

Institutt for samfunnsøkonomi

Eksamensoppgave i SØK3001 Økonometri I / Econometrics I

Faglig kontakt under eksamen: Bjarne Strøm

Tlf.: 73 59 19 33

Eksamensdato: 4. juni 2014

Eksamenstid: 5 timer (09.00-14.00)

Sensurdato: 26. juni 2014

Tillatte hjelpemidler: Flg formelsamling: Knut Sydsæter, Arne Strøm og Peter Berck (2006): Matematisk formelsamling for økonomer, 4utg. Gyldendal akademiske. Knut Sydsæter, Arne Strøm, og Peter Berck (2005): Economists' mathematical manual, Berlin. Enkel kalkulator Casio fx-82ES PLUS, Citizen SR-270x, HP 30S eller SR-270X College

Annen informasjon: Eksamensoppgaven består av 3 oppgaver med delspørsmål som alle skal besvares.

Målform/språk: Norsk og engelsk

Antall sider: 7

Antall sider vedlegg: 2 tabeller

Bokmål

Oppgave 1

I en empirisk undersøkelse av faktorer som påvirker etterspørselen etter varig konsumgoder estimeres først følgende generelle dynamiske spesifisering:

$$c_t = \alpha c_{t-1} + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 u_t + \beta_4 u_{t-1} + \beta_5 R_t + \beta_6 R_{t-1} + \beta_0 + v_t$$

der c_t er logaritmen til etterspørselen etter varig konsum år t , y_t er logaritmen til realdisponibel inntekt år t , u_t er arbeidsledighetsraten regnet i prosent, år t , R_t er realrenta regnet i prosent, år t og v_t er et stokastisk restledd. I undersøkelsen benyttes årlige observasjoner for 42 år.

Estimeringsmetoden som benyttes er minste kvadraters metode (OLS). Resultater for den generelle modellen er gitt i Tabell 1, kolonne I som også rapporterer SSR som er summen av kvadrerte avvik. Tall i parenteser er beregnede t-verdiene.

- Forklar hvordan du kan teste hypotesen $\beta_3 + \beta_4 = 0$, forklar hva denne restriksjonen innebærer for den langsiktige effekten av arbeidsledigheten og test hypotesen ved bruk av resultatene i kolonne II.
- I kolonne III rapporteres resultater for en forenklet versjon av den generelle modellen. Forklar hvilke restriksjoner som er pålagt. Forklar videre hvordan du kan teste om modellen i kolonne III er en gyldig forenkling av den mest generelle modellen og gjennomfør testen ved bruk av opplysningene gitt i Tabell 1.
- Gi en tolkning av de empiriske resultatene i kolonne III. Drøft *spesielt* kort- og langsiktige effekter av de inkluderte variablene samt hvor raskt konsumetterspørselen tilpasses.
- Fra kolonne I ser vi at de estimerte effektene av y_t og R_t ikke er signifikant forskjellig fra null, mens de estimerte effektene av disse to variablene er klart signifikant forskjellig fra null ved bruk av modellen i kolonne III. Drøft mulig årsak til denne forskjellen.

Tabell 1
Venstresidevariabel er c_t

| Forklaringsvariable | Kolonne I | Kolonne II | Kolonne III |
|------------------------------|----------------|----------------|----------------|
| c_{t-1} | 0.62 (3.65) | 0.62 (3.65) | 0.69 (8.95) |
| y_t | 0.87 (1.57) | 0.87 (1.57) | 0.46 (3.79) |
| y_{t-1} | -0.29 (-0.52) | -0.29 (-0.52) | |
| u_t | -0.074 (-2.28) | | |
| u_{t-1} | 0.063 (2.69) | -0.011 (-0.42) | |
| $\Delta u_t = u_t - u_{t-1}$ | | -0.074 (-2.28) | -0.066 (-3.19) |
| R_t | -0.012 (-1.05) | -0.012 (-1.05) | -0.015 (-3.73) |
| R_{t-1} | -0.003 (-0.25) | -0.003 (-0.25) | |
| Constant | -3.64 (-1.81) | -3.64 (-1.81) | -2.69 (-3.22) |
| SSR | 0.20157 | 0.20157 | 0.20609 |

Oppgave 2

En empirisk studie undersøker hvordan relative priser og andre variable påvirker kundenes kjøp av Pepsi eller Coke. Studien bruker data fra dagligvarebutikker i USA om kundenes valg mellom de to merkevarene. Til sammen 1140 kunder kjøpte enten Pepsi eller Coke og 44.7 % av dem kjøpte Coke. En dummyvariabel, *Coke*, tar verdien 1 dersom kunden kjøpte Coke og verdien 0 dersom kunden kjøpte Pepsi. Vi har også følgende opplysninger:

