

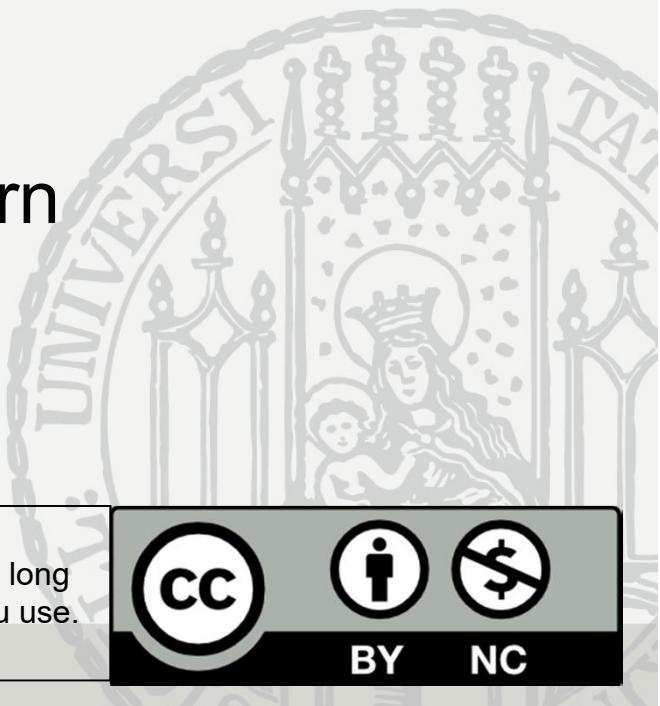
Applied Panel Data Analysis

Using Stata

Josef Brüderl & Volker Ludwig
LMU München & TU Kaiserslautern

April 2019

This presentation is licensed under a CC-BY-NC 4.0 license.
You may copy, distribute, and use the slides in your own work, as long
as you give attribution to the original author on each slide that you use.
Commercial use of the contents of these slides is not allowed.



Contents I

I) Panel Data	06
II) The Basic Idea of Panel Data Analysis	17
III) An Intuitive Introduction to Linear Panel Regression	
– Between Estimation	23
– Within Estimation	32
– Within Estimation With Age/Period Effects	49
IV) The Basics of Linear Panel Regression	
– Linear Panel Models	57
– Fixed- or Random-Effects?	69
V) A Real Data Example: Marriage and Happiness	
– Preparing Panel Data	76
– Describing Panel Data	85
– The Results	88
– Interpreting Results from Panel Regressions	102

Contents II

VI) Modeling Individual Growth

- Growth Curve Models 131
- The Age-Period-Cohort Problem 144
- Group Specific Growth Curves 154

VII) Further Linear Panel Models

- Alternative Within Estimators 162
- The Fixed-Effects Individual-Slopes Model 170
- Mixed-Coefficients Panel Models 189
- The Hybrid Model 204
- Panel Regression with Missing Data 214

VIII) Non-Linear Panel Models: Fixed-Effects Logit

222

IX) Event History Analysis with Repeated Events

229

- Example: Duration of Unemployment

Contents III

X) Limitations of the Within Methodology

– Limitations of Scope	241
– Violations of Strict Exogeneity	247
- Unobserved Time-Varying Confounders	
- Measurement Errors	
- Reverse Causality	
- Panel Attrition	
- Panel Conditioning	
– Three Problems of Causal Analysis	266
- State Dependence: Lagged Dependent Variable Models (LDV)	
- Lagged Effects: Lagged Effects Models (LEM)	
- Direction of Causality: Cross-Lagged Panel Models (CLPM)	

XI) Final Thoughts

References

What This Lecture Aims For

- Introducing basic methods of panel data analysis (PDA)
 - Emphasis on fixed-effects and growth curve methods
 - Complex methods are de-emphasized
- Practical implementation of PDA with Stata
 - Important Stata commands are on the slides
 - The lecture is accompanied by Stata do-files (and data),
whereby all computations can be reproduced
- Presenting and interpreting results
 - The graphical display of regression results is emphasized
 - The era of the regression table is over!
- Further materials on PDA can be found here
 - <https://www.ls3.soziologie.uni-muenchen.de/studium-lehre/lehrmaterialien/index.html>



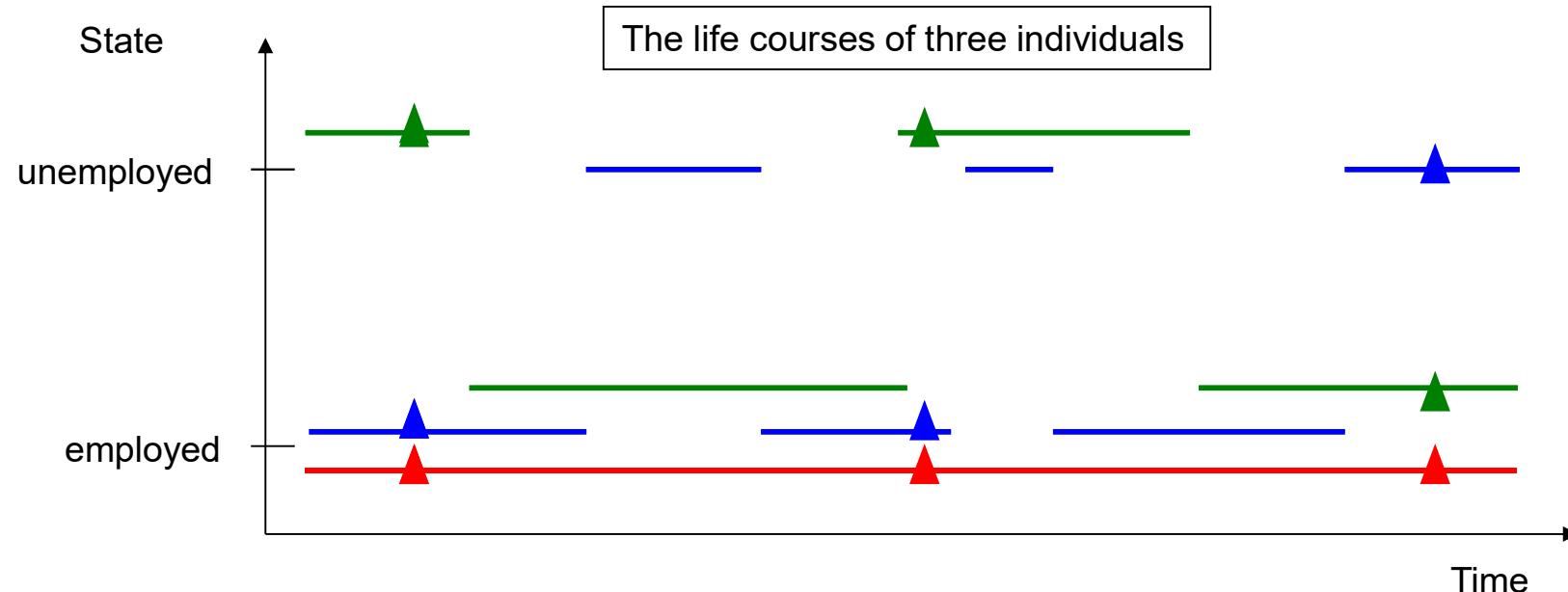
Chapter I: Panel Data

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Hierarchy of Data Structures

- Cross-sectional data
 - “Snapshot” at one time point
- Panel data
 - Repeated measurement
- Event history data
 - Information on the complete life course



Panel Data

Notation: y_{it}

$i = 1, \dots, N$: units

$t = 1, \dots, T$: time

Example: panel data with $T = 2$

id	time	Y	X
1	1	y_{11}	x_{11}
1	2	y_{12}	x_{12}
2	1	y_{21}	x_{21}
2	2	y_{22}	x_{22}
:	:	:	:
N	1	y_{N1}	x_{N1}
N	2	y_{N2}	x_{N2}

These are:

- balanced panel data
- in long format (pooled)

If units are persons and time is years:
→ A row in the panel data is called “person-year” (pyr)

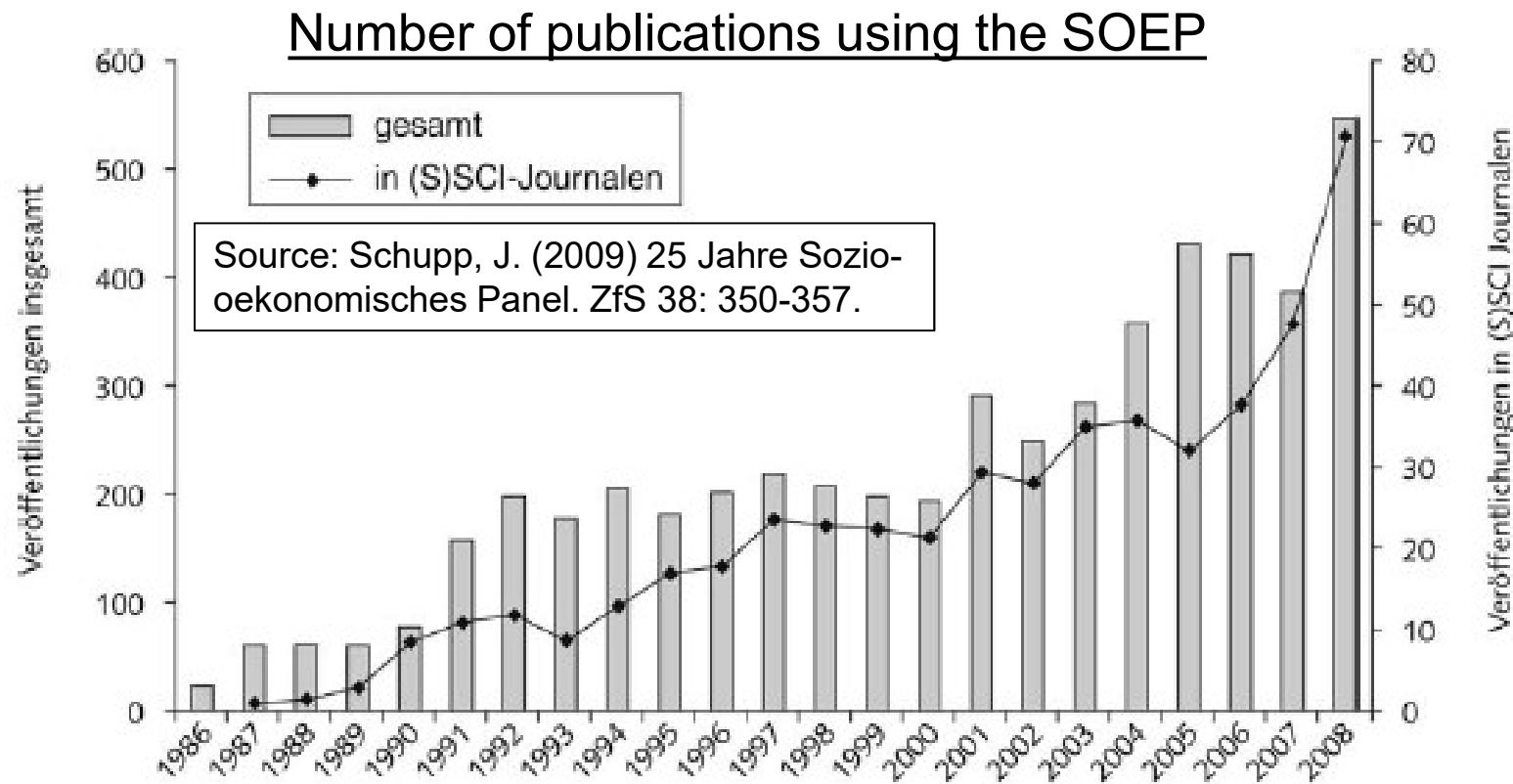
- Repeated measurements of the same variables on the same units
- Macroeconomics, Political Science
 - Unit of analysis: countries
 - N small, T large
 - Cross-sectional time series (**xt**)
- Microeconomics, Sociology
 - Unit of analysis: persons
 - N large, T small
 - (Micro) panel data

→ This lecture emphasizes micro panel data analysis

The Two Major Advantages of Panel Data

- Panel data allow to identify causal effects under weaker assumptions (compared to cross-sectional data)
 - With panel data we know the time-ordering of events
 - Thus we can investigate how an event changes the outcome
- Panel data allow to study individual trajectories
 - Individual growth curves (e.g. wage, materialism, intelligence)
 - One can distinguish cohort and age effects
 - Transitions into and out of states (e.g. poverty)

Usage of Panel Data is on the Rise



According to Young/Johnson (2015) **61%** of all empirical articles published in JMF 2010-2014 used panel data. Methods used:

19% event history methods	16% linear regression
19% fixed effects models	15% logistic regression
22% multilevel models (incl. growth curves)	10% structural equation models

A Few Remarks on Collecting Panel Data

- Cross-sectional survey: retrospective questions
 - Problems with recall
 - Often done for collecting event history data
- Prospective panel survey
 - Panel data
 - Ask for the current status/value
 - Event history data
 - Ask what happened since last interview: between wave retrospective questions (electronic life-history calendar)
 - Ideally using dependent interviewing (preloads) to avoid the seam effect
- The advantages of panel data are threatened by two methodological problems (s. Chapter X)
 - Panel conditioning (panel effect)
 - Panel mortality (attrition)
- More on panel methodology can be found in Lynn (2009)

Important Panel Surveys

- Household panels
 - Panel Study of Income Dynamics (PSID) [since 1968]
 - The role model for all household panels
 - German Socio-Economic Panel (SOEP) [since 1984]
 - Understanding Society (UKHLS) [since 1991]
- Cohort panels
 - British Cohort Studies: children born 1958, 1970, 2000
 - National Longitudinal Survey of Youth (NLSY79): U.S. cohort born around 1960
- Panels on special populations in Germany recently started
 - German Family Panel (pairfam), National Educational Panel Study (NEPS), Survey of Health, Ageing and Retirement in Europe (SHARE), Panel "Arbeitsmarkt und soziale Sicherung" (PASS), TwinLife, Children of Immigrants (CILS4EU), Nationale Kohorte
- Online panel surveys
 - LISS panel: A Dutch online panel survey
 - German internet panel (GIP)
 - GESIS Panel
- Links on German studies you can find here:
<http://www.ratswd.de/forschungsdaten/fdz>

SOEP

Household Panel Study

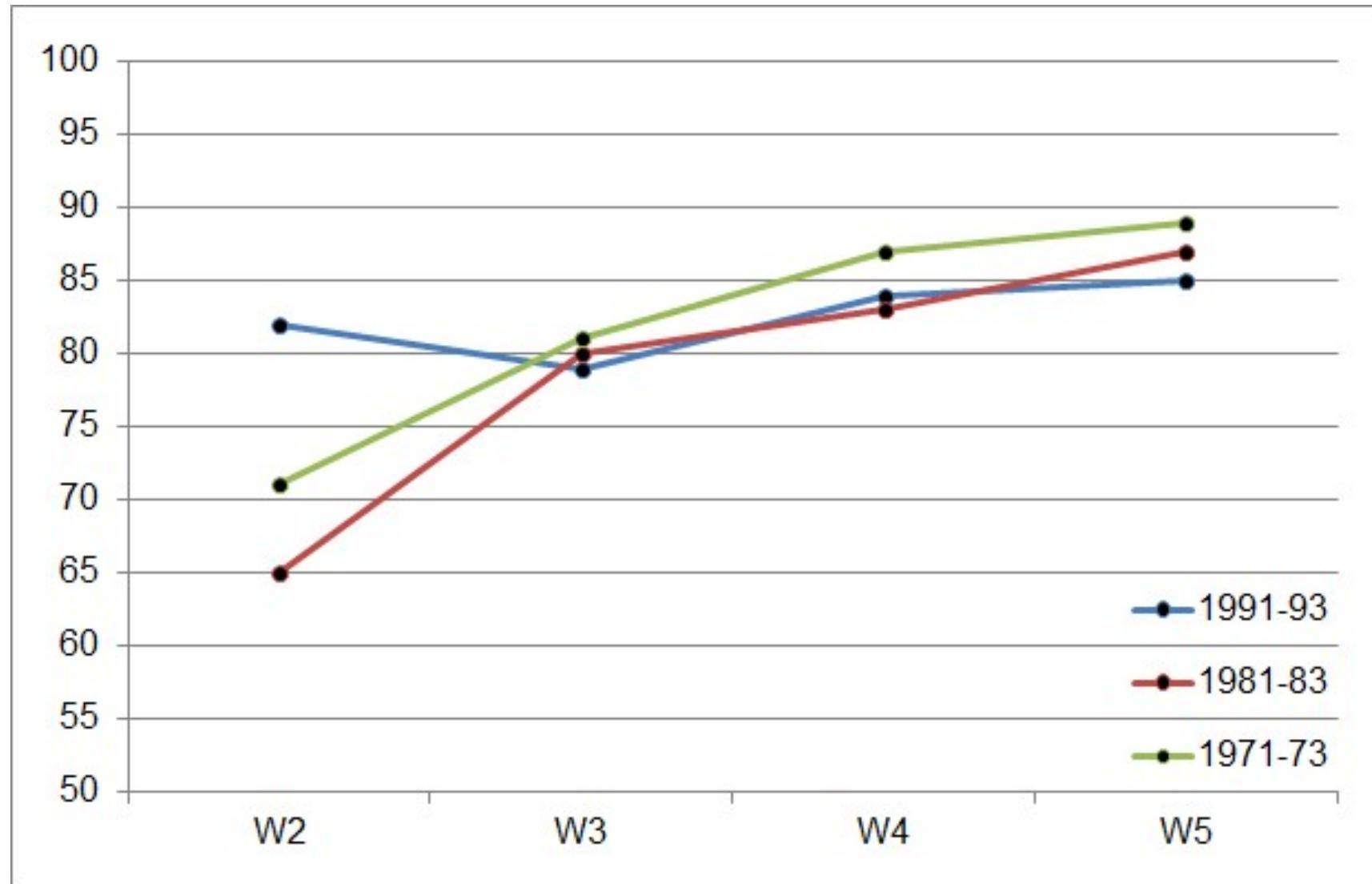
- Sample of households in Germany
- Every person aged 17 or older is interviewed
- For persons under 17 proxies are interviewed
- When a person moves out of the household, he or she is followed
- Persons, households and original households can be identified beyond waves
- First wave 1984 (subsamples A and B)
- Annual interviews (PAPI questionnaire)
- Several refreshment subsamples
 - Meanwhile about 60,000 persons participated in the SOEP

→ More information: <http://www.diw.de/soep>

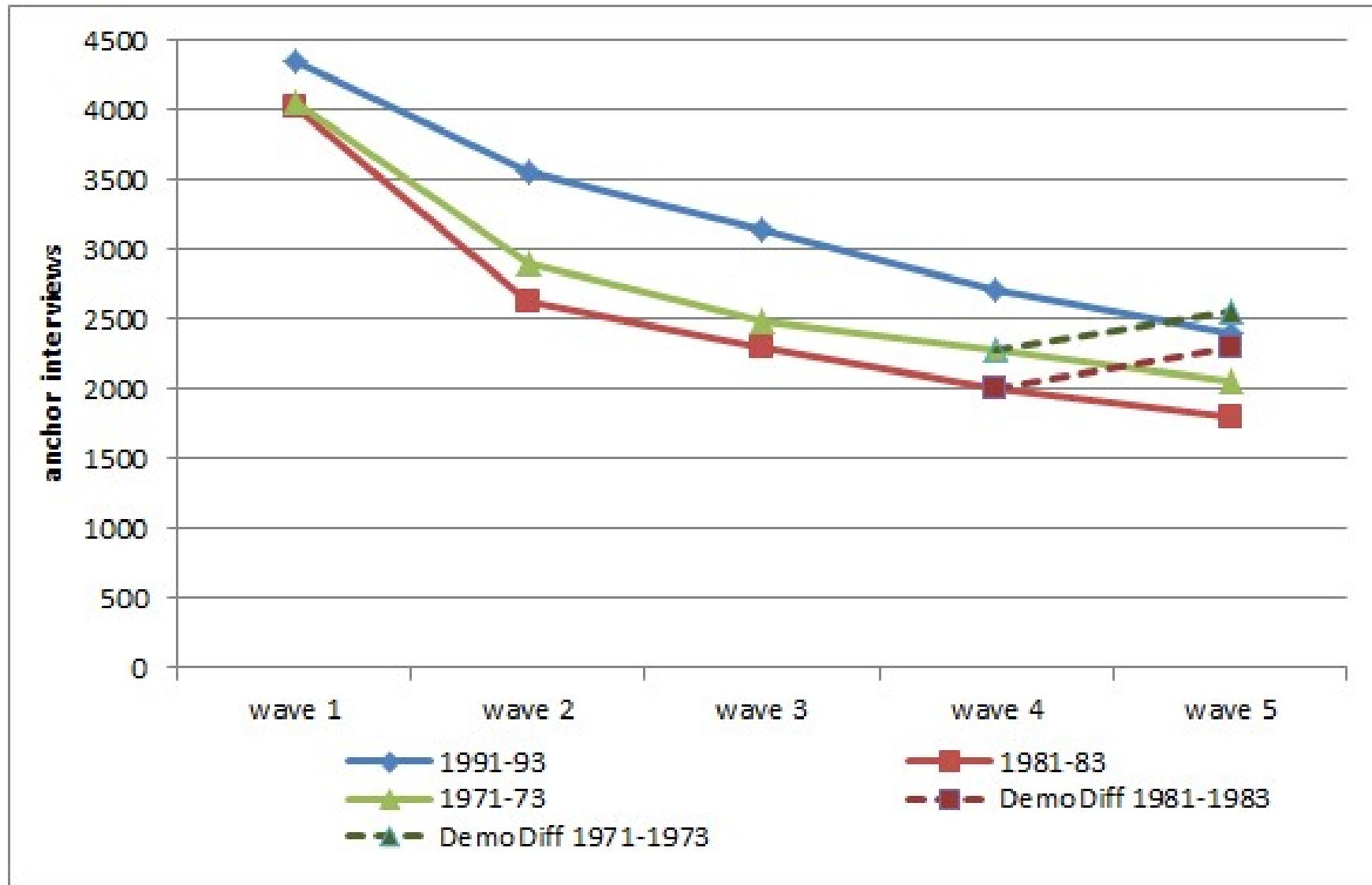
The German Family Panel (pairfam)

- Target population
 - All German residents, who are able to do an interview in German
 - Cohort-sequence design: 1971-73, 1981-83, 1991-93
- Sample
 - Random sample from population registers
 - 343 “Gemeinden” were sampled
 - 42.000 addresses were drawn randomly from the population registers
 - N = 12.402
 - Response rate in wave 1: 37 %
- Interview mode
 - 60 minute CAPI (some parts CASI)
- Multi-actor design
 - Anchor person (AP) and partner, parents, children
- First wave in 2008
 - Waves annually
 - Non-monotonic design: respondents can drop out for one wave
- Data: www.pairfam.de

Response Rate Anchor – Panel Stability



Number of Anchors



Chapter II:

The Basic Idea of Panel Data Analysis

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Panel Research Design

- Social researchers are socialized to the “cross-sectional research design”
 - For identifying the causal effect of a treatment, we compare the outcome of people in the treatment group with the outcome of **different people** in the control group
 - We call this design “**between estimation**”
- Panel data allow for a new research design: the “panel research design”
 - For identifying the causal effect of a treatment, we investigate how the outcome changes, if the **same people** change from the control to the treatment condition over time
 - We call this design “**within estimation**”
- For causal inference, within estimation offers some advantages over between estimation

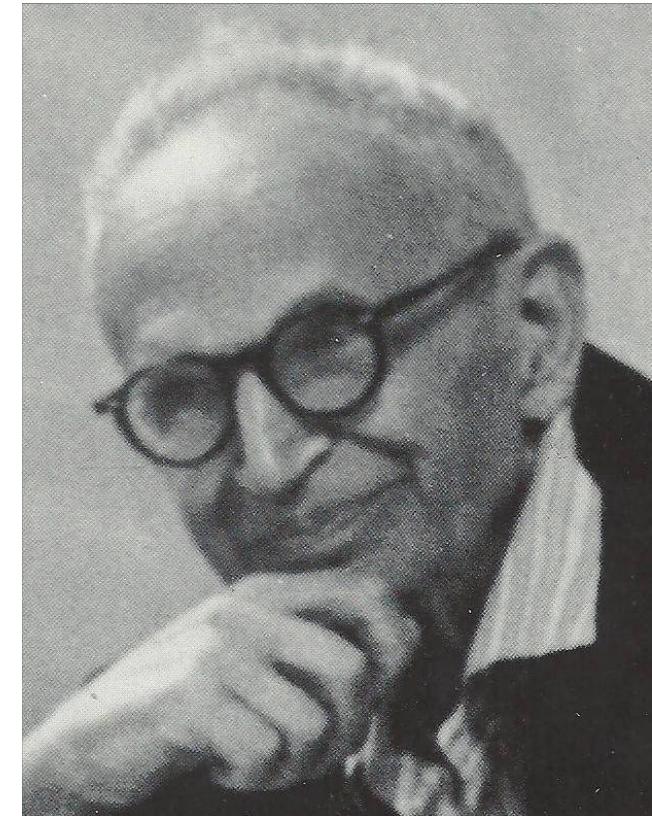
Paul Lazarsfeld on the Panel Design

Princeton „radio project“ (1937-1939)

- Research question
Effect of radio ownership on political attitudes:
Will the Americans become communist?
- Inference from cross-sectional (control group) or panel data?

“Most of the control groups available for social research are ‘self-selected’.”

“If we give radios to a number of farmers and then notice considerable differences without any great external changes occurring at the same time, it is safer to assume that these differences are caused by radio than it would be, if we were to compare radio owners with non-owners.”



Lazarsfeld/Fiske (1938) The “panel” as a new tool for measuring opinion.
Public Opinion Quarterly 2: 596-612.

Between and Within Estimation

- According to the counterfactual approach to causality (Rubin's model) an individual causal effect is defined as

$$\Delta_i = Y_{i,t_0}^T - Y_{i,t_0}^C, \quad T: \text{treatment}, C: \text{control}$$

– However, this is not estimable (fundamental problem of causal inference)

- Estimation with cross-sectional data

$$\widehat{\Delta}_i = Y_{i,t_0}^T - Y_{j,t_0}^C$$

– We compare different persons i and j **(between estimation)**
– Assumption: unit homogeneity (no unobserved heterogeneity)

- Estimation with panel data I

$$\widehat{\Delta}_i = Y_{i,t_1}^T - Y_{i,t_0}^C$$

– We compare the same person over time t_0 and t_1 **(within estimation)**
– Assumption: temporal homogeneity (no period effects, no maturation)

- Estimation with panel data II

$$\widehat{\Delta}_i = (Y_{i,t_1}^T - Y_{i,t_0}^C) - (Y_{j,t_1}^C - Y_{j,t_0}^C)$$

– Now with a control group **(within estimation with control group)**
– Assumption: parallel trends

Between and Within Estimation

- Between estimation works well with experimental data
 - Due to randomization units will differ only in the treatment
- However, with non-experimental data between estimation generally will not work, because the strong assumption of unit homogeneity often will not hold
 - Due to self-selection into treatment
 - Unobserved unit heterogeneity will bias between estimation results
- Within estimation with control group, however, will often work, because the parallel trends assumption is much weaker
 - Unobserved unit heterogeneity will **not** bias within estimation results
 - Only differing time-trends in treatment and control group will bias within estimation results

Difference-in-Differences

- **Fundamental insight:**
Within estimation with control group (difference-in-differences, DiD) is the basic building block of any within estimator (Goodman-Bacon 2018)

Chapter III:

An Intuitive Introduction to Linear Panel

Regression

Section: Between Estimation

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

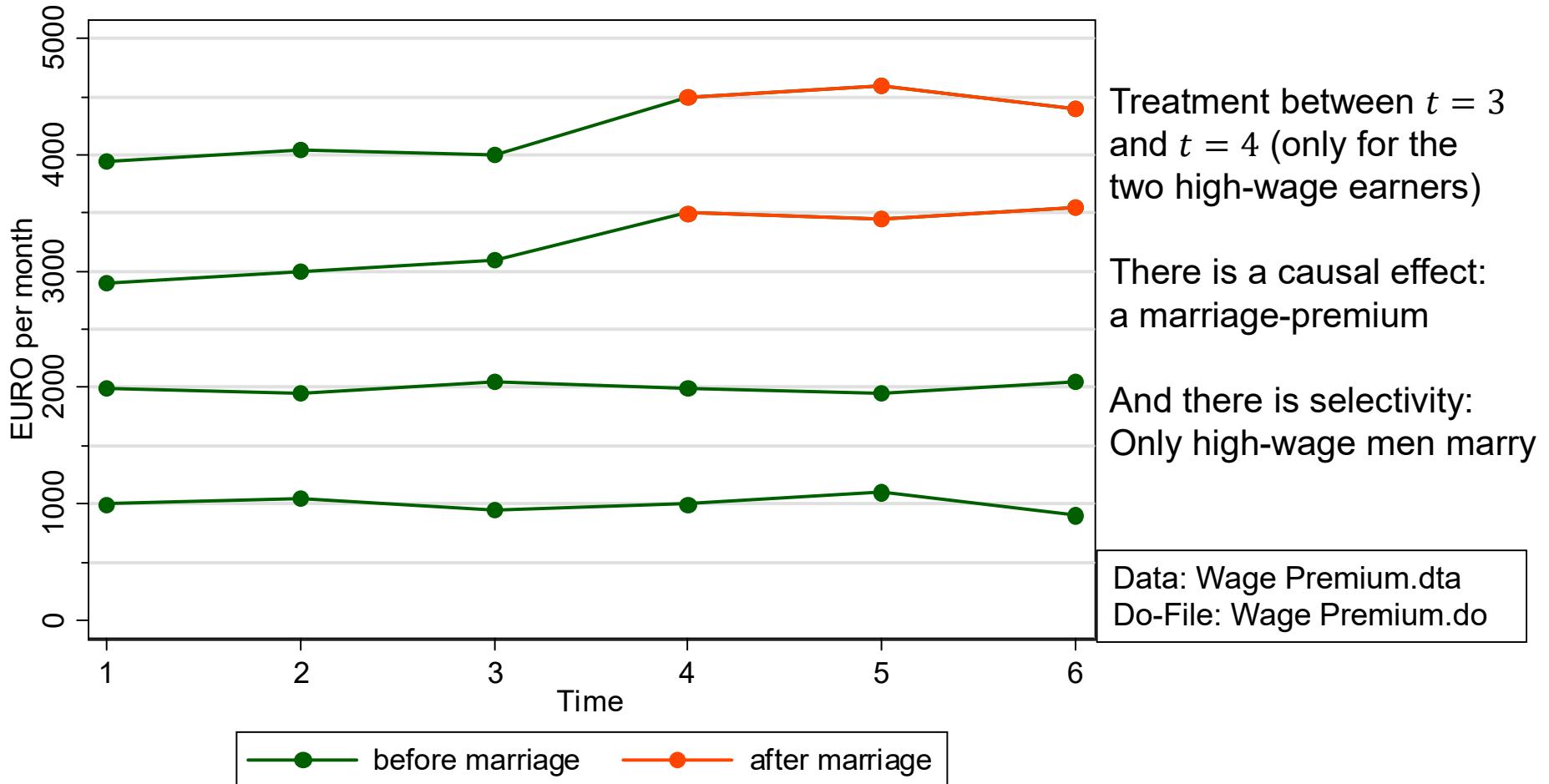


Is There a Marital Wage Premium for Men?

- For demonstrating how between and within estimation work we use a “toy example” with fabricated data (long-format)

. list id time wage marr, separator(6)					Data: Wage Premium.dta Do-File: Wage Premium.do				
	id	time	wage	marr		id	time	wage	marr
1.	1	1	1000	0	13.	3	1	2900	0
2.	1	2	1050	0	14.	3	2	3000	0
3.	1	3	950	0	15.	3	3	3100	0
4.	1	4	1000	0	16.	3	4	3500	1
5.	1	5	1100	0	17.	3	5	3450	1
6.	1	6	900	0	18.	3	6	3550	1
7.	2	1	2000	0	19.	4	1	3950	0
8.	2	2	1950	0	20.	4	2	4050	0
9.	2	3	2050	0	21.	4	3	4000	0
10.	2	4	2000	0	22.	4	4	4500	1
11.	2	5	1950	0	23.	4	5	4600	1
12.	2	6	2050	0	24.	4	6	4400	1

Is There a Marriage-Premium for Men?



In these data we have a problem with self-selection:
Married and unmarried men differ in characteristics other than marriage
(in these data the assumption of unit homogeneity is invalid)

How High is the Marriage-Premium?

- These are observational (non-experimental) data
 - Treatment assignment is not under control of the researcher (no randomization)
 - Instead, men can self-select into treatment (marriage)
 - Therefore, a between approach will be strongly biased (see below)
- A within approach to compute the marriage-premium
 - We have before (pre-treatment, $t = 1, 2, 3$) and after (post-treatment, $t = 4, 5, 6$) measurements
 - This allows for a within approach.
 - Because we also have a control group we can use within estimation with control group
 - Difference-in-differences (DiD)

How High is the Marriage-Premium?

