

Øvelsestime 10

Paneldata

Agenda

- 1) Introduktion til paneldata og PS10
- 2) Gruppearbejde: Diskussion
- 3) Opsamling
- 4) Gruppearbejde: Stata
- 5) Opsamling

Tobias Gabel Christiansen
tgc@econ.ku.dk
17-05-2021

KØBENHAVNS UNIVERSITET



0) Planen for semesteret

Dato	Opgave
Uge 7 - 15/02	PS1
Uge 8 - 22/02	PS2
Uge 9 - 01/03	PS3
Uge 10 - 08/03	OO1
Uge 11 - 15/03	PS4
Uge 12 - 22/03	PS5 + gennemgang af OO1
Uge 13 - 29/03	PS6
Uge 14 - 05/04	Påskeferie
Uge 15 - 12/04	OO2
Uge 15 - 19/04	PS7
Uge 16 - 26/04	PS8 + gennemgang af OO2
Uge 17 - 03/05	OO3
Uge 18 - 10/05	PS9
Uge 19 - 17/05	PS10 + gennemgang af OO3
	Spørgetime

 Her er vi :-)

1) Panel data

Vi har indtil nu set på **tværsnits data**:

$$\underset{(N \times 1)}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad \text{og} \quad \underset{(N \times K)}{X} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}$$

Og **tidsserie data**:

$$\underset{(T \times 1)}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \quad \text{og} \quad \underset{(T \times K)}{X} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{T1} & x_{T2} \end{pmatrix}$$

I dag skal vi se nærmere på **panel data**:

$$\underset{(T \times 1)}{y_i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix} \quad \text{og} \quad \underset{(T \times K)}{X_i} = \begin{pmatrix} x'_{i1} \\ \vdots \\ x'_{iT} \end{pmatrix} \quad (\text{data for observation } i)$$

$$\underset{(NT \times 1)}{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \text{og} \quad \underset{(NT \times K)}{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \quad (\text{data for alle observationer})$$

1) Panel data

- Med Panel data er der altså variation både på tværs af enheder (personer, virksomheder etc.) og på tværs af tid (år, måneder etc.) for den enkelte enhed.
- Fordelene ved Panel data er at:
 - i) Vi kan analysere hvordan effekter varierer over tid.
 - ii) Vi kan kontrollere for uobserverbar heterogenitet hvis denne antages at være konstant over tid.
- Ulempen ved Panel data er det ofte er svært/omkostningsfuldt at indsamle. Kræver at vi følger de samme personer/virksomheder/lande etc. over tid og løbende registrere de faktorer vi er interesserede i.

1) Panel data

Tidsvarierende effekter:

- Med Panel data kan vi analysere om effekterne varierer over tid, ved inkluderer tidsdummier og interaktioner med disse:

$$y_{it} = \beta_0 + \beta_1 dPeriod2_t + \beta_2 x_{it} + \beta_3 (x_{it} \times dPeriod2_t) + u_{it}$$

- β_1 fanger forskelle i skæringen med y-aksen fra periode 1 til periode 2.
- β_3 fanger forskelle i den forventede effekt af at øge x_{it} med en enhed i periode 2 ift. periode 1.

1) Panel data

Uobserverbar heterogenitet:

- Hvis der er uobserverbare forskelle på tværs af vores enheder og disse forskelle korrelerer med vores forklarende variabel hvis OLS estimatoren være biased og inkonsistent.
- Hvis de uobserverbare forskelle er konstante over tid (a_i) kan vi kontrollere for dette og dermed eliminere denne kilde til bias vha. forskellige panel data metoder fx First-differences:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

$$y_{it} - y_{it-1} = \beta_0 - \beta_0 + \beta_1(x_{it} - x_{it-1}) + a_i - a_i + u_{it} - u_{it-1}$$

$$\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta u_{it}$$

- På denne måde eliminerer vi a_i fra modellen så endogenitet der skyldes $cov(a_i, x_{it}) \neq 0$ fjernes også.
- Vi render stadig ind i problemer hvis den tidsvarierende del af fejlledet (u_{it}) korrelerer med de forklarende variable.