P = relative prisforholdet mellom Pepsi og Coke

$Disp_Coke$ = 1 dersom det er reklame for Coke i butikken, 0 ellers

$Disp_Pepsi$ = 1 dersom det er reklame for Pepsi i butikken, 0 ellers

Det presenteres følgende estimerte ligning

$$\hat{Coke} = 0.8902 - 0.4009P + 0.0772Disp_Coke - 0.1657Disp_Pepsi$$

(0.0655) (0.0613) (0.0344) (0.0356)

- Hvordan vil du tolke den estimerte ligningen?
- Sett at prisen på Coke øker med 10% relativt til Pepsi fra en utgangssituasjon med lik pris ($P=1$). Bruk estimeringsresultatene til å beregne effekten på sannsynligheten for at kunden velger Coke i stedet for Pepsi.
- Hvor mye endres sannsynligheten for å kjøpe Coke av at det er reklame for Pepsi i butikken?

Oppgave 3

- a) I en diskusjon om troverdigheten av økonometriske undersøkelser kommer en debattant med følgende utsagn: «Økonometriske undersøkelser har lav troverdighet fordi det alltid vil være relevante forklaringsvariable som er utelatt fra en økonometrisk modell og på grunn av variable som er unøyaktig målt». Drøft dette utsagnet og drøft alternative metoder som kan løse disse problemene.
- b) En undersøkelse i USA rapporterer empiriske resultater for sammenhengen mellom mordraten, antall henrettede fanger og arbeidsledigheten. Under presenteres resultatene fra to regresjonsmodeller som estimerer denne sammenhengen.

Modell 1 benytter et tverrsnittsmateriale der antall mord per 100000 innbyggere i 1990 i delstaten, *Murder* er avhengig variabel. Forklaringsvariable er *Executions* definert ved gjennomsnittlig antall henrettelser i delstaten i perioden 1987-89 samt *Unem* definert ved arbeidsledighetsraten i delstaten i 1990 målt i prosent.

Den estimerte relasjonen er som følger (antall observasjoner er $N=50$ =antall delstater i USA):

$$(1) \quad \bar{Murder} = 0.39 + 0.79 \bar{Executions} + 1.11 \bar{Unem}, \quad R^2 = 0.35$$

(2.27) (0.21) (0.42)

Modell 2 benytter endringen i antall mord per 100000 innbyggere fra 1990 til 1993 som avhengig variabel. Forklaringsvariable er endringen i antall henrettelser i perioden 1987-89 til perioden 1990-1992 samt endring i arbeidsledighetsraten fra 1990 til 1993. Antall observasjoner er 50 også i denne relasjonen.

$$(2) \quad \Delta \bar{Murder} = 0.41 - 0.10 \Delta \bar{Executions} - 0.08 \Delta \bar{Unem}, \quad R^2 = 0.11$$

(0.21) (0.04) (0.16)

- i) Test hypotesen om at hverken antall henrettelser eller arbeidsledigheten påvirker mordraten i hver av de to modellene
- ii) Bruk resultatene til å predikere effekten på mordraten av en økning i antall henrettelser med 10 personer i hver av modellene.
- iii) Forklar hvorfor de to modellene (1) og (2) ser ut til å gi forskjellige resultater når det gjelder effekten av henrettelser og arbeidsledigheten på antall mord begått i delstaten. Hvilken relasjon finner du mest troverdig? Forklar hvorfor.

English**Question 1**

In an empirical study of factors affecting demand for durable consumer goods, the following dynamic specification is first estimated:

$$c_t = \alpha c_{t-1} + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 u_t + \beta_4 u_{t-1} + \beta_5 R_t + \beta_6 R_{t-1} + \beta_0 + v_t$$

where c_t is the logarithm of demand for durable goods year t , y_t is the logarithm of real disposable income year t , u_t is the unemployment rate in per cent, year t , R_t is the real rate of interest in per cent, year t and v_t is a random error term. The study uses annual observations for 42 years. The estimation method is ordinary least squares (OLS). Results for the general model are given in Table 1, column I which also reports SSR which is the sum of squared residuals. Numbers in parentheses are calculated t-values.

- a) Explain how you can test the hypothesis $\beta_3 + \beta_4 = 0$, explain what this restriction implies for the long run effect of unemployment and test the hypothesis using the results in column II.
- b) Column III reports results for a simplified version of the general model. Explain which restrictions that are imposed. Explain further how you can test whether the model in column III is a valid simplification of the most general model and carry out the test using the information given in Table 1.
- c) Give an interpretation of the empirical results in column III. Discuss *in particular* the short- and long run effects of the included variables and how fast consumer demand is adjusted.
- d) From column I we see that the estimated effects of y_t and R_t are not significantly different from zero while the estimated effects of these two variables are clearly significantly different from zero using the model in column III. Discuss the possible reason for this difference.