- DiD is a after-before comparison with control group
 - After-before changes (Δ_i) treatment group
 - $\Delta_4 = 4500 - 4000 = 500$
 - $\Delta_3 = 3500 - 3000 = 500$
 - After-before changes (Δ_i) control group
 - $\Delta_2 = 2000 - 2000 = 0$
 - $\Delta_1 = 1000 - 1000 = 0$
 - To get the average causal effect we take the difference of the averages in treatment and control group
 - In the following we will investigate, whether different statistical regression models can recover this causal effect!
- $$\Delta_i = \frac{1}{3} \sum_{t=4}^6 y_{it} - \frac{1}{3} \sum_{t=1}^3 y_{it}$$
- $$\Delta = \overline{\Delta_{i \in T}} - \overline{\Delta_{i \in C}} = \frac{500 + 500}{2} - \frac{0 + 0}{2} = 500$$

Cross-Sectional Regression

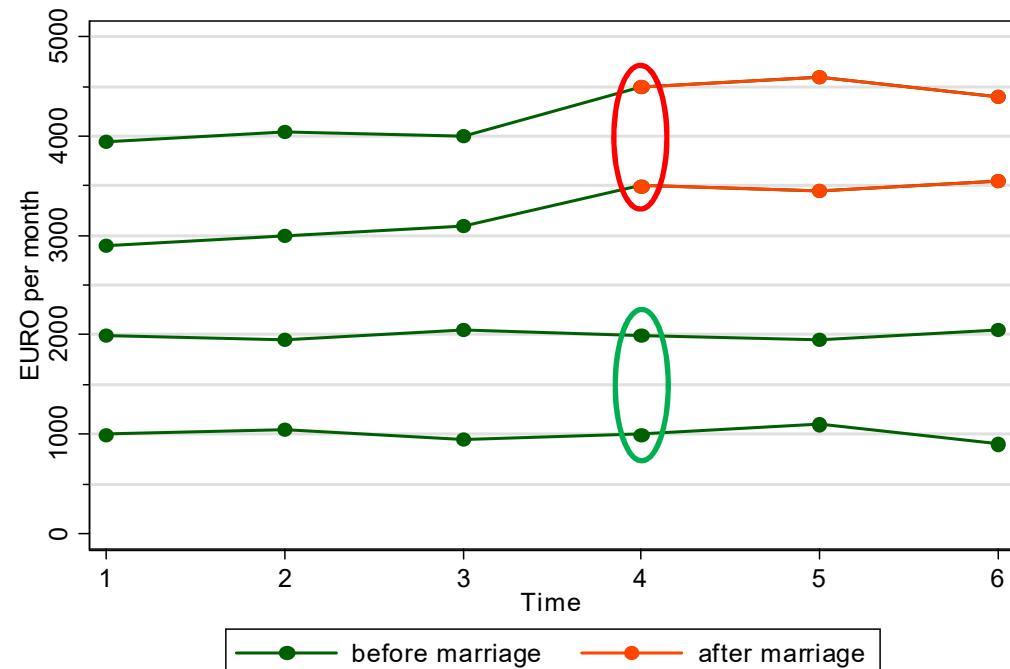
- Result of a cross-sectional regression at $t = 4$:

$$y_{i4} = \alpha + \beta x_{i4} + u_{i4}$$

- This is a between-comparison at $t = 4$: essentially this compares average wages of married and unmarried men at $t = 4$

$$\hat{\beta}_{OLS} = \frac{4500 + 3500}{2} - \frac{2000 + 1000}{2} = 2500$$

- We get a very large marital wage premium
 - **Obviously this is a massively biased result!**
 - The graph shows the information used by the cross-sectional regression



What Is the Problem With Cross-Sectional Regression?

- The most critical assumption of a linear regression

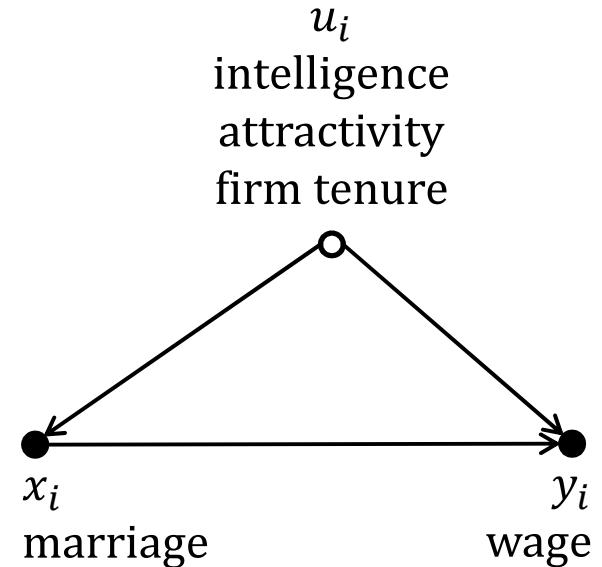
$$y_i = \alpha + \beta x_i + u_i$$

is the **exogeneity assumption**: $E(u_i|x_i) = 0$

- I.e., the error term and the regressor must be statistically independent
 - The exogeneity assumption implies:
 - $E(u_i) = 0$ The (unconditional) mean of the error term is 0
 - $\text{Cov}(x_i, u_i) = 0$ The error term does not correlate with X
 - The exogeneity assumption guarantees **unbiasedness** [$E(\hat{\beta}_{OLS}) = \beta$]
- Unfortunately, in many non-experimental social science research settings the exogeneity assumption will be violated
 - The error term and the regressor are dependent
$$E(u_i|x_i) \neq 0$$
 - Then it is said: the regressor is endogenous (**endogeneity**)
 - $\hat{\beta}_{OLS}$ **will be biased**

What Is the Problem?

- Where does endogeneity come from?
 - There are unobserved confounders (unobservables that affect both X and Y)
 - Then X and the error term are correlated
 - This is called „unobserved heterogeneity“ or „omitted variable bias“
 - [More on this in Chap. X]
- The underlying mechanism: **self-selection**
 - Treatment and control groups are not built by randomization
 - Instead, human beings decide according to unobservables, whether they go into treatment or not

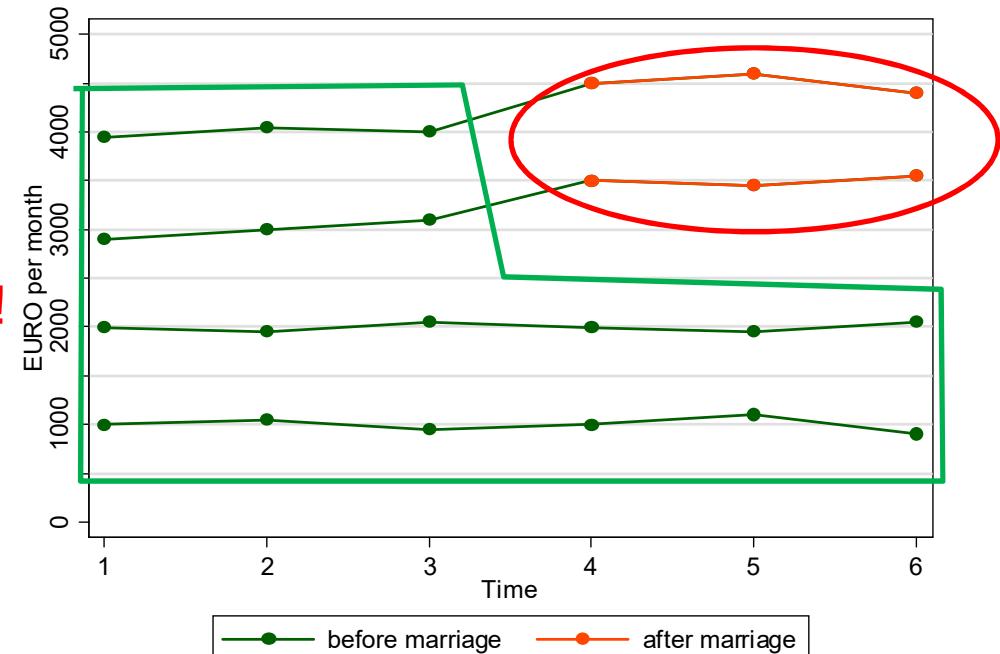


No Solution: Pooled-OLS

- Pool the data and estimate an OLS regression (POLS)

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

- The result is $\hat{\beta} = 1833$
- This is the mean of the red points - the mean of the green points
- The bias is still heavy
- The reason is that POLS also relies on a between comparison
- Panel data per se do not help to identify a causal effect!
- One has to use appropriate methods of analysis to make full advantage of panel data



Chapter III:

An Intuitive Introduction to Linear Panel

Regression

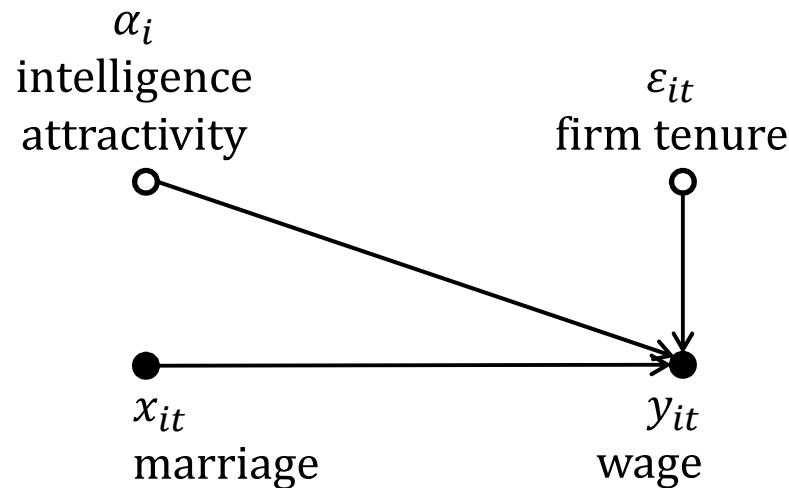
Section: Within Estimation

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



The Error Decomposition

- To make full advantage of panel data use within estimation
 - Within estimators implement a “after-before comparison”
- Starting point: error decomposition $u_{it} = \alpha_i + \varepsilon_{it}$



- α_i : person-specific time-constant error term
 - Assumption: person-specific random variable
- ε_{it} : time-varying error term (idiosyncratic error term)
 - Assumptions: zero mean, homoscedasticity, no autocorrelation

The Error Components Model

- This yields the error components model

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

- Note that the overall constant α has been dropped due to collinearity
- Generally, one assumes that the error components are independent from each other: $E(\varepsilon_{it}|x_{it}, \alpha_i) = 0$
 - We will neglect this subtlety in the following

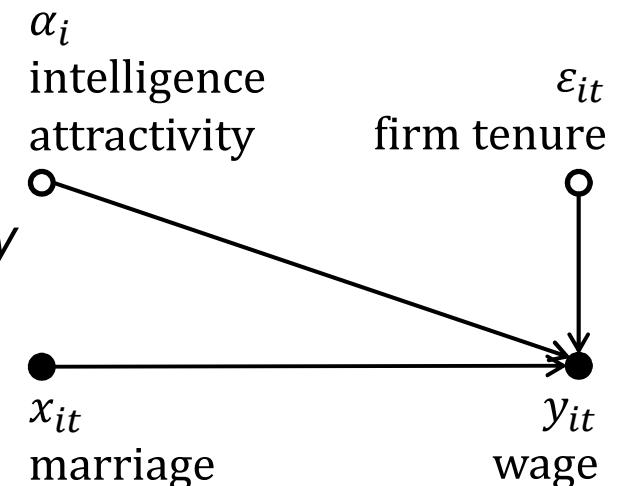
- POLS is unbiased only, if the regressor x_{it} is independent from **both** error components

$$E(\alpha_i|x_{it}) = 0 \quad \text{random-effects assumption}$$

“no (person-specific) time-constant unobserved heterogeneity”

$$E(\varepsilon_{it}|x_{it}) = 0 \quad \text{contemporaneous exogeneity assumption}$$

“no time-varying unobserved heterogeneity”



First-Differences Estimator (FD)

- The random-effects assumption is strong
 - How can we get rid of it?
- By a differencing transformation we can wipe out the α_i

$$\begin{aligned}y_{it} &= \beta x_{it} + \alpha_i + \varepsilon_{it} \\y_{it-1} &= \beta x_{it-1} + \alpha_i + \varepsilon_{it-1}\end{aligned}$$

Subtracting the second equation from the first gives:

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta \varepsilon_{it}$$

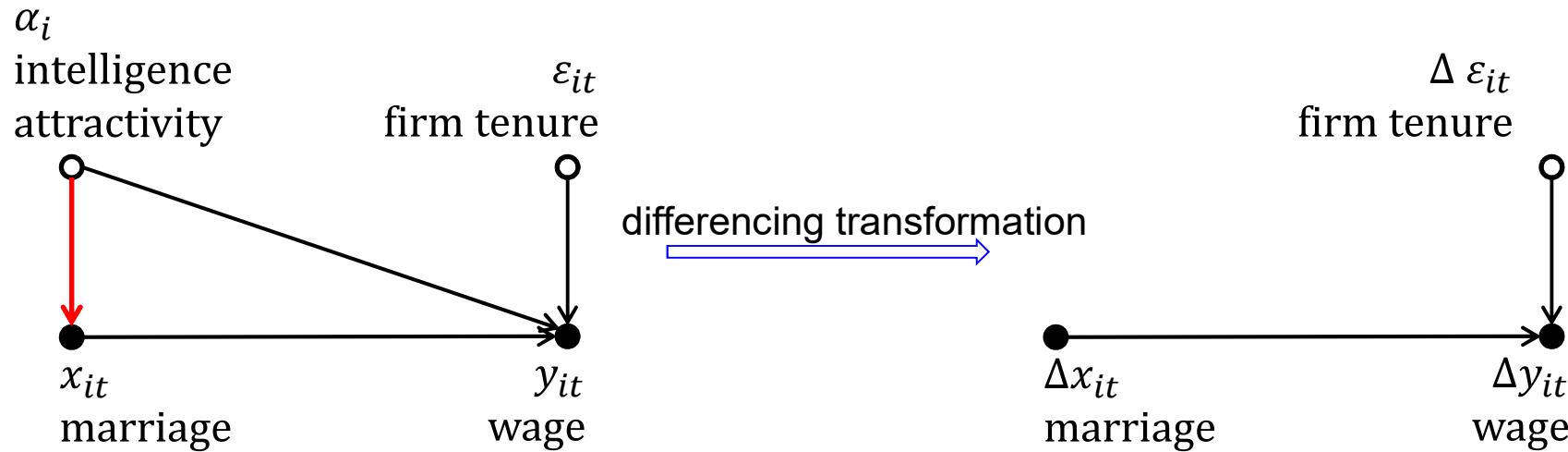
where “ Δ ” denotes the change from $t - 1$ to t .

- Person-specific errors have been “differenced out”. Time-constant unobserved heterogeneity has been wiped out!
- Pooled-OLS applied to these transformed data provides the first-differences estimator.

Assumptions of FD Estimation

- The FD-estimator is unbiased if

$$E(\varepsilon_{is}|x_{it}) = 0 \quad \text{for } t \leq s \quad \text{sequential exogeneity assumption}$$
 - Intuition: otherwise Δx_{it} and $\Delta \varepsilon_{it}$ would be correlated
 - However, because α_i is not in the differenced equation,
 $E(\alpha_i|x_{it}) = 0$ is no longer required for unbiasedness
 - FD identifies the causal effect under weaker assumptions
 - Time-constant unobserved heterogeneity is allowed
 - Only time-varying unobserved heterogeneity must not be



Example: FD-Regression

```
. regress D.(wage marr), noconstant
```

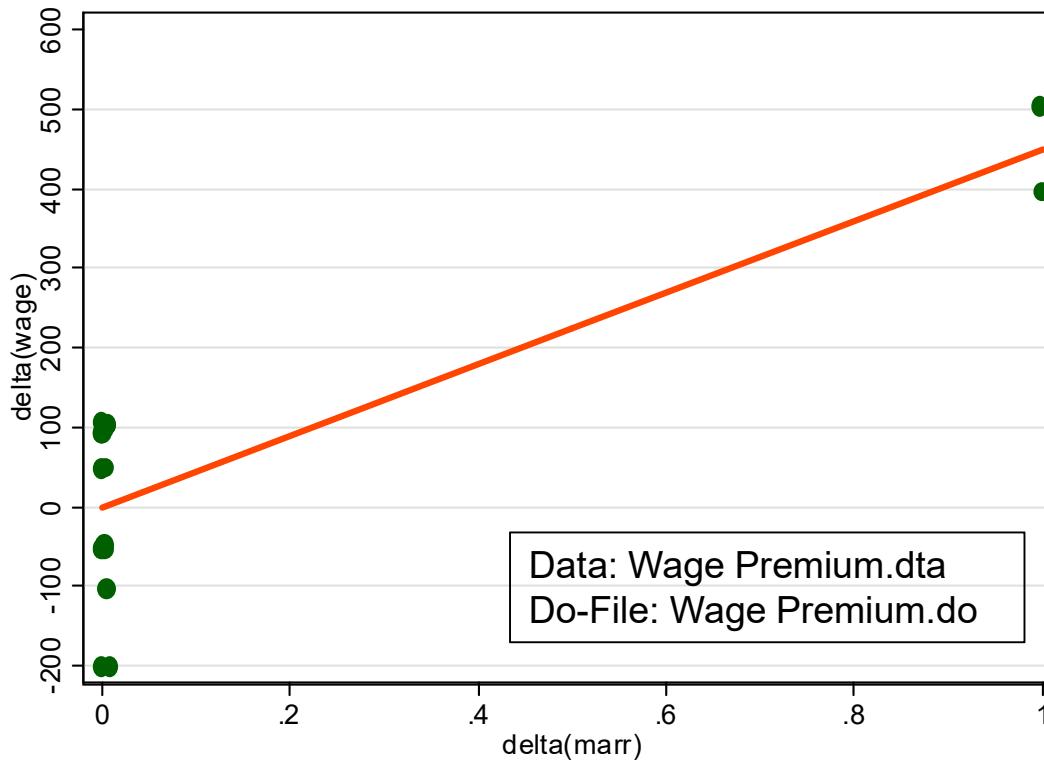
Source	SS	df	MS	Number of obs	=	20
Model	405000	1	405000	F(1, 19)	=	39.97
Residual	192500	19	10131.5789	Prob > F	=	0.0000
Total	597500	20	29875	R-squared	=	0.6778
				Adj R-squared	=	0.6609
				Root MSE	=	100.66

D.wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
marr					
D1.	450	71.17436	6.32	0.000	301.0304 598.9696

Data: Wage Premium.dta
Do-File: Wage Premium.do

- The FD-estimator is 450, which is very close to the true causal effect

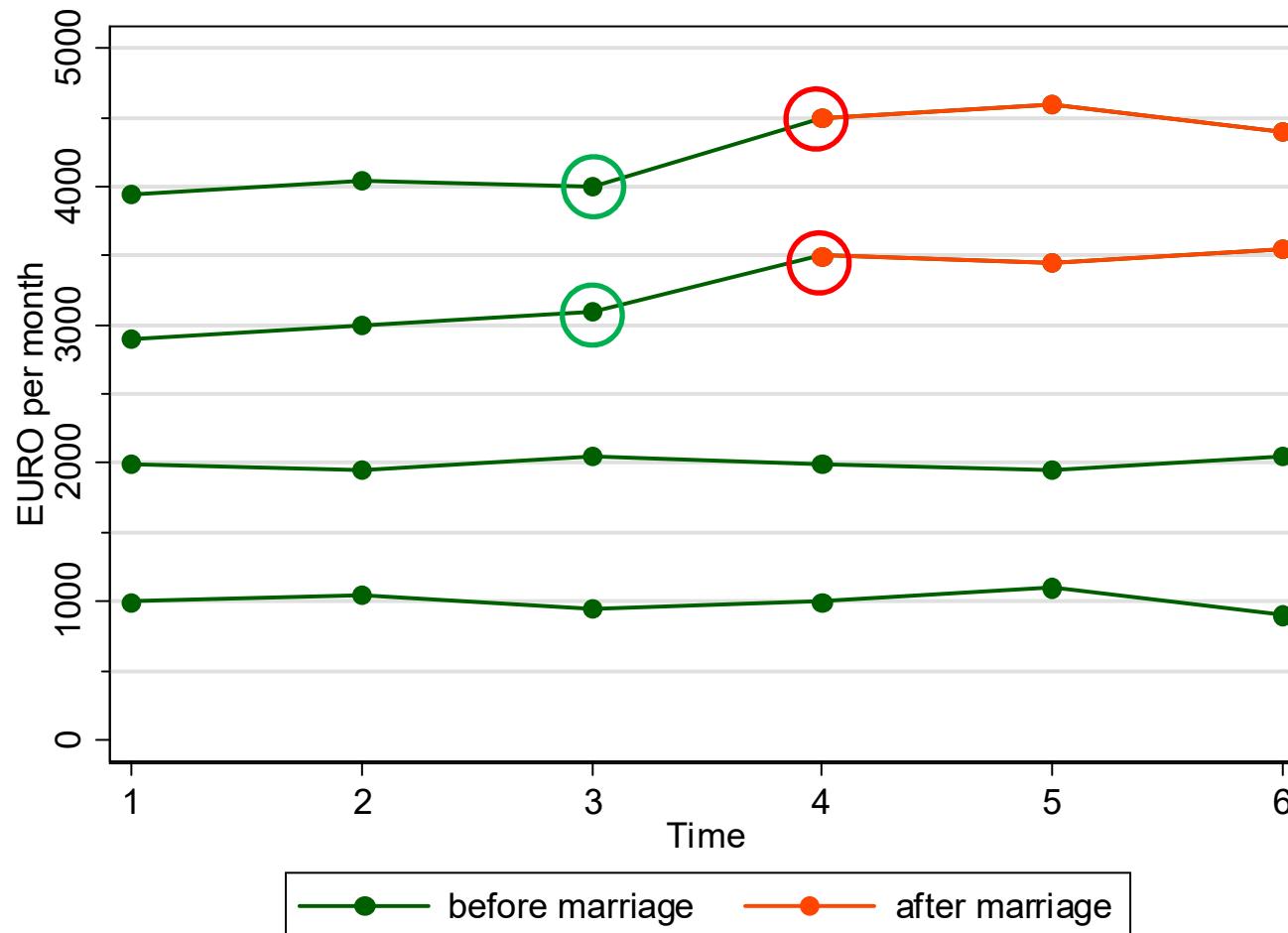
„Mechanics“ of a FD-Regression



- Because there is no constant, the regression line passes the point (0,0)
- The slope is based on only 2 observations
 - $\hat{\beta}$ is simply the average wage change before-after marriage
- With $T > 2$ FD-estimation is obviously inefficient

FD-Regression

- This is the information used by a FD-regression



Fixed-Effects Regression (FE)

- Fixed-effects estimation

- Error components model: $y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$ (1)

- Person-specific means over t: $\bar{y}_i = \beta \bar{x}_i + \alpha_i + \bar{\varepsilon}_i$ (2)

- “Demeaning” the data (within transformation): (1) – (2)

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (3)$$

- Demeaning wipes out person-specific time-constant unobserved heterogeneity! Only within variation is left.
 - Pooled OLS applied to demeaned data provides the fixed-effects estimator

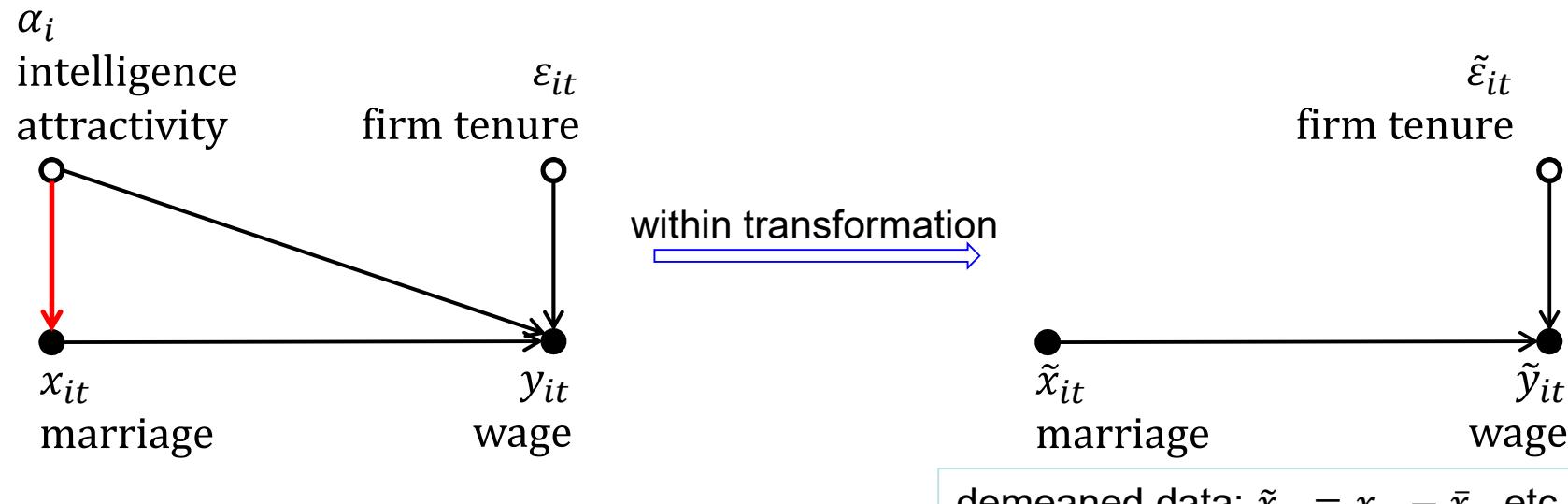
- Note: (2) is called “between regression” (BE)

Assumptions of FE Estimation

- The FE-estimator is unbiased if

$$E(\varepsilon_{is}|x_{it}) = 0 \quad \text{for all } t \text{ and } s \quad \textit{strict exogeneity assumption}$$

- Intuition: otherwise $x_{it} - \bar{x}_i$ and $\varepsilon_{it} - \bar{\varepsilon}_i$ would be correlated
- However, because α_i is not in the demeaned equation,
 $E(\alpha_i|x_{it}) = 0$ is no longer required for unbiasedness
 - FE identifies the causal effect under weaker assumptions
 - Time-constant unobserved heterogeneity is allowed
 - Only time-varying unobserved heterogeneity must not be



Example: Fixed-Effects Regression

```
. xtset id time
    panel variable: id (strongly balanced)
    time variable: time, 1 to 6
        delta: 1 unit
```

```
. xtreg wage marr, fe
```

Fixed-effects (within) regression

Group variable: id

R-sq: within = 0.8982
 between = 0.8351
 overall = 0.4065

corr(u_i, Xb) = 0.5164

Data: Wage Premium.dta
 Do-File: Wage Premium.do

Number of obs = 24

Number of groups = 4

Obs per group: min = 6
 avg = 6.0
 max = 6

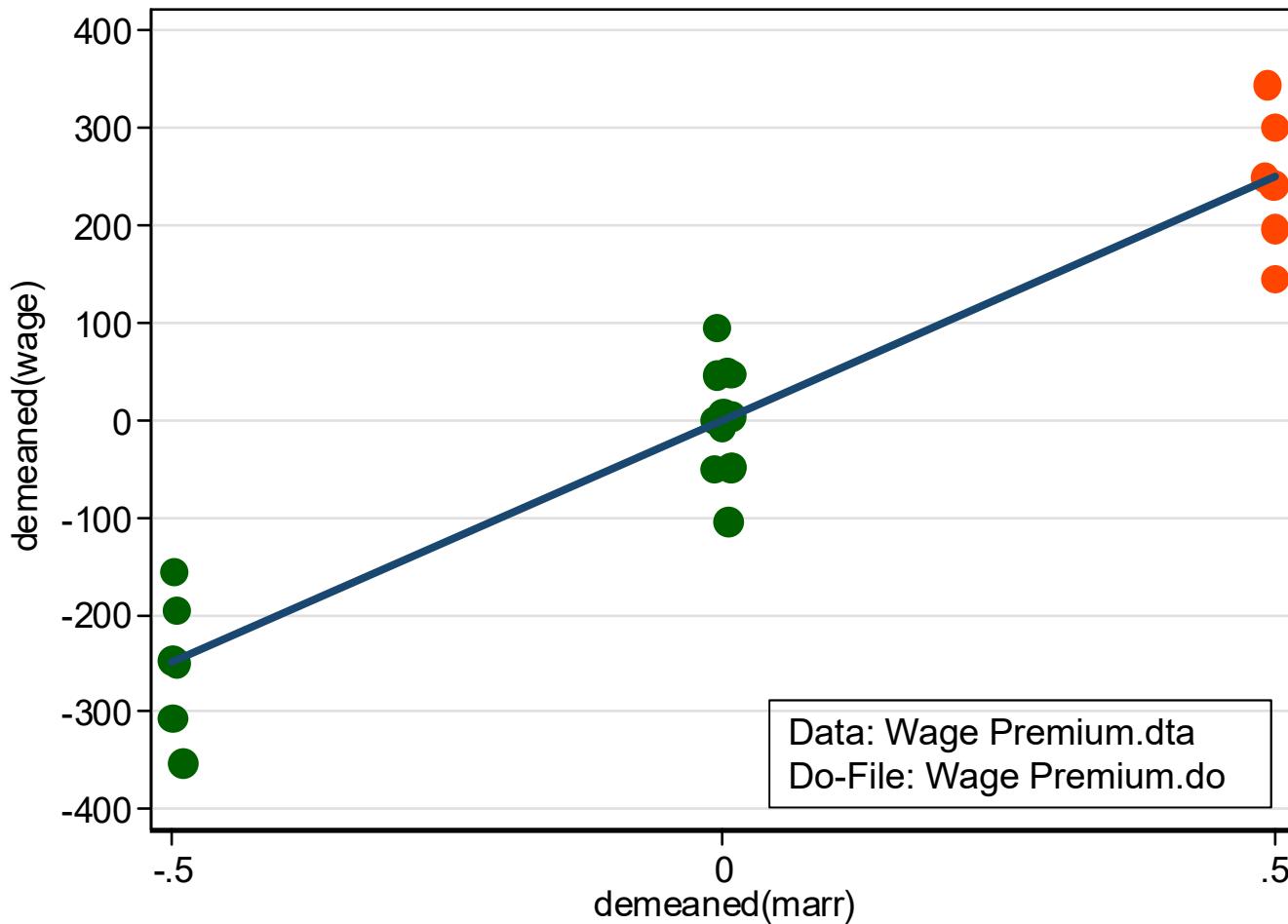
F(1,19) = 167.65
 Prob > F = 0.0000

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
marr	500	38.61642	12.95	0.000	419.1749 580.8251
_cons	2500	16.7214	149.51	0.000	2465.002 2534.998
<hr/>					
sigma_u	1290.9944				
sigma_e	66.885605				
rho	.99732298	(fraction of variance due to u_i)			

Interpreting the FE Output

- The FE model succeeds in identifying the true causal effect!
 - Marriage increases the wage by 500 €
 - The effect is significant (judged by the t-value or the p-value)
 - A constant is reported, since Stata adds back the wage mean for $x_{it} = 0$, which is 2500 here
 - Model fit can be judged by the within R^2 as usual (referring to (3))
 - 90% of the within wage variation is explained by marital status change
 - The between and overall R^2 refer to different models and are not useful here
 - Variance of the error components
 - sigma_u is the estimated standard deviation of $\hat{\alpha}_i$
 - sigma_e is the estimated standard deviation of $\hat{\varepsilon}_{it}$
- Further details: Andreß et al. (2013: 4.1.2.1)

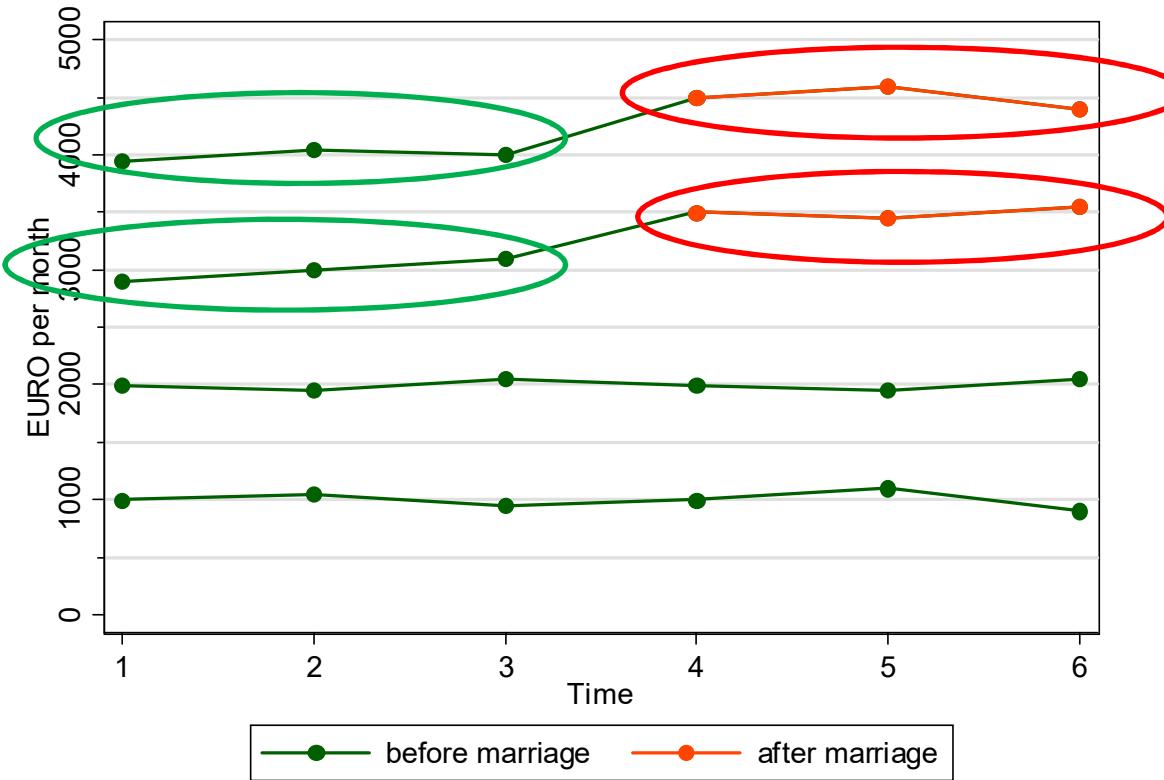
“Mechanics” of a FE-Regression



- Those, never marrying are at $X = 0$. They contribute nothing to the regression.
- The slope is only determined by the wages of those marrying:
It is the difference in the mean wage before and after marriage.

