Applied Fixed Effects Panel Regression using R and Stata

Volker Ludwig [1] Tobias Rüttenauer [2]

[1]TU Kaiserslautern, ludwig@sowi.uni-kl.de, @LudwigVolker1

[2]University of Oxford, tobias.ruttenauer@nuffield.ox.ac.uk, @ruettenauer

March 4, 2021

This course profited a lot from teaching materials by Josef Brüderl and Volker Ludwig
https://www.ls3.soziologie.uni-muenchen.de/studium-lehre/archiv/teaching-marterials/panel-analysis_april-2019.pdf

Part I

Conventional Panel Models

Session I

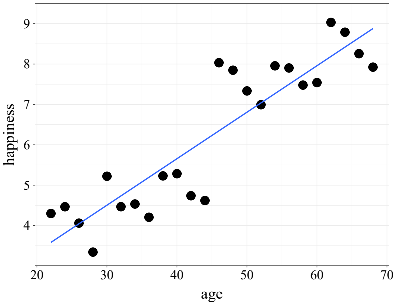
Aim

- Intuitive understanding of panel estimators
- Differences between estimators
- How to decide in practice

Outline

- FE analysis with panel data
- RE, FE, Hybrid / Mundlak framework
- Hausman specification test
- Some practical guidance

Cross-sectional data

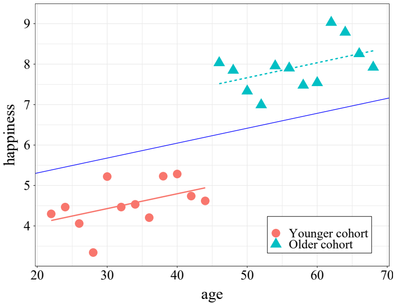


- Conventional Pooled OLS
- Positive correlation between age - happiness

$$y_{it} = \alpha + \beta_1 x_{it} + \upsilon_{it}$$

- Estimator based on complete variance over all observations
- This does not account for any type of clustering, and every observation is treated as an independent case
- Regression minimizes distance to all points

Cross-sectional data

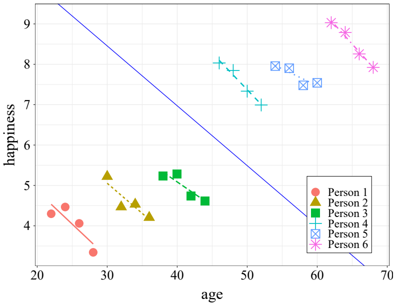


- Conventional Pooled OLS
- Controlling for cohort
- Correlation positive, but weaker

$$y_{it} = \alpha + \beta_1 x_{it} + \beta_2 z_{it} + \upsilon_{it}$$

- Estimator based on variance **within** each cohort
- Regression minimizes distance to points of the same cohort, and discards between-cohort variance
- But still cross-sectional

Advantage of panel data

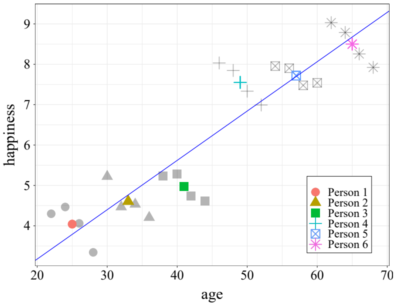


- Within estimator
- Person-fixed OLS
- Controlling for individual person
- Correlation negative

$$y_{it} = \beta_1 x_{it} + \alpha_i + \epsilon_{it}$$

- Estimator based on **within**-person variance only
- Regression minimizes distance to points of the same individual
- This is solely based on changes over time, and discards between-person variance

We can also turn this around



- Between estimator
- BE = POLS - FE
- Using only person-averages
- Correlation close to POLS

$$\bar{y}_i = \alpha + \beta_1 \bar{x}_i + \bar{\upsilon}_i$$

- Estimator based on **between**-person variance only
- Regression minimizes distance to points of individual averages
- This is solely based on differences between individuals, and discards within-person variance

Advantage of panel data

More information

- Observed trajectories over life-course
- Observed order of events
- Between and within variance

Better identification strategies

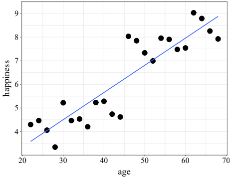
- Correlation of changes rather than states
- Counterfactual based on same individual
- Relaxes some strong assumptions
- Closer to a causal effect

