

Rettevejledning, Økonometri I, Vinter 2020-2021

Opgave 1

- a Datasættet er et balanceret paneldatasæt for 300 regioner for tre år (1995, 2005, 2015). Tabel 1 viser, at mellem 1995 og 2015 stiger antallet af robotter pr. 1000 arbejder fra 1,1 til 2,4. På samme tid falder antallet af beskæftigede uden kompetencegivende uddannelse med næsten 30.000 personer, mens antallet af beskæftigede med kompetencegivende uddannelse stiger med ca. 35.000 personer. I modsætning til dette er aldersfordelingen og andelen af kvinder i arbejdsstyrken ret konstant. Andelen af regioner, der ligger i den nordlige del af landet, er naturligvis konstant over tid. Udover de nævnte variable indeholder datasættet information om fordelingen på brancher og væksten i industrirobotter i andre højindkomstlande.

	1995		2005		2015	
	mean	sd	mean	sd	mean	sd
Besk, ejkomp	102669.076	22743.220	80478.513	37510.952	74364.485	45010.126
Besk, komp	152822.105	31773.734	177169.535	51080.555	189827.729	69566.951
Robot	1.061	0.316	2.179	0.522	2.444	0.576
Nord	0.437	0.497	0.437	0.497	0.437	0.497
Alder	41.362	1.142	41.360	1.135	41.352	1.147
Kvinde	0.500	0.021	0.501	0.021	0.500	0.021
No. obs	300		300		300	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1: Deskriptiv statistik.

- b I figur 1 vises udviklingen i den aggregerede nationale beskæftigelse. Den samlede beskæftigelse af arbejdere uden kompetencegivende uddannelse går fra 31 mio. personer i 1995 til 22 mio. personer i 2015. I den samme periode stiger beskæftigelse af arbejdere med kompetencegivende uddannelse fra 46 mio. personer til 57 mio. personer.

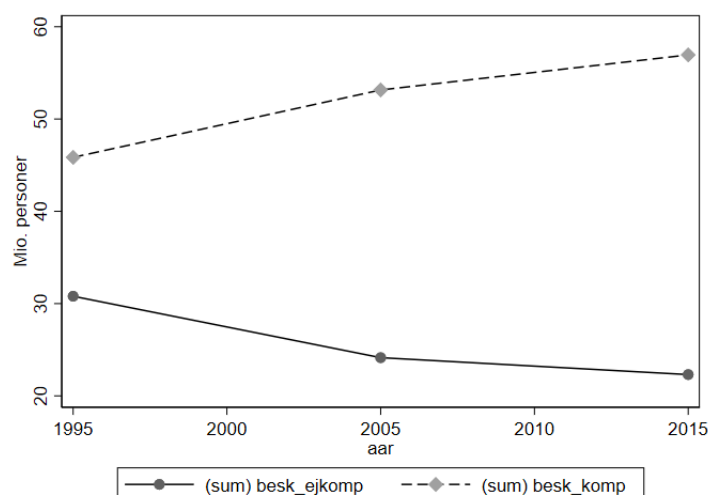


Figure 1: Beskæftigelsen over tid

Opgave 2

- a I tabel 2 er estimerne for estimationen af ligning (1) i opgaveformuleringen rapporteret. For beskæftigede henholdsvis uden og med kompetencegivende uddannelse estimerer vi β_1 til at være -0.364 og 0.141 . Da vi ikke ved, om der er heteroskedasticitet, anvender vi robuste standardfejl.

Vi opstiller nulhypotesen $H_0 : \beta_1 = 0$ mod alternativet $H_1 : \beta_1 \neq 0$. Vi tester mod det dobbeltsidede alternativ, da det teoretisk er uklart, om robotter vil betyde flere eller færre beskæftigede både med og uden kompetencegivende uddannelse. Retningen på denne effekt afhænger af, om typen af arbejder er komplement eller substitut til industrirobotter. Vi finder, at effekterne på beskæftigelsen for både personer uden og med kompetencegivende uddannelse er signifikant på et 1 pct. signifikansniveau.

$$t^{ejkomp} = -12.44, p^{ejkomp} = 0.0000$$
$$t^{komp} = 6.88, p^{komp} = 0.0000$$

