

Eksamen på Økonomistudiet sommer 2018

Økonometri I

Tag-hjem eksamen

May 31, 2018

VEJLEDENDE BESVARELSE

Problem 1

1. Model (1) er en lineær paneldata regressionsmodel som beskriver variation i virksomheders primære resultat ved en lineær funktion af virksomhedskarakteristika i form af antal beskæftigede, alder, beliggenhed og type af direktørskifte i periode 2. Disse variable varierer på tværs af virksomheder og over tid. Dog vil visse variable som fx virksomhedens beliggenhed formentlig ikke ændre sig meget mellem de to perioder. Modellen indeholder også en tidsdummy, $d2_t$, hvor det bemærkes at i -indekset er fraværende. Ergo antages det alle virksomheder optræder i stikprøven i begge perioder, hvilket er kendetegende for et balanceret panel. Modellen indeholder desuden to fejllid, et virksomhedsspecifikt og tidsinvariant fejllid i form af α_i , samt et idiosynkratisk fejllid u_{it} som varierer over tid og virksomheder.
2. For $t = 1$ er modellens skæring lig β_0 , mens skæringen er lig $\beta_0 + \delta_0$ i $t = 2$. Altså angiver δ_0 den forventede procentvise ændring i virksomheders primære resultat mellem de to tidsperioder, alt andet lige. δ_0 vil således inkludere ændringer i makroøkonomiske forhold som påvirker virksomhedernes primære resultat. Samme parameter vil dog også inkludere effekten af at skifte direktør, hvilket alle virksomheder i stikprøven gør i periode 2. Fortegnet for δ_0 kan således være positivt eller negativt alt afhængigt af makroøkonomiske forhold og effekten af en ny direktør. β_1 angiver den forventede procentvise ændring i virksomhedernes primære resultat når et familiemedlem ansættes som direktør i periode 2 i stedet for en professionel direktør, alt andet lige. Ifølge introduktionen til opgaven, kan en større viden om virksomheden eller ikke-pekuniære incitament betyde at en familiedirektør klarer sig bedre, hvorfor $\beta_1 > 0$. Omvendt kan et mindre kvalificeret familiemedlem have negative konsekvenser for virksomheden relativt til en professionel direktør, hvorfor $\beta_1 < 0$.

Table 1: Summary Statistics

	mean	p50	min	max		mean	p50	min	max
<i>Periode 1:</i>					<i>Periode 2:</i>				
preres	12.78	12.45	4.32	74.91	preres	11.20	10.77	1.55	97.52
besk	170.94	70.00	11.00	4060.00	besk	186.27	75.00	6.00	5238.00
alder	6.80	7.00	1.00	45.00	alder	10.80	11.00	5.00	49.00
hovedstad	0.21	0.00	0.00	1.00	hovedstad	0.21	0.00	0.00	1.00
barn1son	0.50	0.50	0.00	1.00	barn1son	0.50	0.50	0.00	1.00
famceo	0.00	0.00	0.00	0.00	famceo	0.56	1.00	0.00	1.00
nborn	1.26	1.00	0.00	4.00	nborn	1.26	1.00	0.00	4.00
<i>Periode 2 og famceo=1:</i>					<i>Periode 2 og famceo=0:</i>				
preres	10.65	10.40	2.79	30.74	preres	11.89	11.32	1.55	97.52
besk	172.71	69.50	9.00	2381.00	besk	203.26	77.00	6.00	5238.00
alder	10.92	11.00	5.00	49.00	alder	10.66	10.00	5.00	47.00
hovedstad	0.14	0.00	0.00	1.00	hovedstad	0.30	0.00	0.00	1.00
barn1son	0.69	1.00	0.00	1.00	barn1son	0.22	0.00	0.00	1.00
famceo	1.00	1.00	1.00	1.00	famceo	0.00	0.00	0.00	0.00
nborn	1.38	1.00	0.00	4.00	nborn	1.11	1.00	0.00	4.00