FE-Regression

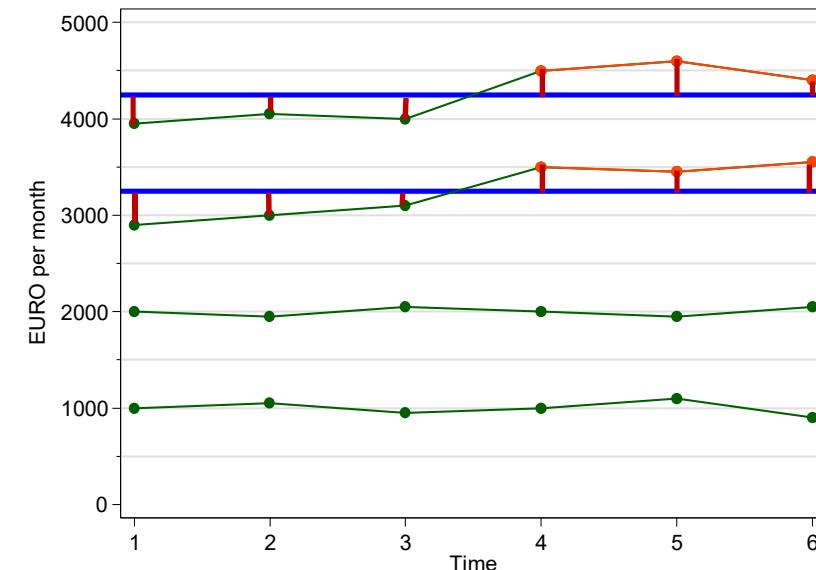
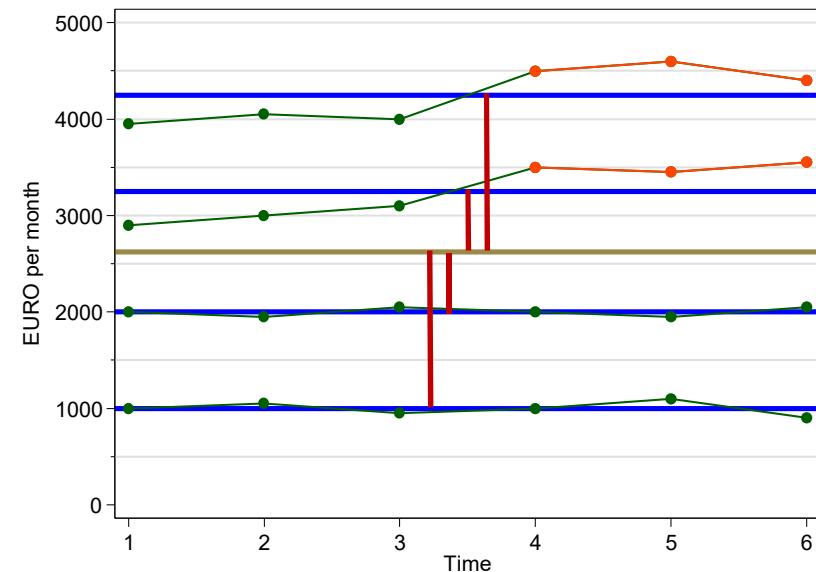
- This is the information used by a FE-regression



- The control group provides no information for FE estimation
 - This is generally true: groups where X does not change contribute nothing to the FE-estimator
 - Note that Stata reports the N in the data, not the N used for FE-estimation!

Between- and Within-Variation

- To identify the causal effect of a marriage ...
 - a between regression (BE) uses the between variation
 - This is heavily affected by self-selection of the high-wage men into treatment
 - The BE marriage premium is estimated to be 4500 €!
 - a within regression (FE) uses only within variation (of the treated only)
 - The causal effect is identified by the deviations from the person-specific means
 - The “contaminated” (Allison 2009) between variation is ignored completely
 - Therefore, self-selection into treatment does not bias results



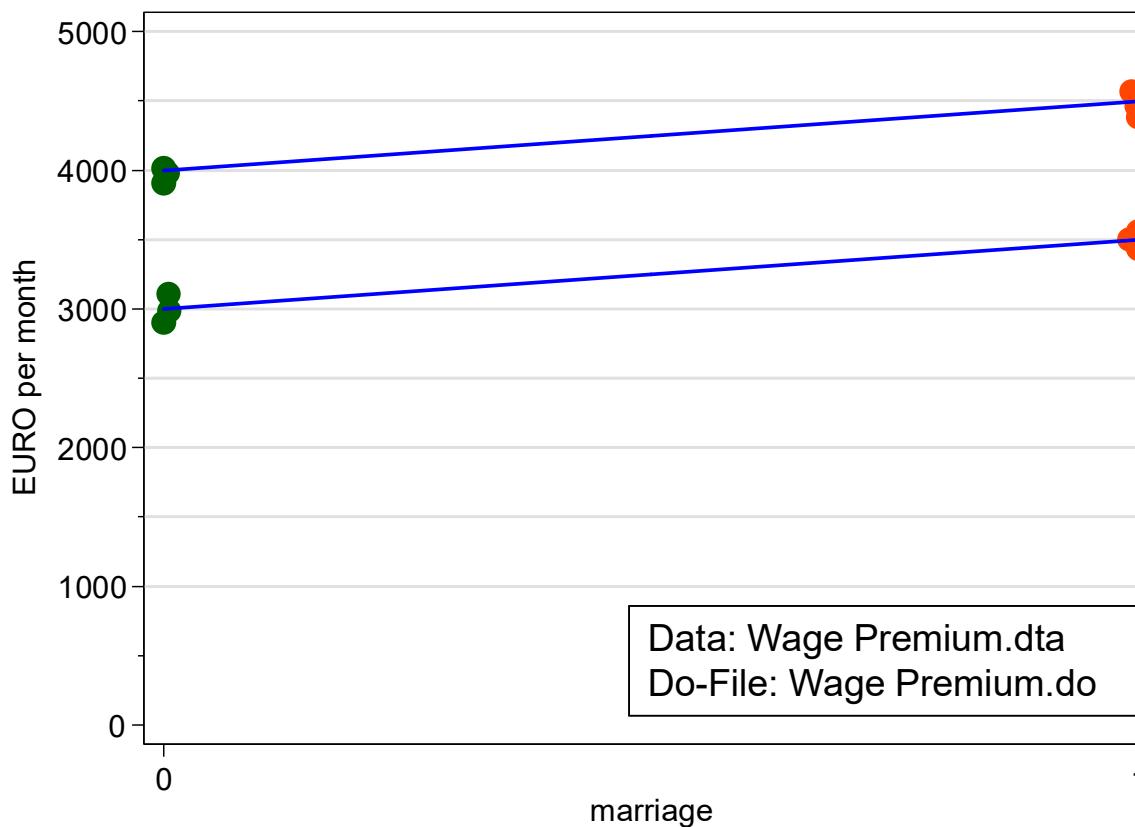
Equivalent FE-estimator I: LSDV

. regress wage marr ibn.id , noconstant						
Source	SS	df	MS	Number of obs = 24		
Model	202500000	5	40500000	F(5, 19) = 9052.94		
Residual	85000	19	4473.68421	Prob > F = 0.0000		
Total	202585000	24	8441041.67	R-squared = 0.9996		
<hr/>						
wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
marr	500	38.61642	12.95	0.000	419.1749	580.8251
id	1	1000	27.30593	36.62	0.000	942.848
	2	2000	27.30593	73.24	0.000	1942.848
	3	3000	33.4428	89.71	0.000	2930.003
	4	4000	33.4428	119.61	0.000	3930.003
<hr/>						

- Least-squares-dummy-variables-estimator (LSDV)
- Practical only when N is small
- We get estimates for the α_i

Data: Wage Premium.dta
Do-File: Wage Premium.do

Equivalent FE-estimator II: Individual Slope Regression



- Estimate a separate regression for every man marrying (blue)
 - In our data the slopes are equal: +500 for both men
- The FE estimator is the (weighted) mean of the individual slopes
 - On average this gives +500, identical to the FE estimator from above

Chapter III:

An Intuitive Introduction to Linear Panel

Regression

Section: Within Estimation With Age/Period
Effects

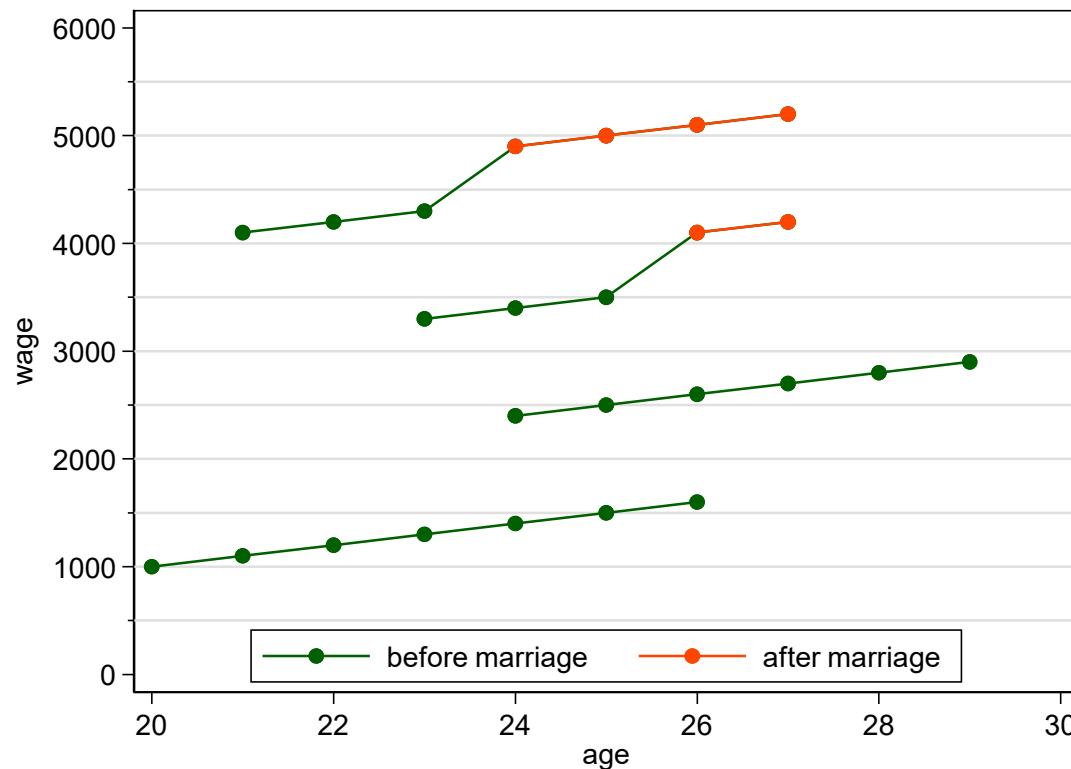
Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Age/Period Effects

- So far we assumed temporal homogeneity
 - Therefore, the control group was not really necessary
 - This differs, however, if we allow for age/period effects
- Only age effects

$$y_{it} = \alpha_i + 500 \cdot marr_{it} + 100 \cdot age_{it}$$



Data: Wage Premium 2.dta
Do-File: Wage Premium 2.do

Modeling the Age Effect

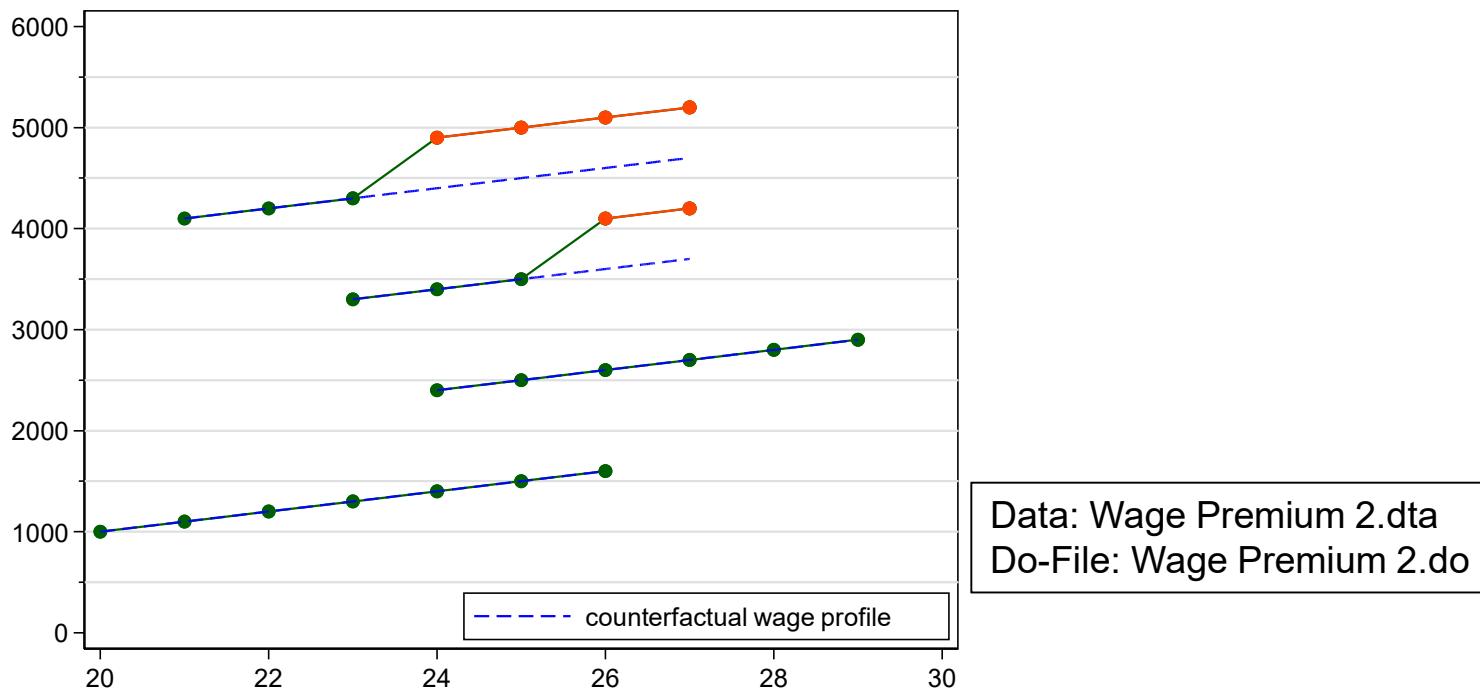
- Now, if we „forget“ to control for „age“, FE is biased

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
marr	808.8235	97.76186	8.27	0.000	604.8959 1012.751
_cons	2857.882	40.79979	70.05	0.000	2772.775 2942.989

- Controlling for age (`xtreg wage marr age, fe`) yields the correct result (+500)
- If age is not controlled, the treatment effect will pick up the age effect (FE makes an after/before comparison)
- Therefore, it is essential for within analyses that one controls age “perfectly”
 - If possible include age dummies (i.e. `i.age`)

Do We Need a Control Group?

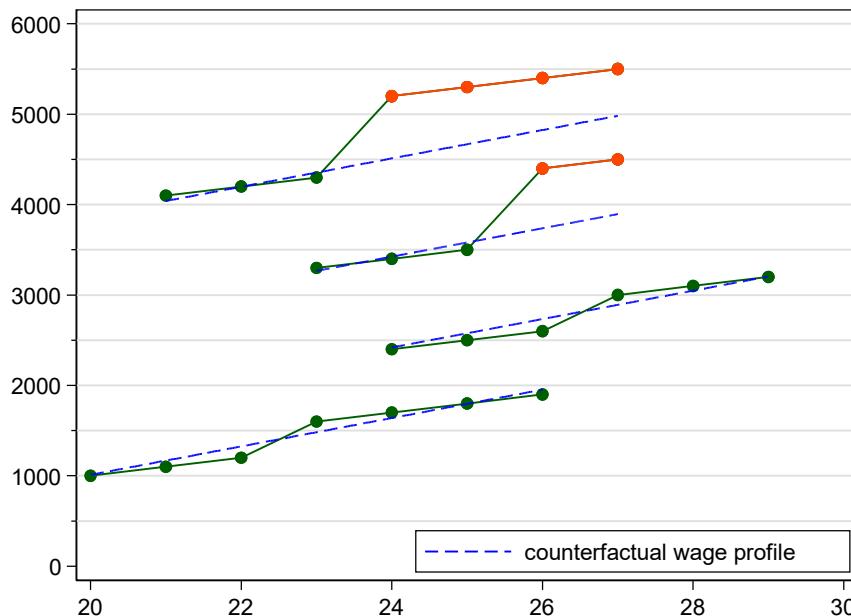
- In this example we would not even need the control group, because we have enough pre-treatment observations for estimating the wage profile



- However, it is inefficient to drop the control group
 - Therefore, we would strongly recommend to always include a control group in within estimation

Period Effects

- Now we add a period effect
 - General wage increase (+300 EUR) at $t = 3.5$

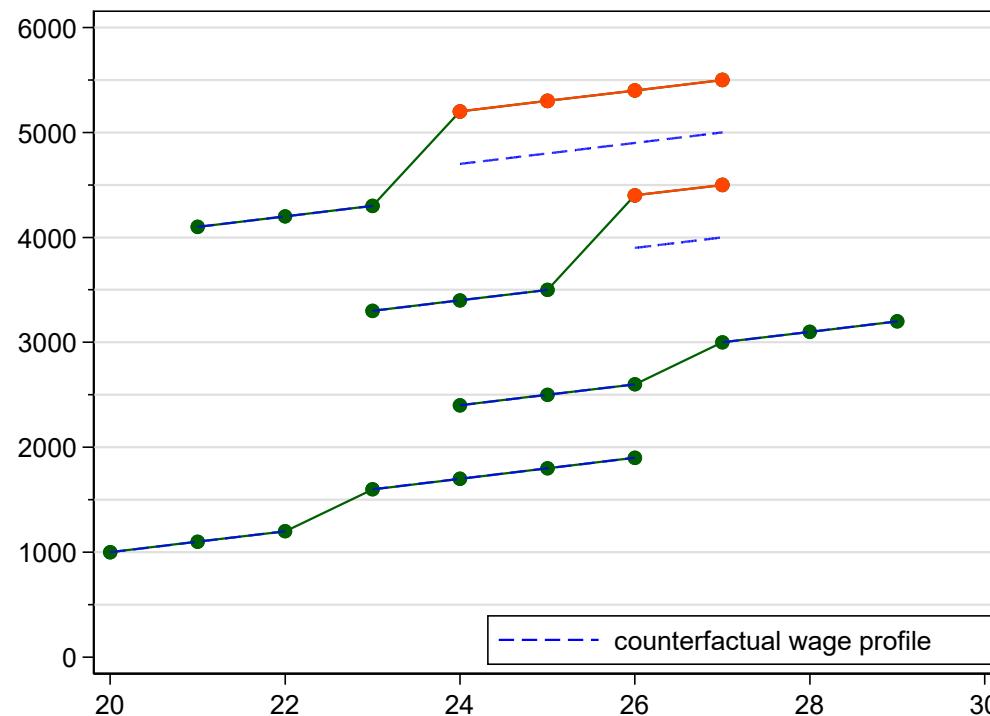


Now, a FE model without period control is biased

xtreg wage1 marr age, fe						
wage1	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
marr	625.3696	54.85666	11.40	0.000	510.5533	740.1859
age	156.547	10.24831	15.28	0.000	135.097	177.9969
_cons	-780.6209	244.4091	-3.19	0.005	-1292.175	-269.0668

Period Effects

- A FE model with period dummy ($t \geq 4$) identifies the true marriage effect (+500 EUR)



- Here the control group is essential for identifying the period effect!
 - Without control group, FE would estimate +800 EUR

Summary: How to Model Time?

- **Main lesson:**
 - Control for both age and period effect
 - Include a control group
- It is, however, not possible to include both i.age and i.time
 - Due to the APC-problem (see below)
 - Therefore, the better strategy is to control for age as good as possible (i.age) and add (selected) period effects

Caution: Two-Way FE-Regression

- Often used: two-way FE-regression

- Including time fixed-effects ($T - 1$ period dummies μ_t)

$$y_{it} = \beta x_{it} + \mu_t + \alpha_i + \varepsilon_{it}$$

- Mixes age and period effect (no substantive interpretation possible)
- Works only, if there are no/constant age effects

CONSTANT (LINEAR) AGE EFFECT

```
xtreg wage1 marr i.time, fe
```

wage1	Coef.
-------	-------

marr	500
------	-----

time	2		100
	3		200
	4		600
	5		700
	6		800
	7		900

QUADRATIC AGE EFFECT

```
xtreg wage2 marr i.time, fe
```

wage2	Coef.
-------	-------

marr	506
------	-----

time	2		100
	3		160
	4		477
	5		457
	6		437
	7		513

Chapter IV:

The Basics of Linear Panel Regression

Section: Linear Panel Models

More details on the statistics can be found in
Brüderl/Ludwig (2015)

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



A Note on the Exogeneity Assumptions

- In this section we generalize to a multivariate regression

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

- Now we have K regressors X_1, \dots, X_K . \mathbf{x}_{it} is the $(1 \times K)$ vector (row vector) of the observed covariates of a person.
- $\boldsymbol{\beta}$ is the corresponding $(K \times 1)$ vector (column vector) of parameters to be estimated (regression coefficients)
- Now the exogeneity assumptions in conditional mean formulation are

$$E(\alpha_i | \mathbf{x}_{it}) = 0$$

$$E(\varepsilon_{it} | \mathbf{x}_{it}, \alpha_i) = 0$$

- The error terms must be independent from all regressors
- For statistical inference, as well as efficiency properties of estimators these (strong) exogeneity assumptions must hold (Wooldridge, 2010: 288)
- For consistency, however, also weaker assumptions of linear independence suffice

$$\text{Cov}(\mathbf{x}'_{it}, \alpha_i) = E(\mathbf{x}'_{it}\alpha_i) = \mathbf{0}$$

$$\text{Cov}(\mathbf{x}'_{it}, \varepsilon_{it}) = E(\mathbf{x}'_{it}\varepsilon_{it}) = \mathbf{0}$$

- In the following we will use the exogeneity assumptions in this weaker form

POLS Estimation

- We start from a multivariate error components model

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

- From the perspective of non-experimental research, the most critical assumptions are exogeneity assumptions on the error terms:
both error terms have to be uncorrelated with the regressors

$$E(\mathbf{x}'_{it}\alpha_i) = \mathbf{0} \quad \text{random-effects assumption}$$

$$E(\mathbf{x}'_{it}\varepsilon_{it}) = \mathbf{0} \quad \text{contemporaneous exogeneity assumption}$$

- Contemporaneous exogeneity requires that idiosyncratic errors are not systematically related to the regressors. This assumption is often reasonable.
 - The random-effects assumption, however, often will be violated because slow-to-change, hard-to-measure traits that are correlated with the regressors are ubiquitous (Firebaugh et al. 2013)
 - E.g., cognitive and non-cognitive ability, genetic disposition, personality, social milieu, peer group characteristics
 - If the random-effects assumption fails, estimates of $\boldsymbol{\beta}$ will be biased (omitted variable bias) (unobserved heterogeneity bias)

FE Estimation

- FE wipes out person-specific time-constant unobservables (fixed-effects) by applying the within transformation

- The model (including time-constant variables \mathbf{z}_i)

$$\begin{aligned}y_{it} &= \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + \alpha_i + \varepsilon_{it} \\ \bar{y}_i &= \bar{\mathbf{x}}_i\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + \alpha_i + \bar{\varepsilon}_i\end{aligned}$$

- “Demeaning” the data (within transformation):

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- Now the α_i have gone from the equation, and POLS will provide consistent estimates of $\boldsymbol{\beta}$, if the assumptions on the next slide hold
 - Note 1: By the within transformation also all time-constant variables \mathbf{z}_i have been eliminated. With FE it is not possible to estimate their effects $\boldsymbol{\delta}$.
 - Note 2: The term “fixed-effects” comes from the older literature, where the α_i are seen as unit-specific parameters to be estimated for each unit (LSDV). The newer literature sees the α_i also as random variables (Wooldridge, 2010: 285f).

FE Assumptions

- No assumption on person-specific time-constant unobserved heterogeneity is needed
 - We no longer need the random-effects assumption
 - Instead, the “fixed-effects assumption” allows for arbitrary correlation between α_i and x_{it}
 - **The FE estimator is consistent even if $E(x'_{it}\alpha_i) \neq 0$**
- For consistent estimates we need
$$E(x'_{is}\varepsilon_{it}) = \mathbf{0}, \text{ for all } s, t = 1, \dots, T \quad \textit{strict exogeneity assumption}$$
 - Covariates in each time period are uncorrelated with the idiosyncratic error in each time period
 - For consistency of estimates strict exogeneity is essential. Therefore, it is very important to discuss the plausibility of this assumption in every FE application [see Chapter X]
- Further FE assumptions (Wooldridge, 2010: 300 ff)
 - Full rank of the matrix of the demeaned regressors (no multicollinearity)
 - Idiosyncratic errors have constant variance across t (homoskedasticity)
 - Idiosyncratic errors are serially uncorrelated (no autocorrelation)

Other Within Estimators

- Least-squares-dummy-variables (LSDV)
 - POLS model, including a dummy for each person
 - Equivalent with FE
 - However, computationally impractical with large N
- First-differences (FD)
 - FE and FD are equivalent for $T = 2$. However for longer panels they will differ. FD is less efficient than FE.
 - However, instead of „strict exogeneity“ only „sequential exogeneity“!
 - Wooldridge (2010: 321ff) gives an extended discussion of the pros and cons of FD and FE. He finally favors FE (as does the literature).
- Difference-in-differences (DiD)
 - DiD implements a before-after comparison with control group
 - DiD is intuitively appealing (and for $T = 2$ equivalent to FD and FE)
 - For $T > 2$ and with controls DiD differs, however, from FD and FE
 - Therefore, generally DiD is not recommended for individual panel data (more on DiD in Chapter VII)

Statistical Inference With Panel Data

- With panel data the idiosyncratic errors are potentially
 - Heteroskedastic (i.e., nonconstant variance)
 - Autocorrelated (i.e., serial correlation in $\{\varepsilon_{it}: t = 1, \dots, T\}$)
- Ignoring this leads to under-estimated S.E.s
 - POLS ignores the panel structure completely
 - FE assumes equi-correlated errors over t , which is a quite unrealistic error structure
- Solution I: Assume a more realistic error structure
 - In Stata with
 - `xtgee`: generalized linear models with unit-specific correlation structure
 - `xtregar`: panel regression with first-order autoregressive error term
 - Drawback: results heavily depend on the assumptions made
 - And: `xtgee` estimates pooled models!
 - Thus, the S.E.s might be improved. However, the effect estimates are probably severely biased!

Panel-Robust S.E.s

- Solution II: panel-robust S.E.s
 - An extension of the Huber-White sandwich estimator
 - They correct for arbitrary serial correlation and heteroskedasticity
 - For formulas see Brüderl/Ludwig (2015: 334)
 - In Stata via `vce(cluster id)`
 - However, panel-robust S.E.s are also biased in finite samples
 - Sometimes they are even smaller than conventional S.E.s
 - But: “Your standard errors probably won’t be quite right, but they rarely are. Avoid embarrassment by being your own best skeptic, and especially, DON’T PANIC!” (Angrist/Pischke, 2009: 327)
 - **The major task with non-experimental data is to get the effect estimates right, a minor task is to get the S.E.s right!**
- Solution III: (panel) bootstrap S.E.s
 - In Stata via `vce(bootstrap)`
 - Draw many samples over i (with replacement) (size N)
 - Calculate the coefficient estimate
 - Compute the variance of all these estimates (50 replications)

Random-Effects Estimation

- There is another popular estimation strategy
 - Random-effects (RE) estimation
- A multivariate error components model (incl. constant)

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + \alpha_i + \varepsilon_{it}$$

- We assume that the α_i are i.i.d. random-effects
 - Usually normal distribution is assumed
- For this model we need **both** exogeneity assumptions
 - No time-constant unobserved heterogeneity

$$E(\mathbf{x}'_{it}\alpha_i) = \mathbf{0} \quad \text{random-effects assumption}$$

- No time-varying unobserved heterogeneity

$$E(\mathbf{x}'_{it}\varepsilon_{is}) = \mathbf{0}, \text{ for all } s, t = 1, \dots, T \quad \text{strict exogeneity assumption}$$

- Then the pooled feasible generalized least squares (FGLS) estimator is consistent and efficient
 - POLS is also consistent but not efficient due to autocorrelated error terms (induced by α_i)

- If the random-effects assumption is violated,
the RE estimates will be biased!

Random-Effects Estimation

- Alternatively RE can be obtained via transformed data

- RE is obtained by applying POLS to the quasi-demeaned data

$$(y_{it} - \theta \bar{y}_i) = \alpha(1 - \theta) + (x_{it} - \theta \bar{x}_i)\beta + z_i(1 - \theta)\delta + \{\alpha_i(1 - \theta) + (\varepsilon_{it} - \theta \bar{\varepsilon}_i)\} \quad (4)$$

- Where $\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{T\sigma_\alpha^2 + \sigma_\varepsilon^2}}$
 - This shows that the RE estimator mixes between and within estimators. The two extreme cases are
 - $\theta = 1$: FE estimator (e.g., $T \rightarrow \infty, \sigma_\alpha^2$ large)
 - $\theta = 0$: POLS
 - RE estimates are between POLS and FE (in the bivariate case)
- As can be seen, with RE we also get estimates of the effects of time-constant regressors (δ)

Example: Random-Effects Regression

```
. xtreg wage marr, re theta

Random-effects GLS regression
Group variable: id
Number of obs      =        24
Number of groups   =         4
Obs per group: min =          6
                           avg =       6.0
                           max =          6

R-sq:  within  = 0.8982
      between = 0.8351
      overall = 0.4065
Wald chi2(1)      =     128.82
Prob > chi2       =    0.0000

Random effects u_i ~ Gaussian
corr(u_i, X)      = 0 (assumed)
theta             = .96138358

-----  
wage |      Coef.    Std. Err.      z    P>|z| [95% Conf. Interval]
-----+-----  
marr |  502.9802  44.31525    11.35  0.000    416.1239  589.8365  
_cons |  2499.255  406.0315     6.16  0.000    1703.448  3295.062
-----+-----  
sigma_u |  706.57936  
sigma_e |  66.885605  
rho    |  .99111885  (fraction of variance due to u_i)
-----
```

Interpreting the RE Output

- The RE model does not succeed in identifying the true causal effect!
 - Marriage increases the wage by 503 €
 - The reason for the bias is that in our data the random-effects assumption is violated
 - However, the bias is low, because $\hat{\theta} = 0.96$
 - This is because σ_{α}^2 is so large in our data
 - The effect is significant (judged by the t-value or the p-value)
 - The “relevant” R^2 would refer to (4) and is not estimable
 - Model fit can “approximately” be judged by the within R^2 where the RE-estimates are plugged into (3)
 - Approximately 90% of the within wage variation is explained by marital status change

Chapter IV:

The Basics of Linear Panel Regression

Section: Fixed- or Random-Effects?

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



The Primary Advantage of Panel Data

- Panel data and within estimation allow to identify causal effects under weaker assumptions: time-constant unobserved heterogeneity does not bias estimates
 - “In many applications the whole point of using panel data is to allow for α_i to be arbitrarily correlated with the x_{it} . A fixed effects analysis achieves this purpose explicitly.”
(Wooldridge, 2010: 300)
 - “The DiD, fixed effects, and first difference estimators (within estimators) offer researchers the capacity to dispense with the random effects assumption and still obtain unbiased and consistent estimates when unit effects [α_i] are arbitrarily correlated with measured explanatory variables. This is widely regarded as the primary advantage of panel data.”
(Halaby, 2004: 516)

Nevertheless RE Models are Often Used

- Unfortunately, sociologist often use RE models
 - Halaby (2004) identifies 31 papers appearing in ASR and AJS between 1990 and 2003 that use panel data for causal analysis. 15 out of these used RE only
 - Giesselmann/Windzio (2014) identify 10 papers appearing in ZfS and KZfSS between 2000 and 2009 that use panel data for causal analysis. 3 out of these used RE only
- RE is biased by time-constant unobserved heterogeneity
 - **Since time-constant unobserved heterogeneity is ubiquitous in non-experimental social research, RE estimates generally will be biased**
 - So why would anybody want to use RE models?
- Two arguments in favor of RE are often brought forward
 - RE allows for estimating effects of time-constant regressors
 - RE is more efficient than FE, if $E(\mathbf{x}'_{it}\alpha_i) = \mathbf{0}$

Effects of Time-Constant Regressors

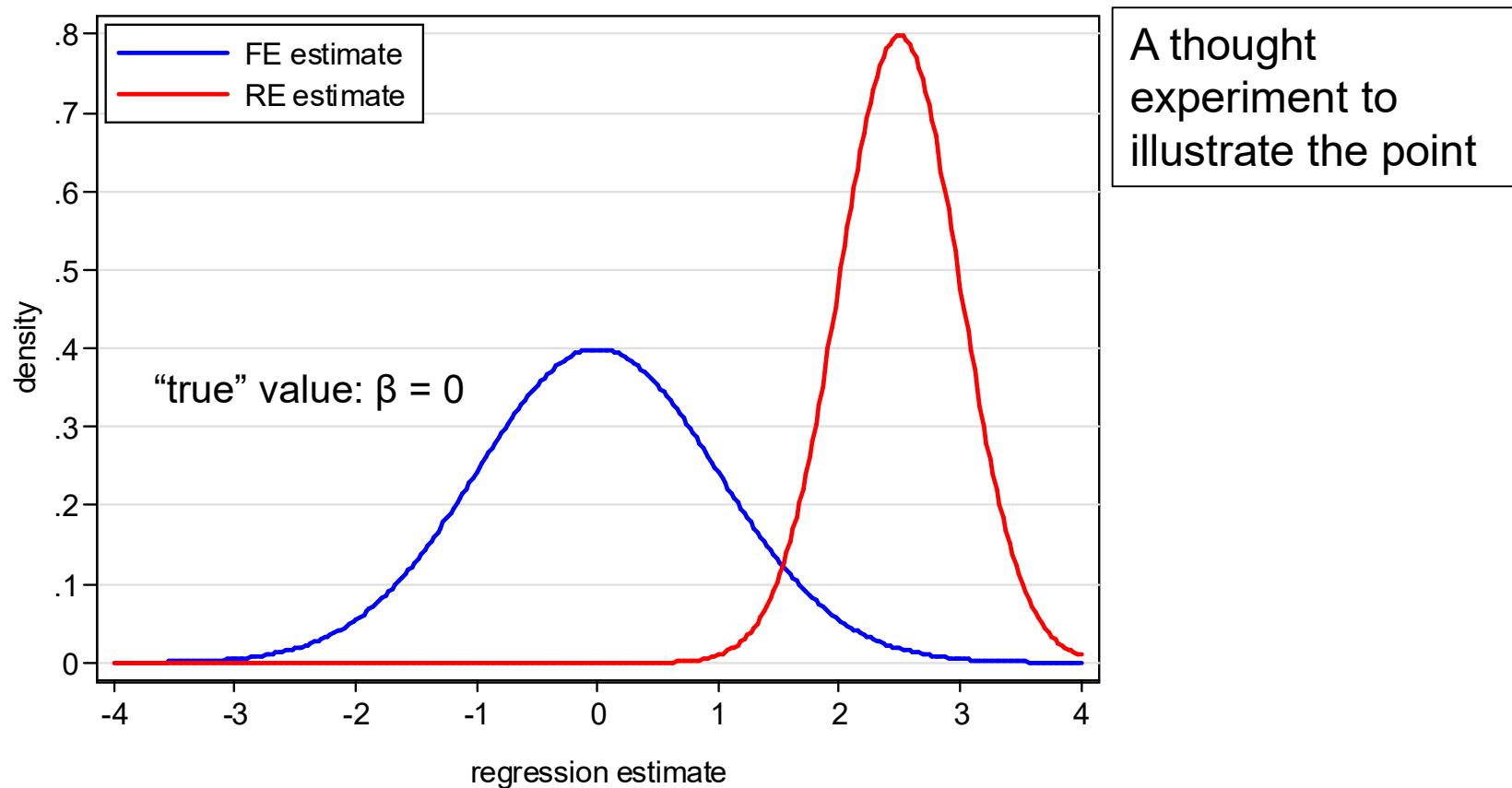
- Within estimators cannot estimate the effects of time-constant variables
 - E.g., sex, nationality, race, social origin, birth cohort, etc.
- Often it is argued that wiping out all time-constant regressors is a shortcoming of FE estimation
 - However, wiping out all time-constant regressors is **not** a shortcoming of FE. In fact it is a major strength, because alongside also all time-constant unobservables are eliminated
- Instead, the shortcoming is with a style of data analysis that by default throws all kinds of controls into a regression (“kitchen-sink-approach”)
 - Using the RE estimator only to report effects of sex, race, etc. is risking to throw away the big advantage of panel data
 - Instead of a thoughtless kitchen-sink-approach we should carefully think about the identification of a single causal effect

Effects of Time-Constant Regressors

- However, there is clearly a restriction of within methodology: it applies only with time-varying variables
 - Something has to “happen”: Only then a before-after comparison is possible
→ Analyzing the effects of events
 - Such questions are the main strength of panel data and within methodology
 - [Event variables can also be metric]
- If one has substantive interest in the effect of a time-constant regressor, one should use group specific growth curves, instead of a simple RE model (see Chapter VI)

Bias versus Efficiency

- Sometimes it is argued, that a biased but efficient estimate (RE) might be preferable to an unbiased but less efficient estimate (FE)
 - This argument is only sound, if the bias is small
 - However, RE is more efficient because it also uses the endogenous between variation. Generally, this is not a good idea, because this will produce a large bias.