Hvis vi kunne give β_1 en kausal fortolkning, vil vi kunne fortolke den som den approksimative procentvise ændring i beskæftigelsen, når der kommer én yderligere industrirobot pr. 1000 arbejder. Det er dog nok tvivlsomt, at vi kan give parameterestimatet en kausal fortolkning, idet det kræver, at antallet af roboter i en region skal være ukorreleret med uobserverbare faktorer. Specielt forestiller vi os, at uobserverbare uddannelseskvalitet (dvs. ud over hvad der er fanget af de to uddannelseskategorier) af arbejdsstyrken vil være korreleret med antallet af robotter. Imidlertid kan en række andre forhold som f.eks. råvarer, der er tilgængelige i regionen, infrastruktur og universiteter i regionen også være korreleret med antallet af industrirobotter per arbejder. Dvs. det virker sandsynligt, at $Cov(robot_{it}, a_i) \neq 0$.

Blandt kontrolvariablene er der en positiv signifikant effekt af *alder* for begge typer af beskæftigede samt en positiv signifikant effekt på beskæftigelsen af personer med kompetencegivende uddannelse af, at regionen er i den nordlige del af landet.

- b For begge typer af beskæftigede finder vi, at den restriktede model forkastes i sammenligning med den urestriktede model. Vi kan formulere testet som $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$ mod alternativet $H_1 : H_0$ er ikke opfyldt. For beskæftigede uden kompetencegivende uddannelse finder vi en F-statistik på 2.61 mod en F-statistik på 7.40 for beskæftigede med kompetencegivende uddannelse. Under H_0 er F-statistikken fordelt med $F(5, 893)$. P-værdierne er henholdsvis 0,0238 og 0.0000. På et 5 pct. signifikansniveau forkaster vi altså de 5 restriktioner.
- c I Breusch-Pagan testet regresseres de kvadrerede residualer mod de forklarende variable. Hvis de forklarende variable har forklaringskraft i denne regression, vil vi forkaste nulhypotesen om homoskedasticitet. For henholdsvis beskæftigede uden og med kompetencegivende uddannelse får vi de to følgende Breusch-Pagan teststatistikker: 63.35 og 73.19, der er χ^2 -fordelt med 6 frihedsgrad. I begge tilfælde afviser vi nulhypotesen om homoskedasticitet med en p-værdi på 0.0000. Baseret på residualplots i figur 2 og figur 3 lader der til at være mindre residualvarians i 1995 sammenlignet med i 2005 og 2015. For beskæftigelsen med personer med kompetencegivende uddannelse er dummyvariablen *robot* og *nord* også variansstyrede.

Opgave 3

- a Når vi tager førstedifferenser af ligning (1) i opgaven, får vi

$$\Delta \log(besk_{it}^{type}) = \alpha_0 + \beta_1 \Delta robot_{it} + \beta_3 \Delta alder_{it} + \beta_4 \Delta kvinde_{it} + \beta_6 d_{it}^{2005} + \Delta u_{it} \quad (1)$$

	Ejkomp. udd, multipel	Komp. udd, multipel	Ejkomp. udd, simpel	Komp. udd, simpel
	b/se	b/se	b/se	b/se
Robot	-0.364*** (0.029)	0.141*** (0.020)	-0.334*** (0.017)	0.132*** (0.012)
Nord	0.051 (0.027)	0.080*** (0.019)		
Alder	0.031** (0.012)	0.036*** (0.008)		
Kvinde	0.255 (0.661)	0.588 (0.476)		
aar1995	-0.047 (0.054)	0.019 (0.039)		
aar2005	0.039 (0.038)	-0.011 (0.027)		
Constant	10.498*** (0.599)	9.917*** (0.418)	11.887*** (0.031)	11.765*** (0.022)
Obs	900	900	900	900
Rsqr	0.296	0.144	0.285	0.108

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Spm 2.1 og 2.2.