Tabell 1
Left hand side variable is c_t

| Explanatory variables | Column I | Column II | Column III |
|------------------------------|----------------|----------------|----------------|
| c_{t-1} | 0.62 (3.65) | 0.62 (3.65) | 0.69 (8.95) |
| y_t | 0.87 (1.57) | 0.87 (1.57) | 0.46 (3.79) |
| y_{t-1} | -0.29 (-0.52) | -0.29 (-0.52) | |
| u_t | -0.074 (-2.28) | | |
| u_{t-1} | 0.063 (2.69) | -0.011 (-0.42) | |
| $\Delta u_t = u_t - u_{t-1}$ | | -0.074 (-2.28) | -0.066 (-3.19) |
| R_t | -0.012 (-1.05) | -0.012 (-1.05) | -0.015 (-3.73) |
| R_{t-1} | -0.003 (-0.25) | -0.003 (-0.25) | |
| Constant | -3.64 (-1.81) | -3.64 (-1.81) | -2.69 (-3.22) |
| SSR | 0.20157 | 0.20157 | 0.20609 |

Question 2

An empirical study investigates how relative prices and other variables affect the customers' purchase of Pepsi or Coke. The study uses data from grocery stores in the US about the customers' choice between the two brands. In all, 1140 customers bought either Pepsi or Coke and 44.7 % of them bought Coke. A dummy variable, *Coke*, takes the value 1 if the customer bought Coke and 0 if the customer bought Pepsi. We also have the following information:

P = the price of Pepsi relative to the price of Coke

$Disp_Coke$ = 1 if there is advertisement for Coke in the shop, otherwise 0

$Disp_Pepsi$ =1 if there is advertisement for Pepsi in the shop, otherwise 0

The following estimated equation is reported:

$$\hat{Coke} = 0.8902 - 0.4009P + 0.0772Disp_Coke - 0.1657Disp_Pepsi$$

(0.0655) (0.0613) (0.0344) (0.0356)

- How will you interpret the estimated equation?
- Say that the price of Coke increases by 10% relatively to the price of Pepsi from an initial situation with equal price ($P=1$). Use the estimation results to calculate the effect on the probability that the customer will chose Coke instead of Pepsi.
- How much will the probability of buying Coke change as a result of advertisement for Pepsi in the shop?

Question 3

a) In a discussion about the reliability of econometric studies a debater gives the following statement: «Econometric studies have low reliability because relevant explanatory variables are always excluded from an econometric model and because of variables that are inaccurately measured». Discuss this statement and discuss alternative methods that can solve these problems.

b) A study in the US reports empirical results for the relation between the murder rate, the number of executed prisoners and the unemployment. Below, results from two regression models estimating this relation are presented.

Model 1 uses cross section data where the number of murders per 100000 inhabitants in 1990 in the state, *Murder* is the dependent variable. Explanatory variables are *Executions* defined as the average number of executions in the state during the period 1987-89 and *Unem* defined as the unemployment rate in the state in 1990, measured in percent.

The estimated relation is given by (number of observations $N=50$ =number of states in the US):

$$(1) \quad \bar{Murder} = 0.39 + 0.79Executions + 1.11Unem, \quad R^2 = 0.35$$

(2.27) (0.21) (0.42)

Model 2 uses the change in the number of murder per 100000 inhabitants from 1990 to 1993 as the dependent variable. The explanatory variables in this model are the change in the number of executions from the period 1987-89 to the period 1990-1992 and the change in the unemployment rate from 1990 to 1993. The number of observations is 50 also in this model.

$$(2) \quad \Delta \bar{Murder} = 0.41 - 0.10\Delta Executions - 0.08\Delta Unem, \quad R^2 = 0.11$$

(0.21) (0.04) (0.16)

i) Test the hypothesis that neither the number of executions nor the unemployment affects the murder rate in each of the two models.

ii) Use the results to predict the effect on the murder rate of increasing the number of executions by 10 persons in each of the models.

iii) Explain why the two models (1) and (2) seems to give different results regarding the effects of executions and the unemployment on the murder rate. Which relation do you find the most credible? Explain why.