1) Problem set 10

- I PS10 skal vi se nærmere på tilbagetrækningsalderen.
- Vi har data for 634 ældre danskere som vi observerer både i 1997 og i 2002 (to observationer for hver person).
- For alle observationerne gælder det at:
 - i) Personerne er født i 1945
 - ii) Personerne var i arbejde i begge år
- Vores afhængige variabel er *tbalder* som måler forventet tilbagetrækningsalder
- Vi bruger en række variable til at forklare den forventede tilbagetrækningsalder gennem en række socioøkonomiske variable (helbred, uddannelsesniveau, arbejdserfaring, størrelsen på offentlig pension, størrelsen på kapitalpension, køn og hvorvidt personen er single).

2) Gruppearbejde - Diskussion

Brug 15 minutter på at diskutere følgende:

$$\begin{aligned} tbalder_{it} = & \beta_0 + \beta_1 enlig_{it} + \beta_2 helbred_{it} + \beta_3 dudd1_{it} + \beta_4 dudd2_{it} + \beta_5 erfar_{it} \\ & + \beta_6 pension_{it} + \beta_7 kpenbelob_{it} + \beta_8 d2002_{it} + \beta_9 kvinde_{it} \\ & + \beta_{10} pension_{it} \cdot kvinde_{it} + \epsilon_{it} \end{aligned} \quad (1)$$

1. Discuss whether the parameters of model (1) can be estimated consistently by the (pooled) OLS estimator using all available observations. What assumptions must be made about the unobserved error terms, a_i and u_{it} ?
2. If all variables are first-differenced, can we then assume that the conditions for consistent estimation, FD.1—FD.4 are satisfied? What assumptions must then be made about the unobserved error terms, a_i and u_{it} ?

Brug for hjælp? Skriv Jeres gruppenummer i Padlet (link i chatten)

2) Gruppearbejde - Opsamling

1. Discuss whether the parameters of model (1) can be estimated consistently by the (pooled) OLS estimator using all available observations. What assumptions must be made about the unobserved error terms, a_i and u_{it} ?

2) Gruppearbejde - Opsamling

1. Discuss whether the parameters of model (1) can be estimated consistently by the (pooled) OLS estimator using all available observations. What assumptions must be made about the unobserved error terms, a_i and u_{it} ?

- Betragt regressionsmodellen:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

- Hvor a_i angiver uobserverbar heterogenitet på tværs af individer der er konstant over tid mens u_{it} ofte refereres til som et idiosynkratisk fejllid (svarer til det fejllid vi kender).
- Det samlede fejllid (composite error) er altså givet ved $\epsilon_{it} = a_i + u_{it}$
- Pooled OLS estimatoren er konsistent under følgende antagelser:

MLR.1 – Lineær i parametre

MLR.2 – Tilfældig stikprøve

MLR.3 – Ingen perfekt multikollinearitet

MLR.4 - $E(\epsilon_{it}|x_{it}) = E(a_i + u_{it}|x_{it}) = 0$

Bemærk: I virkeligheden er $cov(a_i + u_{it}, x_{it}) = 0$ er nok for konsistens.

2) Gruppearbejde - Opsamling

1. Discuss whether the parameters of model (1) can be estimated consistently by the (pooled) OLS estimator using all available observations. What assumptions must be made about the unobserved error terms, a_i and u_{it} ?
- For at POLS estimatoren er konsistent kræver vi altså at:
 1. Det idiosynkratiske fejllid i periode t ikke samvarierer (ikke korrelerer er nok) med de forklarende variable i periode t
$$E(u_{it}|x_{it}) = 0$$
 2. Den uobserverbare heterogenitet ikke samvarierer (ikke korrelerer er nok) med med de forklarende variable
$$E(a_i|x_{it}) = 0$$
- Under disse antagelser kan vi altså glemme panelstrukturen og samle (pool) tværsnitsdataet for hver periode og estimere modellen med OLS (pooled OLS). Dvs. vi bruger blot OLS på data som det er uden at tænke over panelstrukturen.
- I vores model er sidstnævnte antagelse ovenfor dog formegentlig brudt da a_i indeholder ting som præferencer for fritid, der nok både har en effekt på tilbagetrækningsalderen og individets historiske pensionsindbetalinger.
- Selv hvis antagelserne er overholdt vil Pooled OLS estimatoren ikke være efficient da den ikke tager højde for korrelation i fejllidet (ϵ_{it}) på tværs af tidsperioder. Random Effects (RE) estimatoren, som korrigerer for dette, vil være mere efficient.