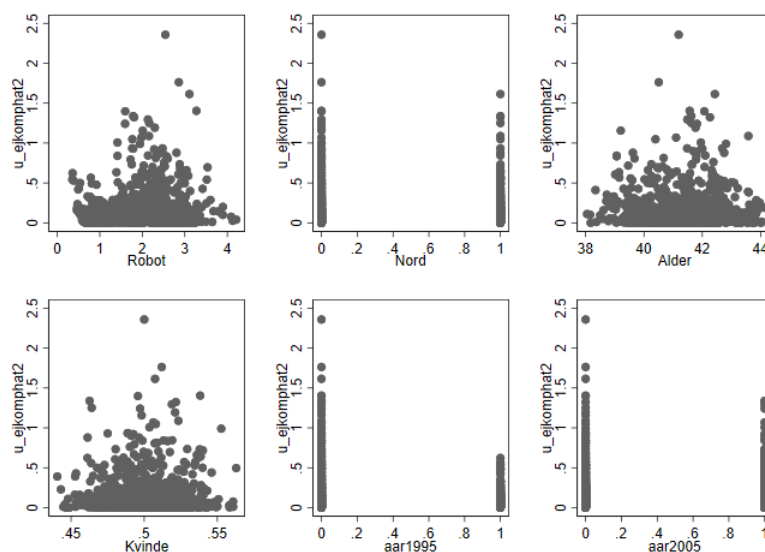


Figure 2: Kvadreret residualplots, ej komp. uddannede

hvor f.eks. $\Delta robot_{it} = robot_{it} - robot_{it-1}$. Den individspecifikke tidskonstante uobserverbare heterogenitet og den tidskonstante variabel $nord_i$ forsvinder, når man tager førstedifferenser. Ydermere bruger vi en tidsperiode på at lave førstedifferenser, så tidsdummyen for 1995 forsvinder ligeledes. Vi kan fortolke β_1 på samme måde i ligning (1) i opgaveformuleringen og i ligning (1) her, dvs. som den approksimative procentvise ændring i beskæftigelsen, når der kommer én yderligere industrirobot pr. 1000 arbejder.

I tabel 3 finder vi, at effekten af industrirobotter på beskæftigede uden kompetencegivende uddannelse nu er -0.855 mod -0.364 i tabel 2. Begge effekter er signifikant

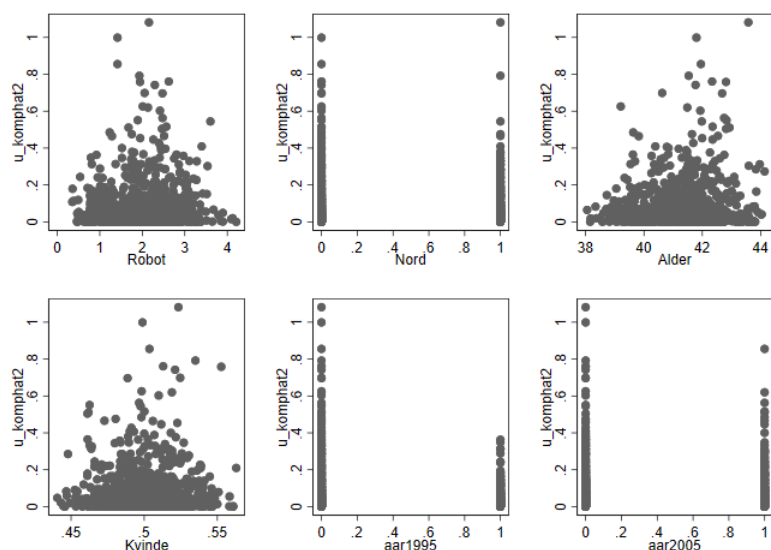


Figure 3: Kvadreret residualplots, komp. uddannede

forskellige fra 0. Den store forskel i estimaterne tyder på, at $robot_{it}$ er positivt korreleret med fixed effecten. Det kan fortolkes som, at der er investeret mere i industrirobotter i regioner med flere beskæftigede uden kompetencegivende uddannelse.

Effekten af industrirobotter på beskæftigede med kompetencegivende uddannelse er nu 0.052 mod 0.141 i tabel 2. Førstedifferens estimatet betyder, at effekten af industrirobotter ikke længere er signifikant forskellig fra 0 på et 5 pct.s signifikansniveau, men dog stadig signifikant forskellig fra 0 på et 10 pct. signifikansniveau.

