`xtlhazard`: Linear discrete time hazard estimation using Stata

Harald Tauchmann[1,2,3]

[1]FAU, [2]RWI, [3]CINCH

May 24th 2019

2019 German Stata Users Group Meeting

work in progress

Outline

Motivation

Hazard models / duration analysis / survival analysis / models for **non-repeated** events & **absorbing states**

- » Modelling (directional) transitions

1. **Continuous time hazard** models
   - » Parametric (Weibull, Gompertz, exponential, ...) models ($\rightarrow$`streg`)
   - » Semi-parametric (Cox) models ($\rightarrow$`stcox`)
   - » **Not considered in this talk**
2. **Discrete time hazard** models
   - » Stacked **binary outcome** models (probit, logit, ...)

Motivation II

- **Unobserved** individual **heterogeneity** ("frailty")
  - Random effects
    - Straightforward (integrating out)
    - No correlation with regressors allowed
  - **Fixed effects**
    - Incidental parameters problem
    - Computationally demanding (possibly intractable)
- **Linear probability** model alternative that allows for **linear fixed effects** estimation?

Does Linear Fixed Effects Estimation Work?

- **Left-hand-side** $y_{i1}, \ldots, y_{iT}$ for unit i in panel of length T
  - $0, 0, \ldots, 0, 0, 0, 0$ (censored)
  - $0, 0, \ldots, 0, 1, 1, 1$ ($\rightarrow$ no info in second, third, ... 1)
  - $0, 0, \ldots, 0, 1$ ($\rightarrow$ effectively $T_i \leq T$ obs. if not cens.)
- **Within-transformed** lhs variable (i observed T_i periods)
  - $0, 0, \ldots, 0, 0, 0, 0$ (censored)
  - $-\frac{1}{T_i}, -\frac{1}{T_i}, \ldots, -\frac{1}{T_i}, \frac{T_i - 1}{T_i}$ (not censored)
  - Transformation has **little effect** on lhs (at least for large T_i)
- **First-differenced** lhs variable (i observed T_i periods)
  - $0, \ldots, 0, 0, 0, 0$ (censored)
  - $0, \ldots, 0, 1$ (not censored)
  - (Besides loosing y_{i1}) transformation has **no effect at all** due to $y_{it-1} = 0$

Does Linear Fixed Effects Estimation Work? II

- **Can transformations that (almost) do not transform the left-hand-side variable eliminate individual heterogeneity?**
- **Implicit** answer of the **literature** seems to be **"yes"**:
  - Miguel et al. (2004, *Journal of Political Economy*)
  - Ciccone (2011, *AEJ: Applied*)
  - Brown and Laschever (2012, *AEJ: Applied*)
  - Cantoni (2012, *Economic Journal*)
  - Harding and Stasavage (2014, *Journal of Politics*)
  - Jacobson and von Schedvin (2015, *Econometrica*)
  - Wang et al. (2017, WP)
  - Bogart (2018, *Economic Journal*)

The Data Generating Process

$$y_{it} = a_i + \mathbf{x}_{it}\beta + \varepsilon_{it}$$

$$\varepsilon_{it} = \begin{cases} 1 - a_i - \mathbf{x}_{it}\beta & \text{if} \quad t = T_i \quad \text{and } i \text{ is not censored} \\ -a_i - \mathbf{x}_{it}\beta & \text{if} \quad t = T_i \quad \text{and } i \text{ is censored} \\ -a_i - \mathbf{x}_{it}\beta & \text{if} \quad t < T_i \end{cases}$$

- a_i unobserved time-invariant individual heterogeneity
- $a_i + \mathbf{x}_{it}\beta \in [0, 1] \; \forall \; it$

Assumption rendering above equation **regression model**:

$$\mathrm{E}\left(\varepsilon_{it} \middle| a_i, \mathbf{x}_{it}, \mathbf{y}_{it^-} = \mathbf{0}\right) = 0 \qquad \text{with} \quad \mathbf{y}_{it^-} \equiv [y_{i0} \ldots y_{it-1}]$$

$$\Rightarrow \mathrm{P}\left(y_{it} = 1 \middle| a_i, \mathbf{x}_{it}, \mathbf{y}_{it^-} = \mathbf{0}\right) = a_i + \mathbf{x}_{it}\beta$$

