Faktor

- en eksamensavis utgitt av ECONnect

Eksamensbesvarelse:
SØK3001 – Økonometri I

Eksamen:	Vår 2011
Antall sider:	20

Om ECONnect:

ECONnect er en frivillig studentorganisasjon for studentene på samfunnsøkonomi- og finansøkonomistudiet ved NTNU. Vi arbeider for økt faglig kompetanse blant våre studenter samt tettere kontakt med næringslivet. Det gjør vi ved å arrangere fagdager, gjesteforelesninger, bedriftspresentasjoner m.m. I dag går det ca. 200 studenter på bachelornivå (1.-3. klasse) og ca. 70 studenter på masternivå (4.-5. klasse). Studentene på masternivå er fordelt på de to linjene samfunnsøkonomi (ca. 50 stk) og finansiell økonomi (ca. 20 stk). Mer om ECONnect og aktuelle arrangementer på www.econnect-ntnu.no.

ECONnect består av følgende personer ved utgivelsestidspunkt:

Ole Christian Grytten(Leder)	ole@econnect-ntnu.no
Daniel Johansson(Bedriftsansvarlig)	daniel@econnect-ntnu.no
Johan Berg Fossen(Fagdagsansvarlig)	johan@econnect-ntnu.no
Mariell Toven(Økonomi/Kandidattreffet)	mariell@econnect-ntnu.no
Georg Næsheim	georg@econnect-ntnu.no
Ellen Normann	ellen@econnect-ntnu.no
Ragnhild Grøv	ragnhild@econnect-ntnu.no
Martine Ødegård (Faktoransvarlig)	martine@econnect-ntnu.no
Inga Friis	inga@econnect-ntnu.no
Caroline Lesiewicz	caroline@econnect-ntnu.no

Post- og besøksadresse:	*Organisasjonsnummer:*	*Hjemmeside:*
ECONnect, NTNU Dragvoll Institutt for samfunnsøkonomi Bygg 7, Nivå 5 7491 Trondheim	NO 994 625 314	www.econnect-ntnu.no

Merk: Eksamensbesvarelsene har i varierende grad feil og mangler, både oppsett og innhold. De vil også kun vise en av flere mulige fremgangsmåter. ECONnect står ikke ansvarlig for selve faginnholdet.

Kommentar til besvarelse SØK3001 vår 2011 – kandidat 10022

Dette er en svært god besvarelse der alle delspørsmål er klart tilfredsstillende besvart. Ved denne eksamenen var det flere meget gode besvarelser, men en grunn til å velge denne som «mønsterbesvarelse» er et høyt presisjonsnivå og svært lite utenomsnakk. Dette gjelder spesielt oppgave 2.

Litt pirk:

Under oppgave 1a) diskuteres ikke hva som påvirker størrelsen og retningen på skjevheten, men dette går imidlertid klart fram under 1c).

Under 1e) er ikke eksempelet med sentrale lønnsoppgjør det beste i denne sammenheng. Nasjonale konjunkturer og ikke minst renteutviklingen ville vært mer nærliggende.

Helt mot slutten av oppgave 2d) diskuteres variansen til IV-estimatoren. Denne diskusjonen gis det ikke mye uttelling for, men helt klart – den trekker ikke ned.

Kåre Johansen

Eksamen SØK3001 Vår 2011

Bokmål
Oppgave 1

I en empirisk undersøkelse benyttes først tverrsnittsdata for å estimere sammenhengen mellom regionale boligpriser og regional inntekt. En av relasjonene som estimeres er gitt ved

(1) $y_i = \beta_0 + \beta_1 x_{1i} + u_i$

der y_i er logaritmen til boligprisen i region i, x_{1i} er logaritmen til inntekt per capita i region i og u_i et stokastisk restledd.

a) Et potensielt problem ved relasjon (1) er at relevante forklaringsvariable er utelatt. Forklar hva slags problem dette skaper ved estimering av effekten av inntekt på boligprisen.

I analysen rapporteres følgende resultater basert på data for 90 regioner:

(2) $\hat{y}_i = \underset{(0.11)}{-1.85} + \underset{(0.17)}{2.28}\, x_{1i}, \quad SSR = 1.899$

(3) $\hat{y}_i = \underset{(0.14)}{-2.14} + \underset{(0.18)}{2.02}\, x_{1i} + \underset{(0.010)}{0.032}\, x_{2i}, \quad SSR = 1.694$

der den nye variabelen i ligning (3), x_{2i}, er gjennomsnittlig sommertemperatur i region i. Begge relasjonene er estimert ved bruk av minste kvadraters metode (MKM). Tall i parenteser under de estimerte parametrene er estimerte standardavvik mens SSR er summen av kvadrerte residualer.

b) Gi en tolkning av resultatene i ligning (3) og test om de partielle effektene av regional inntekt og sommertemperatur er signifikant forskjellig fra null.

c) Diskuter hvorfor den estimerte effekten av regional inntekt reduseres når sommertemperatur inkluderes. Benytt i den forbindelse resultatene fra følgende hjelperegresjon

$$\hat{x}_{2i} = 9.15 + 8.19 x_{1i}$$

Boligprismodellen utvides videre med to nye variable, x_{i3} som er regional arbeidsledighet og x_{i4} som er gjennomsnittlig vintertemperatur. Resultatene for den utvidede modellen er gitt ved

(4) $\hat{y}_i = \underset{(0.17)}{-2.19} + \underset{(0.19)}{2.04}\, x_{1i} + \underset{(0.010)}{0.032}\, x_{2i} + \underset{(0.016)}{0.008}\, x_{3i} - \underset{(0.0043)}{0.0005}\, x_{4i}, \quad SSR = 1.689$

d) Forklar hvordan vi kan teste om relasjon (3) er en gyldig forenkling av relasjon (4) og gjennomfør testen.

I resten av denne oppgaven diskuteres resultater basert på paneldata for de 90 regionene for en periode på 12 år, i alt 1080 observasjoner Alle modellene som estimeres inkluderer et fullt sett av tidsdummyer (faste tidseffekter). Estimerte parametre foran disse dummyvariablene samt konstantleddet rapporteres ikke i ligningene under.

e) Drøft kort argumenter for å inkludere tidsdummyer i paneldatamodellen for regionale boligpriser.

Ved bruk av MKM uten transformasjon av variablene (pooled OLS) får vi følgende resultater:

(5) $\hat{y}_{it} = \underset{(0.19)}{2.06}\, x_{1it} + \underset{(0.0086)}{0.024}\, x_{2it}$

Når vi benytter within-groups transformasjonen (faste regioneffekter) får vi resultatene:

(6) $\hat{y}_{it} = \underset{(0.35)}{1.00}\, x_{1it} + \underset{(0.0027)}{0.002}\, x_{2it}$

f) Sammenlign resultatene i ligning (5) og (6) og diskuter mulige grunner til at disse endres som følge av endret estimeringsmetode (eller transformasjon).

Til slutt estimeres en dynamisk modell for regionale boligpriser der lagget boligpris inkluderes som høyresidevariabel. Resultatet fra denne estimeringen er gitt ved

(7) $\hat{y}_{it} = \underset{(0.065)}{0.37}\, y_{it-1} + \underset{(0.26)}{0.60}\, x_{1it} + \underset{(0.0028)}{0.002}\, x_{2it}$

g) I hvilken grad tyder resultatene i ligning (7) på tregheter i tilpasningen av boligprisen? Finn til slutt de langsiktige effektene av regional inntekt og sommertemperatur.