Kommentar besvarelse i SØK 3001,V2014

Dette er en gjennomgående svært god besvarelse. Den er ryddig og konsis samtidig som den demonstrerer at kandidaten har god forståelse av og behersker bruken av de sentrale økonometriske metoder som inngår i kurset. Noen kommentarer er likevel på sin plass. På oppgave 2 kunne besvarelsen vært noe mer fylldig når det gjelder tolkningen av den lineære sannsynlighetsmodellen og påpekt noen begrensninger ved den. På oppgave 3 iii) burde kandidaten vært noe mer aktiv når det gjelder valg av metode og gjerne henvist aktivt til diskusjonen på delspørsmål 1.

Denne kolonnen er forbeholdt sensor
This column is for external examiner

Oppgave 1

a)

Vi kaller $\beta_3 + \beta_4 = \theta$, og skriver om modellen slik at θ forekommer som parameter foran en variabel. Dette gjøres for å få et estimert standardavvik for θ ut fra STATA, like vi ikke finner direkte fra opprinnelig modell. Dette er gjort i kolonne II ved å legge til $\beta_3 u_{t-1}$, slik at $\beta_4 u_{t-1} + \beta_3 u_{t-1} = \theta u_{t-1}$. Videre har man trukket fra samme størrelse, slik at modellen ikke endres, og man får $\beta_3 u_t - \beta_3 u_{t-1} = \beta_3 \Delta u_t$. Videre har vi fra forutsetningene for OLS, og spesielt MLR.6 om normalfordelte restledd med forventning 0, at test observatoren
$$z = \frac{\hat{\theta} - \theta^{\text{hyp}}}{\text{sd}(\hat{\theta})} \sim N(0,1)$$
, der

θ^{hyp} er forventningen til θ under nullhypotesen $H_0: \beta_3 + \beta_4 = \theta = 0$ ($H_1: \theta \neq 0$). Siden vi ikke kjenner restleddets populasjonsstandardavvik tilnærmer vi $\text{sd}(\hat{\theta})$ gjennom $\text{se}(\hat{\theta})$, og får
$$T = \frac{\hat{\theta} - \theta^{\text{hyp}}}{\text{se}(\hat{\theta})}$$

$\sim t_{n-k-1}$. Denne observatoren er eksakt t -fordelt når MLR.6 gjelder, og vi bruker t -fordelingsstabell til testen. n henviser her til antall år, mens k viser til antall ~~estimerede~~ forklaringsvariable (her 7 stk).

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

Den siste frihetsgraden brukes til α estimere konstantleddet. ~~t fordelingen har alle~~
 $t = -0,42$ følge kolonne II, mens t_{crit} som er kritisk verdi for testen finnes fra tabell. $t_{crit} \approx 2,03$ (ingen eksakt verdi for $df = 42 - 8 = 34$, leses av ca midt mellom 30 og 40). Forhaster H_0 hvis $|t| > t_{crit}$, noe som her ikke er tilfelle, og vi kan ikke forhaste hypotesen og $\beta_3 + \beta_4 = 0$ med signifikansnivå 5%. Valg av dette sign. nivået er vanlig, og representerer 5% sjans for α forhaste H_0 hvis denne er sann. Bruker tosidig test fordi vi i H_1 ikke sier noe om retningen på $\theta \neq 0$, og ser derfor at $t = 0,42$ ikke er i nærheten av α gi forhastning selv med et sign. nivå på 20%.

Langsiktig effekt finnes i likevekt, der $C_t = C_{t-1} = C_t^*$, $u_t = u_{t-1} = u_t^*$ og $R_t = R_{t-1} = R_t^*$;
 $y_t = y_{t-1} = y_t^*$

$$C_t^* = \frac{\beta_1 + \beta_2}{1-d} y^* + \frac{\beta_3 + \beta_4}{1-d} u^* + \frac{\beta_5 + \beta_6}{1-d} R^* + \beta_0 + \sigma_t$$

$\frac{\beta_3 + \beta_4}{1-d}$ representerer altså langsiktig effekt av u_t på C_t , og vi ser at $\beta_3 + \beta_4 = 0$ innebærer ingen langsiktig effekt av økt arbeidsledighet på logaritmen til konsumet. (i prosent prang)

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

b)

Restriksjonene som er pålagt fra kolonne II til III er $\beta_2 = 0, \theta = 0, \beta_6 = 0$. Dette er en simultan hypotese

$H_0: \beta_2 = 0, \theta = 0, \beta_6 = 0$ om at u_{t-1}, y_{t-1} og R_{t-1} samlet ikke har noen ~~effekt~~ signifikant effekt på C_t i II.

Dette kan testes ved en F-test som uttylter at SSR alltid øker ved ekskludering av forklaringsvariable.

Testobservatoren

$$F = \frac{\frac{SSR_r - SSR_{ur}}{q}}{\frac{SSR_{ur}}{n-k-1}} \sim F_{q, n-k-1}$$

under MLR.1 - MLR.6, der q er antall ~~simultane~~ utvalgte variable i det vi kaller restricted model (kolonne III) ift unrestricted model (I - II).