Estimation by pooled OLS

$$y_{it} = \alpha^c + \mathbf{x}_{it}\beta + \varepsilon_{it}^{\text{OLS}}$$

- $\varepsilon_{it}^{\text{OLS}} \neq \varepsilon_{it}$, since a_i not included as regressor

Conditional mean of disturbance:

$$\begin{aligned} \mathrm{E}\left(\varepsilon_{it}^{\text{OLS}} | a_i, \mathbf{x}_{it}, \mathbf{y}_{it^-} = \mathbf{0}\right) &= \left(a_i + \mathbf{x}_{it}\beta\right)\left(1 - \alpha^c - \mathbf{x}_{it}\beta\right) \\ &\quad + \left(1 - a_i - \mathbf{x}_{it}\beta\right)\left(-\alpha^c - \mathbf{x}_{it}\beta\right) \\ &= a_i - \alpha^c \end{aligned}$$

- Renders OLS biased and inconsistent if $\mathrm{Cov}(a_i, \mathbf{x}_{it}) \neq \mathbf{0}$
- First-differences or within-transformation to eliminate a_i?

Estimation by First-Differences Estimation

$$y_{it} = \Delta \mathbf{x}_{it}\beta + \varepsilon_{it}^{\text{FD}} \qquad (y_{it} = \Delta y_{it} \quad \text{due to absorbing state})$$

Conditional mean of disturbance:

$$\begin{aligned} \mathrm{E}(\varepsilon_{it}^{\text{FD}} | a_i, \mathbf{x}_{it}, \mathbf{x}_{it-1}, \mathbf{y}_{it^-} = \mathbf{0}) &= (a_i + \mathbf{x}_{it}\beta)(1 - \Delta \mathbf{x}_{it}\beta) \\ &\quad + (1 - a_i - \mathbf{x}_{it}\beta)(-\Delta \mathbf{x}_{it}\beta) \\ &= a_i + \mathbf{x}_{it-1}\beta \end{aligned}$$

- Taking first-differences
  - **Does not eliminate** a_i
  - Makes $\mathbf{x}_{it-1}$ enter **conditional mean** of disturbance
- **Similar** (yet more involved) result for **within-transformation** (eqiv. for $T = 2$) Within-Transformation
- First-diff. and within estimator biased and inconsistent

First-Differences Estimation with Constant

- **Including constant term** in first-differences estimation improves matters

$$\mathrm{E}(\varepsilon_{it}^{\mathrm{FDC}} | a_i, \mathbf{x}_{it}, \mathbf{x}_{it-1}, \mathbf{y}_{it^-} = \mathbf{0}) \quad = \quad \tilde{a}_i + \widetilde{\mathbf{x}}_{it-1}\tilde{\beta}$$

- Constant captures (estimation sample) mean of a_i
- $\mathrm{E}(\tilde{a}_i|\text{sample}) = 0$, $\tilde{\beta}' \equiv [\tilde{\alpha}^c \ \beta']$, $\widetilde{\mathbf{x}}_{it-1} \equiv [0 \ \mathbf{x}_{it-1}]$, and $\widetilde{\Delta\mathbf{x}}_{it} \equiv [1 \ \Delta\mathbf{x}_{it}]$

Asymptotic Properties of FD Estimation with Constant

Assumption

$\mathrm{Cov}(a_i, \Delta\mathbf{x}_{it}) = \mathbf{0}$, while $\mathrm{Cov}(a_i, \mathbf{x}_{it}) \neq \mathbf{0}$ in the population

$$\begin{aligned}\mathrm{plim}(b^{\mathrm{FDC}}) = \quad & \mathrm{plim}\left(I + \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T_i}\widetilde{\Delta\mathbf{x}}_{it}'\widetilde{\Delta\mathbf{x}}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T_i}\widetilde{\Delta\mathbf{x}}_{it}'\widetilde{\mathbf{x}}_{it-1}\right)\right)\tilde{\beta} \\ & + \mathrm{plim}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T_i}\widetilde{\Delta\mathbf{x}}_{it}'\widetilde{\Delta\mathbf{x}}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=2}^{T_i}\widetilde{\Delta\mathbf{x}}_{it}'\tilde{a}_i\right) \neq \tilde{\beta}\end{aligned}$$

