utvikle summen av prediksjonsfeilene $\sum_{i=1}^{\infty} \gamma^i (D_{t+i}^t - D_{t+i-1}^t)/(1-\gamma)$ at vi kan avlede regresjonen

(5) $$P_t = \beta_0 + \frac{\gamma}{1-\gamma} D_t + v_t$$

der v_t betegner de stokastiske restledd.

Spørsmål 4

(i) Bruk beregningsresultatene for regresjon 2 til å utlede et estimat for diskonteringsfaktoren γ.