2) Gruppearbejde - Opsamling

2. If all variables are first-differenced, can we then assume that the conditions for consistent estimation, FD.1—FD.4 are satisfied? What assumptions must then be made about the unobserved error terms, a_i and u_{it} ?

2) Gruppearbejde - Opsamling

2. If all variables are first-differenced, can we then assume that the conditions for consistent estimation, FD.1—FD.4 are satisfied? What assumptions must then be made about the unobserved error terms, a_i and u_{it} ?

- Betragt regressionsmodellen:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it} \quad (\text{FD.1})$$

- Vi kan fjerne den uobserverbare heterogenitet ved at transformere modellen. En måde er ved at tage første-differencer (FD):

$$\Delta y_{it} = \beta_1 \Delta x_{it} + \Delta u_{it} \quad (\text{FD.2})$$

- Denne transformation fjerner alle observerbare og uobserverbare variable der er konstante over tid, inklusiv a_i .
- Den fjerner dermed også konstantleddet. Ofte inkluderer vi en dummy for periode t i FD-modellen, og tolker dette som ændringen i konstantleddet over tid.
- Koefficienten på den tidsvariende variabel (x_{it}) er uændret. Fortolkningen af parametrene i (FD.2) og (FD.1) er altså den samme (Bemærk: dette gælder også ved FE transformationer)

2) Gruppearbejde - Opsamling

2. If all variables are first-differenced, can we then assume that the conditions for consistent estimation, FD.1—FD.4 are satisfied? What assumptions must then be made about the unobserved error terms, a_i and u_{it} ?
- Estimatoren hvor vi estimerer den transformerede model med OLS kaldes det for FD-estimatoren og denne er konsistent når:

FD.1: Modellen er lineær i parametre.

FD.2: Tilfældig stikprøve.

FD.3: Ingen perfekt multikollinearitet og alle variable vi er interesserede i skal variere over tid.

FD.4: $E[\Delta u_{it} | \Delta x_{it}] = 0$

- Vi behøver altså ikke at antage noget om relationen mellem a_i og x_{it} (udover at a_i er konstant) men kræver at $E[\Delta u_{it} | \Delta x_{it}] = 0$. Dette kan være brudt hvis vi udelader tidsvarierende variable der korrelerer med ændringer i vores forklarende variable.

Bemærk: I virkeligheden er $cov(\Delta u_{it}, \Delta x_{it}) = 0$ nok for konsistens

3) Gruppearbejde - Stataøvelser

Arbejd på Stataøvelser indtil 12:30

Brug for hjælp? Skriv Jeres gruppenummer i Padlet (link i chatten)

4) Gruppearbejde - Opsamling

1. Describe the variables in model (1)

- Variablene for køn og uddannelse er konstante over tid. Derudover stiger erhvervserfaringen lineært med tiden. Effekterne af disse variable kan altså kun estimeres med POLS.
- De resterende variable varierer over tid og effekterne af disse kan estimeres med FD/FE.
- Den forventede tilbagetrækningsalder stiger med ca. 1 år over perioden.
- Vi ser desuden at der er lidt flere med dårligt helbred og lidt flere enlige i 2002 end i 1997.
- De forventede pensionsudbetalinger fra det offentlige stiger en smule, mens de forventede udbetalinger fra privatpension stiger meget.
- Bemærk desuden at nogle ældre står til at få 0 i forventet offentlig pension. Potentiel fejl i data.

```
bysort d2002: outreg2 using deskriptiv.tex, replace tex(frag) sum(detail) dec(2) ///
keep(tbalder enlig helbred dudd1 dudd2 erfar pension kpenbelob kvinde pension) ///
eqkeep(mean sd min max)
```