Two sources of **asymptotic bias** in b^{FDC}

1. '**Ill-scaling bias**' originates from first-differences transformation itself ($\rightarrow$ even in the absence of any unobserved heterogeneity)
2. **Survivor bias** originates from $\mathrm{Cov}(a_i, \mathbf{x}_{it} | \mathbf{y}_{it^-} = \mathbf{0}) \neq \mathrm{Cov}(a_i, \mathbf{x}_{it-1} | \mathbf{y}_{it^-} = \mathbf{0})$ due to selective survival

An Adjusted First-Differences Estimator

$$b^{\text{FDC}}_{\text{adjust}} = \underbrace{\left(I + \left(\sum_{i=1}^{N} \sum_{t=2}^{T_i} \widetilde{\Delta \mathbf{x}}_{it}' \widetilde{\Delta \mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T_i} \widetilde{\Delta \mathbf{x}}_{it}' \widetilde{\mathbf{x}}_{it-1} \right) \right)^{-1}}_{\text{adjustment matrix } \mathbf{W}}$$

$$\times \underbrace{\left(\sum_{i=1}^{N} \sum_{t=2}^{T_i} \widetilde{\Delta \mathbf{x}}_{it}' \widetilde{\Delta \mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=2}^{T_i} \widetilde{\Delta \mathbf{x}}_{it}' y_{it} \right)}_{b^{\text{FDC}}}$$

- Eliminates ‘**ill-scaling bias**’

An Adjusted First-Differences Estimator II

1. Does not suffer from '**ill-scaling bias**'
   - Dominant source of bias of b^{FDC} in many stettings
2. Still subject to **survivor bias**
   - Unless $\mathbf{x}_{it}$ follows **random walk**
   - Unless $\beta = \mathbf{0}$
   - Unless $\text{Var}(a_i) = 0$
   - Yet, OLS also suffers from (different kind of) **survivor bias** even for $\text{Cov}(a_i, \mathbf{x}_{it}) = \mathbf{0}$
3. Computationally very **simple**
4. Never consistent for α
5. Only **exists** if $\mathbf{W}$ is **non-singular**
6. $\text{Var}(\mathbf{b}^{\text{FDC}}_{\text{adjust}}|\mathbf{X}) = \mathbf{W} \times \text{Var}(\mathbf{b}^{\text{FDC}}|\mathbf{X}) \times \mathbf{W}$
   - No serial correlation, just heterosecedasticity

Higher-Order Differences

- Compared to conventional fixed-effects estimators **much stronger assumptions** required
  - Properties of $b^{\text{FDC}}_{\text{adjust}}$ hinge on $\text{Cov}(a_i, \Delta \mathbf{x}_{it}) = \mathbf{0}$
  - May well be violated
  - **Higher-order** differences $\Delta^j \mathbf{x}_{it}$ as possible solution
    Higher-Order
  - Technically fully analogous to $\mathbf{b}^{\text{FDC}}_{\text{adjust}}$
  - Costly in terms of variation in $\mathbf{x}$ that is used for identification

MC Simulation Design

- Five estimators
  1. $b^{\mathbf{OLS}}$ (OLS)
  2. $b^{\mathbf{WI}}$ (within transformation)
  3. $b^{\mathbf{FD}}$ (first-differences w/o constant)
  4. $b^{\mathbf{FDC}}$ (first-differences with constant)
  5. $b^{\mathbf{FDC}}_{\text{adjust}}$ (adjusted first-differences)
- $T = \mathbf{5}$
- $N = \mathbf{4} \cdot \mathbf{10^7}$ (large samp.) or $N = \mathbf{400}$ (small samp.)
- Number of MC replications
  - » 1 (large sample)
  - » 10 000 (small sample)
- Two variants for small sample
  1. $\mathbf{x}_{it}$ and a_i **random**
  2. $\mathbf{x}_{it}$ and a_i **fixed**