$q = 3, n - k - 1 = 34$.

$$F = \frac{\frac{0,20609 - 0,20157}{3}}{\frac{0,20157}{34}} = 0,254$$

Fant finnes analogt til det vi gjorde i t-tabellen til ca 2,88 (eller ikke her tall for teller-df = 34).

$F = 0,254 < F_{crit} = 2,88$, og vi kan ikke forkaste H_0 , altså er III en gyldig forklaring av II.

c) Ser av t-verdiene i III at alle koeffisienter er signifikant forskjellige fra 0 (se 1a). 0,46 angir en elastisitet fordi begge variable er på logaritmisk form, og tolkes som 0,46% ~~endring~~ ~~endring~~ i konsumet når realdisponibel inntekt øker med 1%. $-0,066 \cdot 100$ angir semielastisiteten ved en endring i arbeidsløshetsraten på 1% poeng. Altså vil en ~~1%~~ økning i ledighetsraten gi 6,6% lavere konsum. (Mer nøyaktig verdi = $100 \cdot (e^{P_1} - 1) = 6,39\%$). $-0,015$ gir på tilsvarende måte at ett prosentpoeng høyere realrente gir 1,5% (1,489% mer nøyaktig) lavere konsum. Konstant leddet innebærer negativt konsum ved 0 realdisponibel ~~inntekt~~ inntekt, 0 realrente, osv, og har ingen selvstendig interesse. Dette var de kortsiktige effektene. Effekten på lang sikt finnes tilsvarende som i a, ved å anta at vi når en likevekt der variablene ~~ikke~~ ikke endres fra periode til periode lenger (all effekt er "tatt ut").

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

Får igjen $C_t^*(1-d)$ på venstre side,
 og deler gjennom med $(1-d)$,
 for å finne langsiktige effekter.
 d representerer her graden av
 persistens, eller treghet i konsum-
 tilpassningen, mens $(1-d)$ gjerne kalles
 λ , og representerer tilpassningshastighet.
 $d = 0,69$ er relativt høy persistens,
 det vil si at det tar tid for
 konsumet justeres seg helt etter
 endring i forklaringsvariablene, og
 langtidsseffektene vil derfor være
 klart større enn korttidseffektene.

Vi ender opp med en langsiktig effekt
 av 1% økning i realdisponibel
 inntekt på $\frac{0,46}{(1-0,69)} = 1,484\%$, altså

betydelig større en korttidseffekten,
 grunnet tregheten. Tilsvarende for
 realrente gir langsiktig effekt av
 ett prosentpoeng høyere realrente på
 $\frac{-0,015}{(1-0,69)} = -4,84\%$, altså 4,84%

lavere konsum. I likevekt vil ~~$u_t = u_{t-1}$~~
 noe som gir $\Delta u_t = 0$, og igjen ingen
 langsiktig effekt av endring i
 arbeidsledighetsraten på konsum. Dette
 viser litt merkelig, og gir mistanke
 om at jeg kan ha gjort en

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

Feil jeg ikke har tid til å lete
 mer etter. Fra I tilnes i hvert
 fall at høyere arb.ledighet ifjor,
 alt annet likt skal gi høyere
 konsum i år, noe som virker
 rart.

d)
 En sannsynlig årsak til forskjellen
 i signifikans for disse to fra
 I til III er multikollinearitet, eller
 sterk samvariasjon blant forklarings-
 variablene. Dette slår ut i

$$\widehat{\text{Var}}(\hat{\beta}_j | x) = \frac{\hat{\sigma}^2}{SST_j(1-R_j^2)} \text{ der } R_j \text{ er } R_j^2 \text{ er}$$

R^2 fra regresjonen av variabel x_j
 på alle andre forklaringsvariable.
 Sterk samvariasjon mellom forklarings-
 variable vil da gi høy R_j her,
 noe som gir nevneren over liten
 og std/varians for estimatoren stor.
 Dette er ikke uvanlig for hvilken
 real disponibel inntekt eller realrente,
 som gjerne er relativt sterkt
 korrelert fra år til år. ~~For~~ Inkluderer
 da lag av disse i modellen (1)
 vil det kunne gi høye standard-
 avvik, og vanskelige å gjøre inferens
 på ikke-signifikante koeffisienter, selv
 om variablene samlet sett har

betydelig effekt (uov som synes i III, der multi kollineariteten ikke lenger er så stor).