Oppgave 2

Betrakt følgende modell for et frikonkurransemarked:

(1) $x_t = \beta_0 + \beta_1 p_t + u_{1t}$

(2) $x_t = \alpha_0 + \alpha_1 p_t + \alpha_2 S_t + u_{2t}$

hvor ligning (1) er etterspørselsrelasjonen, ligning (2) er tilbudsrelasjonen, x_t er omsatt kvantum, p_t er pris, S_t er en eksogen variabel som påvirker tilbudet og antas uavhengig av restleddene i både ligning (1) og (2), u_{1t} og u_{2t}.

a) Forklar hvorfor estimering av (1) ved bruk av MKM vil gi en inkonsistent estimator for parameteren β_1.

b) Begrunn at ligning (1) er eksakt identifisert mens ligning (2) ikke er identifisert.

c) Forklar hvordan du kan oppnå en konsistent estimator for β_1 ved bruk av instrumentvariabelmetoden når tilbudsfunksjonen er gitt ved (2).

Anta nå at markedsmodellen er gitt ved

(1*) $x_t = \beta_0 + \beta_1 p_t + \beta_2 D_t + u_{1t}$

(2*) $x_t = \alpha_0 + \alpha_1 p_t + \alpha_2 S_{1t} + \alpha_3 S_{2t} + u_{2t}$

der D_t, S_{1t} og S_{2t} alle er eksogene variable som antas uavhengige av de to restleddene.

d) Forklar hvordan du nå vil gå fram for å estimere parametrene i ligning (1*). Forklar også hvordan du kan teste om de instrumentvariablene som benyttes gir tilstrekkelig informasjon til å identifisere parametrene i ligning (1*) empirisk.

Nynorsk

Oppgåve 1

I ei empirisk undersøking nyttes først tverrsnittsdata for å estimere samanhengen mellom regionale bustadprisar og regional inntekt. Ein av relasjonane som vert estimert er gitt ved

(1) $y_i = \beta_0 + \beta_1 x_{1i} + u_i$

der y_i er logaritmen til bustadprisen i region i, x_{1i} er logaritmen til inntekt per capita i region i og u_i et stokastisk restledd.

a) Et potensielt problem ved relasjon (1) er at relevante forklaringsvariable er utelatne. Forklar kva slags problem dette skaper ved estimering av effekten av inntekt på bustadprisen.

I analysen vert desse resultata rapportert basert på data for 90 regionar:

(2) $\hat{y}_i = \underset{(0.11)}{-1.85} + \underset{(0.17)}{2.28}\, x_{1i}, \quad SSR = 1.899$

(3) $\hat{y}_i = \underset{(0.14)}{-2.14} + \underset{(0.18)}{2.02}\, x_{1i} + \underset{(0.010)}{0.032}\, x_{2i}, \quad SSR = 1.694$

der den nye variabelen i likning (3), x_{2i}, er gjennomsnittlig sommartemperatur i region i. Begge relasjonane er estimert ved bruk av minste kvadraters metode (MKM). Tal i parentesar under dei estimerte parametrane er estimerte standardavvik mens SSR er summen av kvadrerte residualar.

b) Gje ei tolking av resultata i likning (3) og test om dei partielle effektane av regional inntekt og sommartemperatur er signifikant forskjellig frå null.

c) Diskuter kvifor den estimerte effekten av regional inntekt vert redusert når sommartemperatur er inkludert. Bruk i den samanhengen resultata frå denne hjelperegresjonen:

$$\hat{x}_{2i} = 9.15 + 8.19 x_{1i}$$

Bustadprismodellen vert vidare utvida med to nye variablar, x_{i3} som er regional arbeidsløyse og x_{i4} som er gjennomsnittlig vintertemperatur. Resultata for den utvida modellen er gitt ved

(4) $\hat{y}_i = \underset{(0.17)}{-2.19} + \underset{(0.19)}{2.04}\, x_{1i} + \underset{(0.010)}{0.032}\, x_{2i} + \underset{(0.016)}{0.008}\, x_{3i} - \underset{(0.0043)}{0.0005}\, x_{4i}, \quad SSR = 1.689$

d) Forklar korleis vi kan testa om relasjon (3) er ein gyldig forenkling av relasjon (4) og gjennomfør testen.

I resten av denne oppgåva vert resultat basert på paneldata for dei 90 regionane for en periode på 12 år, i alt 1080 observasjonar diskutert. Alle modellane som vert estimert inkluderer eit fullt sett av tidsdummyar (faste tidseffektar). Dei estimerte parametrane til desse dummyvariablane samt konstantleddet vert ikkje rapportert i likningane under.

e) Drøft kort argument for å inkludere tidsdummyar i paneldatamodellen for regionale bustadprisar.

Ved bruk av MKM utan transformasjon av variablane (pooled OLS) får vi desse resultata:

(5) $\hat{y}_{it} = \underset{(0.19)}{2.06}\, x_{1it} + \underset{(0.0086)}{0.024}\, x_{2it}$

Når vi nytter within-groups transformasjonen (faste regioneffektar) får vi resultata:

(6) $\hat{y}_{it} = \underset{(0.35)}{1.00}\, x_{1it} + \underset{(0.0027)}{0.002}\, x_{2it}$

f) Samanlikn resultata i likning (5) og (6) og diskuter moglege grunnar til at desse vert endra på grunn av endra estimeringsmetode (eller transformasjon).

Til slutt vert ein dynamisk modell for regionale bustadprisar estimert der lagga bustadpris er inkludert som høgresidevariabel. Resultatet frå denne estimeringa er gitt ved

(7) $\hat{y}_{it} = \underset{(0.065)}{0.37}\, y_{it-1} + \underset{(0.26)}{0.60}\, x_{1it} + \underset{(0.0028)}{0.002}\, x_{2it}$

g) I kva grad tyder resultata i likning (7) på tregleikar i tilpassinga av bustadprisen? Finn til slutt dei langsiktige effektane av regional inntekt og sommartemperatur.

Oppgåve 2

Betrakt denne modellen for ein marknad med fri konkurranse:

(1) $x_t = \beta_0 + \beta_1 p_t + u_{1t}$

(2) $x_t = \alpha_0 + \alpha_1 p_t + \alpha_2 S_t + u_{2t}$

der likning (1) er etterspurnadsrelasjonen, likning (2) er tilbodsrelasjonen, x_t er omsett kvantum, p_t er pris, S_t er ein eksogen variabel som påverkar tilbodet og vert antatt uavhengig av restledda i både likning (1) og (2), u_{1t} og u_{2t}.

a) Forklar kvifor estimering av (1) ved bruk av MKM vil gje ein inkonsistent estimator for parameteren β_1.

b) Grunngje at likning (1) er eksakt identifisert mens likning (2) ikkje er identifisert.

c) Forklar korleis du kan oppnå ein konsistent estimator for β_1 ved bruk av instrumentvariabelmetoden når tilbodsfunksjonen er gitt ved (2).