Pooled OLS (POLS) estimator

$$y_{it} = \alpha + \beta x_{it} + \upsilon_{it} \quad (1)$$

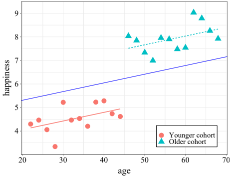
Main assumption for consistency

- $\mathrm{E}(\upsilon_{it}|x_{it}) = 0,\ Cov(x_{it}, \upsilon_{it}) = 0$
  Error (including omitted variables) must not be correlated with x_{it}



Problems

- in observational studies: rarely all confounders observed
- x_{it} is likely endogenous, thus $\hat{\beta}_x$ biased



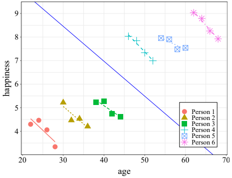
Fixed Effects (FE) estimator

$$y_{it} = \beta x_{it} + \alpha_i + \epsilon_{it} \tag{2}$$
$$FE = POLS - BE \tag{3}$$
$$(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \tag{4}$$

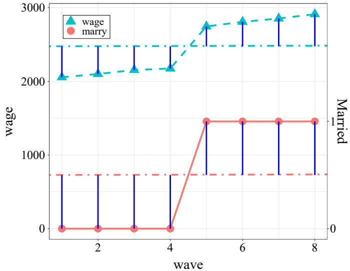
- Two error components $\upsilon_{it} = \alpha_i + \epsilon_{it}$

Main assumption for consistency

- $\mathrm{E}(\epsilon_{it}|x_i, \alpha_i) = 0$
  Idiosyncratic time-variation in ϵ_i must be uncorrelated with variation in x_i across all time periods
- but $\mathrm{E}(\alpha_i|x_i)$ can be any function of x_i
  Time-constant level-differences are allowed to correlate with x_i
- we still get an unbiased estimate of β_x

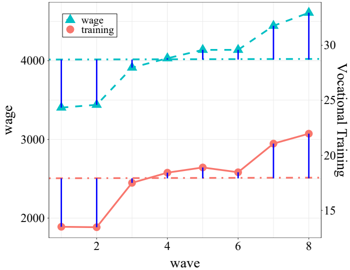


Fixed Effects (FE) estimator



- OLS on demeaned data
- Deviations from person-mean
- Do deviations within the same person correlate?

- Similar for binary and continuous data
- **All** time-constant information is discarded
- including potential confounders

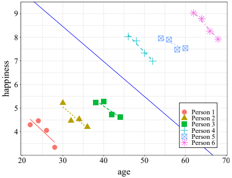


Fixed Effects (FE) estimator

$$y_{it} = \beta x_{it} + \alpha_i + \epsilon_{it} \quad (2)$$
$$FE = POLS - BE \quad (3)$$
$$(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \quad (4)$$

Potential problems

- Inefficient: cannot estimate effect of time-constant x
- One-way FE ignores units without variation in x
- Uncontrolled time-varying confounders still bias $\hat{\beta}_{FE}$
  e.g. economic recession over the 4 waves



Two-ways FE

$$y_{it} = \beta x_{it} + \alpha_i + \zeta_t + \epsilon_{it} \quad (5)$$

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}) = \beta(x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}) + (\epsilon_{it} - \bar{\epsilon}_i - \bar{\epsilon}_t + \bar{\epsilon}) \quad (6)$$

where ζ_t are time fixed effects (analogous to α_i)

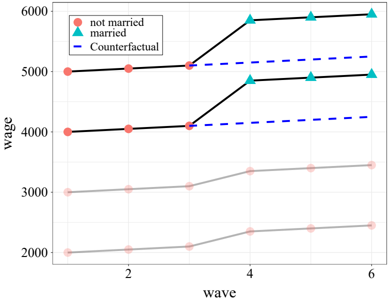
Advantage over oneway FE

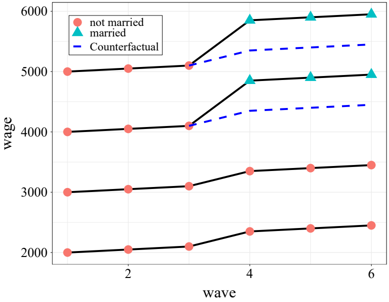
- Removes common time shocks independent of treatment
- Takes back in individuals without variation in x
- Adds a ‘control-group’ to the estimation

Main assumption

- Parallel trends between ‘treatment’ and ‘control’ units

Marriage wage premium





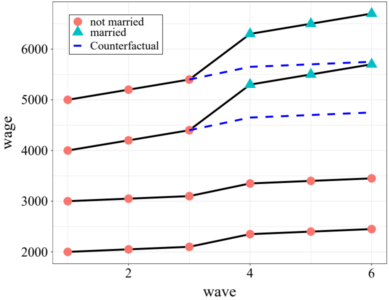
One-way FE

- Discards never-treated
- Adds time-shocks to treatment effect
- Biased marriage effect