3. Tabel 1 viser at det gns. antal beskæftigede er 171 i periode 1 og 186 i periode 2. Virksomheder øger således antallet af ansatte mellem de to perioder. Desuden bemærkes det at antallet af beskæftigede varierer fra 6 til 5238 ansatte i data, hvilket er udtryk for en markant variation i antal beskæftigede på tværs af virksomhederne i stikprøven. Det gennemsnitlige primære resultat er 12.8 mio. kroner i periode 1 og 11.2 mio. kroner i periode 2. Der er altså en mindre gennemsnitlig tilbagegang for virksomhedernes gennemsnitlige profitabilitet mellem de to perioder. Virksomhederne har en gennemsnitlig alder på 6.8 år i periode 1 og 10.8 år i periode 2. Dette er en mekanisk ændring da alle virksomheder blot er 4 år ældre i periode 2. Den afgående direktør har i gennemsnittet 1.3 børn eller 2.5 børn hvis der ses bort fra direktører uden børn. Den førstefødte er i 50% af tilfældene en dreng, hvilket forventes såfremt barnets køn er tilfældigt bestemt. Disse familiekaraktistika varierer på tværs af virksomheder, men ikke mellem de to perioder i data. 56% af virksomheder i periode 2 vælger at ansætte en familiedirektør. Disse virksomheder har i gennemsnit et lavere primært resultat, en lavere beskæftigelse, en højere alder, og er oftere at finde uden for Hovedstadsområdet end de virksomheder som ansætter en professionel direktør. Det ses også at virksomheder som ansætter en familiedirektør typisk har flere børn i familien og kønnet på den fødslefødte langt oftere er en dreng.

Problem 2

Table 2: Estimationsresultater

	(1) OLS	(2) FD	(3) DiD	(4) FD-IV	(5) FD, pos.nborn	(6) FD-IV2
d2	0.0444* (0.0150)					
famceo	-0.0402* (0.0159)	-0.0605* (0.00664)	-0.0701* (0.00798)	-0.0606* (0.0160)	-0.0618* (0.00714)	-0.0707* (0.0151)
lbesk	0.208* (0.00664)	0.229* (0.0138)		0.239* (0.0163)	0.239* (0.0121)	0.238* (0.0122)
hovedstad	0.254* (0.00915)					
alder	-0.0639* (0.00908)					
aldersq	0.000916* (0.000464)	-0.000178 (0.000178)		-0.000930* (0.000140)	-0.000929* (0.000130)	-0.000921* (0.000130)
_cons	1.923* (0.0483)	-0.124* (0.0118)	-0.123* (0.00543)	-0.109* (0.0111)	-0.108* (0.0102)	-0.104* (0.0124)
N	1618	809	809	520	520	520

Standard errors in parentheses

* $p < 0.05$

1. Effekten af at ansætte en familiedirektør beregnes til $\hat{\beta}_1^{OLS} = -0.04$ ifølge kolonne 1 i Tabel 2. Altså forventes det primære resultat i periode 2 at falder med ca. 4 procent hvis virksomhed ansætter et familiemedlem som direktør i stedet for en professionel direktør (referencegruppen). OLS-estimatoren er konsistent såfremt MLR.1-MLR.4 er opfyldt. Det kræver at *famceo* er en eksogen variabel, hvilket betyder at begge fejllid skal være ukorrelerede med *famceo*. Denne betingelse er ikke opfyldt såfremt der er uobserverbare faktorer som har en direkte effekt på virksomhedernes primære resultat og deres beslutning om at ansætte en familiedirektør eller ej.
2. Modellen opskrives i første differenser:

$$\begin{aligned}
\Delta \log pmres_{i2} &= \log pmres_{i2} - \log pmres_{i1} \\
&= (\beta_0 + \delta_0 + \beta_1 famceo_{i2} + \beta_2 \log besk_{i2} + \beta_3 alder_{i2} + \beta_4 alder_{i2}^2 + \beta_5 hovedstad_{i2} + \alpha_i + \epsilon_{i2}) \\
&\quad - (\beta_0 + \beta_2 \log besk_{i1} + \beta_3 alder_{i1} + \beta_4 alder_{i1}^2 + \beta_5 hovedstad_{i1} + \alpha_i + \epsilon_{i1}) \\
&= \delta_0 + \beta_1 famceo_{i2} + \beta_2 \Delta \log besk_{i2} + \beta_3 \times 4 + \beta_4 \Delta alder_{i2}^2 + \Delta u_{i2} \\
&= \tilde{\delta}_0 + \beta_1 famceo_{i2} + \beta_2 \Delta \log besk_{i2} + \beta_4 \Delta alder_{i2}^2 + \Delta u_{i2}
\end{aligned}$$