Anta no at marknadsmodellen er gitt ved

(1*) $x_t = \beta_0 + \beta_1 p_t + \beta_2 D_t + u_{1t}$

(2*) $x_t = \alpha_0 + \alpha_1 p_t + \alpha_2 S_{1t} + \alpha_3 S_{2t} + u_{2t}$

der D_t, S_{1t} og S_{2t} alle er eksogene variablar som vert antatt uavhengige av dei to restledda.

d) Forklar korleis du no vil gå fram for å estimere parametrane i likning (1*). Forklar og korleis du kan teste om dei instrumentvariablane som vert nytta gir tilstrekkeleg informasjon til å identifisere parametrane i likning (1*) empirisk.

English

Question 1

An empirical study uses first cross section data to estimate the relation between regional housing prices and regional income. One of the estimated models is given by

(1) $y_i = \beta_0 + \beta_1 x_{1i} + u_i$

where y_i is the logarithm of the housing price in region i, x_{1i} is the logarithm of per capita income in region i and u_i is a random error term.

a) A potential problem with equation (1) is that relevant explanatory variables are excluded. Explain what kind of problem excluded variables will induce when estimating the effect of income on housing prices.

The analyses reports the following results based on data for 90 regions:

(2) $\hat{y}_i = -\underset{(0.11)}{1.85} + \underset{(0.17)}{2.28}\, x_{1i}, \quad SSR = 1.899$

(3) $\hat{y}_i = -\underset{(0.14)}{2.14} + \underset{(0.18)}{2.02}\, x_{1i} + \underset{(0.010)}{0.032}\, x_{2i}, \quad SSR = 1.694$

where the new variable in equation (3), x_{2i}, is average summer temperature in region i. Both equations are estimated using the method of ordinary least squares (OLS). Numbers in parentheses below the estimated parameters are estimated standard errors and SSR is the sum of squared residuals.

b) Give an interpretation of the results in equation (3) and test whether or not the partial effects of regional income and summer temperature are significantly different from zero.

c) Discuss why the estimated effect of regional income is reduced when summer temperature is included. Utilise here the results from the following auxiliary regression:

$$\hat{x}_{2i} = 9.15 + 8.19 x_{1i}$$

The housing price model is further expanded with two new variables, x_{i3} which is regional unemployment and x_{i4} which is average winter temperature. The results for the expanded model are given by:

(4) $\hat{y}_i = -\underset{(0.17)}{2.19} + \underset{(0.19)}{2.04}\, x_{1i} + \underset{(0.010)}{0.032}\, x_{2i} + \underset{(0.016)}{0.008}\, x_{3i} - \underset{(0.0043)}{0.0005}\, x_{4i}, \quad SSR = 1.689$

d) Explain how we can test whether or not equation (3) is a valid simplification of equation (4) and perform the test.

The remaining part of this question discusses results based on panel data for the 90 regions for a period of 12 years, in total 1080 observations. All models include a complete set of time dummies (time fixed effects). The estimated parameters of these time dummies and the constant term are not reported in the equations below.

e) Discuss briefly arguments in favour of including time dummies in the panel data model for regional housing prices.

Using OLS without making any transformation of the variables (pooled OLS) gives the following results:

(5) $\hat{y}_{it} = \underset{(0.19)}{2.06}\, x_{1it} + \underset{(0.0086)}{0.024}\, x_{2it}$

Using the within-groups transformation (fixed region effects) gives the results:

(6) $\hat{y}_{it} = \underset{(0.35)}{1.00}\, x_{1it} + \underset{(0.0027)}{0.002}\, x_{2it}$

f) Compare the results in equation (5) and (6) and discuss possible reasons why the estimated effects change due to changed estimation method (or transformation).

Finally, we estimate a dynamic model for regional housing prices which includes lagged housing prices as a right hand side variable. The result from this estimation is given by:

(7) $\hat{y}_{it} = \underset{(0.065)}{0.37}\, y_{it-1} + \underset{(0.26)}{0.60}\, x_{1it} + \underset{(0.0028)}{0.002}\, x_{2it}$

g) To what extent do the results in equation (7) indicate sluggish adjustment of the housing price? Finally, find the long run effects of regional income and summer temperature.

Question 2

Consider the following model for a perfectly competitive market:

(1) $x_t = \beta_0 + \beta_1 p_t + u_{1t}$

(2) $x_t = \alpha_0 + \alpha_1 p_t + \alpha_2 S_t + u_{2t}$

where equation (1) is the demand equation, equation (2) is the supply equation, x_t is quantity sold, p_t is the market price, S_t is an exogenous variable that affects supply and is assumed to be independent with the error terms in both equation (1) and (2), u_{1t} and u_{2t}.

a) Explain why estimating equation (1) using OLS will give an inconsistent estimator for the parameter β_1.

b) Explain why equation (1) is exactly identified while equation (2) is not identified.

c) Explain how you can obtain a consistent estimator for β_1 using the instrumental variable method when the supply equation is given by (2).

Assume now that the market model is given by

(1*) $x_t = \beta_0 + \beta_1 p_t + \beta_2 D_t + u_{1t}$

(2*) $x_t = \alpha_0 + \alpha_1 p_t + \alpha_2 S_{1t} + \alpha_3 S_{2t} + u_{2t}$

where D_t, S_{1t} and S_{2t} are all exogenous variables assumed to be independent with the two error terms.

d) Explain how you would now proceed to estimate the parameters in equation (1*). Explain also how you can test whether or not the instrumental variables used give sufficient information to empirically identify the parameters in equation (1*).

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 1

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor

This column is for external examiner

Først i oppgaven vil jeg begynne med å sette opp antagelsene for en klassisk multippel regresjonsmodell på generell form da disse antagelsene og eventuelt brudd på disse vil bli referert til senere i oppgaven. Antagelsene er:

① Populasjonsmodellen er lineær i parameterne,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i$$

② Vi ser på et tilfeldig utvalg fra populasjonen når vi forsøker estimere modellen. Dette innebærer at $cov(u_j, u_i) = 0$ for $j \neq i$

③ Vi har ingen perfekt kollinearitet. Dette vil si at ingen av forklarings variablene kan skrives som eksakte kombinasjoner av andre forkl. variable.

④ Den betingede forventningen til restleddet er null, $E(u_i|X) = 0$, hvor X er en vektor bestående av $x_{1i}, x_{2i}, \dots x_{ik}$. Dette impliserer $E(u_i) = 0$ og $cov(u_i, x_i) = 0$

⑤ Homoskedastisitet. Dette vil si at den betingede variansen til restleddet er konstant $var(u_i|X) = \sigma^2$

⑥ Restleddet er normalfordelt $u_i \sim N(0, \sigma^2)$.

OPPGAVE 1

Relasjon 1 er en enkel regresjonsmodell, de samme antagelsene som over gjelder, men istedenfor at vi krever at vi ikke skal ha perfekt kollinearitet krever vi nå at det skal være variasjon i x_{1i}.

Problemet dersom vi utelater relevante forklaringsvariable er at $cov(u_i, x_i)$ ikke

Den gule kopien beholder du/*Please keep the yellow page*

Tapir Uttrykk 04/10

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 2

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor
This column is for external examiner

ingen vil være null, altså antagelse ④ vil ikke holde. Jeg skal imidlertid først utlede estimatoren til β_1 ved hjelp av MKM som om antagelsene stemmer.