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	d2002 0				d2002 1			
VARIABLES	mean	sd	min	max	mean	sd	min	max
tbalder	60.81	2.62	51.00	70.00	61.89	1.96	57.00	70.00
kvinde	0.45	0.50	0.00	1.00	0.45	0.50	0.00	1.00
enlig	0.19	0.40	0.00	1.00	0.21	0.40	0.00	1.00
dudd1	0.14	0.35	0.00	1.00	0.14	0.35	0.00	1.00
dudd2	0.56	0.50	0.00	1.00	0.56	0.50	0.00	1.00
erfar	33.01	5.05	12.00	42.00	38.01	5.05	17.00	47.00
helbred	0.14	0.35	0.00	1.00	0.18	0.38	0.00	1.00
pension	123.87	34.10	0.00	136.51	133.28	44.73	0.00	157.04
kpenbelob	8.22	22.91	0.00	254.84	211.94	244.96	0.00	1,056.94

4) Gruppearbejde - Opsamling

2. Estimate model (1) by pooled OLS

- Modellen estimeres med POLS og robuste standardfejl (standardfejl udeladt i tabel til højre)
- Under antagelserne fra tidligere (slide 10-11) vil POLS estimatoren være konsistent.
- Vi ser fx at folk med dårligt helbred forventes at gå på pension ca. 1,1 år tidligere end folk med godt helbred (når de andre faktorer holdes fast).
- Desuden, og i overensstemmelse med den deskriptive analyse, ser vi at den forventede pensionsalder stiger med et år i 2002 ift. 1997.

VARIABLES	(1) OLS
enlig	0.167
helbred	-1.095***
dudd1	-0.844***
dudd2	-0.491***
erfar	-0.0146
pension	-0.00846***
kpenbelob	0.00103***
d2002	1.017***
kvindeXpension	0.0140***
Constant	63.28***
Observations	1,268
R-squared	0.163

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4) Gruppearbejde - Opsamling

2. Estimate model (1) by pooled OLS

- Vi kan teste om kvinders og mænds tilbagetrækning påvirkes forskelligt af øget offentlig pension med følgende hypotese:

$$H_0: \beta_{10} = 0 \quad H_a: \beta_{10} \neq 0$$

- Vi kan benytte en (robust) t-test til dette. Under antagelserne fra tidligere vil t-statistikken være asymptotisk normalfordelt under H_0 .
- Vi finder en t-værdi på 3,54. Nulhypotesen afvises derfor på et 5 pct. Signifikansniveau.
- Estimaterne viser at den offentlige pension har signifikant forskellig effekt på kvinder og mænd. Øges den offentlige pension med 1000 DKK fører det til en tidligere (-0,008 år) tilbagetrækning for mænd og senere for kvinder (0,006 år).

VARIABLES	(1) OLS
enlig	0.167
helbred	-1.095***
dudd1	-0.844***
dudd2	-0.491***
erfar	-0.0146
pension	-0.00846***
kpenbelob	0.00103***
d2002	1.017***
kvindeXpension	0.0140***
Constant	63.28***
Observations	1,268
R-squared	0.163

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4) Gruppearbejde - Opsamling

2. Estimate model (1) by pooled OLS

- Vi kan også teste om kvinders tilbagetrækning i det hele taget påvirkes af ændringer i den offentlige pension ud fra følgende hypotese:

$$H_0: \beta_6 + \beta_{10} = 0 \quad H_a: \beta_6 + \beta_{10} \neq 0$$

- Vi kan benytte en (robust) t-test til dette. Det kræver dog at vi reparametriserer modellen, hvilket gøres ved at indsætte $\beta_6 = -\beta_{10}$ i den oprindelige model, omskrive og estimerer den nye model.
- Alternativt – og hurtigere – kan vi benytte en robust F-test ved efter regression at skrive:

`test pension+kvindeXpension=0`

- Gør vi dette finder vi:

`F( 1, 1257) = 3.82`
`Prob > F = 0.0509`

- Med en p-værdi på 0,0509 kan vi altså lige akkurat ikke forkaste nulhypotesen om der ikke er en effekt på et 5% signifikansniveau.
- Bemærk: F-testen med én restriktion svarer til den kvadrerede værdi af t-testen. Havde vi derfor reparametriseret modellen og brugt en robust t-test havde vi fået en t-værdi på $\sqrt{3,82} = 1,95$, hvilket havde givet præcis samme konklusion.