Oppgave 2

a) Ser først at samtlige forklaringsvariabler har signifikante effekter ~~forholdig~~ på sannsynligheten for at kunden velger Cohe, ettersom endring i Cohe i en lineær sannsynlighetsmodell tiltes som endring i sannsynligheten for at dummyvariabelen tar verdi 1. Gitt igjen MLR.1-MLR.6 kan vi igjen bruke T-observatører tilsvarende som i oppg. 1 d. chs for å teste $H_0: \text{Disp-Cohe} = 0$ mot $H_1: \text{Disp-Cohe} \neq 0$. Med $n = 1140$ vil t-fordelingen gå mot standard normalfordeling, slik at vi kan bruke kritisk verdi 1,96 ved en tosidig test med 5% sign. nivå. $T = \frac{0,0772 - 0}{0,0344} = 2,24$

$> 1,96 \Rightarrow H_0$ forkastes. Tilsvarende gir $t = -4,65$ for Disp-Pepsi \Rightarrow forkaster $H_0: \text{Disp-Pepsi} = 0$, og $t = -6,54$ for P \Rightarrow forkaster $H_0: P = 0$ siden $|t| > 1,96$.

Denne kolonnen er
forbeholdt sensor
This column is for
external examiner

-0,1657 tolkes som reduksjonen i sannsynlighet for at kunden kjøper Coke hvis man begynner å reklamere for Pepsi i butikken, alt annet likt. (Altså $-0,1657 \cdot 100\% = -16,57\%$)

0,0772 tolkes som økt sannsynlighet for at kunden kjøper Coke hvis man begynner å reklamere for Coke i butikken, alt annet likt (Ceteris paribus, altså $0,0772 \cdot 100\% = 7,72\%$).

-0,4009 tolkes som reduksjonen i sannsynlighet for at kunden kjøper Coke dersom prisforholdet øker med én enhet, som her betyr 100% økning i prisen på Coke relativt til Pepsi, noe som estimert altså gir 40% lavere sannsynlighet for at kunden kjøper Coke. 100% prisøkning er imidlertid veldig mye, så tilnærmingen er sannsynligvis bedre hvis vi øker p med 0,1 eller 0,01 som betyr hhv 10% eller 1% økning i relativ pris på Coke (med effekt på sanns. for kjøp av Coke på ^{hhv} $-4,009\%$ ^{og} ~~eller~~ $-0,4009\%$). Slike sett kan denne tolkes som en slags relativ priselastisitet.

b)
Som allerede nevnt i a) vil en ~~relativ~~ prisstendring på 10% på Coke relativt til Pepsi bety en endring i P på 0,1, og en estimert effekt på Coke alt annet likt på $0,1 \cdot (-0,4009) = -0,04009 \cdot 100\% = -4,009\%$.
Altså 4,009% lavere sannsynlighet for at kunden velger Coke.

c)
Dette er også kommentert i a), der vi anslo at reklame for Pepsi i butikken alt annet likt vil gi 16,57% lavere sannsynlighet for at kunden kjøper Coke.

Merk at man må være forsiktig med å tolke resultatene fra en slik modell som gyldige veldig langt unna dagens likevekt. En femdobling av prisen på Coke gir f.eks $5 \cdot (-0,4009) = -2,0045 \cdot 100\% = -200,45\%$ mindre sannsynlighet for at kunden kjøper Coke, noe som gir negativ sannsynlighet, og dermed åpenbart ikke gir mening.

Oppgave 3

a)

Antar en sann modell gitt ved

$$(1) y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i,$$

 mens vi utelater x_2 som er en relevant variabel, og ser hva som skjer med MKM-estimatoren for β_1 hvis vi ~~er~~ antar/brukes modellen

$$(2) y_i = \beta_0 + \beta_1 x_{i1} + v_i, \text{ der } v_i = \beta_2 x_{i2} + u_i$$

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{i1} - \bar{x}_1)}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2} \text{ er MKM-estimatoren}$$

 for β_1 fra (2). Har fra (1) at

$$y_i - \bar{y} = \beta_1 (x_{i1} - \bar{x}_1) + \beta_2 (x_{i2} - \bar{x}_2) + (u_i - \bar{u}),$$

 og setter dette inn i $\tilde{\beta}_1$:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^n ((\beta_1 (x_{i1} - \bar{x}_1) + \beta_2 (x_{i2} - \bar{x}_2) + (u_i - \bar{u})) (x_{i1} - \bar{x}_1))}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$$

$$\tilde{\beta}_1 = \beta_1 \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2} + \beta_2 \frac{\sum_{i=1}^n (x_{i2} - \bar{x}_2)(x_{i1} - \bar{x}_1)}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2} + \frac{\sum_{i=1}^n u_i (x_{i1} - \bar{x}_1)}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$$

 der vi har brukt at $\bar{u} \sum_{i=1}^n (x_{i1} - \bar{x}_1) = 0$.