MC Simulation Design II

- a_i iid. continuous $U(0.05, 0.15)$ ($\rightarrow \alpha = 0.1$)
- $\mathbf{x}_{it}$ comprises only one variable, three DGPs:
  1. **stationary**: $x_{it}^{ST} = 0.1 + a_i + \zeta_{it}$, with $\zeta_{it} \sim$ iid. $U(-0.035, 0.035)$
  2. **random walk w/o drift**: $x_{it}^{RW} = x_{it-1}^{RW} + \nu_{it}$, with $x_{i1} = 0.1 + a_i$ and $\nu_{it} \sim$ iid. $U(-0.05, 0.05)$
  3. **trended with increasing variance**: $x_{it}^{TR} = 0.075 + a_i + \eta_{it}$, with $\eta_{it} \sim$ iid. $U(0, 0.025t)$

  » $\text{Cov}(a_i, x_{it}) > 0$ and $\text{Cov}(a_i, \Delta x_{it}) = 0$
  » $a_i + x_{it}\beta \in [0, 1]\ \forall\ i, t = 1 \ldots 5$
  » $\text{P}(y_{it} = 1)$ and $\text{Var}(\Delta x_{it})$ very similar across DGPs
- $\beta = 1$

Large Sample Simulation Results

	$b^{\mathbf{OLS}}$		$b^{\mathbf{WI}}$		$b^{\mathbf{FD}}$		$b^{\mathbf{FDC}}$		$b^{\mathbf{FDC}}_{\text{adjust}}$	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
x_{it}^{ST} **stationary**										
$\hat{\beta}$	1.6671	0.0012	0.9024	0.0025	0.7072	0.0022	0.5008	0.0019	0.9980	0.0037
$\hat{\alpha}$	-0.0345	0.0002	0.1160	0.0005			0.2899	0.0001	0.0955	0.0007
x_{it}^{RW} **follows random walk**										
$\hat{\beta}$	1.4267	0.0009	0.9472	0.0019	1.0011	0.0022	1.0000	0.0018	0.9999	0.0018
$\hat{\alpha}$	0.0134	0.0002	0.1072	0.0004			0.2882	0.0001	0.0951	0.0004
x_{it}^{TR} **trended with increasing variance around trend**										
$\hat{\beta}$	1.5715	0.0012	6.0363	0.0019	4.4998	0.0020	0.6725	0.0019	1.0075	0.0028
$\hat{\alpha}$	-0.0180	0.0002	-0.9154	0.0004			0.2950	0.0001	0.0936	0.0006

Notes: True coefficient values: $\beta = \mathbf{1}$, $\alpha = \mathbf{0.1}$; $N = 4 \cdot 10^7$, $T = 5$; the # of observations for x_{it}^{ST} is 71 748 906, the corresponding #s of observations for x_{it}^{RW} is 71 823 746 and for x_{it}^{TR} being trended 72 218 321. For b^{OLS} the #s of observations are higher by $4 \cdot 10^7$ observations, since the first wave is not eliminated by the within or the first-differences transformation.

- Substantial large sample bias in $b^{\mathbf{OLS}}$, $b^{\mathbf{WI}}$, $b^{\mathbf{FD}}$, and $b^{\mathbf{FDC}}$
- No significant survivor bias in $b^{\mathbf{FDC}}_{\text{adjust}}$
  - Attributable to small value of $\text{Var}(a_i)$
  - Yet, even for much larger values of $\text{Var}(a_i)$ bias of $b^{\mathbf{FDC}}_{\text{adjust}}$ comparatively small

Small Sample Simulation Results (x_{it} and a_i random)