Testing whether RE or FE: Hausman Test

- Use RE models only, if a Hausman test says „it is ok“
 - The intuition: the FE estimates ($\hat{\beta}_{FE}$) are consistent; If the RE estimates ($\hat{\beta}_{RE}$) do not differ too much, one can use RE regression
- $H_0 : \hat{\beta}_{RE} = \hat{\beta}_{FE}$
- $$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})' [\hat{V}(\hat{\beta}_{RE}) - \hat{V}(\hat{\beta}_{FE})]^{-1} (\hat{\beta}_{RE} - \hat{\beta}_{FE})$$
- $$H \stackrel{a}{\sim} \chi^2(k)$$
- If you are **not** able to reject H_0 , then you can use RE

	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	FE	RE	Difference	S.E.
marr	500	502.9802	-2.980234	1.210069

b = consistent under H_0 and H_a ; obtained from xtreg
 B = inconsistent under H_a , efficient under H_0 ; obtained from xtreg
 Test: H_0 : difference in coefficients not systematic

chi2(1) = (b-B)' [(V_b-V_B)^(-1)] (b-B) = 6.07
 Prob>chi2 = 0.0138

Data: Wage Premium.dta
 Do-File: Wage Premium.do

Chapter V: A Real Data Example: Marriage and Happiness

Section: Preparing Panel Data

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



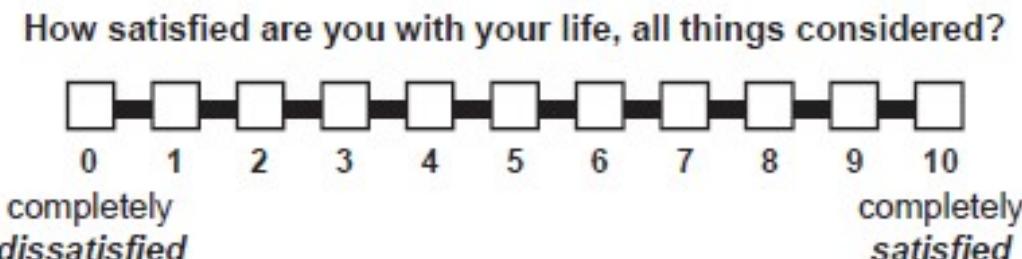
Example: Does Marriage Make Happy?

- In the following we will use a real data example
- The goal is to estimate the causal effect of (first) marriage on happiness
 - More exactly: life satisfaction (or: subjective well-being)
- Data: SOEP 1984-2009 (v26)
 - The data set contains repeated measures for the same persons on the following variables:
 - Life satisfaction, marriage, years married, household income, age, sex, year of interview

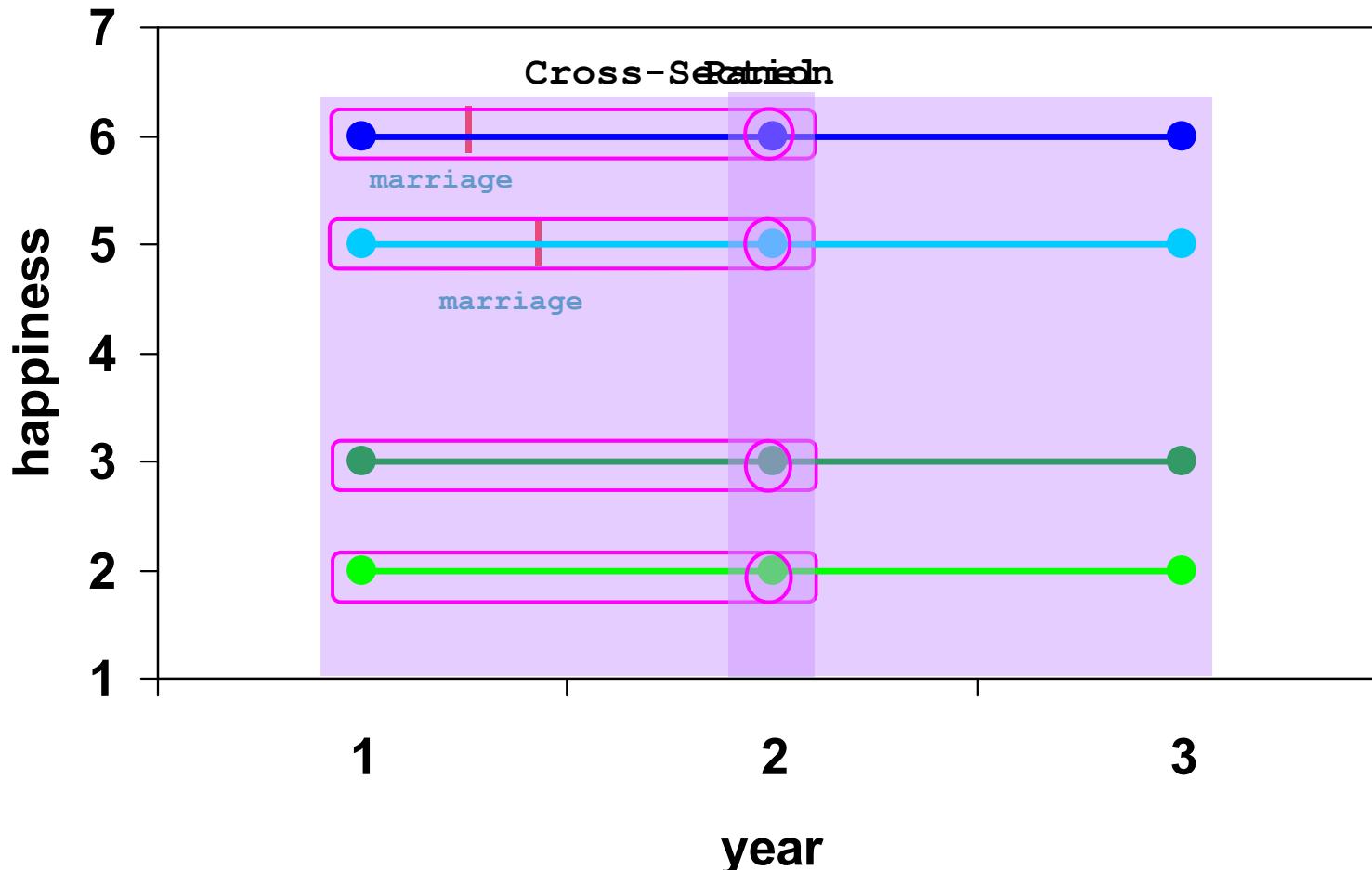
155. In conclusion, we would like to ask you about your satisfaction with your life in general.

 Please answer according to the following scale:

0 means "completely dissatisfied", 10 means "completely satisfied".



The Problem again is Self-Selection:
Happy People are More Likely to Get Married
See also: Stutzer/Frey (2005)



Dilbert also is aware of the problem

Preparing the Data for Panel Analysis

- Retrieving the data: “Happiness 1 Retrieval.do”
 - First “happy” is retrieved from \$P, GPOST, and \$PAGE17
 - The covariates needed are retrieved from \$PEQUIV
- Preparing the data: “Happiness 2 DataPrep.do”
 - Variables are recoded and time-varying covariate “marry” is built
 - The estimation sample is selected
- Analyzing the data: “Happiness 3/4/5 Regressions.do”
 - Data file: “Happiness2.dta”
 - The lecture-package includes an anonymized version of these data (50% sample). Therefore, results are very similar, but not identical to the ones reported in the lecture.

marry:	age:	loghhinc:
marriage dummy	age in years	natural logarithm of annual post-government household income
yrsmarried:	woman:	
years since marriage	dummy for women	

Important Issue: Defining the Estimation Sample

- How to define the estimation sample?
 - Practically very important, but (almost) nothing can be found in literature
- **One should include only those, who potentially can change from the state of not-treated to treated (those who are “at risk”) (Sobel 2012)**
 - Only those, who experience treatment during the observation period provide within information and identify the treatment effect
 - Include also the never-treated persons as a control group
- The already-treated might bias the estimation of the treatment effect
 - The already-treated might improve the precision of the estimated age effect. But if the treatment effect varies over time, then the age effect of the already-treated might be distorted (i.e., old people are in the control, young people in the treatment group)

Defining the Estimation Sample

- For our example, we restrict the estimation sample accordingly
 - Only persons are included, who were single when first observed (persons married when first observed, are discarded altogether!)
 - Person-years after marital separation are excluded
 - Persons with only one person-year are excluded
 - This is not efficient (Chowdhury 1991), so it would be better to leave them in the estimation sample!
- Deleting so many observations differs markedly from what one is used from cross-sectional data analysis!
 - We end up with only 29% of all available person-years

Time-Ordering of Events

- Causality requires that the cause precedes the effect
 - Panel data help to identify the time-ordering of treatment and outcome
 - This has to be taken into regard when preparing the data
- Example: binary treatment (absorbing)
 - An event happens between wave $t - 1$ and t

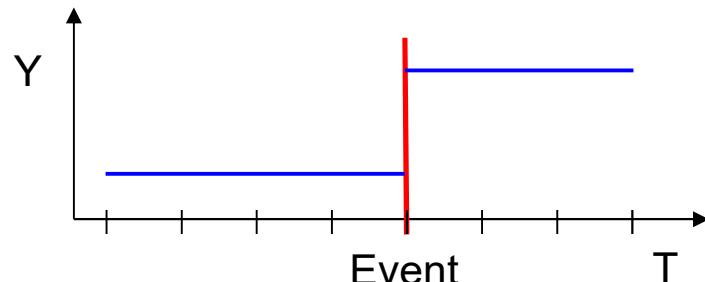
$$x_{i1}, \dots, x_{it-1} = 0$$

$$x_{it}, \dots, x_{iT} = 1$$

- Outcomes have to be measured accordingly
 - y_{i1}, \dots, y_{it-1} measured before event
 - y_{it}, \dots, y_{iT} measured after event
- Happiness example
 - All variables measured at time of interview. Therefore, no problem
 - $m_{it-1} = 0, m_{it} = 1$ if there was a marriage between $t - 1$ and t
 - y_{it-1} is measured then before, and y_{it} after marriage

How to Model a Causal Effect?

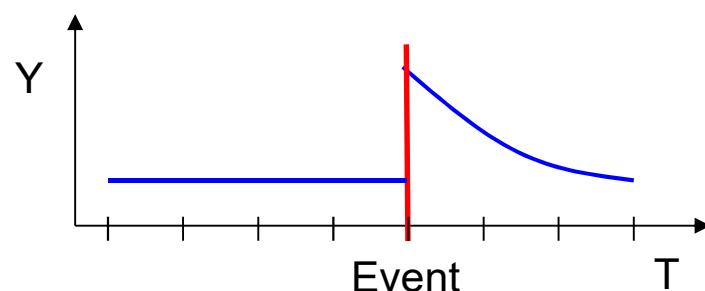
- With panel data we can investigate the time path of a causal effect
 - Termed “impact function” (IF) by Andreß et al. (2013)
 - Different impact functions can be modeled



Step impact function

Immediate and permanent impact

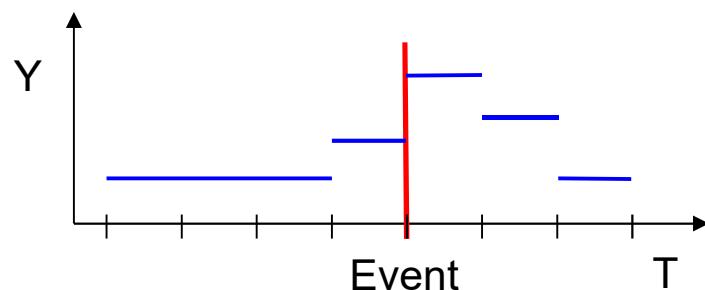
- event dummy $(0,0,0,0,1,1,1)$



Continuous impact function

Immediate, but transitory impact

- event dummy $(0,0,0,0,1,1,1)$
- linear event time $(0,0,0,0,0,1,2)$
- quadratic event time $(0,0,0,0,0,1,4)$



Dummy impact function

Arbitrary impact (including anticipation effect)

- dummy event time
 - 1 dummy $(0,0,0,1,0,0,0)$
 - 0 dummy $(0,0,0,0,1,0,0)$
 - 1 dummy $(0,0,0,0,0,1,0)$
 - 2 dummy $(0,0,0,0,0,0,1)$

Presenting Regression Results Graphically

- Instead of regression tables with many numbers, it is helpful to present the important results graphically (Bauer, 2015)
- Basically there are three types of graphs available
 - Plotting average marginal effects (AME) (**effect plot**)
 - How, does a one unit change in X affect Y? (marginal effect for continuous variables, discrete change for categorical variables)
 - Computed for each observation in the data given their respective values on other variables. Then averaged over all observations.
 - For linear models these are simply the regression coefficients
 - Plotting the predicted values (**profile plot**)
 - What are the predicted values for Y given the values of X?
 - Computed for each observation in the data and then averaged (predictive margins)
 - Plotting AMEs of X conditional on values of Z (**conditional effect plot**)
 - How changes the AME of X over the values of Z?
 - Helpful for interaction effects and for impact functions

Chapter V: A Real Data Example: Marriage and Happiness

Section: Describing Panel Data

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Describing Panel Data I

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

```
. xtset id year  
        panel variable:  id (unbalanced)  
        time variable:  year, 1984 to 2009, but with gaps  
                  delta:  1 unit
```

```

. xtdes, pattern(20)

      id: 103, 202, ..., 8276802
      year: 1984, 1985, ..., 2009

      n = 14634
      T = 26

Distribution of T_i:   min      5%     25%     50%     75%     95%     max
                           2        2        4        7       11       22       26

      Freq.  Percent    Cum. |  Pattern
-----+-----
      808      5.52    5.52 |  .....1111111111
      551      3.77    9.29 |  .....1111
      390      2.67   11.95 |  .....111
      348      2.38   14.33 |  11111111111111111111111111111111
      347      2.37   16.70 |  .....11111111
      337      2.30   19.00 |  .....11
      302      2.06   21.07 |  .....11111
      271      1.85   22.92 |  .....111111
      250      1.71   24.63 |  .....11111111111111111111111111111111
      235      1.61   29.61 |  11.....11111111111111111111111111111111
      222      1.52   31.13 |  111.....11111111111111111111111111111111
      141      0.96   36.96 |  1111.....11111111111111111111111111111111
      8981     61.37 100.00 |  (other patterns)
-----+-----
      14634    100.00 |  XXXXXXXXXXXXXXXXXXXXXXXXX

```

Describing Panel Data II

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

		marry (t+1)		Total
marry (t)	0	1		
0	79,209	3,793	83,002	
	95.43	4.57	100.00	
1	0	24,283	24,283	
	0.00	100.00	100.00	
Total	79,209	28,076	107,285	

Is there enough within variation?

		Mean	Std. Dev.	Min	Max	Observations
marry	overall	.230284	.4210163	0	1	N = 121919
	between		.2694525	0	.9615385	n = 14634
	within		.2671083	-.7312544	1.191823	T-bar = 8.33121
age	overall	29.27678	11.12508	16	97	N = 121919
	between		10.47358	16.5	96.5	n = 14634
	within		4.310836	13.90178	45.19345	T-bar = 8.33121
loghhinc	overall	10.21187	.6562057	0	14.16139	N = 121919
	between		.5664285	6.034016	13.2728	n = 14634
	within		.4252418	2.134135	12.9504	T-bar = 8.33121
woman	overall	.467179	.4989237	0	1	N = 121919
	between		.4992656	0	1	n = 14634
	within		0	.467179	.467179	T-bar = 8.33121

Chapter V:

Section: The Results

- Panel-robust S.E.s
- Step impact function
- Specification issues
- Hausman test
- Continuous impact function
- Dummy impact function

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Panel-Robust S.E.s

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

	FE	FE	FE
S.E.	conventional	panel-robust	bootstrap
marry	0.1668	0.1668	0.1668
	0.0168	0.0226	0.0189
	9.9503	7.3692	8.8254
age	-0.0413	-0.0413	-0.0413
	0.0010	0.0017	0.0017
	-39.3197	-23.9552	-23.6374
loghhinc	0.1245	0.1245	0.1245
	0.0093	0.0123	0.0128
	13.4180	10.1600	9.7087

legend: b/se/t

Conventional S.E.s are too small.

Panel-robust S.E.s are close to the bootstrap S.E.s.

Obviously, with over 14,000 clusters asymptotics works well.

→ In the following we will always use panel-robust S.E.s!

Comparing Results From Step IF

Variable	BE	POLS	RE	FE	FD
marry	0.52***	0.34***	0.20***	0.17***	0.14***
loghhinc	0.50***	0.38***	0.20***	0.12***	0.05**
woman	0.03	0.03	0.03	(omitted)	(omitted)
N	14634	121919	121919	121919	104671
N_clust		14634	14634	14634	14511

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

legend: * p<0.05; ** p<0.01; *** p<0.001

Marriage effect: Heavily biased upwards by BE and POLS
 RE is still biased upwards (median theta is 0.58)
 FD too low due to anticipation (see below)

Income effect: Heavily biased upwards by BE, POLS, and RE; unobservables affecting income and happiness: happy people earn more money

Sex effect: Time-constant variable not estimable with FE and FD

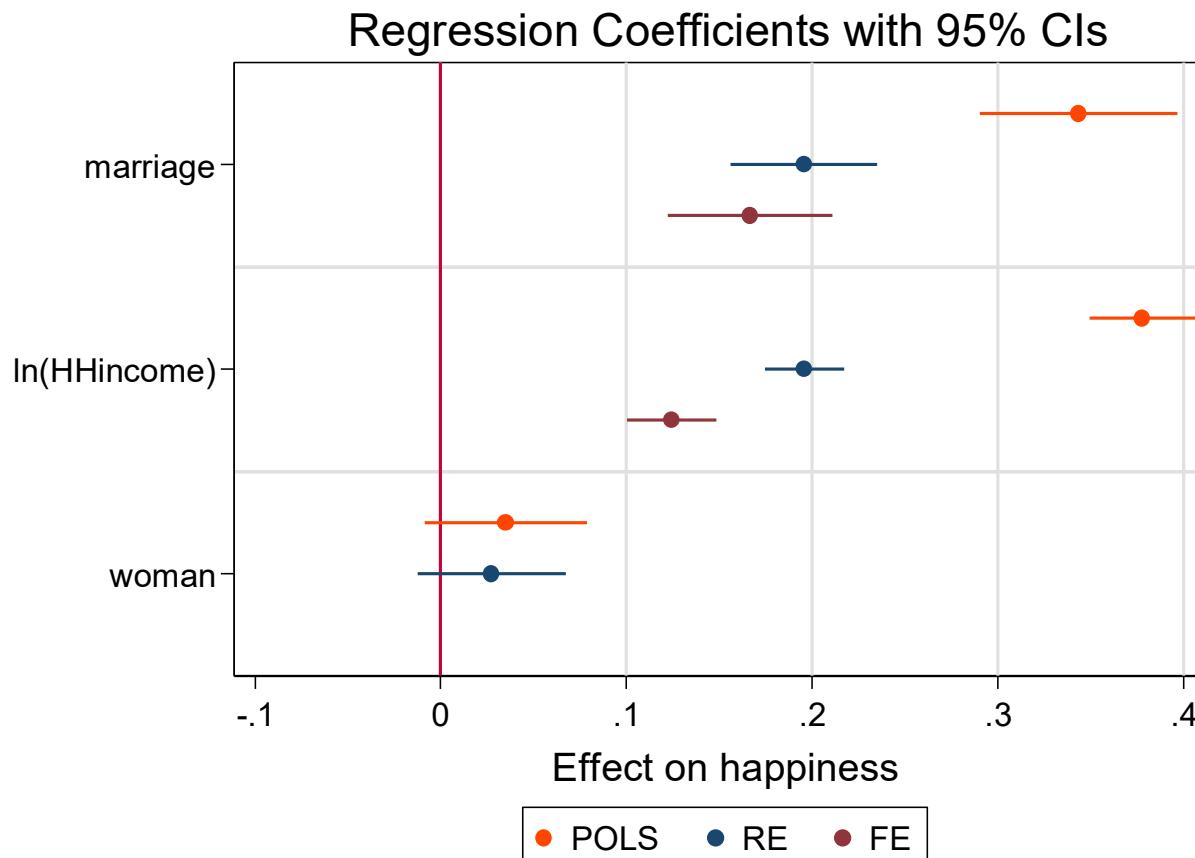
FD: Due to gaps in the data we loose some groups (clusters) and observations.

All models control for age and cohort (see Chapter VI for details)

Comparing Estimation Results Graphically

- Comparing regression coefficients across models is much more effective, if done graphically
 - The “coefplot” package (Jann 2014) is very helpful in this respect

```
ssc install coefplot, replace // Install "coefplot" package (Jann 2014)
coefplot POLS RE FE, keep(marry loghhinc woman) xline(0)
```



Specification Issues: Age Effect, Control Group

FE estimates	(1)	(2)	(3)	(4)
marry	-0.139 0.0139	-0.139 0.0144	0.136 0.0186	0.186 0.0167
age			-0.033 0.0015	-0.039 0.0010
N	49235	121919	49235	121919
N_clust	3793	14634	3793	14634

legend: b/se

- Estimation sample: (1) only those who married, (2) plus control group, no age effect included
 - Marriage effect is not affected, because the control group contributes nothing to the FE estimate
 - The marriage effects in (1) and (2) are obviously heavily biased
- Therefore, (3) and (4) include age as control
 - (3) is without control group: age effect biased upwards, treatment effect is biased downwards

Specification Issues: Optimal Age Specification?

- For simplicity, age is here included linearly
 - This is certainly sub-optimal
 - Below we will show that the age effect is highly non-linear
 - Better: include age dummies (*i.age*)
 - This reduces the marriage effect!

Variable	AGE LINEAR	AGE DUMMIES
marry	0.167***	0.126***
loghhinc	0.125***	0.116***
age	-0.041***	
N	121919	121919
N_clust	14634	14634

legend: * p<0.05; ** p<0.01; *** p<0.001

Deciding Between FE and RE: Hausman Test

```
. xtreg happy marry age loghhinc i.cohort, re
. est store RE
. xtreg happy marry age loghhinc, fe
. est store FE
. hausman FE RE, sigmamore

----- Coefficients -----
|      (b)          (B)          (b-B)      sqrt(diag(V_b-V_B))
|      FE           RE          Difference   S.E.
-----+-----
marry |    .1667519    .1962653    -.0295134    .0068511
age  |   -.0412592   -.040121    -.0011383    .0004667
loghhinc |    .1245142    .1956244   -.0711102    .0039792
-----+-----

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(3) = (b-B)' [ (V_b-V_B)^(-1) ] (b-B)
          =      544.31
Prob>chi2 = 0.0000
```

→ use the FE model

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

Results of a FE Model with Continuous IF

. xtreg happy i.marry c.yrsmarried##c.yrsmarried age loghhinc, fe vce(cluster id)	
Fixed-effects (within) regression	Number of obs = 121919
Group variable: id	Number of groups = 14634
R-sq: within = 0.0162	Obs per group: min = 2
between = 0.0221	avg = 8.3
overall = 0.0152	max = 26
corr(u_i, Xb) = -0.1808	F(5,14633) = 144.05
	Prob > F = 0.0000
	(Std. Err. adjusted for 14634 clusters in id)
-----	-----
	Robust
happy	Coef. Std. Err. t P> t [95% Conf. Interval]
-----+-----	-----+-----
marry 0.2413 0.0231 10.47 0.000 0.1961 0.2865	
yrsmarried -0.0411 0.0065 -6.29 0.000 -0.0539 -0.0283	
c.yrsmarried#c.yrsmarried 0.0017 0.0004 4.71 0.000 0.0010 0.0024	
-----+-----	-----+-----
age -0.0372 0.0021 -17.33 0.000 -0.0415 -0.0330	
loghhinc 0.1306 0.0124 10.57 0.000 0.1064 0.1549	
_cons 6.8787 0.1360 50.57 0.000 6.6121 7.1453	
-----+-----	-----+-----
sigma_u 1.2861966	
sigma_e 1.3325279	
rho .48231322 (fraction of variance due to u_i)	

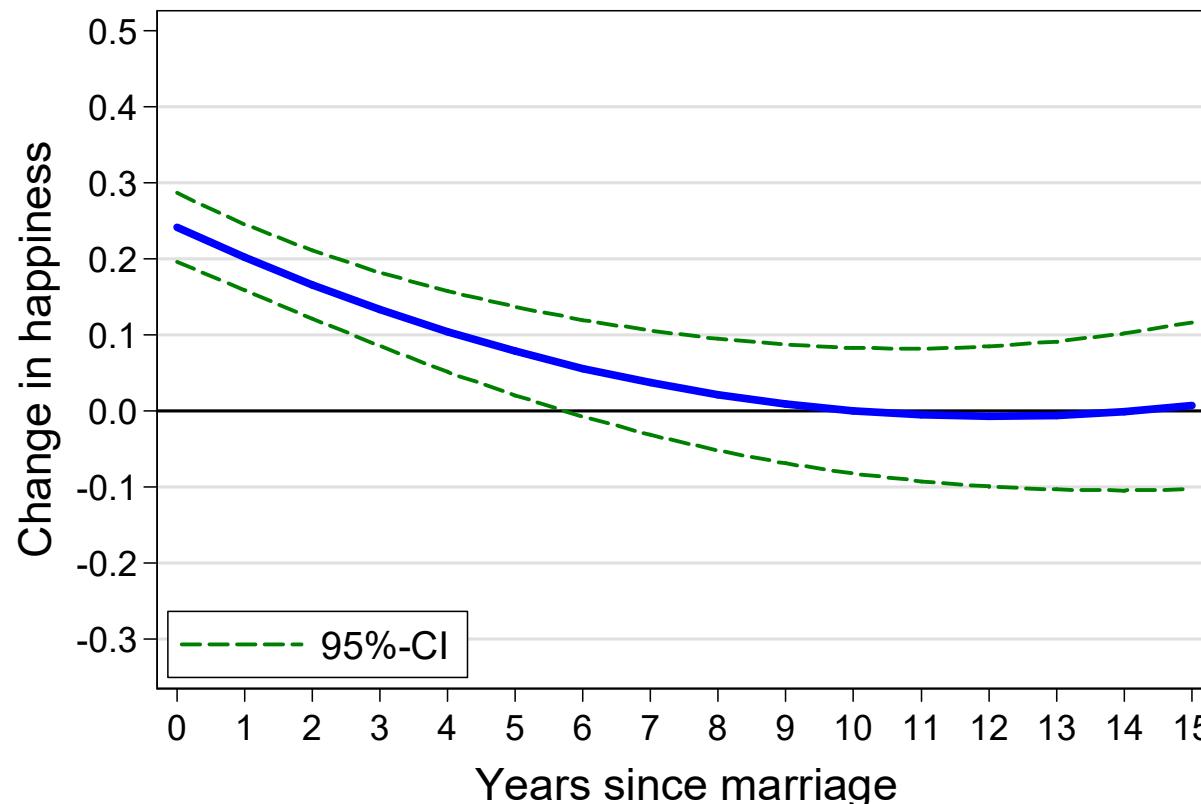
Data: Happiness2.dta
 Do-File: Happiness 3 Regressions.do

Impact Function of Marriage (Conditional Effect Plot)

- Of central interest: time path of the marginal marriage effect
 - Change in happiness due to a marriage (m) over yrsmarried (ym)

$$E(y_{it}) = \beta_1 m_{it} + \beta_2(m_{it} * ym_{it}) + \beta_3(m_{it} * ym_{it}^2) + x_{it}\delta$$

$$\frac{\partial E(y_{it})}{\partial m_{it}} = \beta_1 + \beta_2 ym_{it} + \beta_3 ym_{it}^2$$

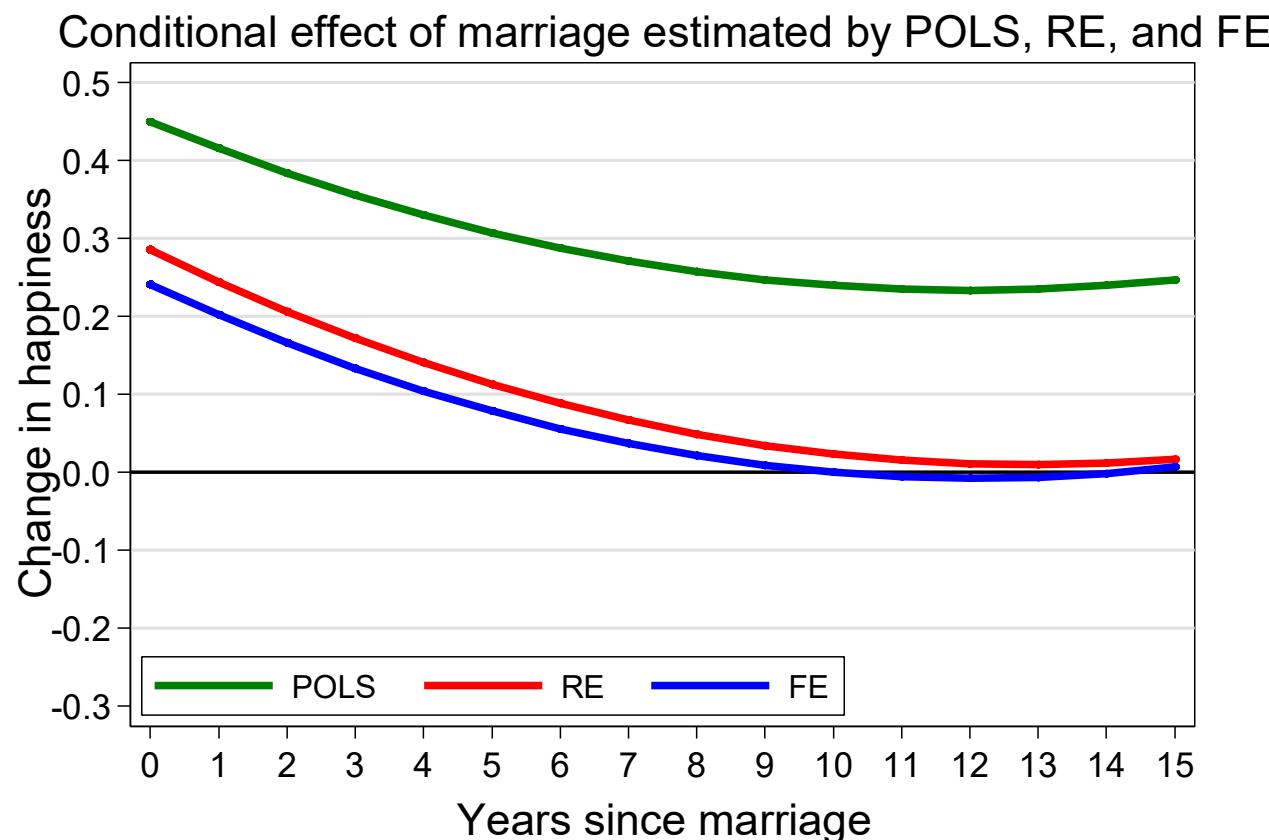


What is the reference point?

- Average happiness of all pyrs before marriage
- Important point: only of those, who eventually marry. Not of the always singles
- After all this is a within estimator!