	Spm 3.1, ejkomp.	Spm 3.1, komp.	Spm 3.2, first stage	Spm 3.2, ejkomp.	Spm 3.2, kom
	b/se	b/se	b/se	b/se	b/se
Delta Robot	-0.855*** (0.022)	0.052* (0.024)		-0.703*** (0.063)	0.159* (0.064)
Delta Alder	0.061 (0.040)	0.021 (0.038)	0.041 (0.066)	0.060 (0.042)	0.020 (0.039)
Delta Kvinde	-0.819 (1.386)	0.828 (1.410)	0.669 (2.371)	-0.752 (1.406)	0.876 (1.420)
aar2005	0.544*** (0.024)	0.035 (0.026)	0.116 (0.091)	0.415*** (0.058)	-0.057 (0.058)
shiftshare			1.833*** (0.205)		
Constant	0.091*** (0.012)	0.034* (0.013)	-0.551*** (0.089)	0.051** (0.019)	0.006 (0.020)
Obs	600	600	600	600	600
Rsqr	0.734	0.051	0.646	0.713	0.015

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Spm 3.1 og 3.2.

b Hvis virksomheder i en region bliver hårdere ramt af international konkurrence f.eks. mellem 1995 og 2005, kunne det tænkes, at virksomhederne som svar på dette foretager

investeringer for at blive mere konkurrencedygtige, herunder indkøber industrirobotter. Dvs. det er muligt, at der er simultanitetsbias, hvor virksomheder i hård international konkurrence reducerer antallet af beskæftigede uden kompetencegivende uddannelse på grund af en mindre efterspørgsel, og for at kunne klare sig i fremtiden (og måske produktinnovere) investerer i industrirobotter. Dermed kan $\Delta robot$ potentielt være korreleret med fejlløbet og altså endogen.

Tabel 3 viser first stage estimaterne i kolonne 3. Her ser vi, at effekten af shift-share instrumentet på $\Delta robot$ er statistisk signifikant på et 1 pct. niveau. Second stage resultaterne er gengivet i kolonne 4-5. Vi ser, at når vi instrumenterer $\Delta robot$, finder vi en mindre negativ effekt for beskæftigede uden kompetencegivende uddannelse og en større positiv effekt for beskæftigede med kompetencegivende uddannelse. Dette stemmer overens med hypotesen om, at virksomheder, der er i hård konkurrence og dermed i forvejen reducerer antallet af beskæftigede uden kompetencegivende uddannelse, kan være mere tilbøjelige til at investere i industrirobotter.

c Vi kan undersøge om $\Delta robot_{it}$ er endogen ved at estimere den følgende ligning

$$\Delta \log(besk_{it}^{type}) = \alpha_0 + \beta_1 \Delta robot_{it} + \beta_3 \Delta alder_{it} + \beta_4 \Delta kvinde_{it} + \beta_6 d_{it}^{2005} + \gamma \hat{\epsilon}_{it} + \Delta v_{it} \quad (2)$$

hvor $\hat{\epsilon}_{it}$ er de prædikterede residualer fra første trin af 2SLS, dvs.

$$\Delta robot_{it} = \pi_0 + \pi_1 \Delta z_{it} + \pi_3 \Delta alder_{it} + \pi_4 \Delta kvinde_{it} + \pi_6 d_{it}^{2005} + \epsilon_{it} \quad (3)$$

Vi finder, at parameteren til $\hat{\epsilon}_{it}$, γ er signifikant forskellig fra 0 på et 5 pct. signifikansniveau for beskæftigede uden kompetencegivende uddannelse, mens effekten ikke er signifikant for de beskæftigede med kompetencegivende uddannelse. Dermed tyder det kun på, at $\Delta robot_{it}$ er endogen i ligning (1), når den afhængige variabel er ændringen i beskæftigelsen af personer uden kompetencegivende uddannelse.

For at shift-share instrumentet er validt kræves det, at $Cov(\Delta robot_{it}, z_{it}) \neq 0$, og at $Cov(z_{it}, \Delta u_{it}) = 0$. Den første antagelse kan undersøges fra første trin estimationerne hvor vi regresserer $\Delta robot_{it}$ mod bl.a. shift-share instrumentet, z_{it} . I det foregående delspørgsmål fandt vi en signifikant parameter til shift-share instrumentet, så den første antagelse er opfyldt. Den anden antagelse kan til gengæld ikke verificeres. Opgave 4 og opgave 5 viser, at for et shift-share instrument betyder det andet krav, at den initiale industrifordeling i 1995 skal være eksogen. Det er et åbent spørgsmål om det er tilfældet, og der er ingenting i denne opgave, der sandynsliggør, at den hverken er eksogen eller ej.