	$b^{\mathbf{OLS}}$		$b^{\mathbf{WI}}$		$b^{\mathbf{FD}}$		$b^{\mathbf{FDC}}$		$b^{\mathbf{FDC}}_{\text{adjust}}$	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
					x_{it} and a_i random					
x_{it}^{ST} **stationary**										
$\hat{\beta}$	1.6755	0.3808	0.9208	0.7885	0.7240	0.7038	0.5133	0.5902	1.0167	1.1728
$\hat{\alpha}$	-0.0356	0.0746	0.1128	0.1549			0.2903	0.0171	0.0923	0.2286
x_{it}^{RW} **follows random walk**										
$\hat{\beta}$	1.4278	0.3004	0.9485	0.6089	1.0068	0.69504	1.0019	0.5862	1.0027	0.5856
$\hat{\alpha}$	0.0138	0.0582	0.1068	0.1195			0.2887	0.0170	0.0954	0.1131
x_{it}^{TR} **trended with increasing variance around trend**										
$\hat{\beta}$	1.5763	0.3654	6.0427	0.6069	4.5072	0.67781	0.6691	0.6155	0.9940	0.9147
$\hat{\alpha}$	-0.0186	0.0733	-0.9167	0.1167			0.2950	0.0187	0.0965	0.1909

Notes: True coefficient values: $\boldsymbol{\beta = 1}$, $\boldsymbol{\alpha = 0.1}$; $N = 400$, $T = 5$; 10 000 replications.

- Very close to large sample simulation results

Small Sample Simulation Results (x_{it} and a_i fixed)

	$b^{\mathbf{OLS}}$		$b^{\mathbf{WI}}$		$b^{\mathbf{FD}}$		$b^{\mathbf{FDC}}$		$b^{\mathbf{FDC}}_{\text{adjust}}$	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
x_{it} and a_i fixed										
x_{it}^{ST} **stationary**										
$\hat{\beta}$	1.6443	0.3826	1.3168	0.7160	0.8548	0.6678	0.5351	0.5790	1.0326	1.1189
$\hat{\alpha}$	-0.0310	0.0743	0.0324	0.1390			0.2853	0.0168	0.0865	0.2161
x_{it}^{RW} **follows random walk**										
$\hat{\beta}$	1.4208	0.3227	1.6595	0.5408	1.5261	0.6514	0.9350	0.5921	0.9807	0.6203
$\hat{\alpha}$	0.0125	0.0627	-0.0344	0.1054			0.2852	0.0166	0.0969	0.1209
x_{it}^{TR} **trended with increasing variance around trend**										
$\hat{\beta}$	1.5638	0.3795	5.9851	0.5921	4.5432	0.6561	0.6581	0.6064	0.9792	0.9023
$\hat{\alpha}$	-0.0172	0.0751	-0.8950	0.1113			0.2903	0.0177	0.0973	0.1855

Notes: True coefficient values: $\boldsymbol{\beta = 1}$, $\boldsymbol{\alpha = 0.1}$; $N = 400$, $T = 5$; 10 000 replications.

- $b^{\mathbf{WI}}$ and $b^{\mathbf{FD}}$ **sensitive** to fixing x_{it} and a_i
- $b^{\mathbf{WI}}$ and $b^{\mathbf{FD}}$ prone to **substantial small sample bias**

The xtlhazard command

- Requires data to be xtset
- Checks whether depvar is consistent with absorbing state

Syntax of xtlhazard

```
xtlhazard depvar indepvars [if] [in] [weight] [, options]
```

Options for xtlhazard

`difference(#)` set order of differencing; difference(1) that is first-differences is the default

`noabsorbing` forces estimation if depvar is inconsitent with model

`tolerance(#)` set tolerance for luinv(); tolerance(3) is the default

`edittozero(#)` use Mata function edittozero() to set matrix entries close to zero to zero; edittozero(0) that is no editing is the default

The xtlhazard command II

Options for xtlhazard cont'd

`vce(vcetype)` vcetype may be robust, cluster clustvar, model [, force], or ols; vce(robust) is the default

`noeomitted` do not consider omitted collinear variables in e(b) and e(V)

`level(#)` set confidence level; default as set by set level

⋮

`ieffect(newvar)` generate variable newvar containing estimated individual fixed-effects

xtlhazard postestimation

- Many standard postestimation commands availavle
- `predict`, `margins`, `test`, `testnl`, `lincom`, `nlcom`, ...