hvor $\tilde{\delta}_0 = \delta_0 + 4\beta_3$. FD-transformationen eliminerer således β_0 og α_i . Desuden defineres ny skæring ved $\delta_0 + 4\beta_3$. Ved OLS-estimation af den transformerede model beregnes effekten af at ansætte en familiedirektør til $\hat{\beta}_1^{FD} = -0.06$. Ifølge FD-metoden forventes det primære resultat i periode 2 at falde med 6 procent, hvis virksomheden ansætter en familiedirektør. Effekten er således mere negativ ved FD-estimation end ved OLS-estimation. FD-estimatoren er konsistent såfremt $cov(famceo, \Delta u) = 0$, hvilket betyder at beslutningen om at ansætte en familiedirektør ikke er systematisk relateret til uobserverbare idiosynkratiske virksomhedsforhold. Fordelen ved FD-estimatoren er at endogenitetsproblem som skyldes uobserverbar heterogenitet fjernes ved hjælp af FD-transformationen. Dette kan være forhold som virksomhedskultur eller *firm performance* som ikke varierer over tid. Hvis FD-estimatoren er konsistent, må OLS-estimatoren være positivt biased, da $\hat{\beta}_1^{FD} < \hat{\beta}_1^{OLS}$. Ergo må en uobserverbar og tidsinvariant virksomhedskultur have en positiv effekt på det primære resultat og være positivt korreleret med ansættelsen af en familiedirektør.

3. Ministeriets model opstilles som:

$$\Delta \log pmres_{i2} = \beta_0 + \beta_1 famceo_{i2} + \epsilon$$

Effekten af at ansætte en familiedirektør beregnes til $\hat{\beta}_1^{DiD} = -0.07$ ifølge kolonne 3 i Tabel 2. Ovenstående model er blot en restrikeret udgave af FD-modellen svarende til hypotesen $H_0 : \beta_2 = \beta_4 = 0$. Det testes over for alternativhypotesen at H_0 ikke er sand. Et robust F-test beregnes til 138. Den relevante kritiske værdi er den 95. percentil i en F-fordeling med (2,805) frihedsgrader. På et 5% signifikansniveau afvises H_0 til fordel for alternativhypotesen. Testproceduren er asymptotisk valid under FD.1-FD.4 og antallet af observationer er gående mod uendelig. Konklusionen er derfor at antal beskæftigede samt virksomhedens alder er vigtige faktorer for virksomhedernes primære resultat og bør derfor indgå i Ministeriets beregningsmetode.

Table 3: Hjælperegressioner

	(1)	(2)	(3)
	1st	1st	OI test
barn1son	0.452* (0.0386)	0.454* (0.0382)	0.00413 (0.00698)
lbesk	-0.0849 (0.0647)	-0.0845 (0.0656)	0.000548 (0.0120)
aldersq	0.000487 (0.000752)	0.000449 (0.000705)	0.00000378 (0.000129)
nborn		0.0601* (0.0226)	-0.0120* (0.00413)
_cons	0.334* (0.0569)	0.217* (0.0695)	0.0213 (0.0127)
N	520	520	520

Standard errors in parentheses

* $p < 0.05$

Problem 3

1. Relevansen af instrumentet undersøges i 1st stage regressionen:

$$famceo_{i2} = \theta_0 + \theta_1 \text{barn1son} + \theta_2 \Delta \log \text{besk}_{i2} + \theta_4 \Delta \text{alder}_{i2}^2 + \epsilon_i$$

Ifølge første kolonne i Table 3 beregnes $\hat{\theta}_1 = 0.45$ med $se(\hat{\theta}_1) = 0.038$. Det er således 45 procent point mere sandsynligt at ansætte en familiedirektør hvis kønnet på den førstefødte er en dreng. Instrumentet indgår signifikant i 1st regressionen og er dermed relevant. Givet at kønnet på den første fødte er tilfældigt bestemt, bør instrumentet ikke være korreleret med Δu . Ergo bør instrumentet være validt. Effekten af at ansætte en familiedirektør estimeres til $\hat{\beta}_1^{FD-IV} = -0.06$, hvilket er tæt på identisk med FD-estimatet. FD-IV-metoden benytter imidlertid kun de observationer hvor den afgående direktør har børn. Ifølge kolonne 5 i Tabel 2, er FD-estimatet stort set uændret når stikprøven begrænses til observationer hvor instrumentet er til rådighed.