Two-ways FE

- Uses never-married as ‘control group’
- True marriage effect

Marriage wage premium



- Same premium as before (500 EUR)
- But steeper trajectory for ever married
- **Parallel trends assumption violated**

- Both one-way and two-ways FE are biased
- One-way FE adds time shocks + trend
- Two-ways FE adds trend

⇒ Solution: Fixed Effects Individual Slopes

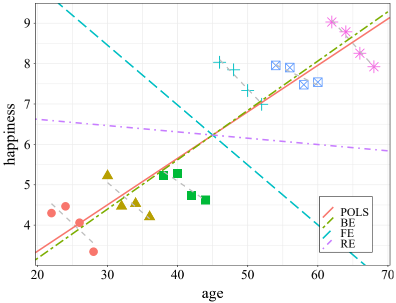
Random Effects (RE) estimator

$$(y_{it} - \lambda \bar{y}_i) = \beta(x_{it} - \lambda \bar{x}_i) + (\epsilon_{it} - \lambda \bar{\epsilon}_i) \quad (7)$$

where $\hat{\lambda} = 1 - \sqrt{\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + T\sigma_\alpha^2}}$, with σ_ϵ^2 denoting the residual variance, and σ_α^2 denoting the variance of the individual effects α_i.

- RE is estimator on the ‘quasi-demeaned’ data
- Weighted average of between and within estimator

- Weights determined by residual variance in FE as share of total residual variance
- T large, σ_α^2 large $\rightarrow$ FE
- σ_α^2 small $\rightarrow$ POLS



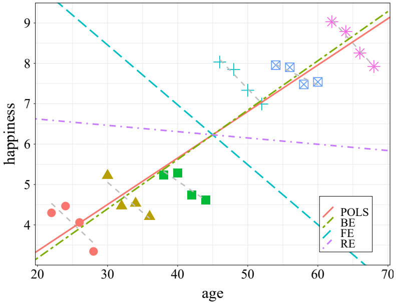
BE-POLS-RE-FE

$$\beta_{POLS} = \omega_{OLS}\beta_{FE} + (1 - \omega_{OLS})\beta_{BE}$$

- where $\omega_{OLS} = \sigma^2_{\tilde{x}}/\sigma^2_x$, with $\tilde{x} = x - \bar{x}_i$

$$\beta_{RE} = \omega_{GLS}\beta_{FE} + (1 - \omega_{GLS})\beta_{BE}$$

- where $\omega_{GLS} = \frac{\sigma^2_{\tilde{x}}}{\sigma^2_{\tilde{x}} + \phi^2(\sigma^2_x - \sigma^2_{\tilde{x}})}$, and $\phi = \sqrt{\frac{\hat{\sigma}^2_{FE}}{\hat{\sigma}^2_{BE}}}$
- here $\omega_{OLS} = 0.026$
  $\hat{\beta}_{POLS}$ close to $\hat{\beta}_{BE}$
- $\omega_{GLS} = 0.509$
  $\hat{\beta}_{RE}$ in the middle of $\hat{\beta}_{BE}$ and $\hat{\beta}_{FE}$
  $\Rightarrow$ most efficient



Random Effects (RE) estimator

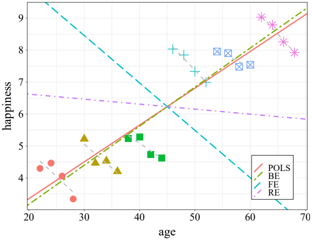
Main assumption for consistency

- $\mathrm{E}(\epsilon_{it}|x_i, \alpha_i) = 0$ (FE assumption) and
- $\mathrm{E}(\alpha_i|x_i) = 0$ (RE assumption)

  In addition to FE assumption, the individual-specific fixed effects must not be correlated with x_i

  Correlated level differences in y and x bias $\hat{\beta}_{RE}$

- RE is most efficient estimator
- important for prediction tasks
- but relies on strong assumption
- $\hat{\beta}$ likely biased in practice



Mundlak / Correlated Random Effects (CRE)