 Tar forventning betinget på $X(x_{i1}, x_{i2})$ på begge sider og får:

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

$$E(\tilde{\beta}_1 | X) = \beta_1 + \beta_2 \delta + \frac{\sum_{i=1}^n E(u_i | X)(x_{i1} - \bar{x}_1)}{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}$$

der vi har $E(u_i | X)$ fra MLR.4, slik at

$E(\tilde{\beta}_1 | X) = \beta_1 + \beta_2 \delta$, og vi ser at estimatoren fra en modell med utelatt variabel er forventningsshjev. Tilsvarende kan det vises at $\tilde{\beta}_1$ er inkonsistent ved å ta sannsynlighetsgrensen til uttrykket nederst på forrige side, der vi setter inn $\frac{1}{n-1}$ foran de to siste leddene, og får

$\text{plim } \tilde{\beta}_1 = \beta_1 + \beta_2 \delta$ når vi utnytter at $\text{Cov}(x_{i1}, u) = 0$ fra MLR.4, og ser at δ er MKM-estimatoren fra regresjonen $x_{i1} = \delta_0 + \delta_1 x_{i2} + u$.

Altså ser vi at utelatt variabel generelt gir ~~shjev~~ forventningsshjev og inkonsistente estimatører så lenge $\beta_2 \neq 0$ (utelatt variabel har effekt på y) og $\delta \neq 0$ (utelatt variabel er korrelert med forklaringsvariable).

Målefeil kan gi problemer på ulike måter. Målefeil i y :

Sann modell $y^* = \beta_0 + \beta_1 x + u$

Vi ~~er~~ har måling av y^* gitt ved ~~y^*~~ $y = y^* + \varepsilon$

Setter inn i sann modell og får

$$y - \varepsilon = \beta_0 + \beta_1 x + u$$

$y = \beta_0 + \beta_1 x + u + \varepsilon = \beta_0 + \beta_1 x + v$, noe som ikke gir skjervhet eller inkonsistens såfremt målefeilen ε i y ikke er korrelert med x , noe som normalt er en rimelig antagelse.

Mer imidlertid at vi får høyere varians siden $\text{Var}(v) = \text{Var}(u) + \text{Var}(\varepsilon)$ (gitt $\text{Cov}(u, \varepsilon) = 0 = \sigma_u^2 + \sigma_\varepsilon^2 > \sigma_u^2$).

Anta nå målefeil i forklaringsvariabel slik at sann modell er

$y = \beta_0 + \beta_1 x^* + u$, mens vi har $x = x^* + e$, der $\text{Cov}(u, e) = 0$, og enten $\text{Cov}(x, e) = 0$ eller $\text{Cov}(x^*, e) = 0$.

Sann modell gir innsett for x^* :

$$y = \beta_0 + \beta_1 (x - e) + u$$

$= \beta_0 + \beta_1 x - \beta_1 e - u = \beta_0 + \beta_1 x + v$, som er den vi estimerer. Hvis $\text{Cov}(x^*, e) \neq 0$ har vi at $\text{Cov}(x, e) = 0$, og $\text{Cov}(v, x) = -\beta_1 \text{Cov}(e, x) = 0$, og ingen skjervhet. $\text{Cov}(x^*, e) = 0$ gir imidlertid $\text{Cov}(x, e) \neq 0$, og $\text{Cov}(v, x) = \text{Cov}(-\beta_1 e - u, x) = -\beta_1 \text{Cov}(e, x) \neq 0$, og brudd

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

På MLR.4. Ser altså at utragnet
 i og for seg stemmer i form
 av at både målefeil i forklarings-
 variable og utelatte forklaringsvariable
~~kan gi skjevhet og inkonsis-~~
 kan gi forventningsshjevhet og
 inkonsistente estimatører. Problemet
 med forventningsshjevhet / inkonsistens
 grunnet målefeil kan løses via
 instrument variabel metoden hvis vi
 kan finne et instrument som er
 korrelert med x (altså målingen),
 men ikke med ~~målefeilen e~~ .
 ~~$(\text{Cov}(x, z) \neq 0, \text{Cov}(z, e) = 0$~~ (restleddet v
 (som inneholder målefeilen e) ($\text{Cov}(z, v) = 0$).
 Dette kan f.eks. forsøkes ved en andre
 måling fra annen kilde, f.eks. ved å
 se ektefelle eller arbeidsgivers oppgi
 en persons inntekt hvis denne
 i utgangspunktet oppga feil inntekt
 med målefeil. Ved utelatte forklarings-
 variable er et alternativ å forsøke
 å anslå retningen på skjevheten
 gjennom å resonnerer rundt sannsynlig
 fortegn på β_2 og δ . En signifikant
 positiv $\hat{\beta}_1$ kan f.eks. være et verdi-
 fullt resultat til tross for skjevhet/inkons.
 hvis man kan argumentere klart
 for at β_2 og δ har ulikt fortegn,
 slik at $E(\hat{\beta}_1) < \beta_1$.