Comparing Models

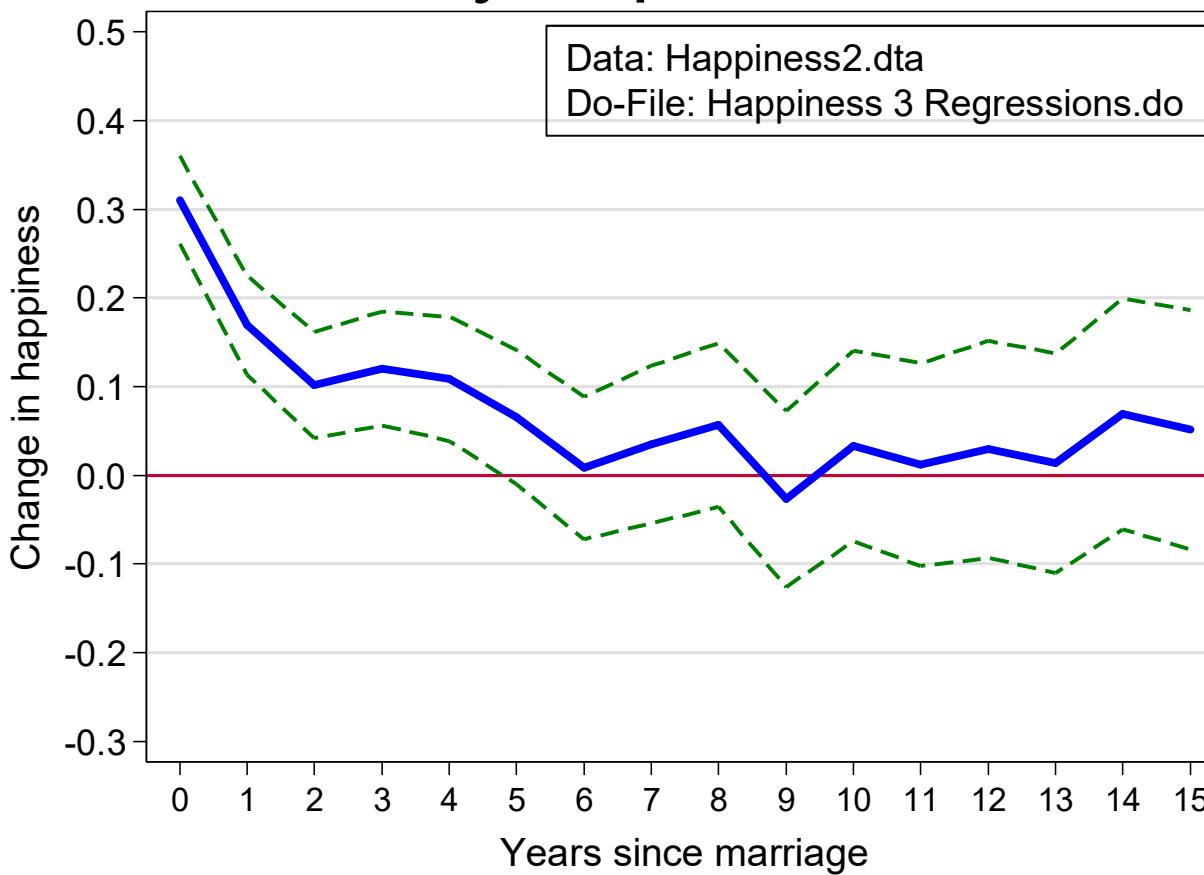
- Comparing the conditional marriage effect over models
 - We see that POLS heavily over-estimates the marriage effect
 - Due to self-selection
 - RE slightly over-estimates the marriage effect
 - RE works quite well with these data (probably due to the long panels)



Dummy Impact Function/Event Studies

- A flexible way to model the causal effect is by event time dummies
 - For this, we have to construct an “event centered” time scale ($_{ym}$)
 - 1 all years before marriage (ref. group)
 - 0 the year of marriage
 - 1 first year after marriage
 - ...
 - 15 15th+ year after marriage
 - Be careful in 0 year: event must have happened before outcome is measured!
 - We collapse the dummies 15 - max due to low case numbers
- The event time dummies are easily included in a regression model via factor notation ($i \cdot _{ym}$) [0-dummy, 1-dummy, ..., 15-dummy]
- Interpretation: the within estimator compares average happiness in a particular year with average happiness in all (!) years before marriage

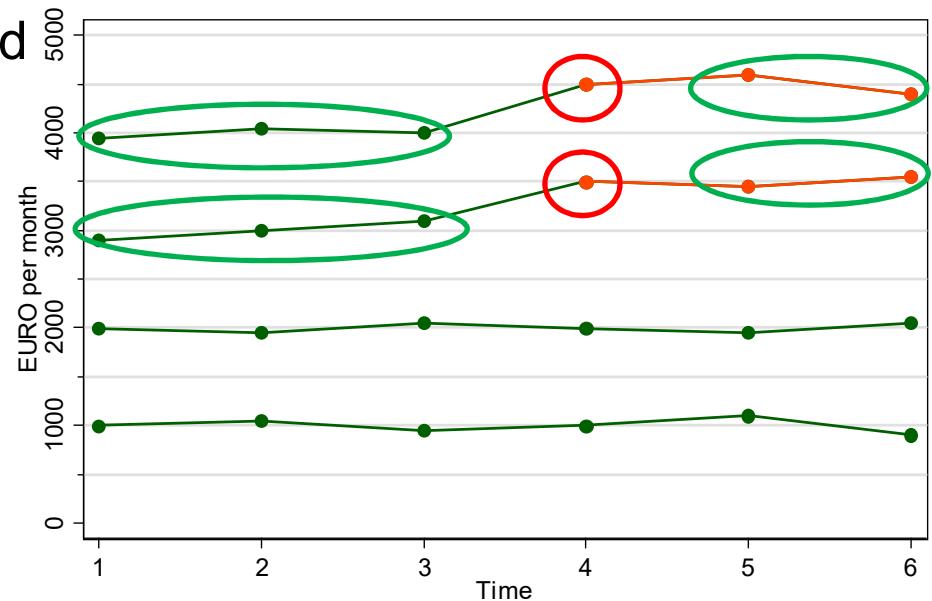
Dummy Impact Function



- The dummy modelling in general supports the results from the parametric modelling above
 - However, we see more details: Compared to the years before marriage
 - Happiness increases by 0.31 in the year of marriage
 - In the first year after marriage, happiness is only higher by 0.17
 - Beginning with the fifth year, happiness is no longer significantly higher

How to Model a Causal Effect?

- Above I argued that one should use
 - Either a step event dummy: 0 before, 1 after the event
 - Or a full set of event time dummies
- Some argue, to use the 0-dummy only
 - The intuitive idea is that this variable captures the change event
 - However, this modeling strategy would make only sense, if the effect is immediate and very short-lived
 - Example: With our fabricated data the within estimator compares red circled minus green circled points
 - The result is +300 €
 - Obviously, this is biased, because the causal effect persists and is therefore in the reference group



Glücksfaktoren

Zehn Glücksbringer und Glückshemmnisse

	Durchschnittlicher positiver Effekt auf die subjektive Lebenszufriedenheit	Durchschnittlicher negativer Effekt auf die subjektive Lebenszufriedenheit	
1	Sehr gute Gesundheit (gegenüber zufriedenstellend)	+0,70	Krankheit (schlechte Gesundheit gegenüber zufriedenstellend) -1,40
2	Ehe/Partnerschaft (gegenüber Singles)	+0,40/0,30	Tod des Partners (verwitwet/ohne Partner gegenüber verheirateten Personen) -0,60
3	Treffen mit Freunden (wöchentlich gegenüber selten oder nie)	+0,25	Arbeitslosigkeit (arbeitssuchend) -0,45
4	Regelmäßiger Sport (wöchentlich gegenüber nie)	+0,10	Soziale und kulturelle Isolation (wenig Kontakt zu Freunden und wenig kulturelle Aktivität) -0,30
5	Eigenheim (gegenüber Situation ohne selbst genutzte Immobilie)	+0,10	Scheidung (geschieden und nicht liiert gegenüber verheiratet) -0,20
6	Autonomie am Arbeitsplatz (Führungsverantwortung oder Selbstständigkeit)	+0,05	Alter (zwischen 40 und 50 Jahre gegenüber 20 bis 30 Jahre) -0,10
7	Gehaltserhöhung (plus 15 Prozent)	+0,05	Behinderung (bei ansonsten guter Gesundheit) -0,05
8	Freizeitaktivität (mindestens einmal pro Monat Kino, Disko oder Ähnliches)	+0,05	Kaufkraftverlust (allgemeiner Preisanstieg um 10 Prozent) -0,05
9	Klassische Kultur (mindestens einmal pro Monat Konzert, Theater oder Oper)	+0,05	Relativer Einkommensverlust (alle anderen Einkommen steigen um fünf Prozent) -0,05
10	Religiosität (mindestens einmal pro Monat Kirchgang bzw. Besuch religiöser Veranstaltungen)	+0,05	Pendeln (15 Kilometer und mehr bis zum Arbeitsplatz) -0,05

Die dargestellten Effekte zeigen, wie stark sich die Lebenszufriedenheit im Zuge verschiedener Ereignisse und Lebensumstände im Mittel verändert. So vermindert Arbeitslosigkeit im Schnitt die subjektive Lebenszufriedenheit um 0,45 Punkte. 0 = ganz und gar unzufrieden, 10 = ganz und gar zufrieden.

Quelle: Eigene Berechnungen auf Basis des SOEP (1992-2010). Anmerkungen: Fixed-Effects-Modell. Sämtliche Effekte sind hochgradig signifikant. Die Effektstärken wurden auf Schritte von 0,05 gerundet.

- FE models
- SOEP, 1992-2010
- The most important factor is good/bad health
- Life-events (marriage, widowhood, unemployment, divorce) are next
 - Here modeled as permanent effects
- Social contacts/isolation comes next
- Age is here mis-specified (see next chapter)
- Money has relatively small effects
 - „Money does not make happy“



Source: Deutsche Post,
Glücksatlas 2012

Chapter V

Section: Interpreting Results from Panel Regressions

- Interpreting panel regression estimates
- Interpreting results from impact functions
- Interpreting effects of continuous variables
- Interpreting interaction effects
- Effects of consecutive life course events

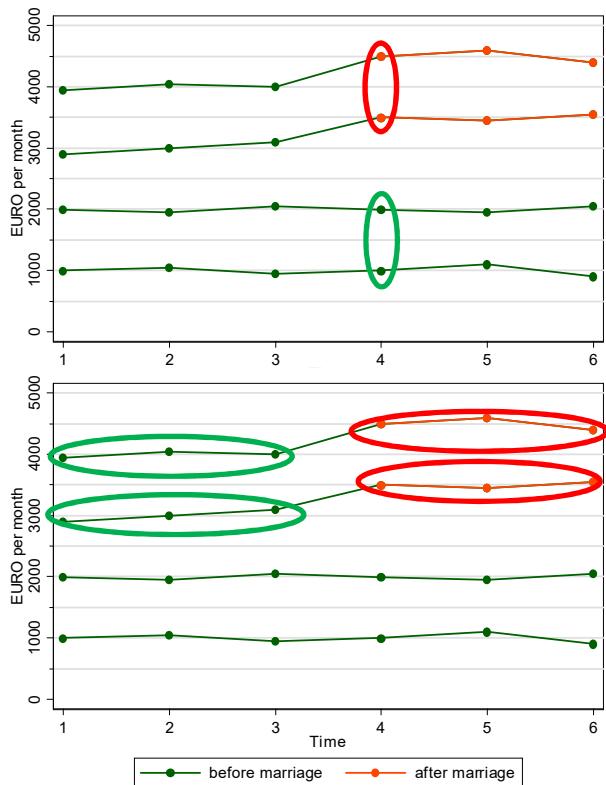
Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Interpreting Regression Estimates

- Regression estimates can be interpreted in two ways
 - I) Descriptive interpretation
 - “People who differ in X by one unit, differ in Y by $\hat{\beta}$ ”
 - II) Causal interpretation
 - “A one unit change in X, causes a change in Y by $\hat{\beta}$ ”
 - Sometimes called “change interpretation”
- The descriptive interpretation is always ok
- The causal interpretation, however, is only warranted, if the exogeneity assumptions are met
 - Unit homogeneity in an cross-sectional regression
 - Parallel trends in a FE regression

Interpreting Panel Regression Estimates



Cross-sectional regression:

Married men earn 2500 € more than unmarried men
(descriptive interpretation)

A marriage causes an increase in men's wage by 2500 €
(causal interpretation)

Fixed-effects regression:

After marriage men earn 500 € more than before marriage
(descriptive interpretation)

A marriage causes an increase in men's wage by 500 €
(causal interpretation)

The descriptive interpretation is always correct.

The causal interpretation is only correct, if the exogeneity assumptions hold.

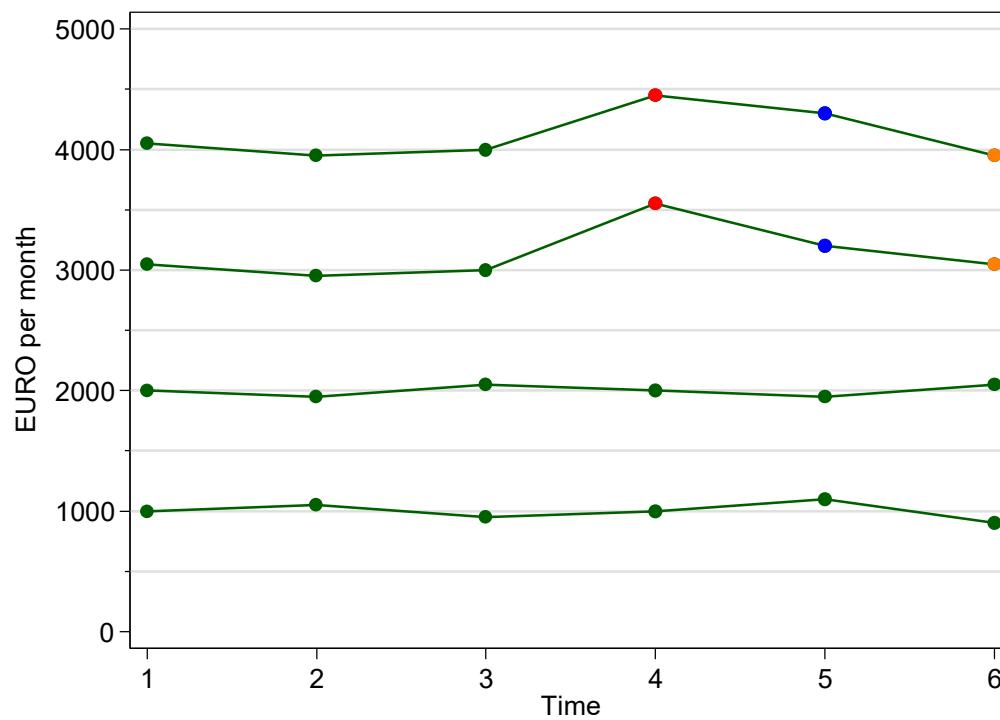
Therefore, researchers who do not believe in causal analysis with non-experimental data (i.e., who believe that exogeneity assumptions generally fail) will stick to the descriptive interpretation.

Confusion About Interpretation

- There is some confusion in the literature on how to interpret results from panel regressions. Some authors provide confused arguments
 - Andreß et al. (2013) argue that
 - I) is appropriate for POLS and FE (modeling the level) and that
 - II) only works for FD (modeling the change)
 - However, all models can be interpreted in levels (descriptive) or change (causal). POLS compares levels between persons. Both FD and FE (!) compare levels within a person over time. The change interpretation for all models works only if assumptions are met.
 - Giesselmann/Windzio (2012) term
 - I) “cross-sectional questions”
 - II) “longitudinal questions”
 - However, both descriptive and causal questions can be answered with either cross-sectional or longitudinal data. But the assumptions needed to get correct answers differ.

Interpreting Results from Impact Functions

- With event time dummies, the within estimator compares the outcome in a particular year with the outcome in the reference years
 - For each unit separately (for sure, only for the treated)
 - The FE estimator is the average of these unit-specific estimates

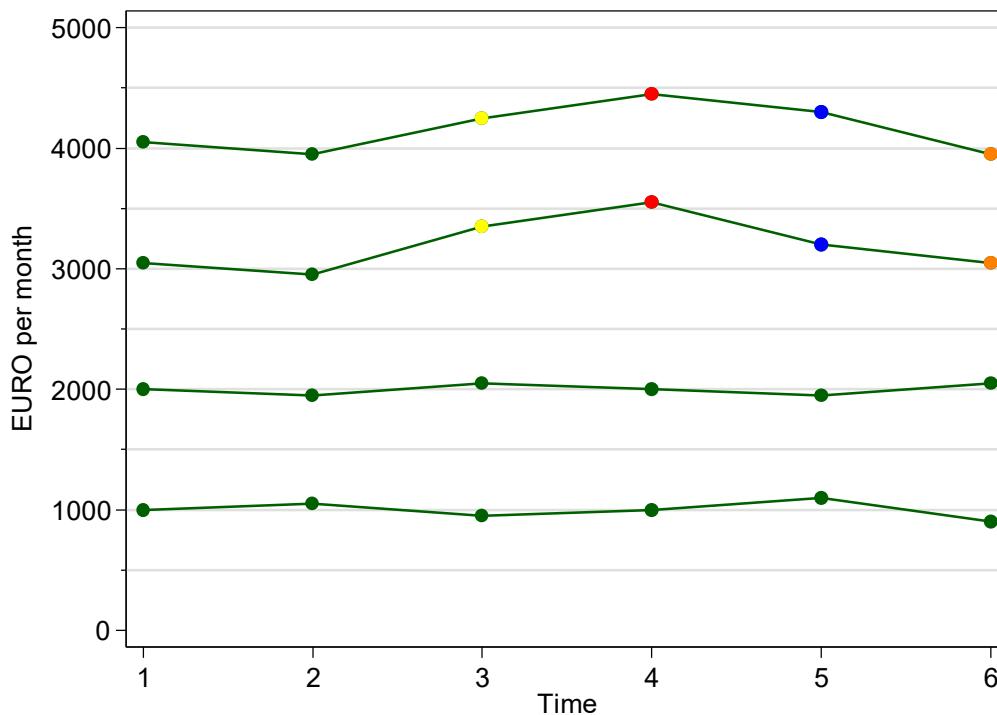


Data: Impact Function.dta
Do-File: Impact Function.do

- Example
 - 0.ym 500
 - 1.ym 250
 - 2.ym 0
 - The effect of 0.ym is the average wage of the red points minus the average wage of the green points (only the treated)
 - The effect of 1.ym is ... blue points ...
 - The effect of 2.ym is ... orange points ...
 - The non-treated contribute nothing to these estimates!

Anticipation Effects

- Anticipation effects generate a problem
 - Often people expect already some time that an event will happen
 - If this expectation affects the outcome, we observe a causal effect already before the event happens (anticipation effect)
 - Because the anticipation effect is part of the reference group, the FE impact function will be biased downwards
 - Remedy: include negative event time dummies that capture the anticipation effect (distributed FE model)



- Example with anticipation effect of +300 EUR at t=-1 (yellow points)
 - Now the results are biased

Event Time	Impact
0 . ym	400
1 . ym	150
2 . ym	-100
 - Remedy: we add a -1-dummy

Event Time	Impact
-1 . ym	300
0 . ym	500
1 . ym	250
2 . ym	0

Anticipation Effects: Marriage Example

- Marriage probably produces anticipation effects
 - To test for these, we expand the distributed FE model up to $t = -1$



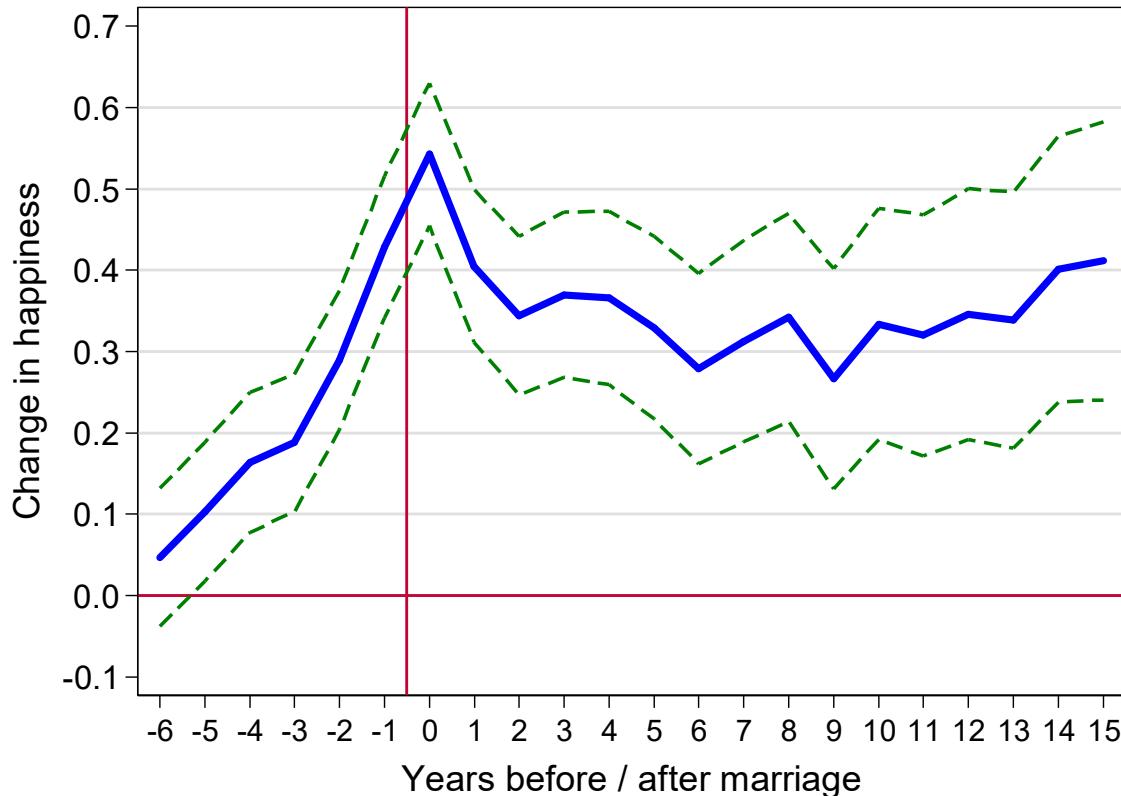
We observe a strong anticipation effect

- Reference group is up to 2 years before marriage
- Already in the year before marriage we observe a significant happiness increase by .27
- Now the marriage effects are somewhat stronger. Right after marriage it increases from .31 to .38
- And the marriage effect is longer significant

– Thus our conclusion changes somewhat

Anticipation Effects: Be Careful!

- But be careful with anticipation effects
 - They easily might capture “non-causal” effects
 - To demonstrate this, we enlarge the distributed FE up to $t = -6$



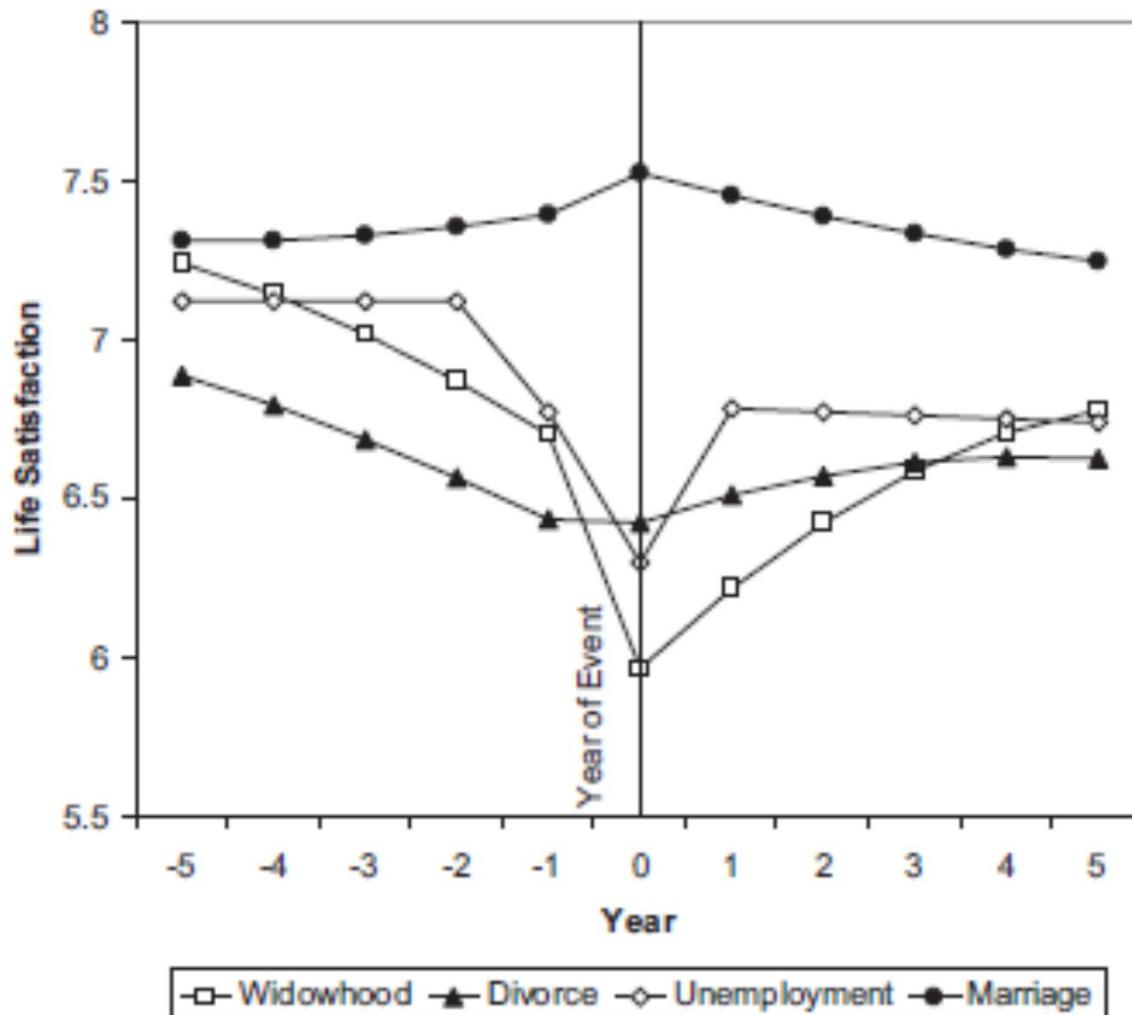
It seems, as if there are extremely strong anticipation (and marriage) effects!
But this is nonsense, and due to the fact that we have omitted an important variable (being in a partnership).

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

Anticipation Effects: Be Careful!

- Generally, anticipation effects might capture:
 - Omitted variables
 - Selection on growth (steeper growth in the treatment group)
 - Reverse causality (the outcome changes before treatment)
- Thus it is difficult to discern, whether the negative time dummies capture a “true” anticipation effect or some “non-causal” effects.
 - Include negative time dummies only if you have strong theoretical reasons for anticipation effects!
- Otherwise negative time dummies can be used as a diagnostic instrument!
 - If there is no theoretical reason for a “true” anticipation effect, then they hint towards severe problems with strict exogeneity.
 - How to deal with these problems, we will discuss in detail later (Chap. X).

The Effects of Life-Events on Happiness



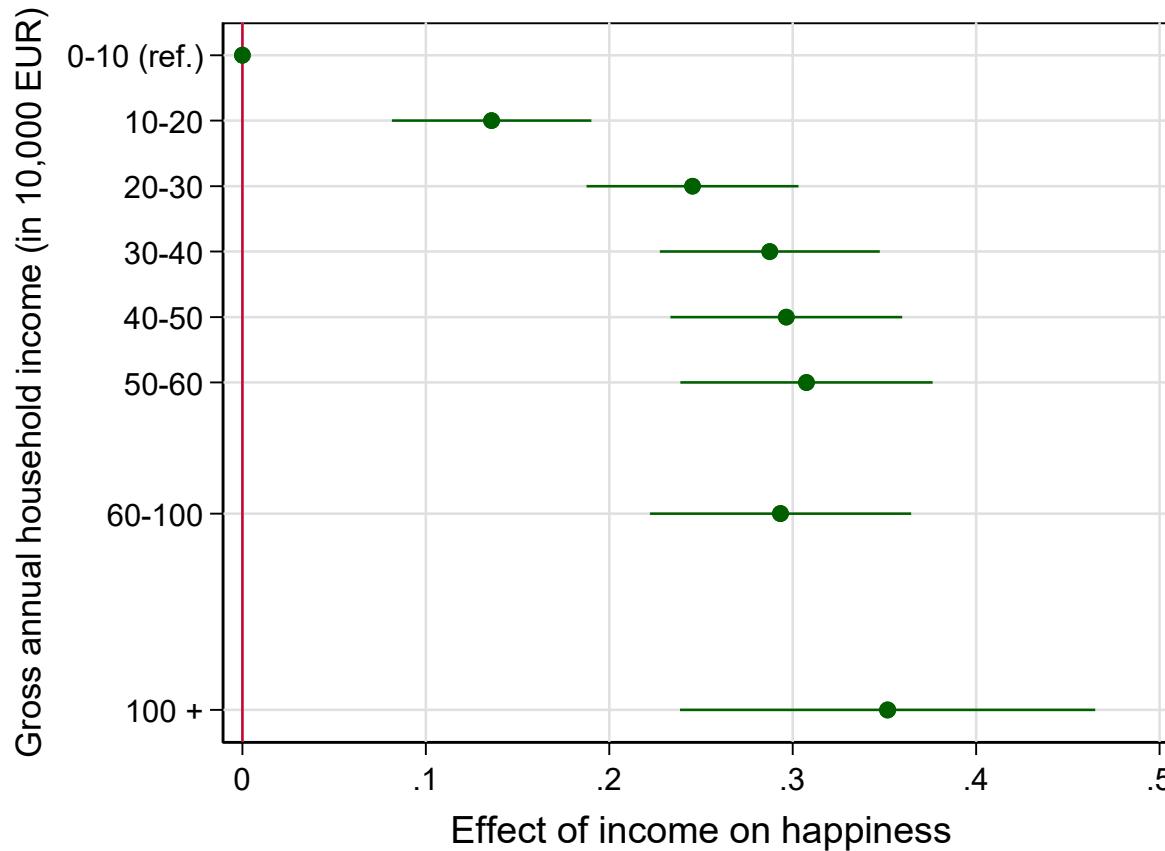
- With this methodology one can investigate the impact of all kinds of life-events
- Diener et al. is a classic in this respect. They use SOEP data and this graph summarizes their results (however, they use RE-modelling!)
- Their result on marriage differs from ours!?
- Unemployment has a sudden and long-lasting negative effect
- Widowhood shows anticipation (health problems of the partner) and a long-lasting negative effect
- Divorce shows anticipation (match deteriorates) but no effect on happiness
 - There is negative selection into divorce

Diener, E. et al. (2006) Beyond the Hedonic Treadmill.
American Psychologist 61 (4): 305-314.

Interpreting Effects of Continuous Treatments

- Within estimator and continuous treatments?
 - Here we cannot compare to “time before event”
 - Instead, for each individual panel a regression line is fitted. The FE estimate is the average of these individual regression lines
 - Descriptive interpretation
 - The outcome difference, with varying treatment
 - Causal interpretation
 - The outcome change, with varying treatment
- Example: income and happiness
 - We now use income in absolute terms (in 10,000 EUR)
 - The FE result is +0.024. The causal interpretation is that a 10,000 EUR increase in hhincome would increase happiness by .024
 - Obviously a very small effect
 - However, this is biased because of the linear modelling. In truth the income effect is highly non-linear
 - This is shown on the next slide by using a categorical income variable

Example: Income and Happiness



- There is a satisfaction threshold at about 40,000 EUR
- This result explains the “Easterlin paradox”
 - Despite GDP increases in the last decades, average happiness in Western Countries did not increase

Interpreting Interaction Effects

- Effect heterogeneity on a group level can be modeled by allowing for an interaction treatment \times group
 - Whereas time-constant group effects are not identified in a within estimation, interactions are identified
 - The result tells us, whether the treatment (X) effect differs between groups ($G: \{0,1\}$). Note that the main effect of G is not identified!
 - Standard interaction parametrization (adding a product term)

$$y_{it} = \beta x_{it} + \delta(G_i \times x_{it}) + \alpha_i + \varepsilon_{it}$$

β (the treatment effect of $G = 0$)

δ (the difference in the treatment effect of $G = 1$)

- Stata code: `i.group##c.treatment`

- Often it is helpful to model interactions as “nested effects”

$$y_{it} = \beta_1(i[G_i = 0] \times x_{it}) + \beta_2(i[G_i = 1] \times x_{it}) + \alpha_i + \varepsilon_{it}$$

where $i[.] = \begin{cases} 1, & \text{if } . = 1 \\ 0, & \text{else} \end{cases}$

$\beta_1 = \beta$ (the treatment effect of $G = 0$)

$\beta_2 = \beta + \delta$ (the treatment effect of $G = 1$)

- Stata code: `i.group#c.treatment`

- “Product term” and “nested effects” parametrization are equivalent

Example: Income Effect by Sex

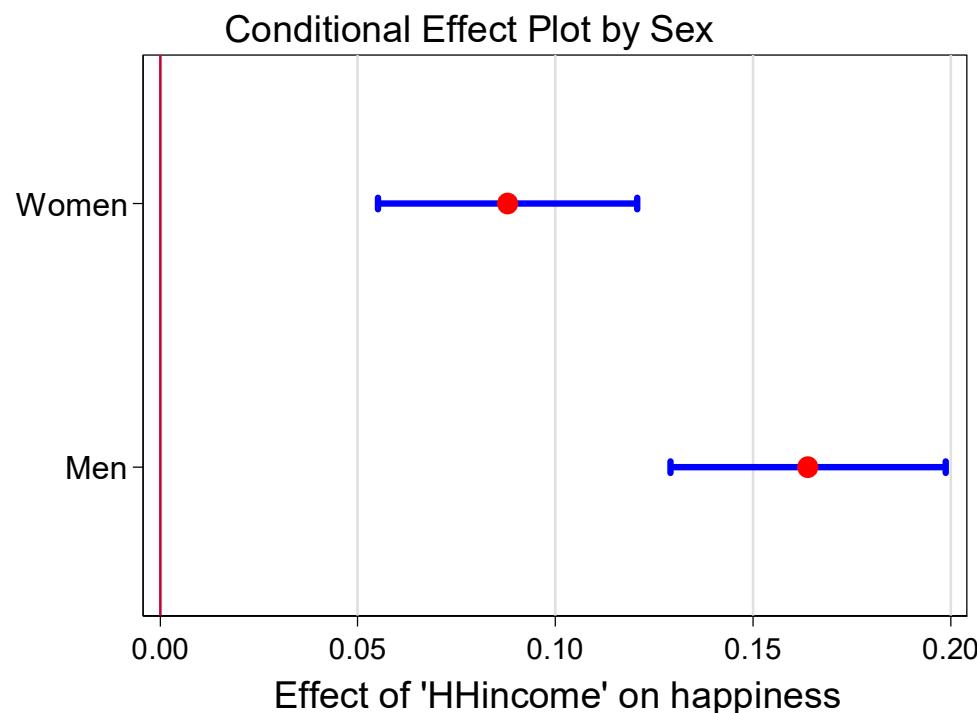
```
. xtreg happy marry i.woman##c.loghhinc age, fe vce(cluster id)
note: 1.woman omitted because of collinearity
```

		Robust				
	happy	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	marry	.1678467	.0226257	7.42	0.000	.1234976 .2121959
	1.woman	0	(omitted)			
	loghhinc	.1639288	.0177442	9.24	0.000	.1291478 .1987097
	woman#c.loghhinc					
	1	-.0759617	.0242279	-3.14	0.002	-.1234514 -.0284721
	age	-.0413842	.0017228	-24.02	0.000	-.0447611 -.0380072
	_cons	7.007378	.1270878	55.14	0.000	6.75827 7.256486

- The standard interaction specification tells us that men have an income effect of +0.164
- Women's income effect is significantly lower by -0.076

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

Example: Income Effect by Sex

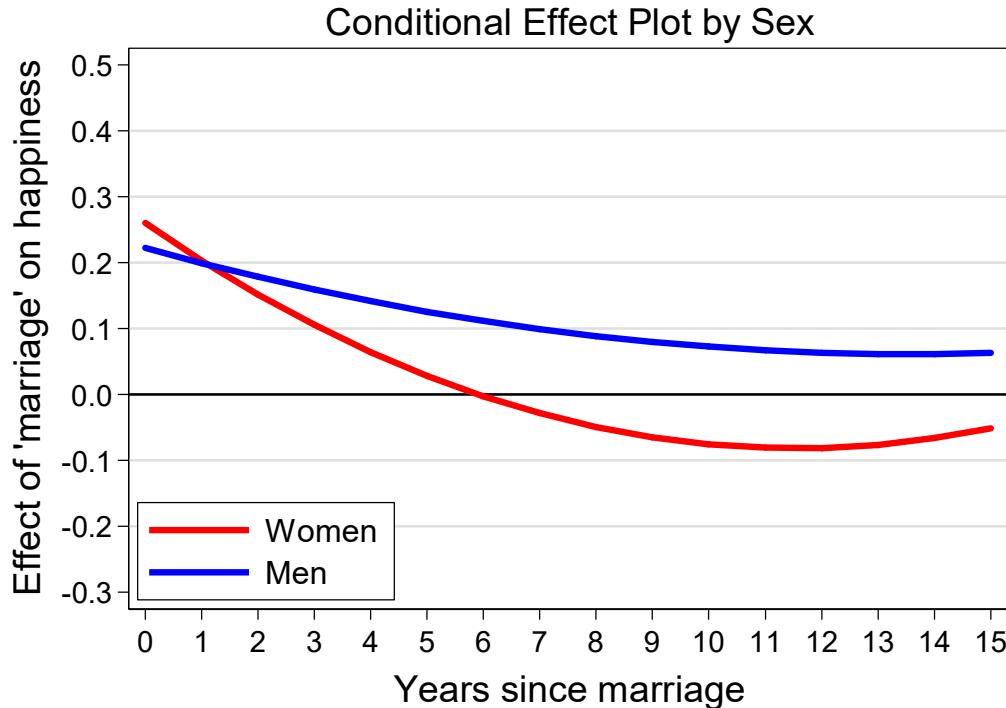


The nested effects specification is often more informative

- Because it gives us both treatment effects directly
- We see that more money makes men much more happier (+0.164) than women (+0.088)!

Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

Example: Marriage Effect by Sex



It is also possible to model interactions of a group variable and an impact function

- Stata code:
i.group#(impact function)
- The “honeymoon effect” is stronger for women
- However, the happiness decline afterwards is faster for them!
- The difference in the marriage effects is significant

```
. test (1.marry          # 0.woman = 1.marry          # 1.woman) ///
>      (c.yrsmarried     # 0.woman = c.yrsmarried    # 1.woman) ///
>      (c.yrsmarried # c.yrsmarried # 0.woman = c.yrsmarried # c.yrsmarried # 1.woman)

( 1) 0b.woman#1.marry - 1.woman#1.marry = 0
( 2) 0b.woman#c.yrsmarried - 1.woman#c.yrsmarried = 0
( 3) 0b.woman#c.yrsmarried#c.yrsmarried - 1.woman#c.yrsmarried#c.yrsmarried = 0

F(  3, 14633) =      2.93
Prob > F =      0.0322
```

Data: Happiness2.dta

Do-File: Happiness 3 Regressions.do

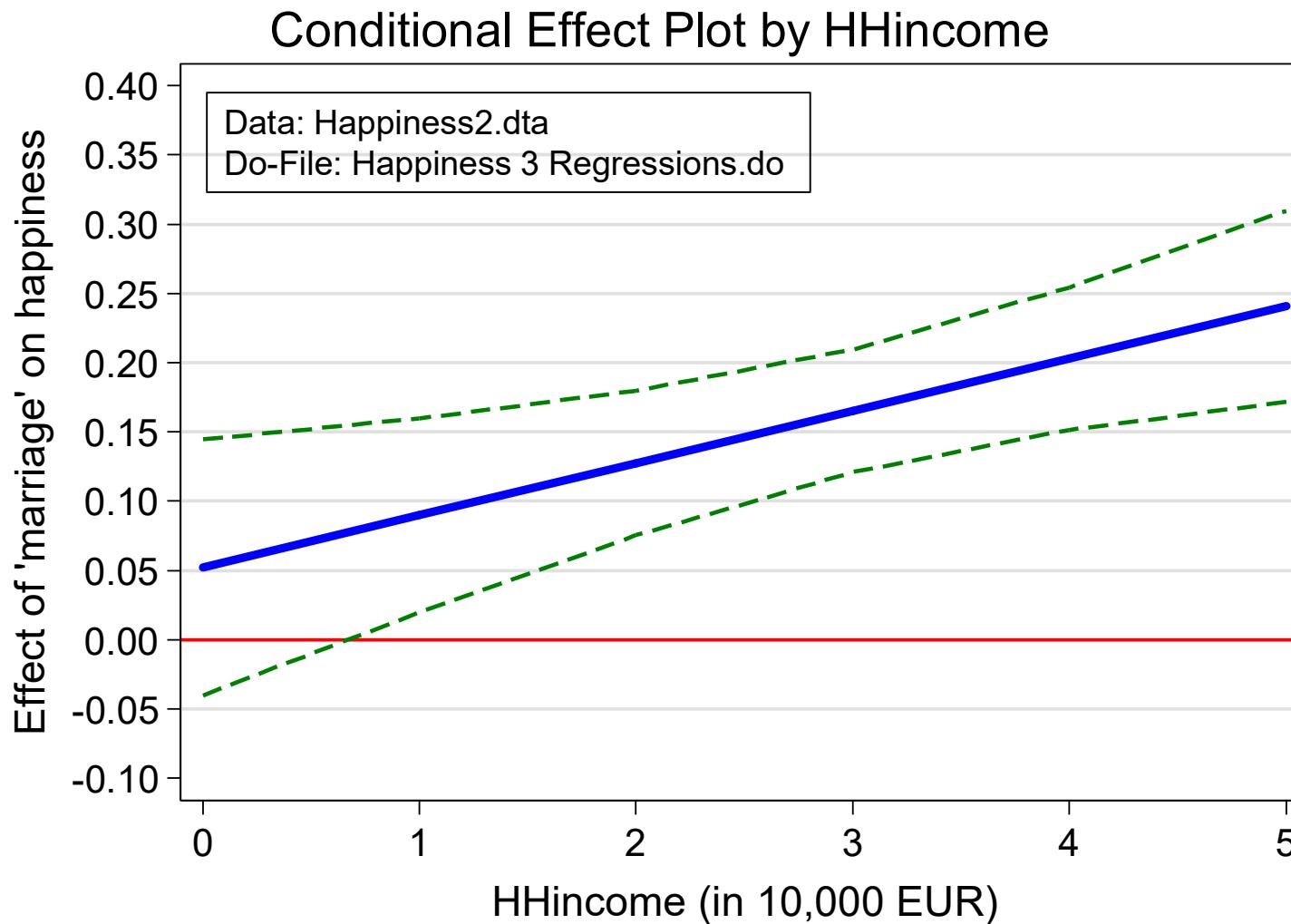
Interaction of Two Time-Varying Variables

- In the same way we can specify the interaction of two time-varying variables
 - Now the interaction can be interpreted symmetrically, because both main effects are in the model
 - Thus, we have to decide which variable is the “treatment”, and which one is the “moderator”
 - Interpretation: The “treatment effect” varies by level of “moderator”
- Example: How varies the marriage effect by HHincome?
 - HHincome measured in absolute terms (in 10,000 EUR and top coded at 50,000 EUR)

```
. xtreg happy i.marry##c.hhinct age, fe vce(cluster id)
```

		Robust				
		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	1.marry	.0522104	.0470784	1.11	0.267	-.0400692 .14449
	hhinct	.0556164	.0068092	8.17	0.000	.0422696 .0689632
	marry#c.hhinct	.037676	.0136458	2.76	0.006	.0109286 .0644234

Example: Marriage Effect by HHincome



Understanding Within-Interaction

- Giesselmann/Schmidt-Catran (2018) started a discussion on how to interpret interaction effects of two time-varying variables in FE models
 - They argue that the standard procedure does not provide a „genuine within estimator“
 - Because it is partly based also on between variation (see the decomposition on the next slide)
 - Instead, they suggest first demeaning X and Z and then enter the product term in FE (double-demeaning)
 - This interaction estimator rests only on within variation
- Now, they argue that only the double-demeaned interaction estimator is unbiased if random-effects assumption is violated
 - This, however, is not true (as simulations show, see below): under violation of random-effects assumption all three components of FE-interaction yield unbiased estimates
 - Some authors even tend to use only the between interaction term, because it is more easily interpretable

The Algebra of FE Interaction

- The FE model including an interaction

$$y_{it} = \beta x_{it} + \gamma z_{it} + \delta x_{it}z_{it} + \alpha_i + \varepsilon_{it}$$

- Demeaning yields

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + \gamma(z_{it} - \bar{z}_i) + \delta(x_{it}z_{it} - \overline{(xz)_i})$$

- Expanding the interaction term only

$$\begin{aligned} x_{it}z_{it} - \overline{(xz)_i} &= x_{it}z_{it} - \frac{\sum_{t=1}^{T_i} x_{it}z_{it}}{T_i} = \\ &[\bar{x}_i + (x_{it} - \bar{x}_i)] [\bar{z}_i + (z_{it} - \bar{z}_i)] - \frac{\sum_{t=1}^{T_i} [\bar{x}_i + (x_{it} - \bar{x}_i)] [\bar{z}_i + (z_{it} - \bar{z}_i)]}{T_i} = \\ &\bar{x}_i\bar{z}_i + \bar{x}_i(z_{it} - \bar{z}_i) + \bar{z}_i(x_{it} - \bar{x}_i) + (x_{it} - \bar{x}_i)(z_{it} - \bar{z}_i) \\ &- \frac{\sum_{t=1}^{T_i} \bar{x}_i\bar{z}_i + \bar{x}_i(z_{it} - \bar{z}_i) + \bar{z}_i(x_{it} - \bar{x}_i) + (x_{it} - \bar{x}_i)(z_{it} - \bar{z}_i)}{T_i} = \\ &\bar{x}_i\bar{z}_i + \bar{x}_i(z_{it} - \bar{z}_i) + \bar{z}_i(x_{it} - \bar{x}_i) + (x_{it} - \bar{x}_i)(z_{it} - \bar{z}_i) - \bar{x}_i\bar{z}_i - \bar{x}_i \frac{\sum_{t=1}^{T_i} (z_{it} - \bar{z}_i)}{T_i} \\ &- \bar{z}_i \frac{\sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)}{T_i} - \frac{\sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)(z_{it} - \bar{z}_i)}{T_i} = \\ &\bar{x}_i(z_{it} - \bar{z}_i) + \bar{z}_i(x_{it} - \bar{x}_i) + (x_{it} - \bar{x}_i)(z_{it} - \bar{z}_i) - \frac{\sum_{t=1}^{T_i} (x_{it} - \bar{x}_i)(z_{it} - \bar{z}_i)}{T_i} \end{aligned}$$

The „Meaning“ of FE Interaction

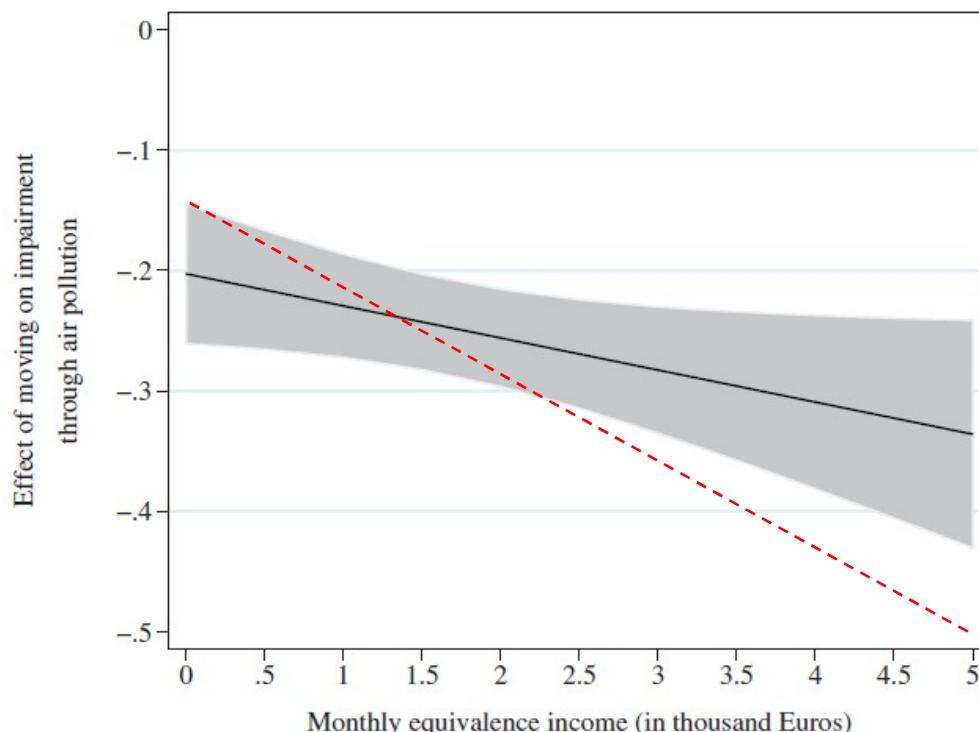
- Thus there are three terms that go into FE interaction
 - $mZ * dmX + mX * dmZ + [dmZ * dmX - m(dmZ * dmX)]$
 - m: person-specific mean of a variable; dm: demeaned variable
 - Thus, FE-interaction is based on
 - a) between variation (terms 1 and 2)
 - b) within variation (term 3)
 - Simulation: $\delta = -1$
 - Random-effects assumption is violated in this scenario
 - Nevertheless, FE-interaction is unbiased
 - Further, all three terms yield the correct (though inefficient) estimator
 - Therefore, one should use the standard FE interaction estimator (in this scenario)

Simulation results			
Variable		FE	factored FE
X		1.003053	1.003092
		0.0042	0.0042
Z		1.001401	1.001405
		0.0042	0.0042
XZ		-1.000044	
		0.0024	
mZdmX			-0.995511
			0.0040
mXdmZ			-1.006347
			0.0040
dmZdmX			-0.997639
			0.0047

legend: b/se

FE Interaction: Using Only the Between Term

- Alternative Strategy
 - Some authors (e.g. Schober/Stahl, ESR 2016) use $\bar{z}_i x_{it}$ instead of $z_{it} x_{it}$
 - This estimates the interaction based solely on between variation (term 1)
- Best/Rüttenauer (2018) estimate the effect of moving on perceived impairment by air pollution
 - Data: SOEP 1986, 1994, 1999, 2004, 2009, 2014; N=12,037
 - Outcome: perceived impairment by air pollution (scale 1 to 5)
 - Treatment: relocation of the household between waves (1. move)
 - Moderator: HH equivalence income in thousand



- The authors also present the result for “mean HHincome”
 - The interaction effect triples (see red line in graph)
 - This might be seen as some kind of robustness check?

Example: Marriage Effect by HHincome

- We repeat the example with the person-specific mean HHincome as moderator ($\bar{hhinc}_i \cdot marr_{it}$)
- TO BE DONE!

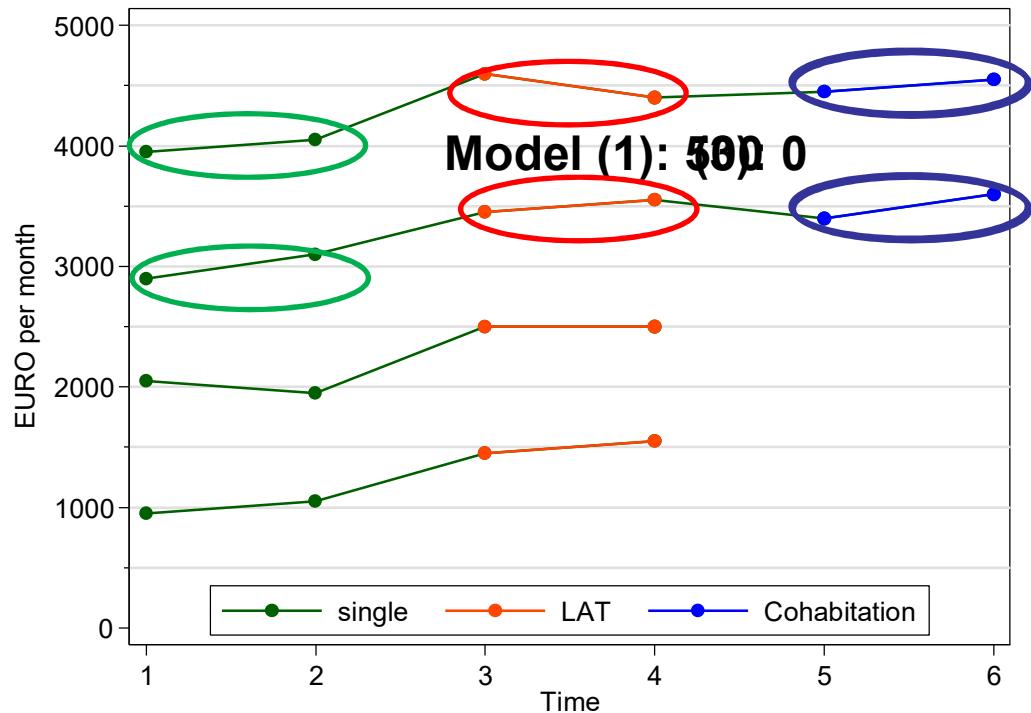
Effects of Consecutive Life Course Events

- Until now we always investigated only one specific transition
- Often researchers analyze several transitions in one analysis (single estimation sample strategy)
 - Effects of different relationship types on happiness/wages
 - Effects of different parity births on wage (motherhood wage penalty)
- When interpreting results from such a single estimation sample strategy one has to be careful
 - Within estimation compares the outcome in the reference state with the outcome in the focus state
 - Only for those, who were observed in both states
 - I.e., the reference group differs for each state
 - This makes it problematic to compare the effects of different states
- Therefore, it is recommended to construct special estimation samples for every single transition of interest
 - **Multiple estimation samples strategy**

Example: Effect of Relationship Type on Wage

- Categorical treatment variable: three relationship types
 - Single, LAT, cohabitation (ref.: single)
 - Those starting a relationship first enter a LAT, eventually later a cohab
 - Single estimation sample strategy
 - LAT-effect: wage difference single-LAT
 - For all those, who enter a relationship
 - Cohab-effect: wage difference single-cohab
 - Only for those, who eventually enter a cohab
 - Problem 1: The cohab-effect is not a causal effect, because there is no observed transition from single to cohab
 - Problem 2: We cannot interpret LAT-effect minus cohab-effect!
 - Because these effect estimates are based on different populations.
 - If the two populations have different LAT-effects (effect heterogeneity) the LAT-cohab difference will be biased
- Multiple estimation samples strategy
 - The LAT-effect is estimated with all single/LAT person-years
 - The cohab-effect is estimated with all LAT/cohab person-years

No Effect Heterogeneity – No Problem

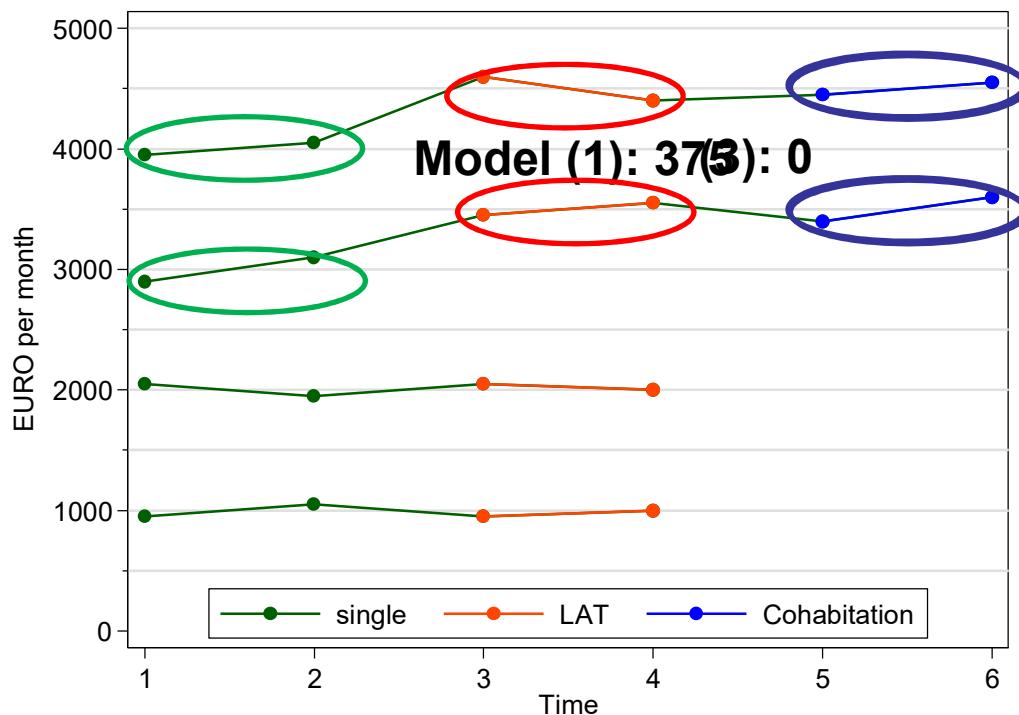


- If there is no effect heterogeneity the single est. sample strategy works
- (1) single est. sample strategy
 - The difference of the LAT/cohab effects is zero
 - The causal effect of changing from LAT to cohabitation is zero
- (2)/(3) multiple est. sample strat.
 - The result in (3) shows that in fact we can interpret the difference in (1)

FE regress		(1) all pyrs	(2) single/LAT	(3) LAT/cohab
single		(ref.)	(ref.)	
LAT		500***	500***	(ref.)
cohabitation		500***		0
pers-yrs		20	16	12
persons		4	4	4

Data: Cat Treatment.dta
Do-File: Cat Treatment.do

Effect Heterogeneity - Problem



- (1) single est. sample strat.
 - It seems, as if the effect of a change from LAT to cohabitation is +125 €
 - However, because people with a low LAT-effect do not enter in a cohabitation, this is biased
- (3) uses the proper estimation sample
 - We see that the effect of a change from LAT to cohabitation is in fact zero!

FE regress (1) all pyrs (2) single/LAT (3) LAT/cohab			
single	(ref.)	(ref.)	
LAT	250**	250**	(ref.)
cohabitation	375**		0
pers-yrs	20	16	12
persons	4	4	4

Data: Cat Treatment.dta
Do-File: Cat Treatment.do

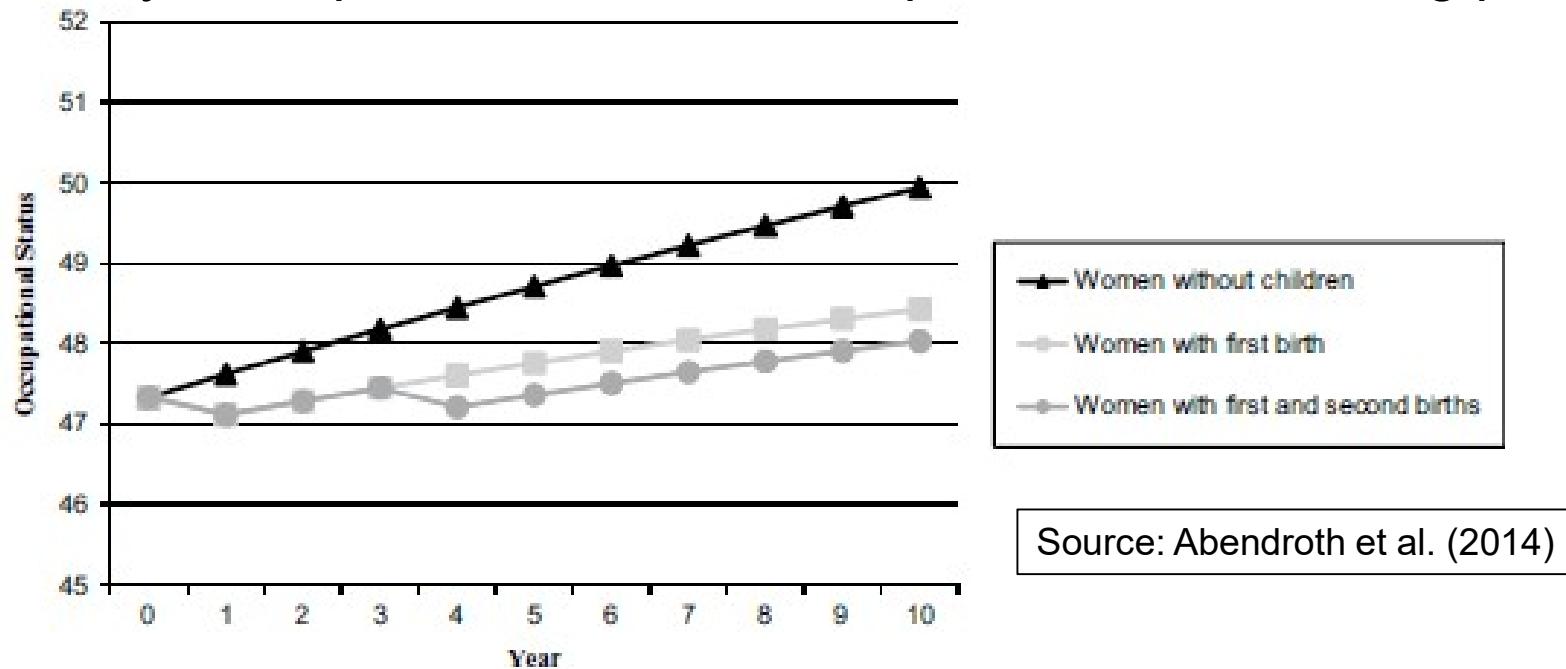
Another Example: Motherhood Status Penalty

- The same problem occurs with research on cumulative (dis)advantage
 - Abendroth et al. (2014) estimate effects of first, second, and third births on women's status trajectories
 - Data: ECHP 1994 to 2001, 13,615 women aged 18 to 40 in 13 countries
 - Outcome: ISEI status score
 - Treatment: parity specific birth dummy and linear event time variable
 - Status trajectory is modeled with linear and quadratic age terms
 - Estimation sample:
All women entering the labor market 1994 – 2001 (some had already children)
 - Estimation strategy:
they use all pyrs and estimate a FE regression with all parities in one model
(single estimation sample)

	OLS	FE
Parity-Specific Birth		
First birth	-.379*	-.394*
	(.181)	(.182)
Second birth	-.509**	-.398*
	(.179)	(.183)
Third birth	-.585*	-.498
	(.304)	(.314)
Time since Birth		
First birth		-.112*
		(.061)
Second birth		-.018
		(.048)
Third birth		.006
		(.057)

Another Example: Motherhood Status Penalty

- For demonstrating cumulative disadvantage and the effect of a 2nd birth they add up the coefficients and produce the following plot



- The impact of the first birth increases over time (cumulative disadvantage)
 - No problem here (only differential attrition perhaps, see below)
- It seems as if the impact of the second birth is much smaller
 - However, this interpretation might be problematic for the reasons given above
 - To compare the effects of the first and second birth, one would have to restrict the estimation sample to those, who eventually have at least two kids

Chapter VI:

Modeling Individual Growth

Section: Growth Curve Models

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

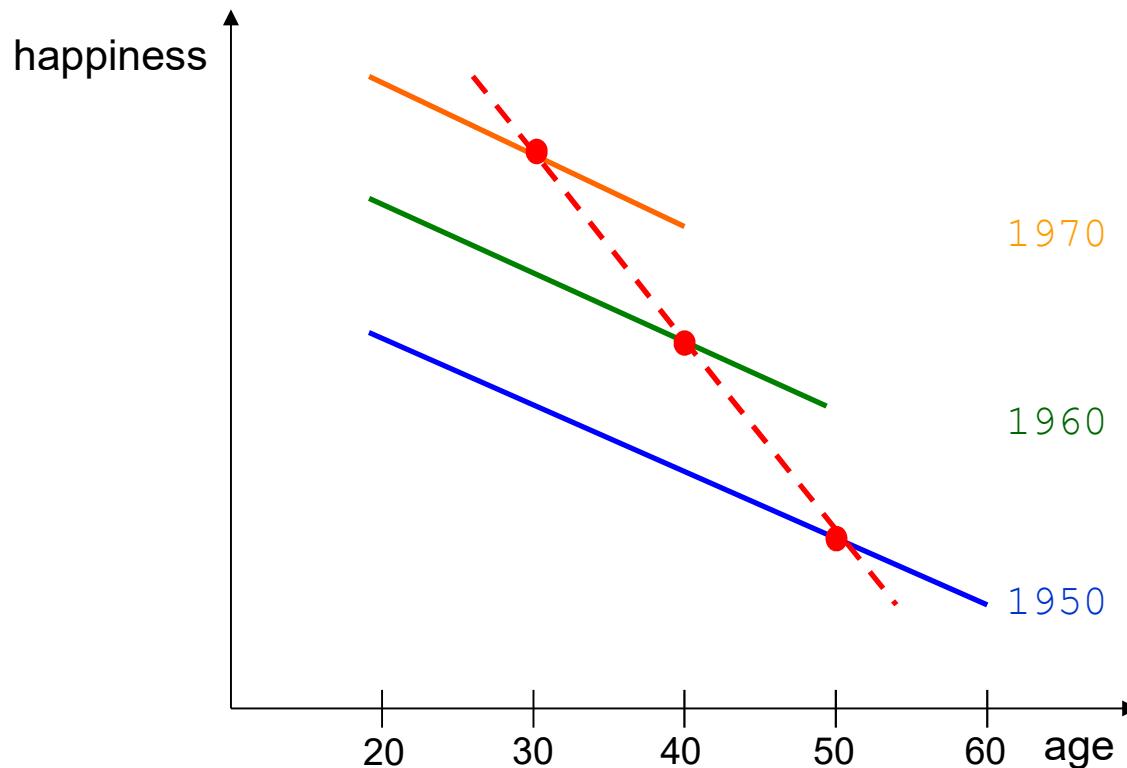


Modeling Individual Growth

- How develops an outcome over age?
 - (Multi-cohort) panel data allow to separate age and cohort effects!
 - Thereby we can model individual growth
 - **This is the second major advantage of panel data**
- Growth curve (GC)
 - Development of the outcome over age (A) net of cohort (C)
$$y_{it} = \beta x_{it} + \gamma A_{it} + \delta C_i + \alpha_i + \varepsilon_{it}$$
 - This is a linear growth curve, but it is easy to model quadratic, cubic, etc. growth curves
 - Most flexible: age (group) dummies, splines (cf. Wunder et al. 2013)
 - Can age have a “causal” effect?
 - In a narrow sense of causality, age can not be seen as a cause, because it is not manipulable
 - However, in a mechanism based approach age is seen as a correlate of intervening mechanisms that potentially can be manipulated

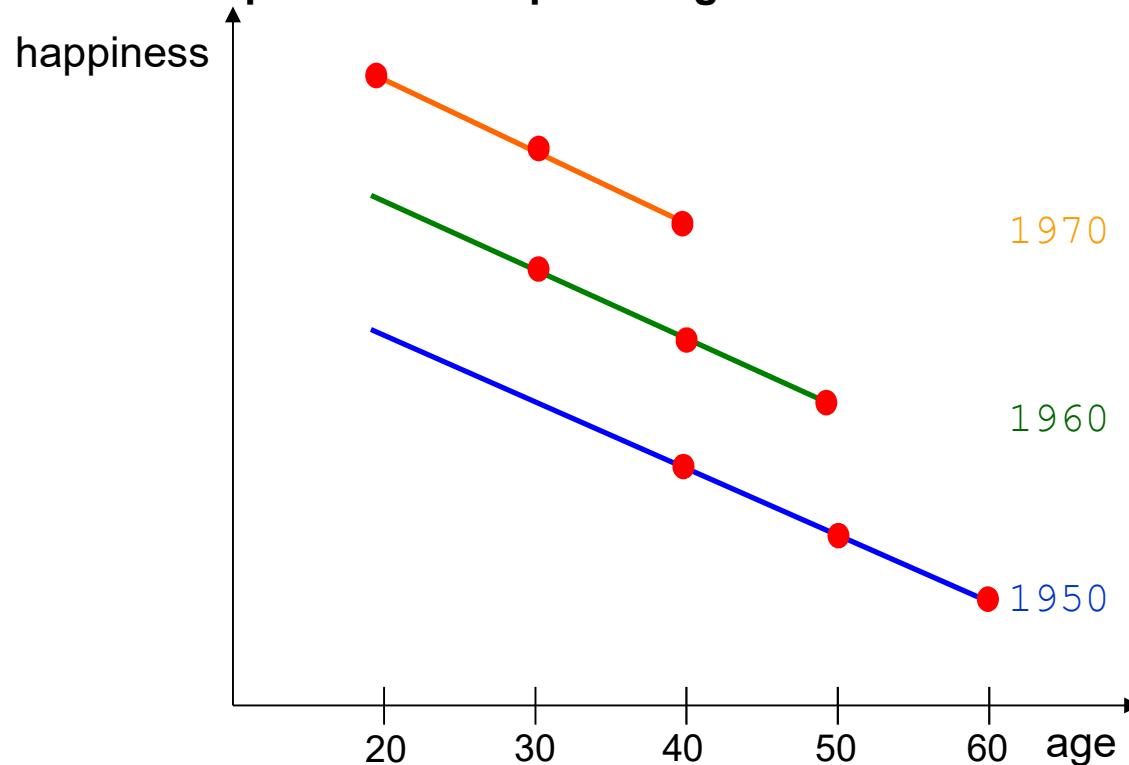
Cross-Sectional Data

- Thought experiment: estimating a growth curve from cross-sectional data
 - We assume that happiness declines linearly with age (age effect)
 - We assume that younger cohorts are happier (cohort effect)
 - There is no period effect
- Cross-sectional survey in the year 2000
 - The estimated growth curve is red dashed: It mixes age and cohort effect
 - **It is impossible to separate age and cohort effect!**



Panel Data

- Thought experiment: estimating a growth curve from panel data
 - We assume that happiness declines linearly with age (age effect)
 - We assume that younger cohorts are happier (cohort effect)
 - There is no period effect
- Panel survey in the years 1990, 2000, and 2010
 - The estimated growth curves (not shown) are identical with the true ones
 - **It is possible to separate age and cohort effect!**



How to Estimate Growth Curves?