We can also estimate the within effect in RE framework

$$y_{it} = \alpha + \beta x_{it} + \gamma \bar{x}_i + \xi_{it} \tag{8}$$

- we split up the individual effect $\alpha_i = \gamma \bar{x}_i + \eta_i$
- and thus only control partially for time-constant heterogeneity by adding the person-specific means $\bar{x}_i$
- $\hat{\beta}$ thus gives us the within estimate for x
- and for consistency of $\hat{\beta}_x$ we only need $\mathrm{E}(\epsilon_{it}|x_i, \bar{x}_i) = 0$: equals the FE assumption for variables additionally included as person-specific means
- usually, include mean for all time-varying $\boldsymbol{x}$ (except t)

(Chamberlain, 1982; Mundlak, 1978)

Hausman test

$$H = (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0)^\intercal (N^{-1}\boldsymbol{V}_{\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0})^{-1} (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0), \tag{9}$$

where $\hat{\boldsymbol{\beta}}_1$ is consistent, and $\hat{\boldsymbol{\beta}}_0$ is efficient.
$N^{-1}\boldsymbol{V}_{\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0} = \mathrm{Var}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0)$
with RE being fully efficient: $\mathrm{Var}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0) = \mathrm{Var}(\hat{\boldsymbol{\beta}}_1) - \mathrm{Var}(\hat{\boldsymbol{\beta}}_0)$
(Hausman, 1978)

- $\hat{\boldsymbol{\beta}}_1$ is consistent, and $\hat{\boldsymbol{\beta}}_0$ efficient
- H_0: $\hat{\beta}_{FE} = \hat{\beta}_{RE}$
- The test shows us if the two estimates differ significantly

⇒ Use FE if Hausman test significant, and H_0 rejected

Obviously, **not helpful if both estimates are biased**

Artificial Regression Test

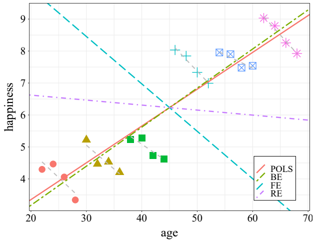
We can also use the CRE to perform a Hausman test

$$y_{it} = \alpha + \beta x_{it} + \gamma \bar{x}_i + \xi_i, \qquad (10)$$

- Estimated via RE
- RE estimator consistent if H_0: $\gamma = 0$,
- in this case, $\gamma \bar{x}_i$ can be omitted, reducing (10) to RE
- With more than one covariate, we can just perform a joint Wald χ^2 test on all or a subset of $\hat{\gamma}$

⇒ Identical to conventional HT, but allows for a variety of different (robust) standard errors

Some practical guidance



Research question

- Let theory decide
- Between or within question?
- Descriptive or causal relation?

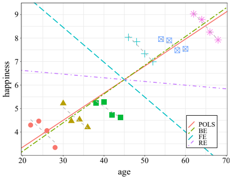
'Older people are happier'

- Descriptive statement
- **Between** comparison ⇒ BE

'Getting older makes happier'

- Causal statement
- **Within** comparison ⇒ FE
- Why would one use between variance for this statement?

Some practical guidance



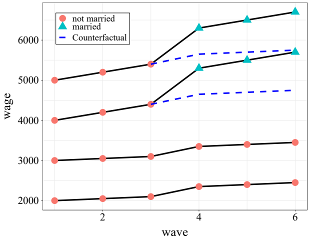
Caution with RE and POLS

- Both mix **within** and **between** variance
- Both rely on strong assumptions
- Very likely to be biased in practice
- Substantive interpretation of results?
- Causal questions likely to require within variance only

One should always

- check how close the coefficients are to BE and FE
- test for consistency (Hausman test)

Some practical guidance



FE estimator

- Only **within** variance
- Weaker assumptions
- Correlation based on changes in x and y
- Closer to a causal effect

Usually, one should

- use two-ways FE estimators
- check the amount of within variance in the data
- test the **parallel trends assumption**
- Consider time-varying confounders

Further readings

Extensive slides by Josef Brüderl and Volker Ludwig
`https://www.ls3.soziologie.uni-muenchen.de/studium-lehre/archiv/teaching-marterials/panel-analysis_april-2019.pdf`
See also Brüderl and Ludwig (2015)

Books

- Intuitive: Allison (2009)
- Comprehensive and formal: Wooldridge (2010)
- For R users: Croissant and Millo (2019)
- General introduction (e.g. for teaching):
  Angrist and Pischke (2015); Firebaugh (2008)

Part II

Fixed Effects Individual Slopes

Session II

Aim

- Extend standard FE methods to cover situations with heterogeneous slopes (Wooldridge, 2010)
- Detect bias due to heterogeneous slopes and eliminate the bias