Denne kolonnen er
forbeholdt sensor

 This column is for
external examiner

Videre er bruk av en proxy for utelatt variabel en mulighet for å fange opp i alle fall deler av den utelatte effekten. Slike f.eks IQ kan brukes som proxy for mye uønskede "innate ability" eller medfødte evne forskjeller. Har man paneldata tilgjengelig kan man eliminere effekten av utelatte variabelshverhet på flere måter. Hvis man har utelatt en aggregert variabel som kun varierer over tid vil tidsdummyer ha denne effekten. Mer vanlig er imidlertid uobserverbar heterogenitet, der utelatt variabel kun varierer mellom, og kanskje innad hos enhetene (igjen tenk f.eks ability). Om dette er tilfellet kan restleddet dekomponeres $u_{it} = \eta_i + \varepsilon_{it}$ der

$$E(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2 & \text{hvis } j=i \text{ og } t=s \\ 0 & \text{ellers} \end{cases}$$

$$E(\eta_i, \eta_j) = \begin{cases} \sigma_\eta^2 & \text{hvis } j=i \\ 0 & \text{ellers} \end{cases}$$

$$E(\eta_j, \varepsilon_{it}) = 0 \text{ for alle } j, i \text{ og } t.$$

$$E(\varepsilon_{it} | X) = 0, \text{ og vi mistenker at}$$

$$E(\eta_i | X) = 0$$

Denne kolonnen er forbeholdt sensor
This column is for external examiner

Er dette tilfellet kan vi bruke first difference eller fixed effects / within groups transformasjon der vi transformerer bort all variasjon mellom ~~gruppe~~ enheter. Vi kan også inkludere enhets spesifikke dummies som vil ha samme effekt. Dette løser endogenitets problemet, men utnytter ikke variasjonen i data settet mellom enheter. Har vi uløst variabel som varierer i begge dimensjoner er saken vanskeligere, så vi kan si at utsagnet har et poeng, men at mange slike problemer kan løses økonometrisk.

b)

i)

Brukes F-testen med R^2 istedet for SSR, siden forklart varians / $R^2 = 0$ for restricted model her:

$$F = \frac{\frac{0,35}{2}}{\frac{(1-0,35)}{47}} = 12,65 \text{ for første relasjon.}$$

$$F = \frac{\frac{0,11}{2}}{\frac{(1-0,11)}{47}} = 2,9 \text{ for andre relasjon.}$$

Denne kolonnen er forbeholdt sensor
This column is for external examiner

Fcrit fra tabell for begge tester er ca 3,2 med sign. nivå 5%.

Forkaster altså selvfølgelig $H_0: Exec=0, Uuem=0$ men beholder H_0 men forkaster ikke $H_0: \Delta Exec=0, \Delta Uuem=0$ på 5% sign. nivå.

ii) Modell (1) predikerer $\approx 10.0,79 = 7,9$ flere mord pr 100000 innbyggere $\$$ ved $\$$ hning på 10 kensetelser. Tilsvarende fra modell (2) er $10 \cdot (-0,1) = -1$, altså ett færre mord pr 100000 innbyggere.

iii) Alle variabler er undersøkt over et kort tidsrom, slik at modell (1) er sårbar for trender. Hvis alle variabler hver for seg trender samme retning i måleperiodene kan det gi ~~tilsynelatende~~ ~~klar~~ stor samvariasjon uten ~~at~~ at det sier noe om kausalitet. Trender i ~~begge~~ ^{ikke} ~~begge~~ ^{men} i ~~alle~~ ^{men} de to variablene, så jeg finner modell (2) mest troverdig. Dette understøttes av forventet fortegn på $\Delta Executions$, $\Delta Uuem$ ~~er~~ ~~jeg~~ ~~ville~~ ~~jeg~~ forventet med positivt fortegn i den grad denne skulle ha effekt på ^{morder}, men denne er altså ikke signifikant i (2), noe $\Delta Executions$ er (t-tester som i oppg. 7)

Denne kolonnen er
forbeholdt sensor

This column is for
external examiner

R^2 på 0,35 i første modell viser
også unimelig høyt med kun
disse to faktorene hvis vi tenker
kausaltitet.