Ved MKM ønsker vi å finne verdier på β_0 og β_1 som minimerer summen av det kvadrerte avviket mellom faktisk y og predikert y, i dette tilfellet bolig prisen. Dette avviket kalles residualene. Presist formulert er problemet:

$$\min \sum_{i=1}^{n} (y_i - \hat{y}_i) = \sum_{i=1}^{n} (\hat{u}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1})^2$$

Dersom vi deriverer denne mhp. $\hat{\beta}_0$ og $\hat{\beta}_1$ finner vi to førsteordenskrav, som også kalles normalligningene:

$$\partial / \partial \hat{\beta}_0 = 0 \Rightarrow 2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1})(-1) = 0 \qquad ①$$

$$\partial / \partial \hat{\beta}_1 = 0 \Rightarrow 2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1})(-x_{i1}) = 0 \qquad ②$$

Vi ser at de to betingelsene impliserer hhv $\sum \hat{u}_i = 0$ og $\sum u_i x_{i1} = 0$. Videre fra relasjon ① har vi

$$\sum_{i=1}^{n} y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} \qquad / \cdot \tfrac{1}{n}$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 \qquad / \quad \text{hvor } \bar{y} = \tfrac{1}{n}\sum_{i=1}^{n} y_i = \text{gj.snitt til } y.$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$$

Dersom vi nå setter inn for $\hat{\beta}_0$ i minimeringsproblemet finner vi

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

Deriverer vi denne mhp. $\hat{\beta}_1$ og løser for $\hat{\beta}_1$, finner vi.

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 3

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor

This column is for external examiner

$$\partial/\partial\hat{\beta}_1 = 2\sum_{i=1}^{n}\left((y_i-\bar{y})-(x_{i1}-\bar{x}_1)\hat{\beta}_1\right)\left(-(x_{i1}-\bar{x}_1)\right) = 0$$

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

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

Dette vil være estimatoren til $\hat{\beta}_1$ i modell 1). Men anta nå at den samme modellen var gitt ved:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$$

Slik at $y_i - \bar{y} = \beta_1(x_{i1}-\bar{x}_1) + \beta_2(x_{i2}-\bar{x}_2) + (u_i-\bar{u})$. Dersom vi setter dette inn i uttrykket for estimatoren ~~til~~ $\hat{\beta}_1$, finner vi:

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

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

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

brukte at $\bar{u} = 0$.

Vi ser derfor at uttrykket for $\hat{\beta}_1$ nå ~~tre~~ inkluderer to ledd. For å forklare hva slags problemer utelatt variabel skaper skal jeg nå finne forventning og konsistens.

Tar først forventningen, og antar at antagelse (4) $E(u_i|x_{i1}, x_{i2})=0$ holder.

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

√

$$= \beta_1 + \beta_2 \hat{\delta}$$

Hvor $\hat{\delta}$ er MKM estimatoren på relasjonen $x_{i2} = \alpha + \delta x_{i1} +$ støy.
Vi ser derfor at hvis vi utelater en relevant forklaringsvariabel, dvs at $\beta_2 \neq 0$, så vil dette skape et problem for estimatoren til β_1, effekten av inntekt pr. capita dersom det er en sammenheng mellom x_{i2} og x_{i1}, noe som vil gi seg utslag i $\hat{\delta} \neq 0$. Vi ser derfor at å utelate relevante forklar variabl vil føre til forventningsskjeve estimater dersom β_2 og $\hat{\delta} \neq 0$.

Hva så med konsistens?

$$\text{plim}(\hat{\beta}_1) = \beta_1 + \beta_2 \frac{\text{plim}\, 1/n \sum (x_{i2}-\bar{x}_2)(x_{i1}-\bar{x}_1)}{\text{plim}\, 1/n \sum (x_{i1}-\bar{x}_1)^2} + \frac{\text{plim}\, 1/n \sum (u_i-\bar{u})(x_{i1}-\bar{x}_1)}{\text{plim}\, 1/n \sum (x_{i1}-\bar{x}_1)^2}$$

$$= \beta_1 + \beta_2 \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)} + \frac{\text{cov}(u, x_1)}{\text{var}(x_1)}$$

√

$$= \beta_1 + \beta_2 \frac{\text{cov}(x_1, x_2)}{\text{var}(x_1)}$$

Her har vi antatt at de empiriske momentene konvergerer mot sine teoretiske momenter, samt at $\text{cov}(u, x_1) = 0$. Det vi da ser er at dersom $\text{cov}(x_1, x_2) \neq 0$ vil vi ei heller få konsistente estimatorer ved å utelate en relevant forkl. variabel.

Oppsummert er problemet ved at relevante forklaringsvariable er utelatt at den estimerte effekten av inntekten på boligprisen vil være forventningsskjev og ikke konsistent.

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 5

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor

This column is for external examiner

(B)

Når det gjelder realleddet i denne relasjonen, så har det liten mening i seg selv da x_{1i}, innt. pr. capita ikke vil være null.

Videre kan vi tolke effekten av inntekt pr. capita slik at om denne øker med 1% i region i vil boligprisene i region i øke med 2,02%. Dette fordi begge, dvs y_i og x_{1i} er på logaritme form slik at forholdet de imellom tolkes som elastisiteter.

√

Sommertemperatur er derimot ikke logaritmisk og tolkningen av denne effekten er at dersom gjennomsnittlig sommertemperatur i en region i øker med 1 grad, vil boligprisene i regionen øke med $100 \cdot 0{,}032 = 3{,}2\%$.

√

For å teste om de partielle effektene er signifikant forskjellig fra null, velger jeg å bruke en t-test.

For å først forklare noe av bakgrunnen for denne, så kan vi begynne med at hvis antagelse 1-6 holder vil $\hat{\beta}_j$ være normalfordelt med forventning lik ~~β~~ β_j og varians lik $Var(\hat{\beta}_j)$. Videre vil da

$$\frac{\hat{\beta}_j - \beta_j}{sd(\hat{\beta}_j)}$$

være standardnormalfordelt med forventning lik null og varians lik 1. Dette er også omtrent den testobservatoren vi bruker ved t-testen, forskjellen er at vi ikke kjenner standardavviket, dette må estimeres, slik at testobservatoren da er:

√

$$\frac{\hat{\beta}_j - \beta_j}{\frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}}} \sim t_{n-k-1} \quad , \text{ hvor } \frac{\hat{\sigma}}{\sqrt{SST_j(1-R_j^2)}} \text{ er det estimerte std. avviket.}$$

som er t-fordelt med n-k-1 frihetsgrader. I testen vil da $\hat{\beta}_j$ være

NTNU

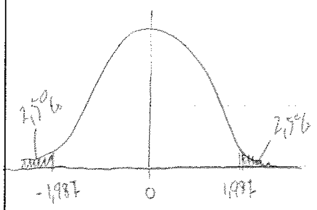
Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6 - 11 Side/*Page:* 6

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor
This column is for external examiner

estimert verdi mens β_j vil være den hypotetiske verdien. Det t-testen da gjør er å se hvor mange standardavvik fra sin hypotetiske verdi $\hat{\beta}$ er, for så å definere en kritisk verdi for når vi ikke lenger kan anta at den estimerte effekten er lik den hypotetiske. Denne kritiske verdien finner vi fra t-fordelingen. Gitt antall frihetsgrader må vi bestemme oss for et signifikansnivå og så finner vi kritisk verdi fra tabellen. Signifikansnivået vil jeg sette på 5%. Dette betyr at det er en 5% sjanse for å forkaste H_0 dersom denne er sann. I vårt tilfelle har vi $90-2-1 = 87$ frihetsgrader, noe som gir en kritisk verdi lik $\approx 1,987$ dersom vi kjører en to-sidig test. Det at vi kjører en to-sidig test betyr at vi



sammenligner absoluttverdien til testobservatoren med den kritiske verdien, slik at vi forkaster H_0 både dersom observert verdi er større enn 1,987 og mindre enn -1,987. Siden vi kjører et signifikansnivå på 5% vil det tilsvarende arealet til summen utgjøre 5% av arealet til fordelingen. Vi er nå klare til å kjøre testene.