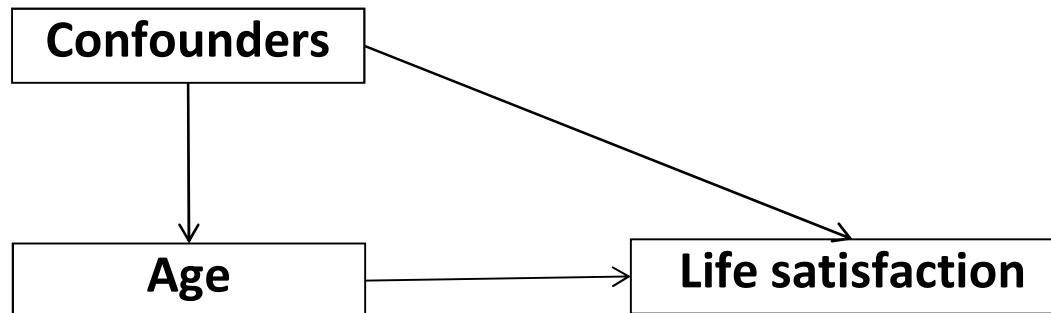
- The naive strategy is to estimate the following equation by POLS:

$$y_{it} = \gamma A_{it} + \delta A_{it}^2 + \alpha_i + \varepsilon_{it}$$

- This, however, is open to several sources of bias
 - Confounder bias
 - Endogenous selection bias
 - Overcontrol bias
- More details can be found in Kratz/Brüderl (2019)
- In the following we will discuss the essentials by using an example:
 - How evolves life satisfaction over age?

Estimation of Growth Curves

- I: Control for confounders



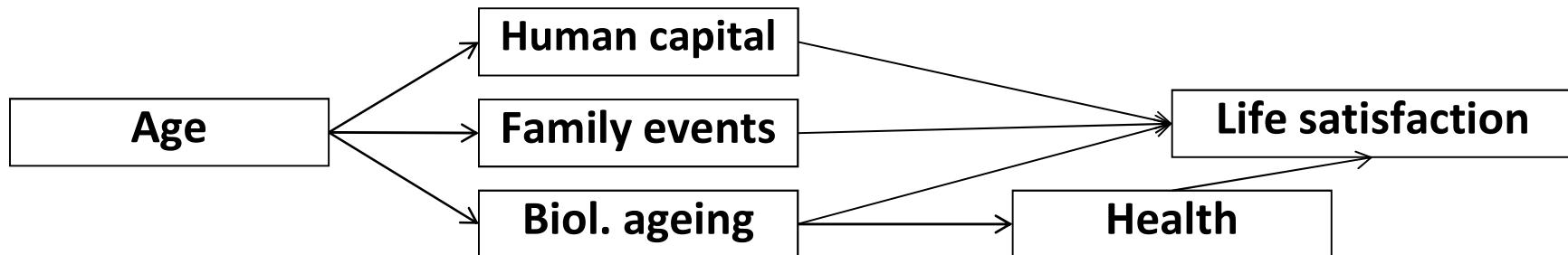
- How can it be that something affects age? After all, age is not manipulable
 - But in a sample some mechanisms might affect the sample composition at different ages (composition effects)
- Observed confounders
 - Cohort effects might distort the growth curve estimates
 - Older cohorts (born shortly before WWII) are happier
 - In the SOEP these are observed in high ages only
 - Period effects, method effects, etc.
 - Composition effects (woman, East German, foreigner, etc.)

Estimation of Growth Curves

- II: Endogenous selection bias
 - Selection might distort the growth curve estimates
 - Due to differential mortality (e.g., happier people survive longer)
 - Due to differential attrition (e.g., happier people attrite less)
 - Selectivity might bias estimates of covariate effects and that might indirectly bias the age effect (Frijters/Beaton 2012)
- Methodological lesson I
 - Control at least for observables
 - Even better: estimate FE growth curves
 - FE analysis provides a within estimate of the age effect
 - How does happiness change, if a respondent grows one year older?
 - Therefore, sample composition/selection does not bias the age effect
 - There is a better chance to have unbiased covariate effects
 - This point has been overseen in most of the literature that estimates GCs by RE (or POLS) methods
 - Many of the GCs reported in the literature are biased

Estimation of Growth Curves

- III: Overcontrol might distort the growth curve estimates
 - Generally: When estimating causal effects, one should control confounders, but not intervening mechanisms
 - Otherwise you will get “overcontrol bias”
 - Overcontrol is the consequence of a “kitchen sink approach” that unfortunately is very common in social research based on regressions



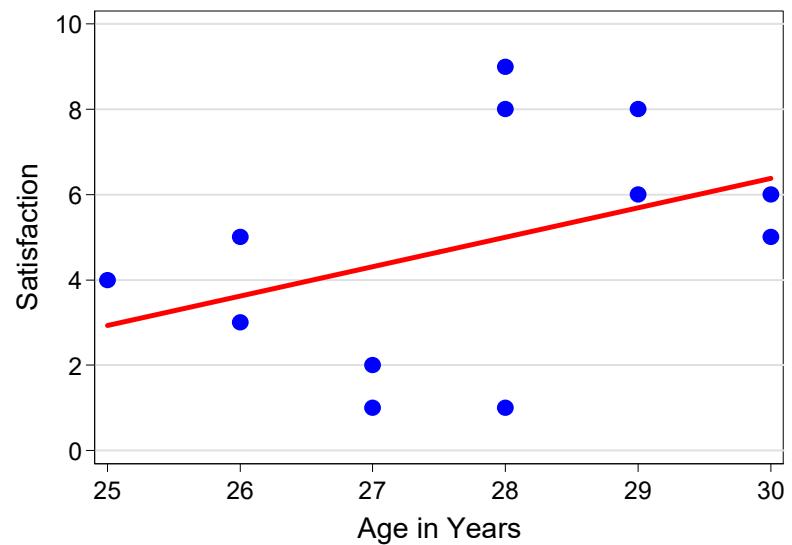
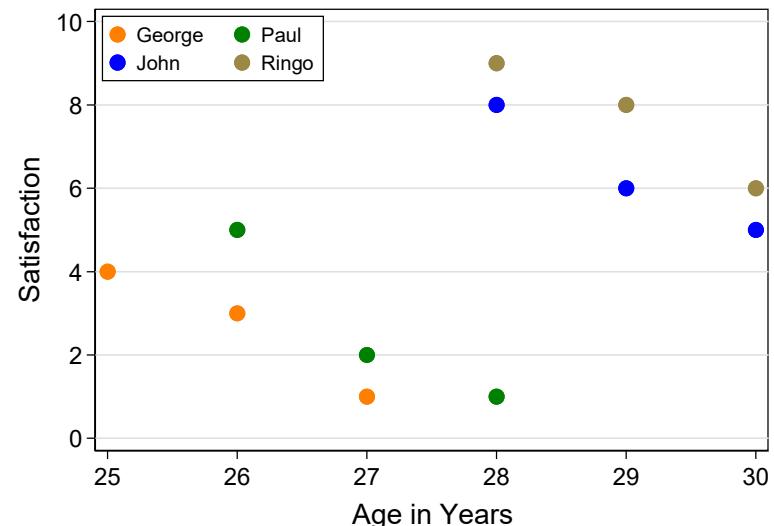
- Methodological lesson II
 - Do not control for intervening mechanisms
 - First, get the gross causal effect of age by controlling for confounders only
 - In a second step, try to get closer on the intervening mechanisms by adding them to the model

Example: Life satisfaction of Four Englishmen

- An extreme example on how misleading POLS GCs can be
- How does satisfaction develop with age?
 - It seems, as if one grows happier with age
 - $\hat{\gamma}_{POLS} = +0.69$

Data: Beatles.dta
Do-File: Beatles.do

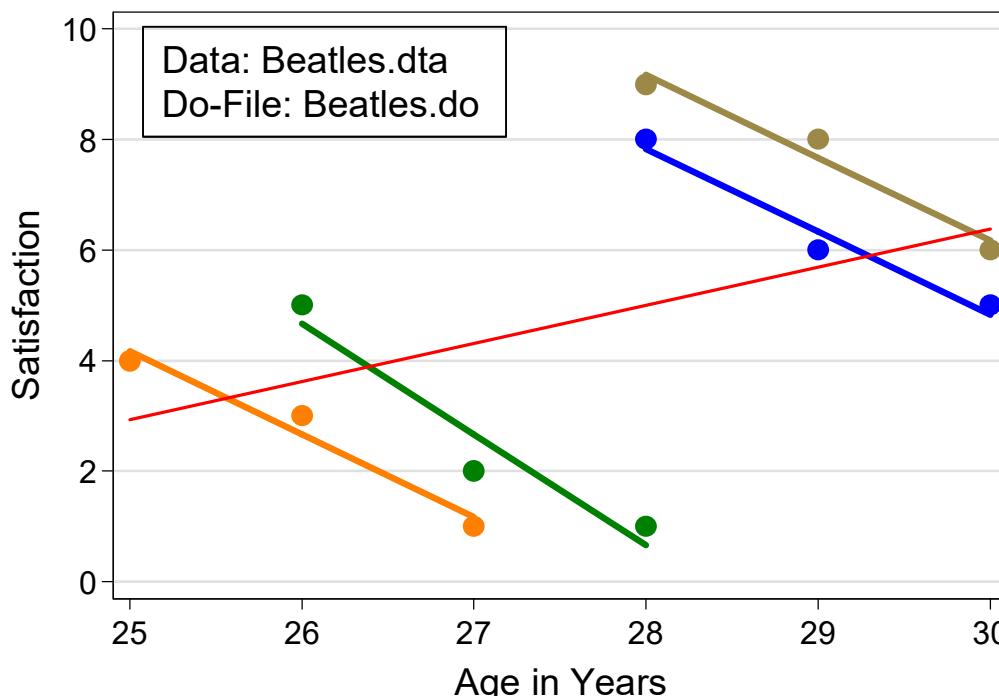
Source: Kohler/Kreuter (2009) Data Analysis Using Stata. Stata Press. Pp. 242 ff.



Example: Life satisfaction of Four Englishmen

- However, between estimation is completely misleading here
 - Because there is heterogeneity between the men (perhaps a cohort effect?)
- Within estimation shows that there is actually a negative age effect!
 - Estimate separate regression lines for each man: All men grow less happy
 - Green ($\hat{\gamma} = -2.0$), orange, blue, brown ($\hat{\gamma} = -1.5$)
 - The FE estimate is the average of these four slopes

$$\hat{\gamma}_{FE} = -\frac{2.0+1.5+1.5+1.5}{4} = -1.625$$



Happiness Example: Specification of Growth Curves

Robust						
happy	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-0.1343	0.0128	-10.50	0.000	-0.1594	-0.1092
c.age#c.age	0.0025	0.0003	7.71	0.000	0.0019	0.0032
c.age#c.age#c.age	-0.0000	0.0000	-7.95	0.000	-0.0000	-0.0000

. margins, at(age=(18(4)98))

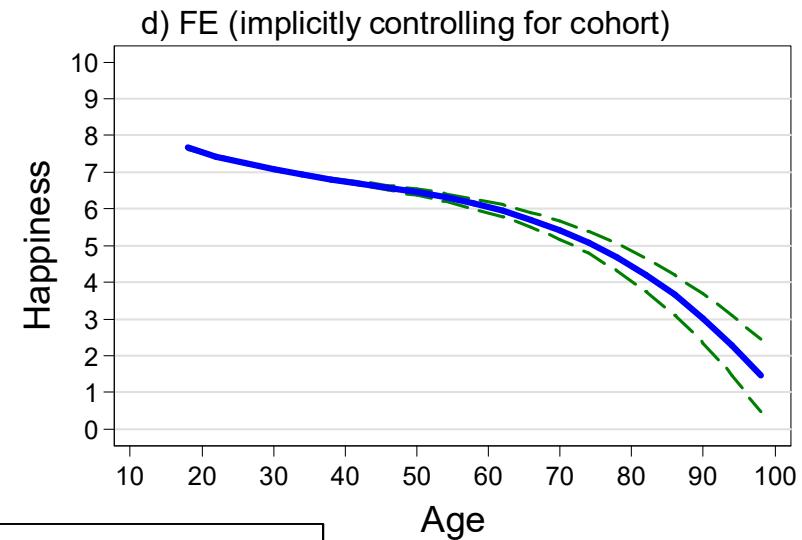
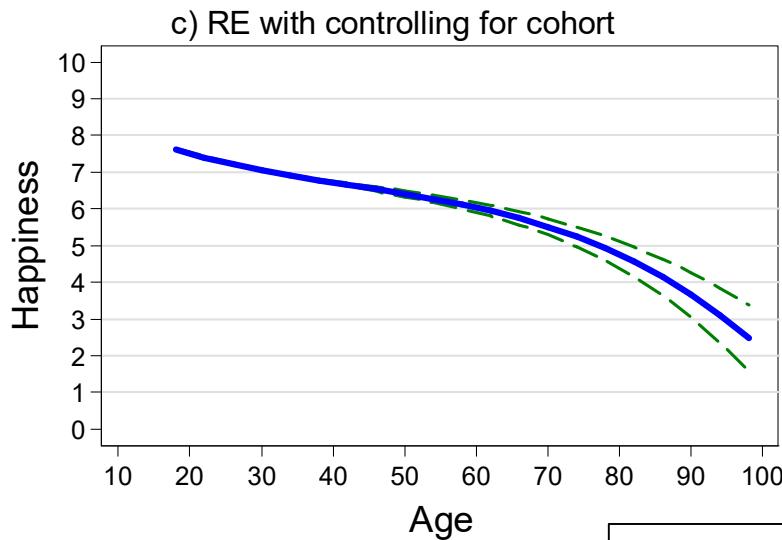
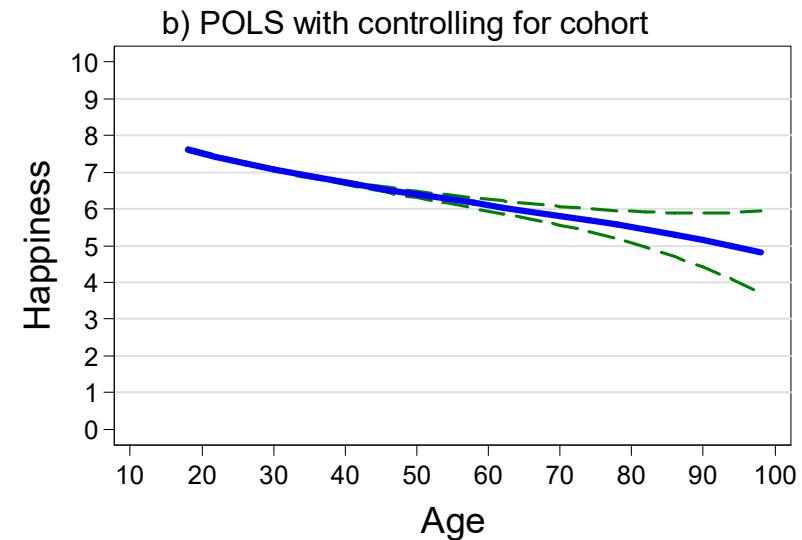
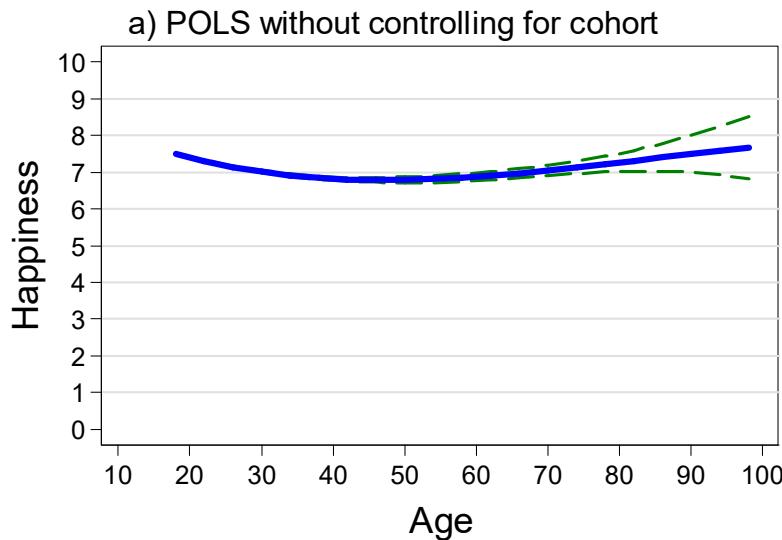
. marginsplot, recast(line) recastci(rline)

Data: Happiness2.dta

Do-File: Happiness 4 Regressions.do

- After explorative analyses we decided to use a cubic specification
 - Wunder et al. (2013) argue that a cubic polynomial does not adequately represent happiness growth. They suggest semi-parametric splines. Below we will use the most flexible age dummy specification.
- Cohort is controlled by introducing 90 birth year dummies
 - With `i.cohort` (not in the FE model!)
- GC estimation results can best be visualized by a profile plot
 - Please note that the FE profile plot rests on adding an arbitrary constant!

Happiness Example: Comparing Different Specifications



Data: Happiness2.dta
Do-File: Happiness 4 Regressions.do

Happiness Example: Comparing Different Specifications

- See previous slide
 - Panel a): POLS without controlling for cohort
 - This is a grossly mis-specified model! We find a U-shaped GC. However, this is an artifact, since older cohorts are happier.
 - Nevertheless, many studies in the economics literature report such an U-shape and take it seriously! (Wunder et al. 2013)
 - Panel b): POLS with controlling for cohort
 - Now we see declining happiness
 - Nevertheless, this is still biased upwards, because happier people live longer (self-selection into higher ages)
 - Panel c): RE with controlling for cohort
 - The late age decline becomes steeper
 - Nevertheless still a small upward bias
 - Panel d): FE (implicitly controlling for cohort)
 - Within estimation shows that the late age decline is even more sharp
 - Before age 60 happiness declines slowly.

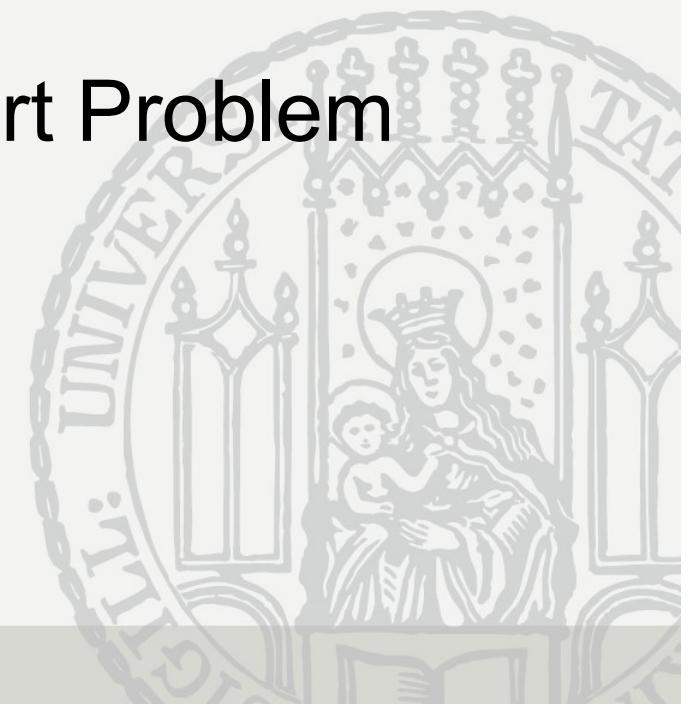


Chapter VI:

Modeling Individual Growth

Section: The Age-Period-Cohort Problem

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Period Effects

- Until now we ignored period effects
 - I.e., we assumed that they are zero
 - Age (in years), period (interview year), cohort (birth year)
 - Allowing for period (P) effects produces the following model

$$y_{it} = \beta x_{it} + \gamma A_{it} + \delta C_i + \mu P_t + \alpha_i + \varepsilon_{it}$$

- Because $P_t = C_i + A_{it}$ we arrive at

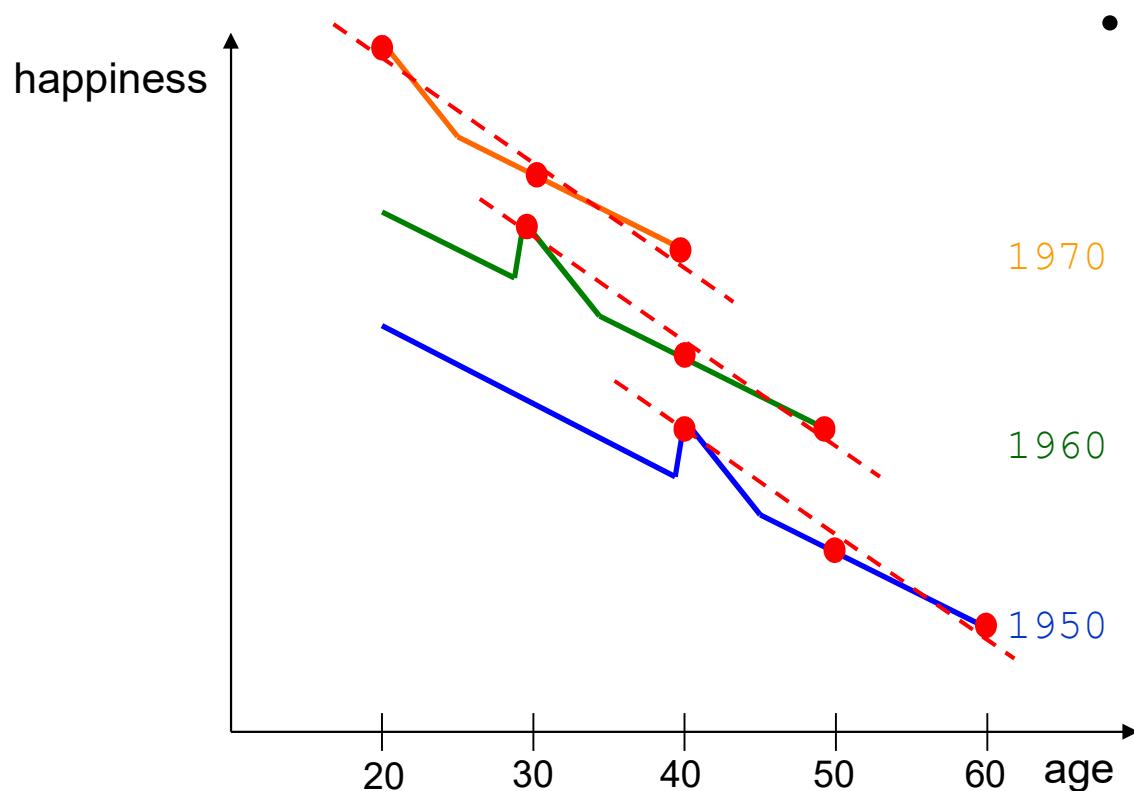
$$y_{it} = \beta x_{it} + \gamma A_{it} + \delta C_i + \mu(C_i + A_{it}) + \alpha_i + \varepsilon_{it}$$

$$\Rightarrow y_{it} = \beta x_{it} + (\gamma + \mu)A_{it} + (\delta + \mu)C_i + \alpha_i + \varepsilon_{it}$$

- If $\mu \neq 0$ both age and cohort effect estimates will be biased
- In an extreme case it could be that an age effect is totally due to period effects. For instance, if economic circumstances become worse and worse each year, it will appear as if happiness declines with age.
- Thus, generally it is not a good idea to simply ignore period effects
- Only if there are plausible arguments for the absence of period effects this is a sensible strategy

Period Effect Biassing the Age Effect

- Thought experiment: estimating a growth curve from panel data
 - We assume that happiness declines linearly with age (age effect)
 - We assume that younger cohorts are happier (cohort effect)
 - We assume that winning the world championship 1990 made people temporarily happier (period effect)



- Panel survey in the years 1990, 2000, and 2010
 - The estimated growth curves are too steep
 - Ignoring the period effect biases the age effect
 - Adding a 1990-dummy would produce unbiased results, because the period effects are then correctly specified

The Age-Period-Cohort (APC) Problem

- The fundamental identification problem

- This is the model with all three terms

$$y_{it} = \beta x_{it} + \gamma A_{it} + \delta C_i + \mu P_t + \alpha_i + \varepsilon_{it}$$

- We have linear dependence because $P_t = C_i + A_{it}$
 - Due to collinearity it is impossible to estimate all three effects
 - The model is not identified
 - It does not matter, whether the terms enter linearly or as dummies

- Variant: two-way FE model

$$y_{it} = \beta x_{it} + \gamma A_{it} + \mu_t + \alpha_i + \varepsilon_{it}$$

- Here only age and period terms are specified. However, C_i is person-specific and therefore included in the fixed-effects
 - The period term is here specified as year (wave) dummies μ_t . One year dummy has to be dropped (base category)
 - This model is not identified. With FE estimation already an age-period specification is not identified
- The same problem one has with a single cohort panel
 - An age-period specification (POLS) is not identified

The APC Problem: Solutions

- One has to impose (sensible) APC restrictions
 - Equality constraints
 - Example period dummies: at least a second period dummy has to be dropped. But which one?
 - Stata default: the last period dummy is dropped additionally
 - Hand made: use a RE-model (with cohort restriction) and look for years with similar outcome values, and use these as base
 - Problem: RE estimated with cohort restriction!
 - The age effect could be specified non-linear
 - For instance, include only a quadratic age term
 - This method often produces nonsense
 - Specify at least one term as a step function
 - At least one term has to be grouped
 - Often 5-year cohorts (or “war cohort”, “baby boomers”, etc.)
 - Terms could be grouped with different interval length
 - For instance, 5-year periods and 10-year age groups

Happiness Example: The Stata Default Solution

Letting Stata decide which restriction to use:

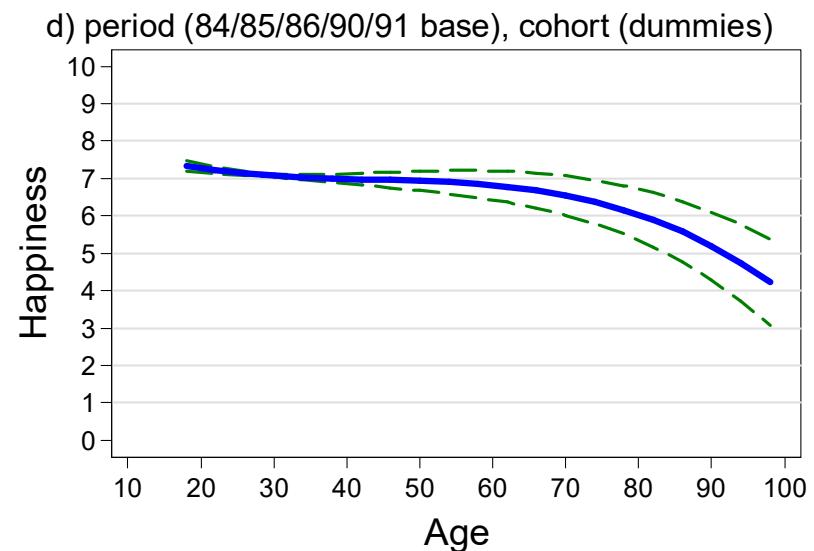
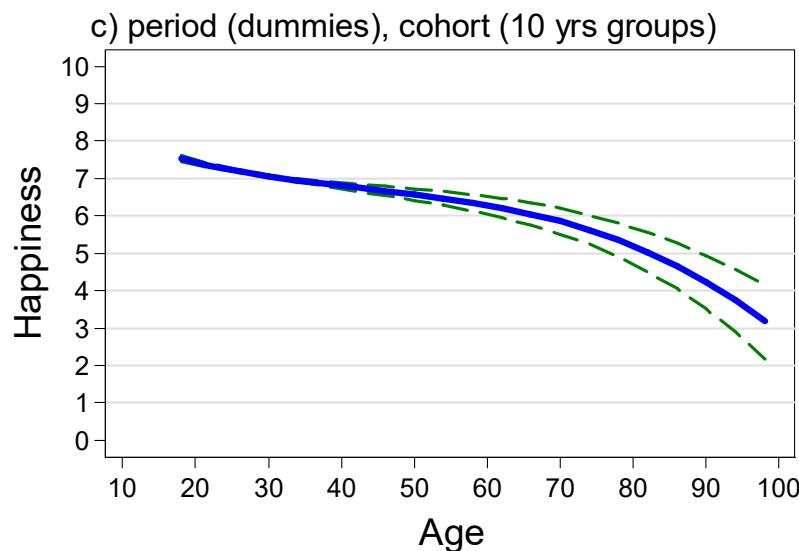
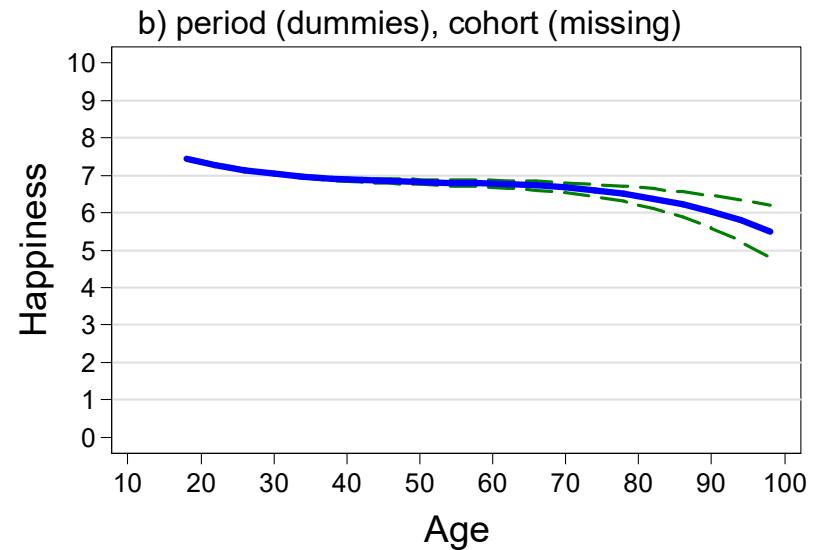
Robust						
happy	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	-0.1373	0.0129	-10.64	0.000	-0.1626	-0.1120
c.age#c.age	0.0026	0.0003	7.76	0.000	0.0019	0.0032
c.age#c.age#c.age	-0.0000	0.0000	-8.01	0.000	-0.0000	-0.0000
year						
1984	0.0000	(base)				
1985	-0.0478	0.0454	-1.05	0.292	-0.1367	0.0411
1986	0.0502	0.0452	1.11	0.267	-0.0384	0.1388
...						
2007	0.0733	0.0207	3.54	0.000	0.0327	0.1138
2008	0.0690	0.0199	3.47	0.001	0.0301	0.1080
2009	0.0000	(omitted)				

Data: Happiness2.dta
Do-File: Happiness 4 Regressions.do

Happiness Example: APC Restrictions

- Unfortunately, the estimated age effect is often very sensitive to the restrictions chosen
- On the next slide we demonstrate, how different APC restrictions affect a cubic RE happiness GC
 - We use RE, because then we have full flexibility in specifying the restrictions (e.g., we can drop “cohort” from the model)
 - Panel a): age-cohort model (period dummies missing)
 - Mis-specified because there are period effects
 - Panel b): age-period model (cohort dummies missing)
 - Mis-specified because there are cohort effects
 - Panel c): age-period-cohort model (cohort grouped)
 - Panel d): age-period-cohort model (some period dummies miss.)
 - Our favorite model: 1984/85/86/90/91 are the base. A RE model (without cohort) showed that these years have the highest and very similar happiness levels
 - Here we see that the growth curve is essentially flat until 70

Happiness Example: Comparing Different APC Restrictions

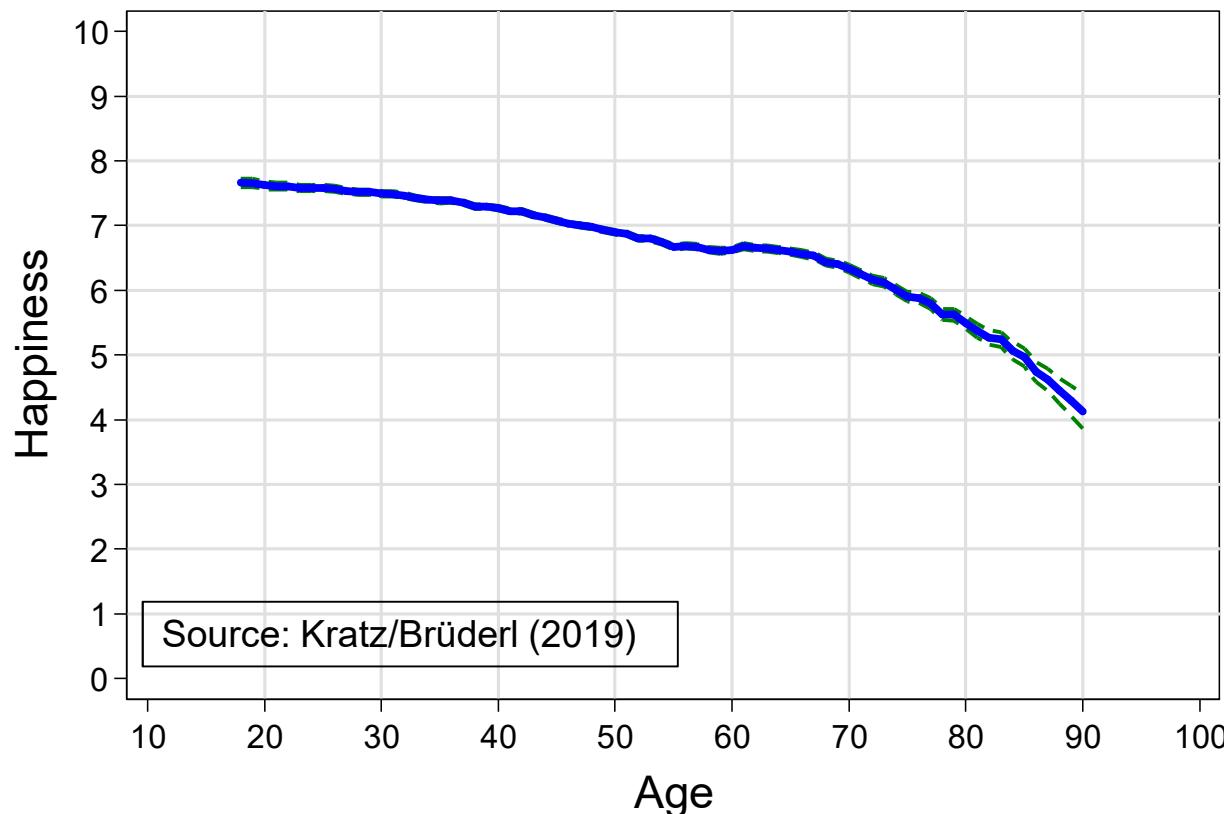


The APC Problem: A Better Solution

- Obviously, APC restrictions are problematic
- “Substantive” solution for the APC problem
 - For similar arguments see Wunder et al. (2013)
 - Do not use the proxy variable “period”, but specify the events that produced the period effects directly
 - Under good economic circumstances (GDP) happiness should be higher
 - Fall of the wall and German unification: happiness increase 1990/91
 - Germany became soccer world champion: increase in 1991
- Recommendation
 - Age and cohort effects are in most applications most “important”
 - Include those in all panel regressions (age and cohort dummies)
 - Then start thinking about potential period effects
 - Include only the relevant “period variables”