Opgave 4

a

$$\hat{\pi}_1 = \frac{\sum_{i=1}^n x_i (z_i - \bar{z})}{\sum_{i=1}^n z_i (z_i - \bar{z})}$$

$$\hat{\pi}_0 = \frac{1}{n} \sum_{i=1}^n x_i - \hat{\pi}_1 \frac{1}{n} \sum_{i=1}^n z_i$$

	Spm 3.3, ejkomp.	Spm 3.3, komp.
	b/se	b/se
Delta Alder	0.060 (0.040)	0.020 (0.038)
Delta Kvinde	-0.752 (1.375)	0.876 (1.406)
aar2005	0.415*** (0.056)	-0.057 (0.058)
Delta Robot	-0.703*** (0.062)	0.159* (0.063)
epshat	-0.172** (0.066)	-0.121 (0.067)
Constant	0.051** (0.019)	0.006 (0.020)
Obs	600	600
Rsqr	0.736	0.055

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Spm 3.3.

b Vi kan skrive instrumentet som $z_i = s_i g_1 + s_{2i} g_2 = s_i g_1 + (1 - s_i) g_2 = g_2 + (g_1 - g_2) s_i$.

$$\begin{aligned}
\hat{\pi}_1 &= \frac{\sum_{i=1}^n x_i (z_i - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \\
&= \frac{\sum_{i=1}^n x_i (g_2 + (g_1 - g_2) s_i - g_2 - (g_1 - g_2) \bar{s})}{\sum_{i=1}^n (g_2 + (g_1 - g_2) s_i - g_2 - (g_1 - g_2) \bar{s})^2} \\
&= \frac{\sum_{i=1}^n x_i (g_1 - g_2) (s_i - \bar{s})}{\sum_{i=1}^n ((g_1 - g_2) (s_i - \bar{s}))^2} \\
&= \frac{\sum_{i=1}^n x_i (g_1 - g_2) (s_i - \bar{s})}{\sum_{i=1}^n ((g_1 - g_2) (s_i - \bar{s}))^2} \\
&= \frac{(g_1 - g_2) \sum_{i=1}^n x_i (s_i - \bar{s})}{(g_1 - g_2)^2 \sum_{i=1}^n ((s_i - \bar{s}))^2} \\
&= \frac{1}{g_1 - g_2} \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2}
\end{aligned}$$

$$\begin{aligned}
\hat{\pi}_0 &= \frac{1}{n} \sum_{i=1}^n x_i - \hat{\pi}_1 \frac{1}{n} \sum_{i=1}^n z_i \\
&= \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{g_1 - g_2} \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} \frac{1}{n} \sum_{i=1}^n (g_2 + (g_1 - g_2) s_i) \\
&= \bar{x} - \left[\frac{g_2}{g_1 - g_2} + \bar{s} \right] \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2}
\end{aligned}$$

c Den prædikterede værdi er dermed

$$\begin{aligned}
\hat{x}_i &= \hat{\pi}_0 + \hat{\pi}_1 z_i \\
&= \bar{x} - \left[\frac{g_2}{g_1 - g_2} + \bar{s} \right] \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} + \frac{1}{g_1 - g_2} \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} (g_2 + (g_1 - g_2)s_i) \\
&= \bar{x} - \left[\frac{g_2}{g_1 - g_2} + \bar{s} \right] \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} + \left[\frac{g_2}{g_1 - g_2} + s_i \right] \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} \\
&= \bar{x} + (s_i - \bar{s}) \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2}
\end{aligned}$$