<u>inntekt pr. capita</u>

Dersom vi skal se om denne er signifikant vil ~~vi~~ vi ha følgende hypoteser:

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Dette gir følgende testobservator $t_{\hat{\beta}_1} = \left| \frac{2,02 - 0}{0,18} \right| = 11,22$

Vi ser at $|t_{\hat{\beta}_1}| > t_{krit}$, $11,22 > 1,987$ slik at vi forkaster H_0, at inntekt pr. capita ikke har noe å ~~f~~ si for boligprisen.

<u>sommertemp.</u>

Dersom vi nå kaller estimatoren for denne effekten for ~~$\hat{\beta}_2$~~ $\hat{\beta}_2$, vil hypotesene være:

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6 - 11 Side/*Page:* 7

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor
This column is for external examiner

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

✓ Dette gir følgende testobservator $t_{\hat{\beta}_2} = \left|\frac{0{,}032}{0{,}010}\right| = 3{,}2$
Vi ser at $|t_{\hat{\beta}_1}| > t_{krit}$, $3{,}2 > 1{,}987$ slik at vi kan forkaste H_0, at gj.snittlig sommertemp ikke har noe å si for boligprisen. Konklusjonen er derfor at begge de partielle effektene er signifikant forskjellig fra null.

(c)

For å gjøre dette går vi tilbake til oppgave 1 a) hvor vi fant at den betingede forventningen til estimatoren for effekten av inntekt per capita var gitt ved:

✓ $$E(\hat{\beta}_1 | x_{i1}, x_{i2}) = \beta_1 + \beta_2 \hat{\delta}.$$

Vi har nå i oppgave 1 b) slått fast, ved en t-test at β_2, effekten av gj.snitt. temperaturer på husprisen er positiv og signifikant. Videre ser vi fra hjelpe-regresjonen at det er en klar positiv sammenheng mellom gj.snittlig sommer temp. i en region og inntekt per capita i regionen, slik at $\hat{\delta} > 0$. Vi har derfor at $\beta_2 \hat{\delta} = 0{,}032 \cdot 8{,}19 = 0{,}262$. Dette ser vi også at er det samme som avviket mellom $\hat{\beta}_1$ i relasjon 2) og $\hat{\beta}_1$ i relasjon 3), $2{,}28 - 2{,}02 = 0{,}26$. ✓
Slik at grunnen til at den estimerte effekten av regional inntekt reduseres når sommertemperaturer inkluderes er "utelatt varibel skjevhet", vi har utelatt en relevant forklaringsvaribel, i 2) som inkluderes i 3).

3,0

(D)

For å teste dette vil jeg bruke en F-test. F testens byggeklosser er summen av residualene kvadrert, SSR, eller $\Sigma \hat{u}_i^2$. Det som er poenget er at om du inkluderer flere forklaringsvariable vil SSR gå ned. Ved å bruke relasjon 4) som eksempel, og bytte ut leddene med estimatorene kan vi:

$$\Sigma(\hat{u}_i)^2 = \Sigma(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_1 x_{i2} - \hat{\beta}_3 x_{i3} - \hat{\beta}_4 x_{i4})^2$$

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 8

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor
This column is for external examiner

Vi skal nå teste om 3) er en gyldig forenkling av 4) og vi ser fra SSR til 4) at denne nødvendigvis vil øke dersom $\hat{\beta}_3$ og $\hat{\beta}_4$ settes lik null. Spørsmålet er bare om økningen er stor nok, da det vil være liten endring dersom $\hat{\beta}_3$ og $\hat{\beta}_4$ er små. For å finne ut hva som er stor nok danner vi F-testen sin testobservator som e gitt ved:

√

$$F = \frac{\frac{SSR_0 - SSR_4}{q}}{\frac{SSR_4}{n-k-1}}$$

Denne vil være F-fordelt under $H_0: \beta_3 = \beta_4 = 0$ med q, n-k-1 frihetsgrader, hvor q er antall restriksjoner som i dette tilfellet er 2. Videre er det og verdt å merke seg at vi antar at ①-⑥ gjelder for at F-testen skal være gyldig. Men vi trenger også her en testobservator, og denne finner vi fra F-fordelingen. Vi har i vårt tilfelle 2, 90-4-1 frihetsgrader, noe som gir en kritisk verdi på ca. 3,10 dersom vi velger et 5% signifikans-nivå. Dersom vi nå skal teste om 3) er en gyldig forenkling vil hypotesene være:

√

$H_0: \beta_3 = \beta_4 = 0$
$H_1: \beta_3$ eller $\beta_4 \neq 0$

Vi får så følgende testobservator

√

$$F = \frac{\frac{1{,}694 - 1{,}689}{2}}{\frac{1{,}689}{90-4-1}} = \underline{0{,}12}$$

Vi ser at $F_{OBS} < F_{KRIT}$, $0{,}12 < 3{,}10$, slik at vi kan ikke forkaste H_0, noe som ifølge testen betyr at 3) er en gyldig forenkling av relasjon 4)

E)

Begrunnelsen for å inkludere tidsdummyer kan og sees på som en metode for å unngå utelatt variabel skjevhet. For å se dette bedre kan vi nå anta at restleddet består av to deler:

$$u_{it} = \gamma_t + \varepsilon_{it}$$

Hvor γ_t er en komponent som varierer over tid, men ikke mellom regioner, mens ε_{it} varierer både over tid og mellom regioner. ε_{it} antas å være uavhengig av forklaringsvariablene med en betinget forventning lik null og konstant betinget varians. γ_t er det derimot verre med. For å illustrere, så kan vi anta at γ_t fanger opp forhold som for eksempel sentrale lønnsoppgjør. Disse bestemmes i stor grad nasjonalt, og har dermed selt en effekt også på regionale lønninger. Det er derfor urimelig å anta at regionale lønninger er uavhengig av de sentrale lønnsoppgjørene, og dermed at x_{it} er uavhengig av γ_t. Videre er det og rimelig at de sentrale lønnsoppgjørene har en effekt på boligprisene. For å bruke terminologien fra oppg. 1a) vil "$\beta_2 \delta$" $\neq 0$ og vi vil dermed få et utelatt relevant variabel problem. Dette løser vi da ved å inkludere tidsdummyer, for på denne måten vil γ_t, som da vi har illustrert med sentrale lønnsoppgjør, flyttes fra restleddet og til konstantleddet. Igjen i restleddet er da ε_{it} som vi har antatt har de ønskede egenskapene til et restledd.