VARIABLES	(1) OLS
enlig	0.167
helbred	-1.095***
dudd1	-0.844***
dudd2	-0.491***
erfar	-0.0146
pension	-0.00846***
kpenbelob	0.00103***
d2002	1.017***
kvindeXpension	0.0140***
Constant	63.28***
Observations	1,268
R-squared	0.163
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

4) Gruppearbejde - Opsamling

3. Estimate model (1) by the first-differenced (FD) estimator.

- Nå vi estimerer modellen med FD-estimatoren udgår alle variable der ikke varierer over tid eller som varierer lineært med tiden.

$$\Delta tbalder_{it} = \delta_0 + \delta_1 \Delta enlig_{it} + \dots + \delta_{10} \Delta (pension_{it} \times kvinde_{it}) + \epsilon_{it}$$

- Vi kan derfor fx ikke estimere effekten af ændringer i erfaring (β_5) da denne variable vokser lineært med tiden for alle observationer.
- "Konstantleddet" for FD-estimatoren fanger ændringen i konstantleddet fra 1997 til 2002. Med andre ord fanger dette niveauforskellene på tværs af de to perioder. Dette er præcis det samme som vores difd2002 dummy gør (β_8), hvorfor denne udlades pga. perfekt kollinearitet
- Vi kan stadig estimere interaktionsleddet (β_{10}) selvom køn ikke varierer over tid. Det skyldes at pensionen ændrer sig over tid, og vi estimerer dermed hvordan denne ændring afhænger af køn.
- Læg desuden mærke til at vi kun har 634 observationer nu (vi mister 634 observationer fordi vi tager differensen)

```
sort lbnr d2002
xtset lbnr d2002

foreach var in tbalder enlig helbred dudd1 dudd2 erfar pension kpenbelob d2002 kvinde kvindeXpension {
    by lbnr: generate dif`var'=`var'[_n]-`var'[_n-1]
    generate D`var'=d.`var'
}

regress diftbalder difenlig difhelbred difdudd1 difdudd2 diferfar difpension ///
        difkpenbelob difd2002 difkvinde difkvindeXpension, robust
```

VARIABLES	(1) FD
difenlig	-0.457
difhelbred	-0.528***
o.difdudd1	-
o.difdudd2	-
o.diferfar	-
difpension	-0.00334
difkpenbelob	0.000252
o.difd2002	-
o.difkvinde	-
difkvindeXpension	0.0114**
Constant	1.048***
Observations	634
R-squared	0.026
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

4) Gruppearbejde - Opsamling

3. Estimate model (1) by the first-differenced (FD) estimator.

Vi kan teste om ændringer i kvinders og mænds tilbagetrækning påvirkes forskellige af øget offentlig pension med følgende hypotese:

$$H_0: \beta_{10} = 0 \quad H_a: \beta_{10} \neq 0$$

Vi kan benytte en (robust) t-test til dette. Under antagelserne fra tidligere vil t-statistikken være asymptotisk normalfordelt under H_0 .

Vi finder en t-værdi på 2,34. Nulhypotesen afvises derfor på et 5 pct. Signifikansniveau.

Øget offentlig pension har altså en signifikant forskellig effekt på mænd og kvinder.