Research Question of Brown and Laschever (2012)

Peer Effects in Retirement of School Teachers?

Identification

- Two unexpected **pension reforms** exerting **heterogenous incentives** for retirement
- **Incentives** for others teachers as **instrument for peer retirement** while controlling for own incentives

Data

- Short yearly **panel** (1999-2001)
- **Individual teacher** level (LA Unified School District)
- No longer observed after retirement ($\rightarrow$**absorbing state**)

Result

- Significant positive peer effects

Research Question of present Application

Does Method used for Estimation Matter?

- Focus on **reduced form model**
- Focus on specification that includes **teacher fixed effects**
- Comparing results of **Brown and Laschever (2012)** who use $b^{\mathbf{WI}}$ to results from $b^{\mathbf{FD}}$ and $b^{\mathbf{FDC}}_{\text{adjust}}$
  - $b^{\mathbf{FD}}$ and $b^{\mathbf{FDC}}$ coincide because of year dummies

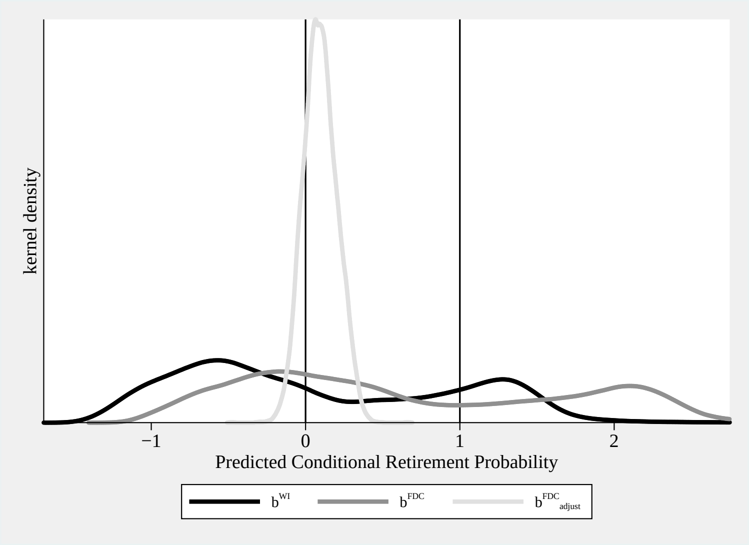
Results for Key Reduced Form Coefficients

	$b^{\mathbf{WI}}$ ‡		$b^{\mathbf{FDC}}$		$b^{\mathbf{FDC}}_{\text{adjust}}$	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
change in pension wealth of peers ($t-1$)	0.003 **	0.001	0.003 **	0.001	-0.007	0.095
change in pension wealth of peers ($t-2$)	0.002 *	0.001	0.002	0.001	-0.004	0.054
change in own pension wealth	0.033 ***	0.011	-0.003	0.009	-0.005	0.041
change in own peak value	-0.002	0.002	-0.002 *	0.001	-0.005 *	0.003

Notes: 21 290 observations, 8 320 teachers, and 586 school clusters for within-transformation estimation. 12 968 observations, 7 088 teachers, and 578 school clusters for first-differences estimation. *N* redundant observations in the within-transformed model.