F)

For å diskutere dette kan vi nok en gang se på restleddet. Anta nå at dette er på formen:

$$u_{it} = \eta_i + \varepsilon_{it}.$$

ε_{it} antas å ha de samme egenskapene som under E) mens η_i nå er en regionspesifikk komponent som varierer mellom regioner, men ikke over tid. Videre er det rimelig å anta at denne individspesifikke komponenten vil være avhengig av de inkluderte forklaringsvariable. For eksempel er geografisk beliggenhet noe som ikke endres over tid, men som kan ha stor betydning for

Denne kolonnen er forbeholdt sensor

This column is for external examiner

temperatur. Videre kan det og tenkes at en region har en befolkning som på en eller annen måte har evner som gjør at inntekten deres er forskjellig fra andre regioner. Slike evner kan antas å være konstante over tid og vanskelig å kvantifisere, slik at de vil typisk være inkludert i η_i og dermed være korrelert med x_{i1}, regionale lønninger. Vi har dermed nok en gang et utelatt relevant variabel problem, noe som vil gi skjevhet ved å bare estimere ved vanlig MKM.

Hva skjer så om vi benytter within-groups transformasjonen?

Vi begynner med å definere modellen i form av gj.snitt over tid.

$$\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it} \qquad \bar{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$$

$$\bar{X}_{1i} = \frac{1}{T}\sum_{t=1}^{T} X_{it1} \qquad \bar{\eta}_i = \frac{1}{T}\sum_{t=1}^{T} \eta_i$$

$$\bar{X}_{2i} = \frac{1}{T}\sum_{t=1}^{T} X_{it2}$$

$$\bar{Y}_i = \beta_0 + \beta_1 \bar{X}_{i1} + \beta_2 \bar{X}_{i2} + \eta_i + \bar{\varepsilon}_i$$

Dersom vi nå trekker dette fra den opprinnelige modellen finner vi:

✓

$$Y_{it} - \bar{Y}_i = \beta_0 - \beta_0 + \beta_1(X_{it1} - \bar{X}_{i1}) + \beta_2(X_{it2} - \bar{X}_{i2}) + \eta_i - \eta_i + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

$$Y_{it} - \bar{Y}_i = \beta_1(X_{it1} - \bar{X}_{i1}) + \beta_2(X_{it2} - \bar{X}_{i2}) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

✓

Vi har nå "fixed effects" modellen, hvor vi da ser at den region-spesifikke komponenten η_i er transformert bort. Gitt at ε_{it} oppfyller kriteriet om betinget forv. lik null, vil vi nå få forventningsrette og konsistente estimatorer. Grunnen til at estimatene endret seg er derfor at vi i 5) har utelatt variabel skjevhet, mens det har vi ikke i 6).

Videre kan vi og merke oss at ved å gjennomføre t-tester som i oppgave 1 b) finner vi at effektene av både inntekt per capita og sommertemperatur er signifikant forskjellige fra null i pooled OLS regresjonen med verdier på testobservatoren på hhv 10,84 og 2,79. Hvis vi derimot ser på modellen etter at vi har foretatt within groups transformasjonen, er fortsatt effekten av inntekt per capita signifikant med en t-verdi på 2,86, mens effekten av sommertemperatur-

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 4

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor
This column is for external examiner

βra

er nå ikke lenger signifikant med en t-verdi på 0,74. Dette er imidlertid ikke overraskende da variablene i fixed-effects modellen kun inngår som individuelle gj.snitt over tid. Det er derfor lite trolig at gj.snitt-temperaturene på et sted varier så mye over 12 år at dette skal gi en effekt, noe estimeringen også tyder på. Videre er også t-verdien til inntektseffekten redusert, men her vil det være rimelig å anta at variasjonen over tid vil være stor nok til å gi signifikante effekt, noe estimeringen bekrefter. Det at temperaturene ikke lenger er signifikant illustrerer ulempen ved å bruke within-transformasjonen, nemlig at vi kun utnytter variasjon over tid men ikke mellom individ, eller i dette tilfellet regioner. Slik at å estimere regionspesifikke variabler vil ikke være mulig etter en within transformasjon.

Denne kolonnen er forbeholdt sensor

This column is for external examiner

9)

For å undersøke dette vil begynne med en generell modell, anta at likevekts-sammenhengen mellom y_{it} og x_{it1}, x_{it2} er gitt ved

$$y_{it}^* = \beta_1 x_{it1} + \beta_2 x_{it2}$$

Men siden det er treghet i tilpasningen vil vi ha at:

$$\text{i)} \quad y_{it} - y_{it-1} = \lambda (y_{it}^* - y_{it-1})$$
$$\Updownarrow$$
$$\text{ii)} \quad y_{it} = \lambda y_{it}^* + (1-\lambda) y_{it-1}$$

Slik at ~~tilp~~ tilpasningen på tidspunkt t er en vektet sum av likevektssammenhengen og tilpasningen i forrige periode. λ kan her tolkes som justeringshastigheten. Vi ser at om denne er høy vil ~~et~~ stort en stor del av avviket mellom likevekt og tilpasning i forrige periode ~~jevnes~~ jevnes ut. (ii.) I det ekstreme tilfellet at $\lambda = 1$, ser vi at $y_{it} = y_{it}^*$ og vi er med en gang tilbake i likevekt.

For å finne en estimerbar relasjon setter vi y_{it}^* inn i ii). dette gir:

$$y_{it} = \lambda\beta_1 x_{it1} + \lambda\beta_2 x_{it2} + (1-\lambda) y_{it-1}.$$

Det er denne relasjonen som er estimert i 7) og vi ser at vi har $0{,}37\, y_{it-1}$. Dette betyr at $(1-\lambda) = 0{,}37 \Rightarrow \lambda = 1 - 0{,}37 = \underline{0{,}63}$. Så ja, det er treghet i tilpasningen siden ~~den~~ en $\lambda < 1$ vil bety treghet, men denne er ikke ekstremt stor.

Enig

På lang sikt vil vi være i likevekt slik at $y_{it} = y_{it-1}$. dersom vi setter dette inn i 7) finner vi:

$$y_{it}^* (1 - 0{,}37) = 0{,}60\, x_{1it} + 0{,}002\, x_{it2}$$
$$y_{it}^* = {}^{0{,}60}/_{0{,}63}\, x_{it} + {}^{0{,}002}/_{0{,}63}\, x_{it2}$$
$$y_{it}^* = 0{,}95\, x_{it1} + 0{,}0032\, x_{it2}$$

Denne kolonnen er forbeholdt sensor

This column is for external examiner

Slik at på lang sikt vil en 1% økning i inntekten føre til en 0,95% økning i boligprisene, mens en økning i gj.snittlig sommertemp med 1 grad vil føre til en økning i boligprisene på 100 · 0,0032 = 0,32%.
Det kan imidlertid være verdt å merke seg at i relasjon 7) er de respektive t-verdiene til inntekt og temperatur, 2,31 og 0,715 slik at det her bare er inntektens (effekt) som er signifikant forskjellig fra null på et 5% nivå (jfr. testene i oppgave 1 b).