2. Tilfældet med flere instrumentvariable end endogene variable er også kendt som overidentifikation. IV-estimatoren vil være mere efficient i dette tilfælde end situationen med eksakt identifikation. En udvidet 1st regression viser at antal børn øger sandsynligheden for at ansætte en familiedirektør, jf. kolonne 2 i Tabel 3. Dermed er antal børn et relevant instrument. Validiteten af instrumentet er dog en mere problematisk antagelse, da valget af børn formentlig afhænger af uobserverbare faktorer som kan være relateret til uobserverbar *firm performance*. Effekten af at ansætte en familiedirektør estimeres også her til at være 7 procent. En hjælperegression af IV-residualerne på alle eksogene variable har en $R^2 = 0.017$. Testet for overidentifikation beregnes til 8.9. Sammenlignet med den 95. percentil i en χ^2 -fordeling med en frihedsgrad afvises testet. Ergo er mindst et instrument korreleret med u .

Problem 4

1. *Svar:* OLS-estimatorens grænsesandsynlighed er:

$$\begin{aligned}\text{plim } \hat{\beta}_1^{OLS} &= \beta_1 + \frac{\text{cov}(D_2, a)}{\text{var}(D_2)} + \frac{\text{cov}(D_2, u_2)}{\text{var}(D_2)} \\ &= \beta_1 + \frac{\text{cov}(D_2, a)}{\text{var}(D_2)} \\ &> \beta_1\end{aligned}$$

idet $\text{cov}(D_2, a) > 0$ og $\text{cov}(D_2, u_2) = 0$. OLS-estimatoren vil således overestimere effekten af programdeltagelse. Dermed er OLS ikke en konsistent estimator, dvs. $\text{plim } \hat{\beta}_1^{OLS} \neq \beta_1$.

2. *Svar:* Opskriv regressionsmodellen i første differenser:

$$\begin{aligned}\Delta y_{i2} &= y_{i2} - y_{i1} \\ &= (\beta_0 + 2\theta + \beta_1 D_{i2} + a_i + u_{i2}) - (\beta_0 + \theta + a_i + u_{i1}) \\ &= \theta + \beta_1 D_{i2} + \Delta u_{i2}\end{aligned}$$

FD-estimatorens grænsesandsynlighed er derfor:

$$\begin{aligned}\text{plim } \hat{\beta}_1^{FD} &= \beta_1 + \frac{\text{cov}(D_2, \Delta u_2)}{\text{var}(D_2)} \\ &= \beta_1 - \frac{\text{cov}(D_2, u_1)}{\text{var}(D_2)} \\ &> \beta_1\end{aligned}$$

idet $\text{cov}(D_2, u_2) = 0$ og $\text{cov}(D_2, u_1) < 0$. Ergo er FD-estimatoren positivt biased og vil derfor overestimere effekten af programmet. Dette skyldes at programdeltagelsen er negativt korreleret med det idiosynkratiske chok i periode 1, hvilket medfører en positiv korrelation mellem programdeltagelse og det idiosynkratiske chok udtrykt i første differenser.

3. *Svar:* Ved sammenligning af OLS og FD-estimatorerne ses det at:

$$\begin{aligned}\text{plim } \hat{\beta}_1^{FD} &= \beta_1 - \frac{\text{cov}(D_2, u_1)}{\text{var}(D_2)} > \beta_1 + \frac{\text{cov}(D_2, a)}{\text{var}(D_2)} = \text{plim } \hat{\beta}_1^{OLS} \\ &\text{hvis og kun hvis } |\text{cov}(D_2, u_1)| > |\text{cov}(D_2, a)|\end{aligned}$$

FD-estimatoren er således mere biased i forhold til OLS-estimatoren når kovariansen mellem programdeltagelse og den uobserverbare heterogenitet er større end kovariansen mellem programstatus og det idiosynkratiske fejllid i $t = 1$. Bemærk at denne betingelse kun gælder såfremt $D_{i2} = \Delta D_{i2}$ og dermed $\text{var}(D_2) = \text{var}(\Delta D_2)$. Dvs. deltagelse i programmet må kun finde sted i periode 2.