Hvis vi i stedet havde benyttet andelene som instrumenter, ville den prædikterede værdi være

$$\begin{aligned}
\hat{x}_i &= \hat{\pi}_0 + \hat{\pi}_1 z_i \\
&= \left[\bar{x} - \bar{s} \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} \right] + s_i \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2} \\
&= \bar{x} + (s_i - \bar{s}) \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2}
\end{aligned}$$

hvor $\hat{\pi}_0 = \bar{x} - \bar{s} \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2}$ og $\hat{\pi}_1 = \frac{\sum_{i=1}^n x_i (s_i - \bar{s})}{\sum_{i=1}^n (s_i - \bar{s})^2}$. Konklusionen er dermed, at den prædikterede værdi er identisk, om man bruger shift-share instrumentet eller det initiale andele. Dermed vil 2SLS ved brug af de to instrumenter give identiske resultater for parameter estimatet i hovedligningen, hvor x er endogen.

Opgave 5

- a Tabel 5 viser, at eftersom $Cov(x_i, u_i) \neq 0$, så er OLS biased. Den sande værdi er $\beta_1 = 2$, men gennemsnittet af OLS estimaterne er 4,772. Til gengæld klarer 2SLS sig godt og estimerer i gennemsnit β_1 til 1,995. Histogrammet viser også, at IV estimatoren er fordelt omkring den sande værdi. Disse resultater holder for tilfældet med to brancher ($J = 2$).

	2 brancher		5 brancher	
	mean	sd	mean	sd
OLS	4.772	0.071	5.673	0.069
IV shift-share	1.995	0.169	1.995	0.264
IV shares	1.995	0.169	2.033	0.258
No. simulations	500		500	

Table 5: Spm. 5.1-5.3.

- b Tabel 5 viser, at for $J = 2$ er 2SLS estimatoren med shift-share instrument og 2SLS estimatoren med *shares* numerisk ækvivalente. Dette resultat er implikationen af, at den prædikterede værdi er den samme, når man har, at $J = 2$, som vi så i opgave 4.3. Man kan også se i figur 4, at de to sæt estimater er helt sammenfaldende.
- c Når $J > 2$ er 2SLS med shift-share instrumentet og 2SLS med shares ikke længere numerisk ækvivalente. Det skyldes, at med $J > 2$ vil OLS estimationen reelt vægte de enkelte shares via π -parametrene anderledes *shift*-variablene gør. Med andre ord inkluderer vi nu flere instrumenter, når vi benytter *shares*, end med shift-share instrumentet, der er et enkelt instrument.

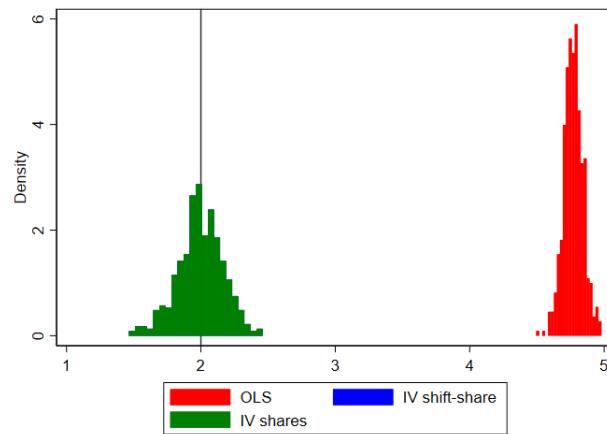


Figure 4: Spm. 5.1-5.2: Simulation med 2 brancher

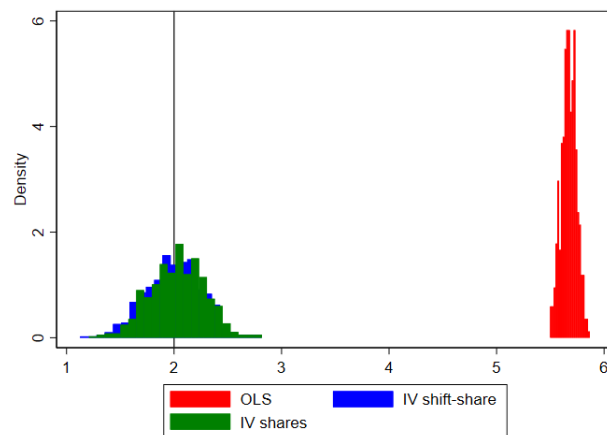


Figure 5: Spm. 5.3: Simulation med 5 brancher