OPPGAVE 2

a)

For å gjøre dette begynner jeg med å finne ligningen for p_t på redusert form:
Setter $x_t = x_t$ og løse for p_t

$$\beta_0 + \beta_1 p_t + u_{1t} = \alpha_0 + \alpha_1 p_t + \alpha_2 S_t + u_{2t}$$

$$p_t(\beta_1 - \alpha_1) = \alpha_0 + \alpha_2 S_t - \beta_0 + u_{2t} - u_{1t}$$

$$p_t^* = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} S_t + \frac{u_{2t} - u_{1t}}{\beta_1 - \alpha_1}$$

Det vi nå ser er at p_t^* har u_{1t} i sin ligning, noe som gjør at p_t i relasjon 1 vil være korrelert med restleddet. Dersom vi bruker formelen/samme fremgangsmåte for estimering av relasjon 1), som vi brukte i oppgave 1a) vil vi finne at:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{t=1}^{T}(p_t - \bar{p})(u_{1t} - \bar{u}_t)}{\sum_{t=1}^{T}(p_t - \bar{p})^2} \quad \Big| \text{ siden } x_t - \bar{x}_t = \beta_1(p_t - \bar{p}) + (u_t - \bar{u})$$

Dersom vi nå finner sannsynlighetsgrensen til denne har vi at:

$$\text{plim}(\hat{\beta}_1) = \beta_1 + \frac{\text{plim}\, \frac{1}{T}\sum_{t=1}^{T}(p_t - \bar{p})(u_{1t} - \bar{u}_t)}{\text{plim}(p_t - \bar{p})^2}$$

$$= \beta_1 + \frac{\text{cov}(p, u_1)}{\text{var}(p)}$$

Her har vi og benyttet antagelsen om at de empiriske momentene vil konvergere i sannsynlighet mot sine teoretiske. Dersom $\hat{\beta}_1$ skulle ha vært konsistent måtte $\text{cov}(p, u_1) = 0$, men vi har sett at dette ikke er tilfellet da p_t har u_{1t} i sin ligning. Nærmere bestemt vil kovariansen mellom de to være:

NTNU

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6 -11 Side/*Page:* 15

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor
This column is for external examiner

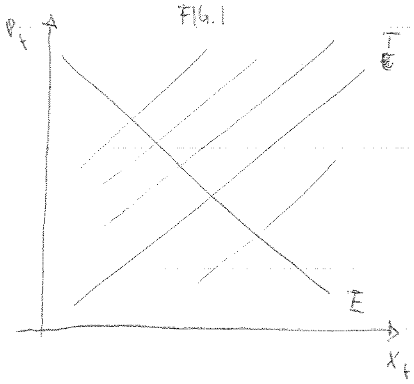
$$\text{cov}\left(\frac{u_{2t}-u_{1t}}{\beta_1-\alpha_1},\ u_{1t}\right) = -\frac{\sigma_{u1}^2}{\beta_1-\alpha_1}$$

Hvor vi da har benyttet antagelsene om at S_t er eksogen og uavhengig av restleddene samt at restleddene er innbyrdes uavhengig, $\text{cov}(u_{1t}, u_{2t}) = 0$.

B)

Vi starter med hvorfor 1) er eksakt identifisert. Dette er fordi tilbudsrelasjonen inneholder den eksogene variablen S_t som vil skifte tilbudskurven, og gi bevegelse langs etterspørselskurven. Dette har vi illustrert grafisk i figur 1. Ettersom S_t endrer seg vil altså tilbudskurven skifte. Alle skift er her assosiert med skift i S_t. Som sagt bidrar dette til bevegelse langs etterspørselskurven og vi kan bruke likevektsmekanismen til å finne punkter som kan identifisere/vise hvor etterspørselskurven ligger når S_t endres.

FIG. 1



Grunnen til at 2) ikke er identifisert er at etterspørselsfunksjonen ikke inneholder eksogene variable som vil skifte kun den og ikke tilbudskurven, da prisen inngår i begge relasjoner. En dvelse slik som illustrert i fig 1. vil derfor ikke være mulig. Det vil kunne forekomme noen små skift gjennom endringer i restleddet, men dette er slag vi ikke kan identifisere og hjelper oss slik sett ingen ting.

c)

Bra

Det vi da kan gjøre er å bruke S_t som et instrument for p_t. Vi ser fra P_t^* ligningen at S_t inngår samt at vi har antatt at den er uavhengig av u_1 altså har vi:

$\text{cov}(p_t, S_t) \neq 0$

$\text{cov}(u_{1t}, S_t) = 0$

Dette er de to kriteriene en instrumentvariabel må oppfylle, og i dette tilfellet, S_t

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6 - 11 Side/*Page:* 16

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor

This column is for external examiner

Må oppfylle dersom den skal være et gyldig instrument for p_t.

Metoden er som følger:

Ta forventningen til begge sider av 1):

$$E(x_t) = \beta_0 + \beta_1 E(p_t) + E(u_t)$$

Trekk så dette ifra relasjon 1)

$$x_t - E(x_t) = \beta_0 - \beta_0 + \beta_1(p_t - E(p_t)) + (u_{1t} - E(u_t))$$

$$x_t - E(x_t) = \beta_1(p_t - E(p_t)) + (u_{1t} - E(u_{1t}))$$

Multipliser så med $(s_t - E(s_t))$

$$(x_t - E(x_t))(s_t - E(s_t)) = \beta_1 (p_t - E(p_t))(s_t - E(s_t)) + (u_{1t} - E(u_{1t}))(s_t - E(s_t))$$

Tar vi nå forventningen til dette og bruker at $cov(u_{1t}, s_t) = 0$ får vi:

$$cov(x, s) = \beta_1 cov(p, s)$$

$$\beta_1 = \frac{cov(x,s)}{cov(p,s)}$$

Dersom vi nå erstatter de teoretiske momentene med de empiriske, har vi instrumentvariabel estimatoren:

$$\hat{\beta}_1^{IV} = \frac{1/n \sum_{t=1}^{T}(x_t - \bar{x})(s_t - \bar{s})}{1/n \sum_{t=1}^{T}(p_t - \bar{p}_t)(s_t - \bar{s})}$$

$$\hat{\beta}_1^{IV} = \frac{\sum_{t=1}^{T}(x_t - \bar{x})(s_t - \bar{s})}{\sum_{t=1}^{T}(p_t - \bar{p}_t)(s_t - \bar{s})}$$

For å vise at denne er konsistent setter vi nå inn for $x_t - \bar{x} = \beta_1(p_t - \bar{p}) + (u_{1t} - \bar{u}_1)$

Dette gir

$$\hat{\beta}_1^{IV} = \frac{\sum_{t=1}^{T}\left(\left(\beta_1(p_t - \bar{p}) + (u_{1t} - \bar{u}_1)\right)(s_t - \bar{s})\right)}{\sum (p_t - \bar{p}_t)(s_t - \bar{s})}$$

$$\hat{\beta}_1^{IV} = \beta_1 \frac{\sum_{t=1}^{T}(p_t - \bar{p}_t)(s_t - \bar{s})}{\sum_{t=1}^{T}(p_t - \bar{p}_t)(s_t - \bar{s})} + \frac{\sum_{t=1}^{T}(u_{1t} - \bar{u}_1)(s_t - \bar{s})}{\sum_{t=1}^{T}(p_t - \bar{p}_t)(s_t - \bar{s}_t)}$$

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 17

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor

This column is for external examiner

$$\hat{\beta}_1^{IV} = \beta_1 + \frac{\Sigma(u_{1t}-\bar{u}_1)(s_t-\bar{s})}{\Sigma(p_t-\bar{p})(s_t-\bar{s})}$$