Outline

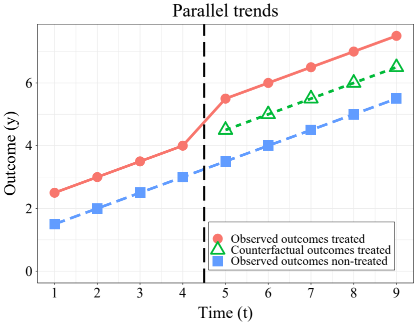
- FE bias due to heterogeneous slopes
- Estimation of FEIS estimator
- Specification test for FEIS vs. FE
- Implementation
  - Stata: `xtfeis` (Ludwig 2015)
  - R: `feisr` (Rüttenauer and Ludwig, 2020)
- Monte Carlo results

The problem with heterogeneous slopes

- Leading case: effect of some event (binary treatment) x_{it} on continuous outcome y_{it}, controlling for time z_{it}

$$y_{it} = \beta x_{it} + \alpha_{1i} + \alpha_{2i} z_{it} + \epsilon_{it}. \quad (11)$$

DGP: $y_{it} = 1 + \beta \cdot x_{it} + 0.5 \cdot t + 1 \cdot treat_i$



FE returns $\hat{\beta} = 1$

```
xtreg y x t, fe
```

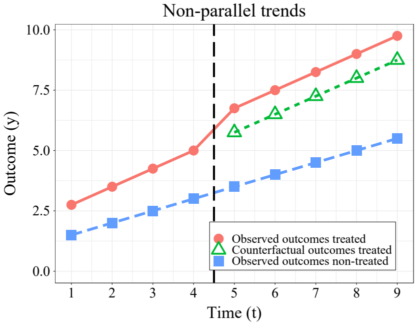
```
plm(y ~ x + t, model
= "within", effect =
"individual")
```

The problem with heterogeneous slopes

- Leading case: effect of some event (binary treatment) x_{it} on continuous outcome y_{it}, controlling for time z_{it}

$$y_{it} = \beta x_{it} + \alpha_{1i} + \alpha_{2i} z_{it} + \epsilon_{it}. \tag{12}$$

DGP: $y_{it} = 1 + \beta \cdot x_{it} + 0.25 \cdot t + 1 \cdot treat_i + 0.25 \cdot treat_i \cdot t$



FE returns $\hat{\beta} = 2.01$

```
xtreg y x t, fe
```

```
plm(y ~ x + t, model = "within", effect = "individual")
```

Estimation of standard FE

- 3 ways to control for α_{1i}
  - Least Squares Dummy Variable (LSDV): include N person dummies
  - estimate by Pooled OLS

$$y_{it} = \beta x_{it} + \sum_{i=1}^{N} \alpha_{1i} d_i + \alpha_2 z_{it} + \xi_{it} \tag{13}$$

Estimation of standard FE

- 3 ways to control for α_{1i}
  - Time-demeaning (FE): subtract person-specific average for each variable
  - estimate by Pooled OLS

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \alpha_2 \ddot{z}_{it} + \ddot{\xi}_{it}, \tag{14}$$

where, for some variable w, $\ddot{w}_{it} = w_{it} - \bar{w}_i$.

Estimation of standard FE

- 3 ways to control for α_{1i}
  - Correlated Random Effects (CRE): include person-specific average for each indep var in the equation
  - estimate by Generalized Least Squares (GLS)

$$y_{it} = \beta x_{it} + \gamma \bar{x}_i + \alpha_2 z_{it} + \delta \bar{z}_i + \xi_{it} \tag{15}$$

- CRE suggests a simple test for RE heterogeneity bias
  - Artificial Regression Test for FE vs. RE

Bias of standard FE

- Condition for consistency of FE is strict exogeneity of the idiosyncratic error term
- Violated if we estimate

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \alpha_2 \ddot{z}_{it} + \ddot{\xi}_{it}. \tag{16}$$

With $\alpha_{2i} = \alpha_2 + \ddot{\alpha}_{2i}$, we get

$$\ddot{y}_{it} = \beta \ddot{x}_{it} + \alpha_2 \ddot{z}_{it} + \ddot{\alpha}_{2i} \ddot{z}_{it} + \ddot{\epsilon}_{it}. \tag{17}$$

- Strict exogeneity fails: $E(\xi_{it} | x_{it}, z_{it}) \neq 0$ because $Cov(\ddot{\alpha}_{2i}, x_{it}) \neq 0$