- Similar results for $b^{\mathbf{WI}}$ and $b^{\mathbf{FDC}}$
- Instruments turn insignificant and negative for $b^{\mathbf{FDC}}_{\text{adjust}}$
- Results from $b^{\mathbf{FDC}}_{\text{adjust}}$ **conflict with** retirement incentives for peer teachers mattering for own retirement decision, i.e. **peer effects in retirement**

Predicted Conditional Retirement Probabilities



Predicted Conditional Retirement Probabilities II

- Unlike $b^{\mathbf{FDC}}$, predictions from $b^{\mathbf{WI}}$ and $b^{\mathbf{FDC}}_{\text{adjust}}$ centered to sample mean of y_{it}
- All estimators yield some predicted probabilities **outside unit interval**
- Share of **irregular** estimated probabilities heterogeneous
  - » $b^{\mathbf{WI}}$: 77.9%
  - » $b^{\mathbf{FDC}}$: 71.8%
  - » $b^{\mathbf{FDC}}_{\text{adjust}}$: 19.2%
- Something seems to be wrong with $b^{\mathbf{FDC}}$ and $b^{\mathbf{WI}}$

Results for Age Coefficients

	$b^{\mathbf{WI}}$ ‡		$b^{\mathbf{FDC}}$		$b^{\mathbf{FDC}}_{\text{adjust}}$	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
change in pension wealth of peers ($t-1$)	0.003 **	0.001	0.003 **	0.001	-0.007	0.095
change in pension wealth of peers ($t-2$)	0.002 *	0.001	0.002	0.001	-0.004	0.054
change in own pension wealth	0.033 ***	0.011	-0.003	0.009	-0.005	0.041
change in own peak value	-0.002	0.002	-0.002 *	0.001	-0.005 *	0.003
⋮						
age $\geq$ 54 years	-0.154 ***	0.013	-0.179 ***	0.015		
age $\geq$ 55 years	-0.123 ***	0.013	-0.163 ***	0.015	-0.016	0.029
age $\geq$ 56 years	-0.140 ***	0.012	-0.174 ***	0.014	-0.013	0.011
age $\geq$ 57 years	-0.138 ***	0.013	-0.173 ***	0.014	0.001	0.010
age $\geq$ 58 years	-0.127 ***	0.012	-0.163 ***	0.014	0.008	0.014
age $\geq$ 59 years	-0.099 ***	0.014	-0.132 ***	0.015	0.030 ***	0.010
age $\geq$ 60 years	-0.051 ***	0.015	-0.076 ***	0.017	0.056 **	0.022
age $\geq$ 61 years	-0.024	0.017	-0.038 **	0.019	0.034	0.028
age $\geq$ 62 years	0.027	0.020	0.023	0.021	0.060 ***	0.020
age $\geq$ 63 years	-0.009	0.021	0.001	0.023	-0.022	0.031
age $\geq$ 64 years	-0.055 ***	0.021	-0.054 ***	0.021	-0.052 *	0.030
age $\geq$ 65 years	0.000	0.025	-0.009	0.026	0.037	0.046
age $\geq$ 66 years	-0.025	0.026	-0.024	0.026	-0.017	0.034

Notes: 21 290 observations, 8 320 teachers, and 586 school clusters for within-transformation estimation. 12 968 observations, 7 088 teachers, and 578 school clusters for first-differences estimation. *N* redundant observations in the within-transformed model.

Results for Age Coefficients II

- $b^{\textbf{FDC}}_{\text{adjust}}$ does not yield a very distinct pattern for baseline hazard
- $b^{\textbf{FDC}}$ and $b^{\textbf{WI}}$ yield a **steady and steep decrease in the baseline retirement hazard** for teachers in their 50th
- This pattern is **in no way mirrored** by the unconditional **sample retirement rates**
- According to $\widehat{\beta^{\textbf{WI}}}$ baseline retirement hazard **decreases by 83 percentage points** between the age of 53 and the age of 60
  - Seems to make little sense
- $b^{\textbf{FDC}}$ and $b^{\textbf{WI}}$ almost certainly yield **misleading results regarding the baseline retirement hazard**

Conclusions

- **Conventional fixed-effects** estimators (within-transformation, first-differences) **inappropriate** for **discrete-time linear hazard** model
  - Bias may well exceed bias of OLS
- **Adjusted first-differences** as alternative
  - Unobserved individual **heterogeneity** is **not eliminated**
  - Corrects for incorrect ‘scaling’ of $b^{\mathbf{FDC}}$
- `xtlhazard` implements adjusted first (and higher-oder) differences estimation in stata