Problem 5

Parametrene ρ_1 og ρ_2 angiver om programdeltagelsen i $t = 2$ afhænger af den uobserverbare heterogeneitet og det idiosynkratiske fejllid i $t = 1$, jf. ligning (4).

I tilfældet $\rho_1 = 0, \rho_2 = 0$ er D en eksogen variabel og både OLS og FD forventes at være middelfrette og konsistente estimatorer. Tabel 4 bekræfter dette da gennemsnittet af de 1000 simulerede $\hat{\beta}_1^{OLS}$ og $\hat{\beta}_1^{FD}$ -estimatorer i begge tilfælde er lig den sande værdi, $\beta_1 = -3$. Det bemærkes at $sd(\hat{\beta}_1^{FD})$ er lavere end $sd(\hat{\beta}_1^{OLS})$, hvilket viser at FD-estimatoren er mere effektiv end OLS-estimatoren i dette tilfælde. Dette resultat afhænger dog af variansantagelserne for de to fejllid, $Var(a)$ og $Var(u)$, og det kan derfor ikke konkluderes at FD-estimatoren altid er mere effektiv end OLS-estimatoren.

OLS-estimatoren forventes at være konsistent i tilfældet $\rho_1 = -1, \rho_2 = 0$, da programdeltagelse ikke afhænger af a . Omvendt forventes FD-estimatoren at være positivt biased, idet $Cov(D_2, \Delta u_2) > 0$, jf. besvarelsen til Opgave 4.2. Tabel 4 bekræfter disse forventninger, da gennemsnittet af de simulerede $\hat{\beta}_1^{OLS}$ -estimatorer er lig den sande værdi, $\beta_1 = -3$, hvorimod FD-estimatorerne i gennemsnit er lig -0.44 . FD-estimatoren indeholder således en positiv bias svarende til $-0.44 - (-3) = 2.56$. Det bemærkes at FD-estimatoren har den mindste varians. OLS-estimatoren foretrækkes dog, da den er konsistent.

I tilfældet $\rho_1 = 0, \rho_2 = 1$ er D positivt korreleret med den uobserverbare heterogeneitet, men ukorreleret med u . OLS-estimatoren forventes derfor at være positivt biased, jf. besvarelsen til Opgave 4.1. FD-estimatoren forventes derimod at være konsistent, da denne metode netop eliminerer endogenitetsproblemer som skyldes uobserverbar heterogenitet. Ifølge Tabel 4 har OLS-estimatoren en positiv bias svarende til $-0.27 - (-3) = 2.73$, mens FD-estimatorerne i gennemsnit er lig den sande værdi, $\beta_1 = -3$. FD-estimatoren er således konsistent i dette tilfælde.

I det sidste tilfælde er D en endogen variabel fordi den varierer med begge fejllid. Derfor forventes begge estimatorer at være biased og inkonsistente, jf. Opgave 4. Simulationseksperimentet bekræfter dette, da gennemsnittet af de simulerede $\hat{\beta}_1^{OLS}$ -estimatorer er lig -0.86 , hvilket svarer til en bias på 2.14. FD-estimatoren er mindre biased, da gennemsnittet af FD-estimatorerne er tættere på den sande værdi, $\beta_1 = -3$. Ifølge besvarelsen til Opgave 4.3 er dette et udtryk for at $|cov(D_2, u_1)| < |cov(D_2, a)|$. Dvs. at D er mere korreleret med den uobserverbare heterogenitet end det idiosynkratiske fejllid for $t = 1$. Dette medfører en større bias i OLS-estimatoren.

Table 4: Simulationsresultaterne

		Mean	Std.dev.	Min	Max
$\rho_1 = 0, \rho_2 = 0$:	OLS	-3.001	0.237	-3.802	-2.162
	FD	-2.991	0.220	-3.672	-2.296
$\rho_1 = -1, \rho_2 = 0$:	OLS	-2.996	0.237	-3.894	-2.156
	FD	-0.441	0.180	-0.996	0.116
$\rho_1 = 0, \rho_2 = 1$:	OLS	-0.267	0.200	-0.885	0.384
	FD	-2.994	0.215	-3.713	-2.311
$\rho_1 = -1, \rho_2 = 1$:	OLS	-0.859	0.219	-1.621	-0.054
	FD	-1.229	0.199	-1.804	-0.543