Bias of standard FE

- Suppose x_{it} depends on slope variable z_{it}
  With $\delta_i = \delta + \ddot{\delta}_i$ (unobserved effects, like α_{2i}) get

$$\ddot{x}_{it} = \delta \ddot{z}_{it} + \ddot{\delta}_i + \nu_{it}, \tag{18}$$

  where ν_{it} is an independent random variable.
  - Bias of the FE estimator is (Rüttenauer and Ludwig, 2020)

$$\mathrm{E}(\hat{\beta}_{FE}) = \beta + \frac{\mathrm{Var}(\ddot{\boldsymbol{z}})\mathrm{Cov}(\ddot{\boldsymbol{\delta}}, \ddot{\boldsymbol{\alpha}}_2)}{\mathrm{Var}(\ddot{\boldsymbol{z}})\mathrm{Var}(\ddot{\boldsymbol{\delta}}) + \mathrm{Var}(\ddot{\boldsymbol{\nu}})} \tag{19}$$

Estimation of FEIS

- 3 ways to control for α_{1i} and α_{2i}
  - Extend LSDV: include N interactions person dummy X slope variable

$$y_{it} = \beta x_{it} + \sum_{i=1}^{N} \alpha_{1i} d_i + \sum_{i=1}^{N} \alpha_{2i} d_i z_{it} + \epsilon_{it} \tag{20}$$

Estimation of FEIS

- 3 ways to control for α_{1i} and α_{2i}
  - General Within-transform (FE-IS): subtract person-specific time-varying estimate for each variable

$$\tilde{y}_{it} = \beta\tilde{x}_{it} + \alpha_{2i}\tilde{z}_{it} + \tilde{\epsilon}_{it}, \tag{21}$$

where, for some variable w, $\tilde{w}_{it} = w_{it} - \hat{w}_{it}$,

and $\hat{w}_{it}$ is the predicted value from person-specific regression of w_{it} on $(1, z_{it})$.

Estimation of FEIS

- 3 ways to control for α_{1i} and α_{2i}
  - Extend CRE: include time-varying predicted values in RE

$$y_{it} = \beta x_{it} + \gamma_1 \bar{x}_i + \gamma_2 \hat{x}_{it} + \alpha_2 z_{it} + \delta \bar{z}_i + \epsilon_{it} \qquad (22)$$

- Extended CRE suggests a simple test for FE heterogeneity bias
  - Artificial Regression Test for FEIS vs. FE

Specification test

- With the CRE estimation approach, we can devise a version of the Hausman test: Artificial Regression Test (ART)
- CRE to estimate FE within effects

$$y_{it} = \beta x_{it} + \gamma \bar{x}_i + \alpha_2 z_{it} + \delta \bar{z}_i + \xi_{it} \tag{23}$$

The RE is a restricted model of the CRE:

With restriction $\gamma = 0$ we get

$$y_{it} = \beta x_{it} + \alpha_2 z_{it} + \delta \bar{z}_i + \xi_{it} \tag{24}$$

- After CRE estimation, we test $H_0 : \hat{\gamma} = 0$

  Using a Wald test, $H \sim \chi^2(K)$.

  If $p < 0.05$, H_0 is rejected, i.e. we use FE.

Specification test

- The CRE approach can also be used to test FEIS vs. FE

$$y_{it} = \beta x_{it} + \gamma_1 \bar{x}_i + \gamma_2 \hat{x}_{it} + \alpha_2 z_{it} + \delta \bar{z}_i + \epsilon_{it} \qquad (25)$$

  The FE is a restricted model of the CRE:

  With restriction $\gamma_2 = 0$ we get the FE estimator
- After CRE estimation, we test $H_0 : \hat{\gamma}_2 = 0$

  Using a Wald test, $H \sim \chi^2(K)$.

  If $p < 0.05$, H_0 is rejected, i.e. we use FEIS.
- ART works even though FE is not efficient! (Arellano 1993)
  - Important side-effect: can use panel-robust standard errors

Estimation and tests using Stata or R

	Stata
Installation	`ssc install xtfeis`
Estimation	`xtfeis y x , slope(t) cluster(id)`
ART	`xtart [FEIS] [, fe re]`
BSHT	`xtbsht FEIS FE, seed(123) reps(100)`

	R
Installation	`install.packages("feisr")`
Estimation	`feis(y ~ x | t, data=df, id="id", robust=TRUE)`
ART	`feistest(FEIS, robust=TRUE, type="all")`
BSHT	`bsfeistest(FEIS, seed=123, rep=100, type="all")`

- Note: alternatives for estimation
  - in Stata: `reghdfe` by Sergio Correia
  - in R: `lfe` by Simen Gaure or `fixest` by Laurent Berge