Error Cond. Mean in Within-Transformed Model

$$
\begin{aligned}
\mathrm{E}\left(\varepsilon_{it}^{\mathrm{WI}} \middle| a_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_i}, \mathbf{y}_{it^-} = \mathbf{0}\right) = \\
(a_i + \mathbf{x}_{it}\beta)\left(\frac{t-1}{t} - \left(\mathbf{x}_{it} - \frac{1}{t}\sum_{s=1}^{t}\mathbf{x}_{is}\right)\beta\right) \\
+ \sum_{T_i=t+1}^{T}(a_i + \mathbf{x}_{iT_i}\beta)\left[\prod_{s=t}^{T_i-1}(1 - a_i - \mathbf{x}_{is}\beta)\right]\left(-\frac{1}{T_i} - \left(\mathbf{x}_{it} - \frac{1}{T_i}\sum_{s=1}^{T_i}\mathbf{x}_{is}\right)\beta\right) \\
+ \left[\prod_{s=t}^{T}(1 - a_i - \mathbf{x}_{is}\beta)\right]\left(- \left(\mathbf{x}_{it} - \frac{1}{T_i}\sum_{s=1}^{T}\mathbf{x}_{is}\right)\beta\right)
\end{aligned}
$$

Error Cond. Mean in Within-Transformed Model II

For $t = T$, conditional mean simplifies to:

$$\mathrm{E}\left(\varepsilon_{iT}^{\mathrm{WI}} | a_i, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, \mathbf{y}_{iT^-} = \mathbf{0}\right) = \left(\frac{T-1}{T}\right) a_i + \frac{1}{T}\left(\sum_{s=1}^{T-1} \mathbf{x}_{is}\right)\beta$$

For $T = 2$, we get

$$\mathrm{E}\left(\varepsilon_{i2}^{\mathrm{WI}} | a_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}, y_{i1} = 0\right) = \frac{1}{2} a_i + \frac{1}{2}\mathbf{x}_{i1}\beta$$

which coincides with result for $\mathrm{E}\left(\varepsilon_{it}^{\mathrm{FD}} | a_i, \mathbf{x}_{it}, \mathbf{x}_{it-1}, \mathbf{y}_{it^-} = \mathbf{0}\right)$.

Estimator based on Higher-Order Differences

$$
\begin{aligned}
b^{\mathrm{JDC}}_{\mathrm{adjust}} &= \left(I + \left(\sum_{i=1}^{N} \sum_{t=j+1}^{T_i} \widetilde{\Delta^j \mathbf{x}_{it}}' \widetilde{\Delta^j \mathbf{x}_{it}} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=j+1}^{T_i} \widetilde{\Delta^j \mathbf{x}_{it}}' \widetilde{(\mathbf{x}_{it} - \Delta^j \mathbf{x}_{it})} \right) \right)^{-1} \\
&\quad \times \left(\sum_{i=1}^{N} \sum_{t=j+1}^{T_i} \widetilde{\Delta^j \mathbf{x}_{it}}' \widetilde{\Delta^j \mathbf{x}_{it}} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=j+1}^{T_i} \widetilde{\Delta^j \mathbf{x}_{it}}' y_{it} \right)
\end{aligned}
$$

for $j = 2, 3, \ldots$

$$
\begin{aligned}
\Delta^2 \mathbf{x}_{it} &= \Delta \mathbf{x}_{it} - \Delta \mathbf{x}_{it-1} \\
&= \mathbf{x}_{it} - 2\mathbf{x}_{it-1} + \mathbf{x}_{it-2} \\
\Delta^3 \mathbf{x}_{it} &= (\Delta \mathbf{x}_{it} - \Delta \mathbf{x}_{it-1}) - (\Delta \mathbf{x}_{it-1} - \Delta \mathbf{x}_{it-2}) \\
&= \mathbf{x}_{it} - 3\mathbf{x}_{it-1} + 3\mathbf{x}_{it-2} - \mathbf{x}_{it-3} \\
&\vdots
\end{aligned}
$$