Vi finder igen at $\hat{\beta}_6 < 0$ og $\hat{\beta}_6 + \hat{\beta}_{10} > 0$. Effekten for mænd er negativ, mens den er positiv for kvinder.

```
sort lbnr d2002
xtset lbnr d2002

foreach var in tbalder enlig helbred dudd1 dudd2 erfar pension kpenbelob d2002 kvinde kvindeXpension {
    by lbnr: generate dif`var'=`var'[_n]-`var'[_n-1]
    generate D`var'=d.`var'
}

regress diftbalder difenlig difhelbred difdudd1 difdudd2 diferfar difpension ///
difkpenbelob difd2002 difkvinde difkvindeXpension, robust
```

VARIABLES	(1) FD
difenlig	-0.457
difhelbred	-0.528***
o.difdudd1	-
o.difdudd2	-
o.diferfar	-
difpension	-0.00334
difkpenbelob	0.000252
o.difd2002	-
o.difkvinde	-
difkvindeXpension	0.0114**
Constant	1.048***
Observations	634
R-squared	0.026
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

4) Gruppearbejde - Opsamling

3. Estimate model (1) by the first-differenced (FD) estimator.

Vi kan også teste om kvinders tilbagetrækning i det hele taget påvirkes af ændringer i den offentlige pension ud fra følgende hypotese:

$$H_0: \beta_6 + \beta_{10} = 0 \quad H_a: \beta_6 + \beta_{10} \neq 0$$

Igen benytter vi en robust F-test ved efter regression at skrive:

Gør vi dette finder vi:

```
test difpension+difkvindeXpension=0
```

```
F( 1, 628) = 4.04
Prob > F = 0.0450
```

Med en p-værdi på 0,045 kan vi altså lige akkurat forkaste nulhypotesen om pension ikke har på et 5% signifikansniveau.

```
sort lbnr d2002
xtset lbnr d2002

foreach var in tbalder enlig helbred dudd1 dudd2 erfar pension kpenbelob d2002 kvinde kvindeXpension {
    by lbnr: generate dif`var'=`var'[_n]-`var'[_n-1]
    generate D`var'=d.`var'
}

regress diftbalder difenlig difhelbred difdudd1 difdudd2 diferfar difpension ///
        difkpenbelob difd2002 difkvinde difkvindeXpension, robust
```

VARIABLES	(1) FD
difenlig	-0.457
difhelbred	-0.528***
o.difdudd1	-
o.difdudd2	-
o.diferfar	-
difpension	-0.00334
difkpenbelob	0.000252
o.difd2002	-
o.difkvinde	-
difkvindeXpension	0.0114**
Constant	1.048***
Observations	634
R-squared	0.026
Robust standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

4) Gruppearbejde - Opsamling

4. Estimate the model by the Fixed Effects (FE) estimator.

- Vi estimerer nu modellen med FE:

$$tbalder_{it} - \overline{tbalder_i} = \gamma_0 + \gamma_1(enlig_{it} - \overline{enlig_i}) + \dots + v_{it}$$

- Det kan gøres manuelt ved selv at danne variablene:

```
foreach var in tbalder enlig helbred dudd1 dudd2 erfar pension kpenbelob d2002 kvinde kvindeXpension {
  by lbnr: egen m_`var'=mean(`var')
  generate wt_`var'=`var'-m_`var'
}
```

- Og derefter estimerer modellen med de transformerede variable med OLS.
- Alternativt kan vi benytte Stata's indbyggede kommando der udføres ved at skrive "xtreg y x, fe".
- Fordelen ved at benytte Stata's indbyggede kommando er at denne tager højde at vi bruger frihedsgrader på at de-mean variablene (Standardfejlene der fås ved manuelt at implementere FE-estimatoren tager ikke højde for dette)
- Den nemmeste måde at forstå denne frihedsgrads-korrektion er ved at indse af FE-estimatoren svarer til Least-Squares Dummy Variable Estimatoren som inkluderer en dummy for hver af vores observationer. Her kontrollerer vi altså direkte for α_i ved at inkludere en dummy for hvert i (bortset fra en enkelt der fungerer som referencegruppe). Her er det klart at vi har færre frihedsgrader da vi skal estimere flere parametre (en for hver dummy).

```
xtreg tbalder enlig helbred dudd1 dudd2 erfar ///
      pension kpenbelob d2002 kvinde kvindeXpension, fe robust
```