Monte Carlo Simulations

'Elwetritsch' - our High Performance Cluster at TUK

Basic setup

- Generate panel data with $N = 300$ and $T = 10$
- DGP

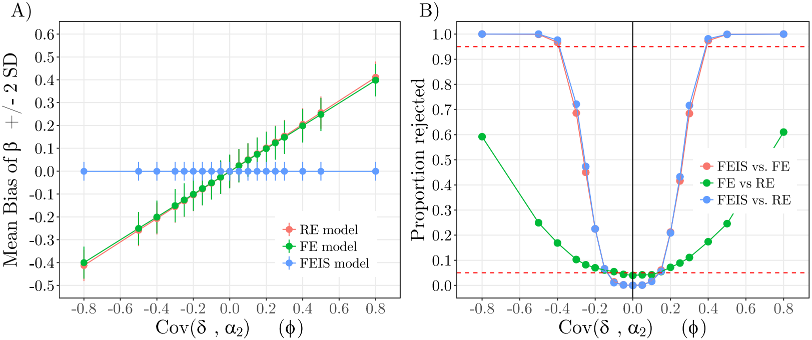
$$y_{it} = \beta x_{it} + \alpha_{1i} + \alpha_{2i} z_{it} + \epsilon_{it}, \tag{26}$$

$$x_{it} = \theta \alpha_{1i} + \delta_i z_{it} + \nu_{it}, \tag{27}$$

  - where ϵ_{it}, ν_{it} are Gaussian,
  - α_{1i} is a normally dist random variable,
  - $\theta \in \{0, 1\}$ specifies bias due to α_{1i}
  - α_{2i} and δ_i drawn from a bivariate normal dist with $\phi = \mathrm{Cov}(\boldsymbol{\delta}, \boldsymbol{\alpha}_2)$
  - ϕ specifies bias due to α_{2i}
  - Parameters for $\mathrm{Var}(\boldsymbol{\delta})$, $\mathrm{Var}(\boldsymbol{z})$ and $\mathrm{Var}(\boldsymbol{\nu})$ are set to fixed values
  - True $\beta = 1$
- Estimate RE, FE, FEIS and ARTs in 1,000 replications,
- Compute mean bias of $\hat{\beta}$ and rejection rate (at 5 % level)

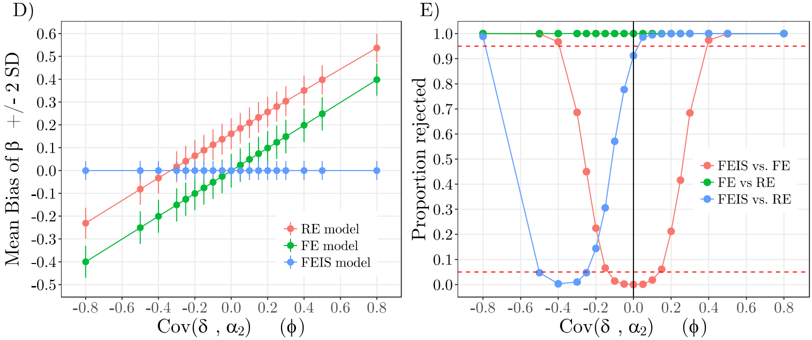
Simulation results: Bias in RE and FE

- Bias due to α_{2i}, no bias due to α_{1i}



Simulation results: Bias in RE and FE

- Bias due to α_{2i} and α_{1i}



Summary

- FE biased if heterogenous slopes of some variable related to the causal variable
- Can use `xtfeis` Stata or `feisr` in R to estimate unbiased FEIS and test for bias due to α_{2i}
- Standard Hausman test for FE versus RE has no power to detect bias due to α_{2i}
  - Might choose wrong estimator
  - If bias due to α_{1i} and α_{2i} have opposite sign and cancel each other out, FE and RE give similar estimates
- Simulations show the ART for FEIS versus FE (or RE) has good size and power to detect the bias
  - Can be applied with clustered s.e.
  - alternative: bootstrapped Hausman test (BSHT)