Nå er vi klare til å se om denne er konsistent:

$$\text{plim}(\hat{\beta}_1^{IV}) = \beta_1 + \frac{\text{plim}\,1/T\left(\sum_{t=1}^{T}(u_{1t}-\bar{u}_1)(s_t-\bar{s})\right)}{\text{plim}\,1/T\left(\sum_{t=1}^{T}(p_t-\bar{p})(s_t-\bar{s})\right)}$$

$$= \beta_1 + \frac{\text{cov}(u,s)}{\text{cov}(p,s)}$$

B/a

$$= \underline{\underline{\beta_1}}$$

Her har vi først brukt at de empiriske momentene konvergerer mot sine teoretiske momenter samt antagelsen om at $\text{cov}(u,s)=0$. Med dette har vi da vist at vi kan oppnå en konsistent estimator for β_1 i etterspørselsfunksjonen ved bruk av instrumentvariabelmetoden.

b)

Siden D_t er antatt å være uavhengig av restleddet er det nok en gang p_t og dermed estimatoren $\hat{\beta}_1$ som er problemet. Dette ser vi ved å løse systemet for p_t:

$$X_t = x_t$$

$$\beta_0 + \beta_1 p_t + \beta_2 D_t + u_{1t} = \alpha_0 + \alpha_1 p_t + \alpha_2 S_{1t} + \alpha_3 S_{2t} + u_{2t}$$

$$p_t(\beta_1 - \alpha_1) = \alpha_0 - \beta_0 + \alpha_2 S_{1t} + \alpha_3 S_{2t} - \beta_2 D_t + u_{2t} - u_{1t}$$

N

$$p_t = \frac{\alpha_0-\beta_0}{\beta_1-\alpha_1} + \frac{\alpha_2}{\beta_1-\alpha_1}S_{1t} + \frac{\alpha_3}{\beta_1-\alpha_1}S_{2t} - \frac{\beta_2}{\beta_1-\alpha_1}D_t + \frac{u_{2t}-u_{1t}}{\beta_1-\alpha_1}$$

Nok en gang ser vi altså at p_t har u_{1t} i ligningen sin slik at $\text{cov}(p,u_{1t}) \neq 0$ og vi har et lignende problem som i a).

Kandidat nr./*Candidate no.* 10022

Dato/*Date:* 7/6-11 Side/*Page:* 18

Emnekode/*Subject* SØK 3001 Antall ark/*Number of pages:* 20

Denne kolonnen er forbeholdt sensor

This column is for external examiner

Forskjellen er nå at vi har to variable som antas å oppfylle kravene til en instrumentvariabel. Disse er S_{1t} og S_{2t}, hvor vi da antar:

$$cov(u_{1t}, S_{1t}) = 0 \qquad cov(p_t, S_{1t}) \neq 0$$
$$cov(u_{1t}, S_{2t}) = 0 \qquad cov(p_t, S_{2t}) \neq 0$$

For å utnytte begge instrumentene må vi bruke metoden som kalles to-stegs MKM. Det vi her utnytter er at om S_{1t} og S_{2t} er uavhengige av u_{1t} vil også en lineær kombinasjon av de to være det. Spørsmålet er da hvilken lineær kombinasjon av de to som forklarer ~~det~~ variasjonen i p_t best. Denne kombinasjonen kan vi da finne ved å estimere følgende ligning med MKM:

i) $p_t = \pi_0 + \pi_1 S_{1t} + \pi_2 S_{2t} + \pi_3 D_t + støy$

Bra

Legg merke til at vi her inkluderer D_t. Dette er fordi vi ønsker å se variasjonen i p_t av en endring i S_{1t} og S_{2t} kontrollert for endringer i D_t. Dette er viktig for at 1*) skal ha en meningsfull tolkning senere når vi ønsker å se på ceteris paribus effekten av en endring i p_t. Estimering av relasjonen over gir et estimert verdi $\hat{p}_t$ som vil være uavhengig av u_{1t} siden både S_{1t}, S_{2t} og D_t er antatt å være uavhengige av u_{1t}. Neste steg er så å sette $\hat{p}_t$ inn i 1*):

ii) $X_t = \beta_0 + \beta_1 \hat{p}_t + \beta_2 D_t + u_{1t}$

Dette er nå en relasjon med bare eksogene forklaringsvariable slik at denne nå kan estimeres med vanlig MKM og gi forventningsrette og konsistente estimater.

Når det gjelder å teste hvorvidt instrumentvariablene gir tilstrekkelig informasjon er det ligning i) vi må fokusere på. Dersom instrumentvariablene skal gi tilstrekkelig informasjon må i det minste π_1 og π_2 være signifikant (sammen) i relasjon i)

Dette kan vi sjekke ut ved en F-test, hvor da modellen med restriksjoner, gitt:

$$H_0: \pi_1 = \pi_2 = 0$$

Vil være gitt ved:

iii) $p_t = \pi_0 + \pi_3 D_t + \text{støy}$.

Testobservatoren vil da bli være:

$$F_{obs} = \frac{\frac{SSR_{iii} - SSR_{ii}}{2}}{\frac{SSR_{ii}}{n-k-1}}$$

Som vi gjorde i oppgave 1 b) sammenligner vi denne med en kritisk verdi fra F-fordelingen, og dersom instrumentene skal gi tilstrekkelig informasjon bør denne forkastes med god margin.

Om ikke en formell test, kan man og få en viss formening om informasjonen instrumentene gir ved å se på den asymptotiske variansen til $\hat{\beta}_1$. Variansen til $\hat{\beta}_1$ vil dersom vi bruker instrumentvariabel metoden være gitt ved:

$$var(\hat{\beta}_1^{IV}) = \frac{\hat{\sigma}^2}{SST_p R_p^2}$$

Se neste ark.

~~SST_p er $\Sigma(p_t - \bar{p})$, mens R_p^2 er R^2 fra ligning ii). R^2 er et mål på andel forklart variasjon i en modell. Dersom S_{1t} og S_{1t} gir lite informasjon om p_t vil R^2 i ligning ii) være lav. Dette ser vi der videre at vil føre til en høy varians på $\hat{\beta}_1^{IV}$. Slik at dersom du finner et $\hat{\beta}_1$ estimator med veldig høy varians etter å ha brukt instrumentvariabel metoden kan dette være et tegn på at instrumentene ikke gir tilstrekkelig informasjon, selv om det er vanskelig å si hva som er tilstrekkelig og ikke.~~

Denne kolonnen er forbeholdt sensor

This column is for external examiner

OK

Dette er strengt tatt ikke direkte relatert til $\hat{\beta}_1^{IV}$ i oppgave 2 d) da variansen, slik den er satt opp er variansen til $\hat{\beta}_1^{IV}$ fra oppgave 2 c). $SST_p = \sum_{t=1}^{T} (p_t - \bar{p})$ mens R_p^2 er den kvadrerte verdien til den empiriske korrelasjonskoeffisienten mellom p_t og s_t. Slik at dersom s_t forklarer lite av variasjonen i p_t vil denne være lav og variansen være høy. Slik at dersom du finner en estimator med veldig høy varians etter å ha brukt instrument-variabel metoden kan dette være et tegn på at instrumentene ikke gir tilstrekkelig informasjon, selv om det er vanskelig å si hva som er tilstrekkelig og ikke. Men som sagt, variansen er bare en indikator og ikke et grunnlag for bedømning. En test som F-testen beskrevet over eller lignende bør derfor heller brukes for å teste informasjonen til instrumentene.