VARIABLES	(1) FE
enlig	-0.457
helbred	-0.528***
o.dudd1	-
o.dudd2	-
erfar	0.210***
pension	-0.00334
kpenbelob	0.000252
o.d2002	-
o.kvinde	-
kvindeXpension	0.0114**
Constant	53.83***
Observations	1,268
Number of lbnr	634
R-squared	0.190

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4) Gruppearbejde - Opsamling

4. Estimate the model by the Fixed Effects (FE) estimator.

- Estimerer identiske med FD-estimatoren. Skyldes af FD og FE estimatoren er identiske når vi kun har to tidsperioder.
- For flere end to tidsperioder varierer de to estimators.
- Bemærk dog at vi ved FE estimation får et estimat for effekten af øget erfaring (β_5).
- Det skyldes at de FE-transformerede observationer for denne variabel er -2,5 i 1997 og 2,5 i 2002. Variablen varierer derfor i modsætning til FD-transformationen hvor ændringen i erfaring er lig 5 for alle observationer.
- Variationen vokser dog lineært med tiden, hvorfor dette svarer til at inkludere en tidsdummy. Vi kan derfor udelade "erfar" og inkluderer "d2002" i stedet uden at det ville ændre noget (de to er perfekt korrelerede). Se fx også at effekten på tbalder fra 1997 til 2002 målt ud fra "erfar" er lig $5 \cdot 0,21 = 1,05$ som er lig koefficienten på "d2002" hvis denne inkluderes i stedet.

Bemærk: Konstantleddet ved FE er et gennemsnit af de individuelle fixed effects (a_i) der implicit estimeres og bør ikke tolkes / tillægges mening.

```
xtreg tbalder enlig helbred dudd1 dudd2 erfar ///
      pension kpenbelob d2002 kvinde kvindeXpension, fe robust
```

VARIABLES	(1) FE
enlig	-0.457
helbred	-0.528***
o.dudd1	-
o.dudd2	-
erfar	0.210***
pension	-0.00334
kpenbelob	0.000252
o.d2002	-
o.kvinde	-
kvindeXpension	0.0114**
Constant	53.83***
Observations	1,268
Number of lbnr	634
R-squared	0.190

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

4) Gruppearbejde - Opsamling

5. Make a summary of the relevant results in the form of a table.

- For at gøre tabellen mere overskuelig er navnene på FD-variablene (fx difenlig) omdøbt til de originale navne (fx enlig).
- Hvis den uobserverede heterogenitet (α_i) korrelerer med de forklarende variable er POLS ikke konsistent, mens FE/FD er konsistente.
- Pga. de store forskelle på tværs af POLS og FE/FD estimaterne er der noget der tyder på at dette er tilfældet. Vi foretrækker derfor FE/FD-estimatorerne.
- Det ses fx at partielle effekt af helbred er væsentlig mindre ved FE/FD estimation.
- Overordnet kan det konkluderes at størrelsen af offentlige pensioner har en beskeden effekt på tilbagetrækningsalderen. Dog viser resultaterne at mænd og kvinder har forskellig tilbagetrækningsadfærd.

VARIABLES	(1) OLS	(2) FE	(3) FD
enlig	0.167	-0.457	-0.457
helbred	-1.095***	-0.528**	-0.528**
dudd1	-0.844***		
dudd2	-0.491***		
erfar	-0.0146		
pension	-0.00846***	-0.00334	-0.00334
kpenbelob	0.00103***	0.000252	0.000252
d2002	1.017***		
kvinde	-2.716***		
kvindeXpension	0.0140***	0.0114***	0.0114***
Constant	63.28***	53.83***	1.048***
Observations	1,268	1,268	634
R-squared	0.163	0.190	0.026
Number of lbnr		634	

Robust standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Tak for i dag

- Eksamenen er tirsdag d. 15 juni
- Husk det er en 12 timers hjemmeksamenen (overblik fremfor udenadslæren)
- Spørgetime torsdag d. 10 juni kl. 10.00-13.00
- Send spørgsmålene til mig på mail senest onsdag d. 9 juni kl 15:00