Limitations

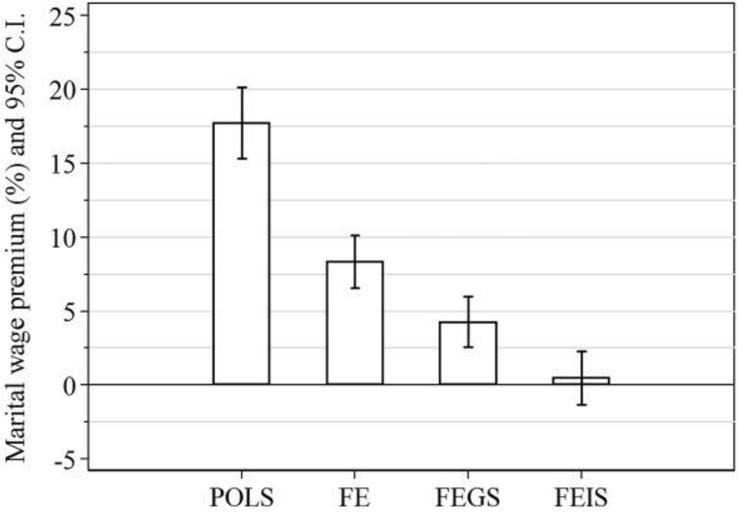
- FEIS still is not the magic bullet
- Like FE, extended FEIS biased in situations with
  - measurement error on the treatment variable (or other covariates) (Griliches and Hausman, 1986)
  - true DGP including a Lagged Dependent Variable (Nickell, 1981; Phillips and Sul, 2007)
  - with variation of treatment timing and variation of the treatment effect over time left unspecified (Meer and West, 2016; Goodman-Bacon, 2018)

Extensions

- FEIS is more general (and more efficient) than the Random Trend estimator (Second Differencing)
- Can be extended to all sorts of multi-level data structures (Rüttenauer and Ludwig, 2020)
  - children in families, students in schools, workers in firms, persons in countries
  - data with more than two levels possible
- Unit-specific slopes possible also for poisson (FEIS poisson) (Correia et al., 2020)

The male marital wage premium

Study by Ludwig and Brüderl (2018)



Effect of preschool on cognitive ability

Rüttenauer and Ludwig (2020), replication of Deming (2009)

	Replication (1)	FE (2)	FEIS (3)
Head Start			
Ages 5–6	0.143	0.133	0.350**
	(0.085)	(0.087)	(0.115)
Ages 7–10	0.132*	0.117	0.319***
	(0.059)	(0.060)	(0.096)
Ages 11–14	0.054	0.029	0.241*
	(0.061)	(0.061)	(0.102)
Other Preschool			
Ages 5–6	−0.081	−0.105	−0.095
	(0.084)	(0.083)	(0.132)
Ages 7–10	0.046	0.029	0.009
	(0.064)	(0.061)	(0.120)
Ages 11–14	−0.023	−0.040	−0.060
	(0.069)	(0.066)	(0.120)
Pretreatment index		0.056	
		(0.034)	
R^2	.050	.020	.031
Adjusted R^2	−.099	−.115	.027
Number of observation	4,687	4,646	4,646
Number of groups:			541

References I

Allison, P. D. (2009). *Fixed Effects Regression Models*, volume 160 of *Quantitative Applications in the Social Sciences*. Sage, Los Angeles.

Angrist, J. D. and Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton Univ. Press, Princeton.

Brüderl, J. and Ludwig, V. (2015). Fixed-Effects Panel Regression. In Best, H. and Wolf, C., editors, *The Sage Handbook of Regression Analysis and Causal Inference*, pages 327–357. Sage, Los Angeles.

Chamberlain, G. (1982). Multivariate Regression Models for Panel Data. *Journal of Econometrics*, 18(1):5–46.

Correia, S., Guimarães, P., and Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, 20(1):95–115.

Croissant, Y. and Millo, G. (2019). *Panel Data Econometrics with R*. John Wiley and Sons, Hoboken, NJ.

Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Firebaugh, G. (2008). *Seven Rules for Social Research*. Princeton University Press, Princeton, N.J. and Woodstock.

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. *NBER Working Paper*, 25018.

Griliches, Z. and Hausman, J. A. (1986). Errors in Variables in Panel Data. *Journal of Econometrics*, 31(1):93–118.

Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6):1251–1271.

Ludwig, V. and Brüderl, J. (2018). Is There a Male Marital Wage Premium? New Evidence from the United States. *American Sociological Review*, 83(4):744–770.

Meer, J. and West, J. (2016). Effects of the Minimum Wage on Employment Dynamics. *Journal of Human Resources*, 51(2):500–522.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, 46(1):69.

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects. *Econometrica*, 49(6):1417.

Phillips, P. C. and Sul, D. (2007). Bias in Dynamic Panel Estimation with Fixed Effects, Incidental Trends and Cross Section Dependence. *Journal of Econometrics*, 137(1):162–188.

Rüttenauer, T. and Ludwig, V. (2020). Fixed Effects Individual Slopes: Accounting and Testing for Heterogeneous Effects in Panel Data or Other Multilevel Models. *Sociological Methods & Research*, OnlineFirst.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Mass.