Our Final Result on Happiness Growth

- Using all SOEP pyrs 18-90 (v29, 1984-2012); $N \cdot T = 470,022$, $N = 57,758$
- FE model (controls implicitly for all time-constant confounders)
 - Fully flexible specification through age dummies
 - Period variables: unemployment rate, GDP growth rate
 - Panel conditioning: wave 1-3 dummies



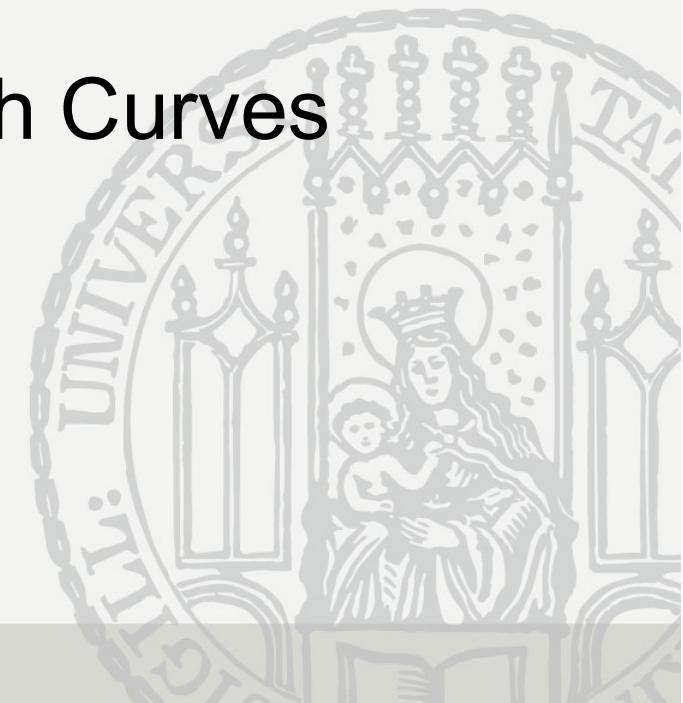
- Happiness decline up to age 60
- There is a little “pension hump” after age 60
(Wunder et al. 2013)
- Happiness declines sharply after age 65
 - Terminal decline 3-5 years before death:
“Death is lurking around the corner”
(Gerstorf et al. 2010)

Chapter VI:

Modeling Individual Growth

Section: Group Specific Growth Curves

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Group Specific Growth Curves (GCs)

- Interact age with a time-constant group variable ($G: \{0,1\}$)
 - Here the question is not on group-specific age effects (see Chap.V)
 - But on the group difference in the outcome and how it develops over age
 - Therefore, we need models with a group main effect
 - It makes sense to center age
- Example: linear GC [(uncentered) age is A_{it}]

$$y_{it} = \alpha + \vartheta G_i + \beta A_{it} + \delta(G_i \times A_{it}) + \alpha_i + \varepsilon_{it}$$

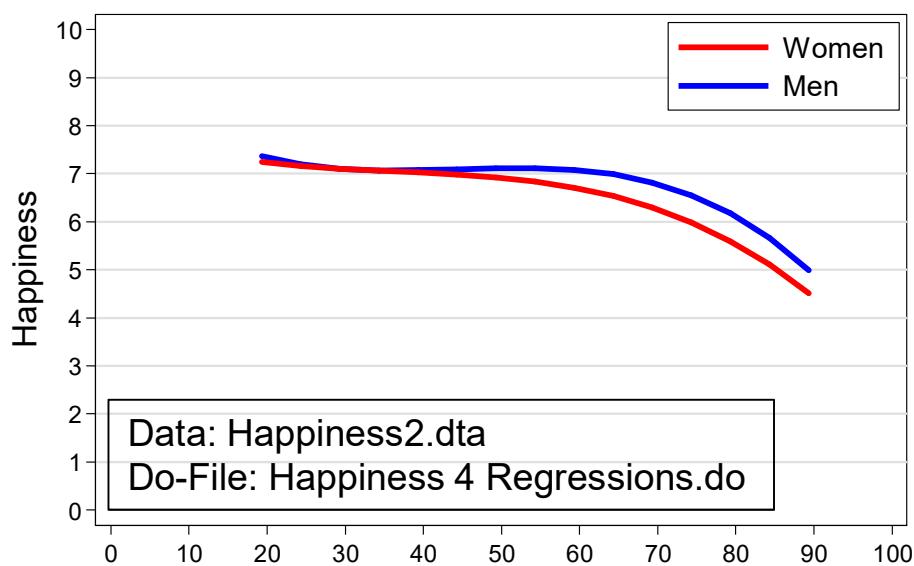
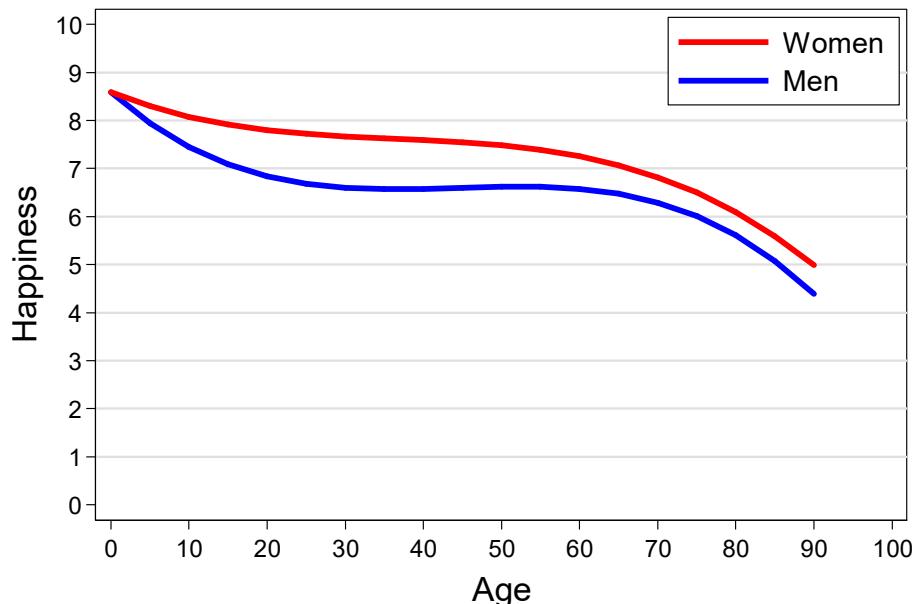
ϑ the group difference at $A = 0$

δ the change in the group difference with every year

- Stata code: i.group##c.age

- Group specific growth curves provide much more information on the effect of a time-constant variable than a simple group effect
 - It is a waste of panel data to estimate only group effects!!!
 - Instead estimate group specific growth curves!

Happiness Example: FE Estimation



Data: Happiness2.dta
Do-File: Happiness 4 Regressions.do

With FE we obtain a strange result

- It seems, as if women are much more happy (more than one scale point!) (upper Fig.)
- However, this is nonsense: In FE there is no main effect of “sex”
 - Therefore, FE forces the restriction of equal happiness at age 0.
 - Therefore, the difference between the GCs is not interpretable
 - If we center “age” (mean=29.3), a different restriction applies and now men seem to be happier (lower Fig.)
- Thus, group-specific GCs can not be estimated with FE!
- Unfortunately there is no way around this problem (“you can't trick the data”)
 - Separate FE regressions for men and women also provide no well defined constant
 - A Hybrid model (see below) has a main sex effect, but this is a between constant. Within results are identical with FE.

Happiness Example: RE Estimation

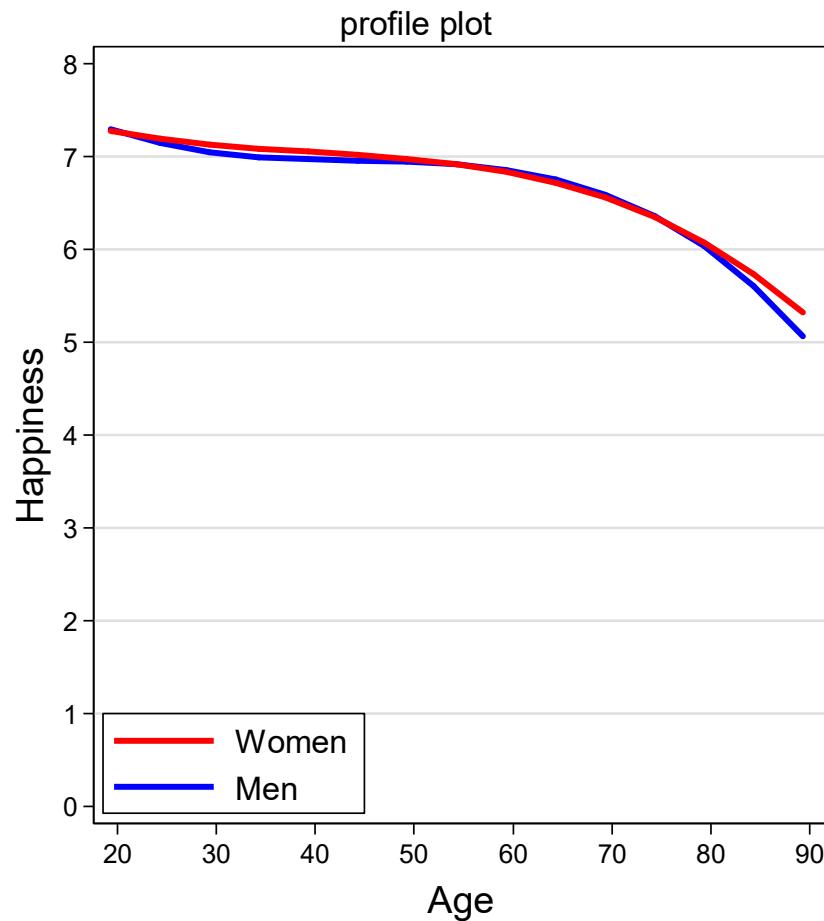
- Therefore, in the following we estimate RE-GCs.
 - Fortunately with long panels RE-GCs are generally close to FE-GCs
 - This can be checked by comparing the FE-GC and the RE-GC
 - Note, the group effect might be affected by unobserved heterogeneity
 - Again this reflects the fact, that there is no within variation to estimate the group effect. We have to use the between variation. Nevertheless, RE-GCs make much better use of the between variation in panel data
- The results (see next slides)
 - Age should be centered (agec)
 - Thus the sex main effect is the happiness difference at age 29.3
 - Profile plot: women are happier around age 30 and at higher ages
 - To assess the significance of the sex differences one has to use a conditional effect plot
 - Conditional effect plot: women are significantly happier than men from about age 25 to age 40! Difference is +0.09 points at about age 35.
 - Obviously, this is much more information than what a simple sex effect (+0.03) from a RE model tells us!

Happiness Example: Sex Specific RE-GCs

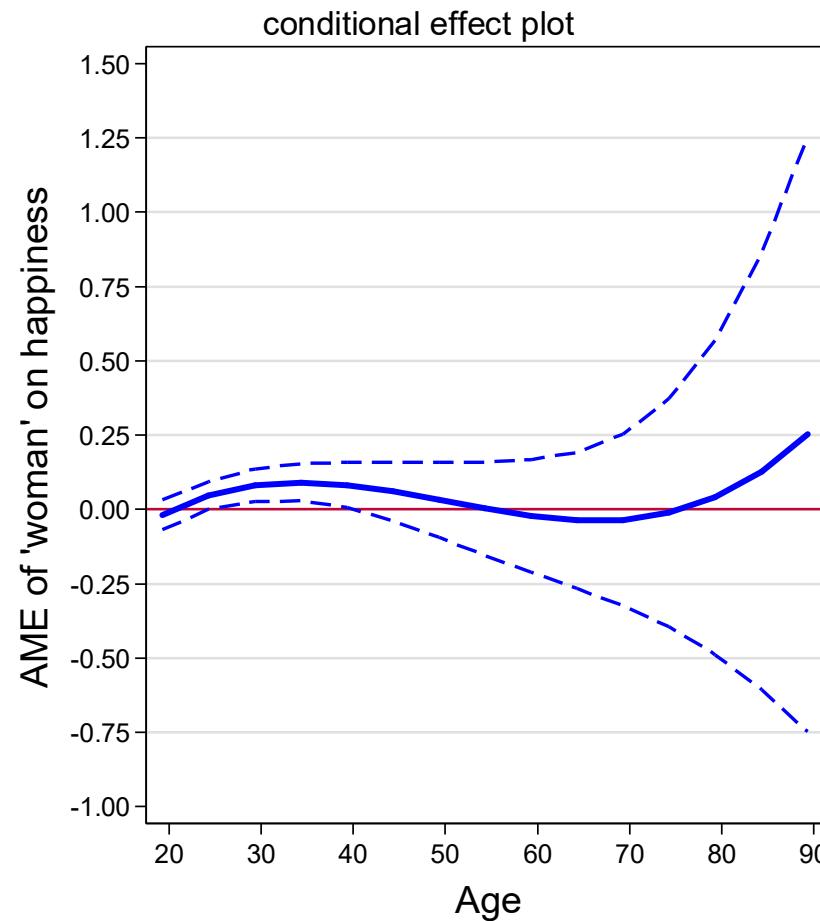
. xtreg happy i.woman i.woman#(c.agec##c.agec##c.agec) ///	
> ib(last).cohort i(1987/1989 1992/2009)bn.year, re vce(cluster id)	
Random-effects GLS regression	Number of obs = 121919
Group variable: id	Number of groups = 14634
R-sq: within = 0.0199	(Std. Err. adjusted for 14634 clusters in id)
-----	-----
	Robust
happy	Coef. Std. Err. z P> z [95% Conf. Interval]
-----+-----	-----+-----
1.woman .0812636 .0280571 2.90 0.004 .0262727 .1362545	
woman#c.agec	
0 -.0143229 .0062536 -2.29 0.022 -.0265797 -.0020661	
1 -.010067 .0062785 -1.60 0.109 -.0223725 .0022386	
woman#c.agec#c.agec	
0 .0008397 .0001548 5.43 0.000 .0005363 .001143	
1 .0003486 .0001542 2.26 0.024 .0000462 .0006509	
woman#c.agec#c.agec#c.agec	
0 -.0000192 4.16e-06 -4.61 0.000 -.0000273 -.000011	
1 -.0000114 2.89e-06 -3.94 0.000 -.000017 -5.71e-06	

Age-cohort-period(84/85/86/90/91 base) specification

Happiness Example: Sex Specific RE-GCs



```
margins woman, at(agec=(-10(5)60))  
marginsplot, recast(line) noci
```

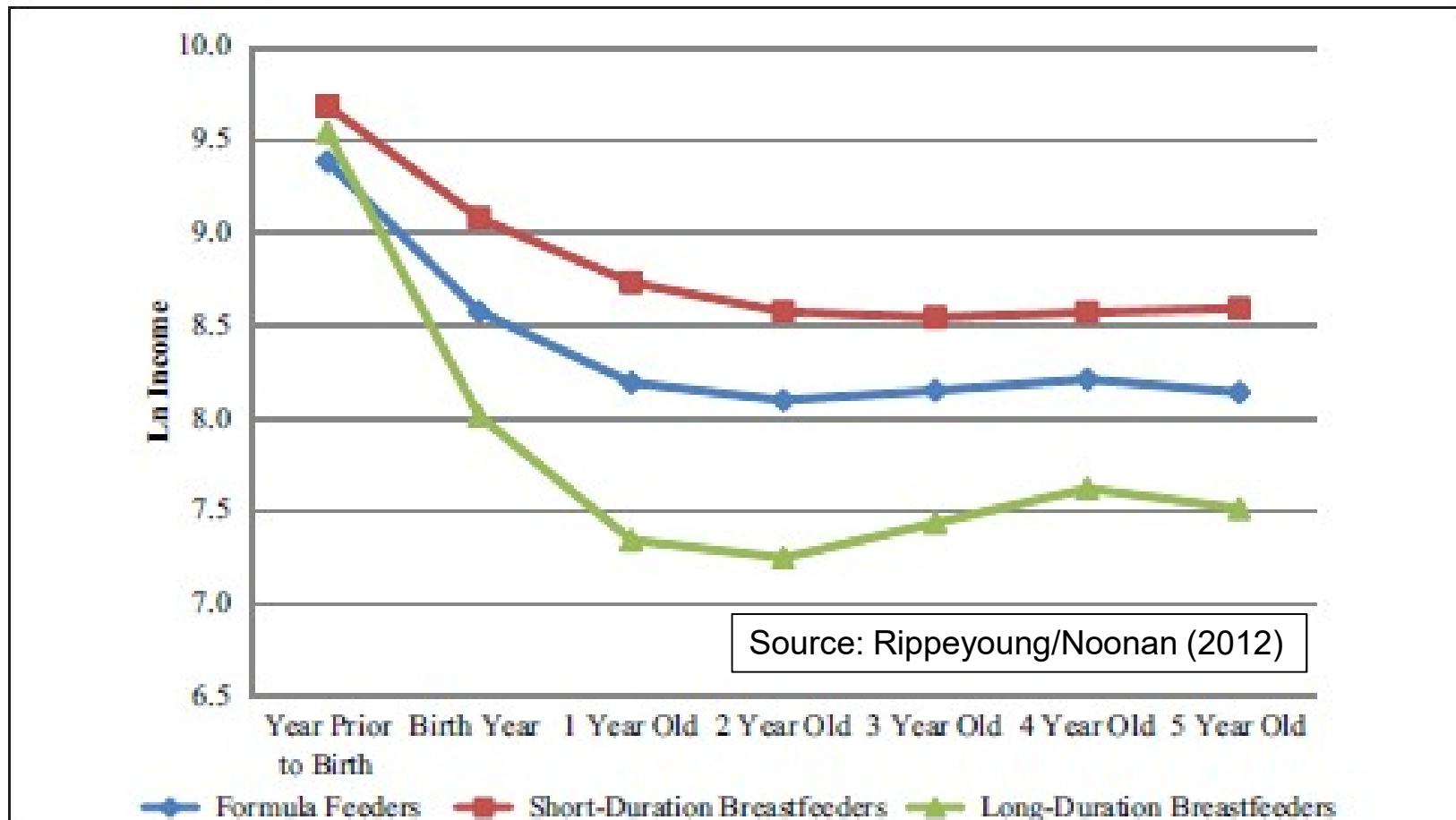


```
margins , at(agec=(-10(5)60)) dydx(woman)  
marginsplot, recast(line) recastci(rline)
```

Data: Happiness2.dta
Do-File: Happiness 4 Regressions.do

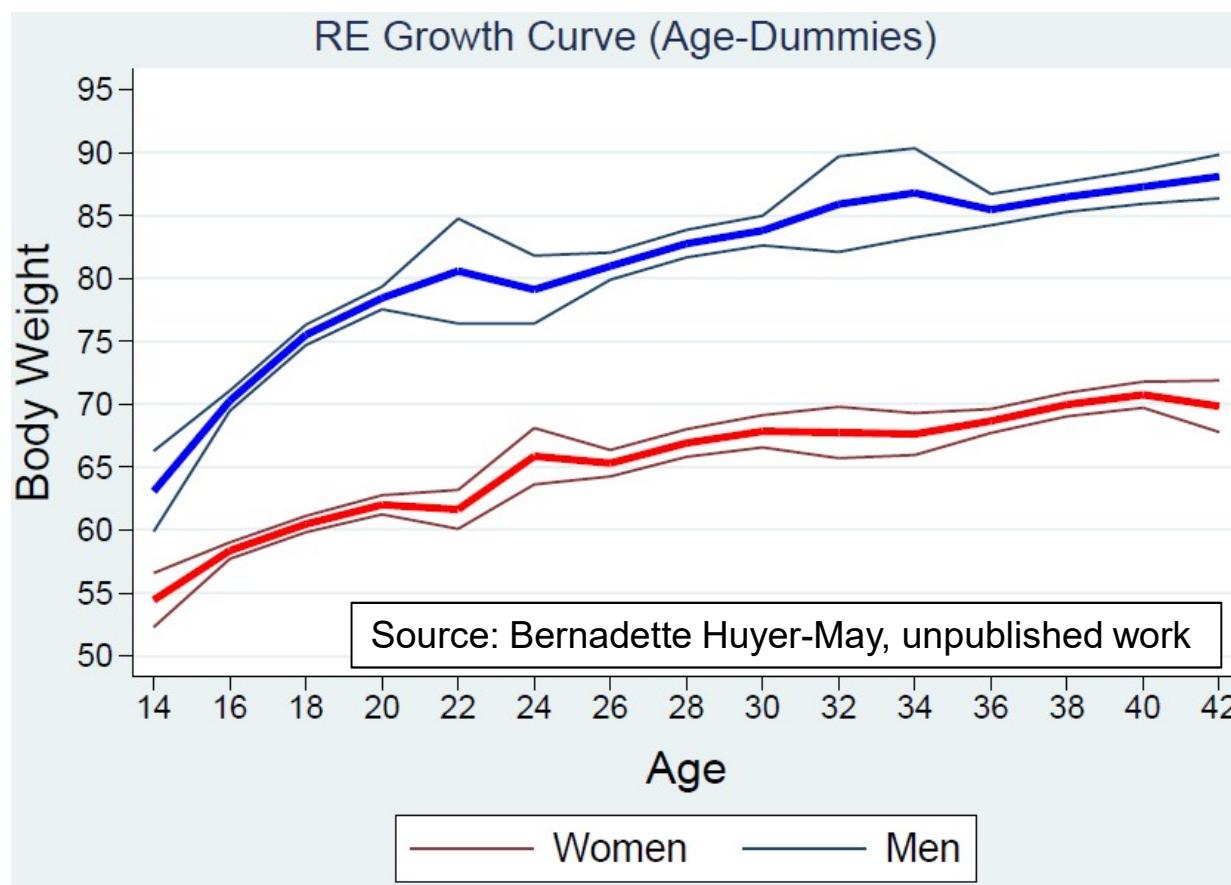
Example: The Cost of Breastfeeding

- Group specific GCs are mostly used in a purely descriptive manner. However, in a complex social world this type of description is very informative
- Example: Mothers' income trajectories after the birth of the first child
 - Cubic growth curves: time since childbirth (year prior birth=0)
 - By infant feeding type



Example: Body Weight Growth

- Pairfam (v5) data on body weight
 - Pairfam has three cohorts that do not yet overlap
 - Note the smooth trajectories (data seem to be valid!)
 - There is a clear cohort effect visible: The younger cohorts are on an ever higher trajectory (secular increase in height [and thereby weight])



Chapter VII:

Further Linear Panel Models

Section: Alternative Within Estimators

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



LSDV: areg

- **areg** is a LSDV estimator
 - Person dummies are not estimated, but absorbed
 - R^2 is variance explained including the person dummies
 - I.e., it is very large (it is not the within R^2)
 - **areg** S.E.s are larger than **xtreg** S.E.s
 - **areg** assumes that the number of persons does not increase with sample size (only then the estimates of person dummies are consistent)
 - **xtreg** does not rest on this assumption
 - **areg** allows for panel weights that change over waves
 - **xtreg** does not!
- **Recommendation**
 - Use **areg** only, if you want to do weighted FE estimation
 - This is detailed below

Individual Slope Regression

- The FE estimator is the weighted average of the slopes of the individual regressions
 - Weighted by the length of the individual panels
 - Only true for the bivariate model

Data: Happiness2.dta Do-File: Happiness 5 Regressions.do

```
. keep if pycount == 26 //only panels with 26 observations (equal weight!)
```

```
. xtreg happy age, fe //benchmark FE model
```

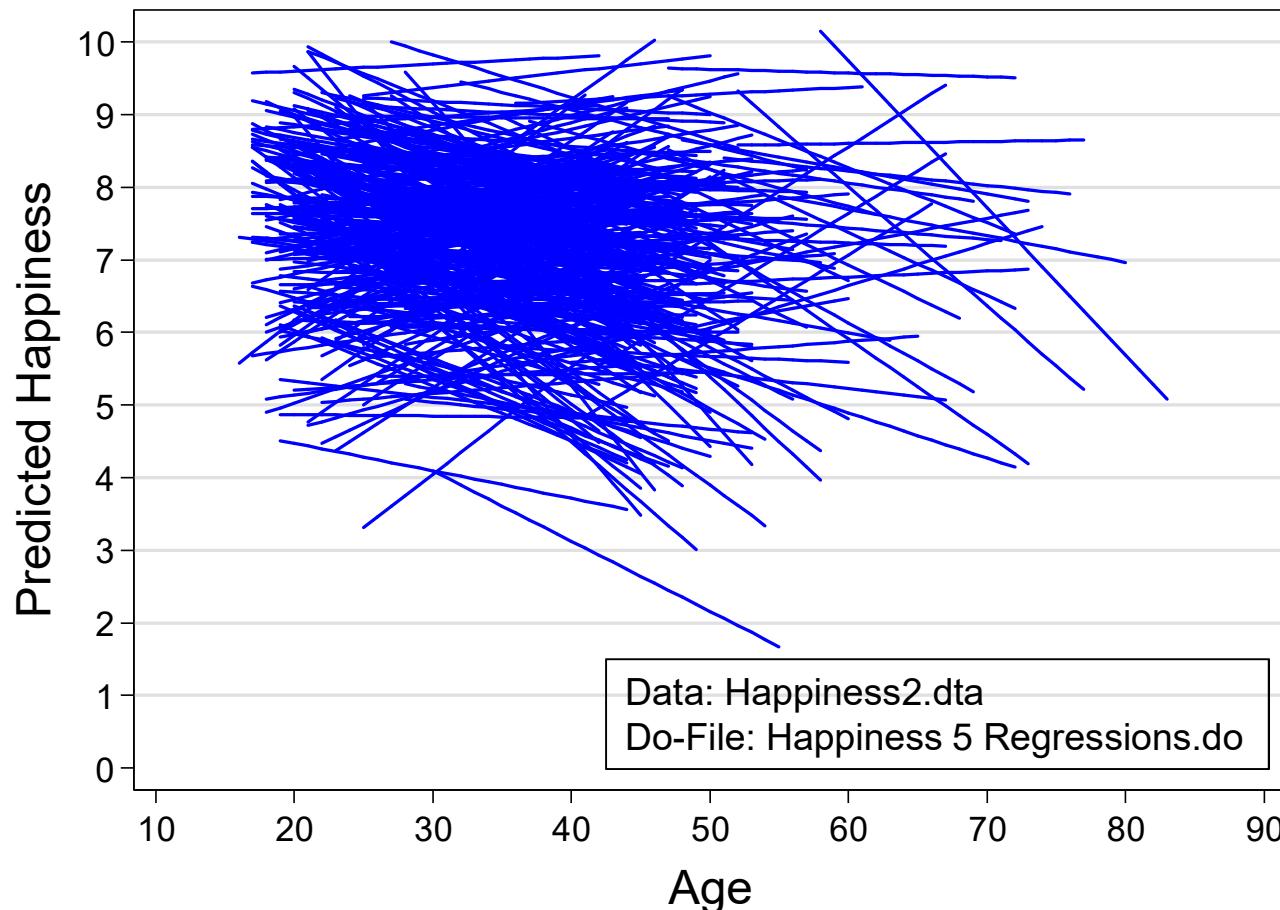
	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
age	-.0236349	.0017556	-13.46	0.000	-.0270764 -.0201935
_cons	8.078896	.0659529	122.49	0.000	7.949613 8.208179

```
. * Now we run 348 regressions!
. statsby cons=_b[_cons] slope=_b[age], by(id):      regr happy age
. summ slope cons                                //perfectly identical
```

Variable	Obs	Mean	Std. Dev.	Min	Max
slope	348	-.0236349	.0647752	-.2413675	.1480342
cons	348	8.099078	2.612797	-.3589744	21.8841

Individual Slope Regression

- Plot the 348 regression lines (spaghetti plot)
 - One can see the negative overall tendency
 - However, there are also some lines that increase (effect heterogeneity)!



Difference-in-Differences (DiD)

- DiD can be implemented in a regression framework

$$y_{it} = \alpha + \delta_1 p_{it} + \delta_2 d_i + \delta_3 (p_{it} \times d_i) + x_{it}\beta + \varepsilon_{it}$$

- Post-treatment period dummy: p_{it}
- Treatment dummy: d_i
- α : pre-treatment mean outcome in control group
- δ_1 : post-treatment change in mean outcome in control group
- δ_2 : pre-treatment difference in mean outcome in treatment group
- δ_3 : post-treatment difference in change in mean outcome in treatment group (this is the treatment effect)
- DiD models only group differences. Therefore, it can be applied to aggregate pseudo-panel data (trend data on the group level).
- This makes DiD a very useful tool for policy analysis with data from natural experiments (treatment and control groups are assigned without randomization)
- However, with individual panel data DiD makes less sense
 - Controls enter as levels (not differences) and therefore their effects will be biased and this in turn can bias the treatment effect estimate

Difference-in-Differences (DiD)

- Without controls DiD is equivalent to FE

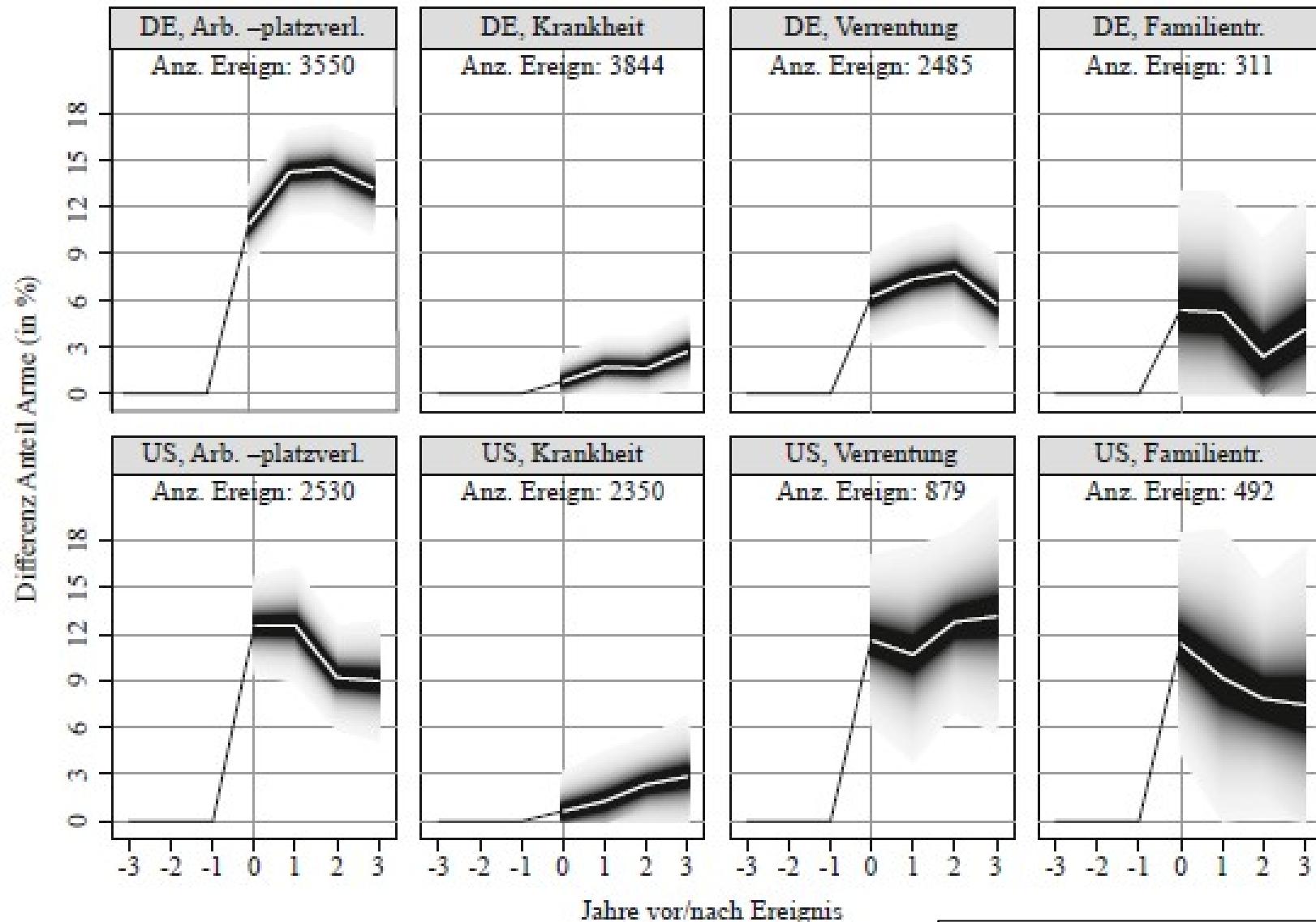
Data: Wage Premium.dta Do-File: Wage Premium.do						
<pre>. gen treat = id >= 3 . gen post = time >= 4 . gen posttreat = post * treat . regr wage post treat posttreat</pre>						
wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
post	2.32e-13	318.4598	0.00	1.000	-664.2956	664.2956
treat	2000	318.4598	6.28	0.000	1335.704	2664.296
posttreat	500	450.3702	1.11	0.280	-439.4558	1439.456
_cons	1500	225.1851	6.66	0.000	1030.272	1969.728

- With controls this is generally no longer true
- Further, with non-policy (“natural”) treatments DiD is not straightforward, because treatment does not occur at a fixed time point.
 - Then one lacks a clear before-after for the control group. One has to use some kind of matching to construct the control group (DiD-matching).

DiD-Matching Example: Poverty Risks

- Kohler et al. (2012) investigate how welfare states protect people from poverty after critical life events
 - Germany (conservative welfare state), U.S. (liberal welfare state)
 - Outcome: % in poverty (below 60% median equivalence income)
 - Treatment: job loss, sickness, retirement, family dissolution
 - Method: DiD (3 years before, 4 years after)
 - Treatment group: 3 years not poor, then having a critical life event
 - Control group: exact matching (all those, who were also not poor in these 3 years, and who did not have the event in the next four years)
 - Data: SOEP and PSID, 1984 - 2009
 - Results: see next slide
 - Criticism:
 - Groups are selected according to the outcome (3 years not poor). Thus, in fact this is not DiD, it is only the post-treatment difference
 - The construction of the control group is sub-optimal: exact matching without any controls. Better would be Propensity-Score DiD-matching.

DiD-Matching Example: Poverty Risks



Source: Kohler et al. 2012

Chapter VII:

Further Linear Panel Models

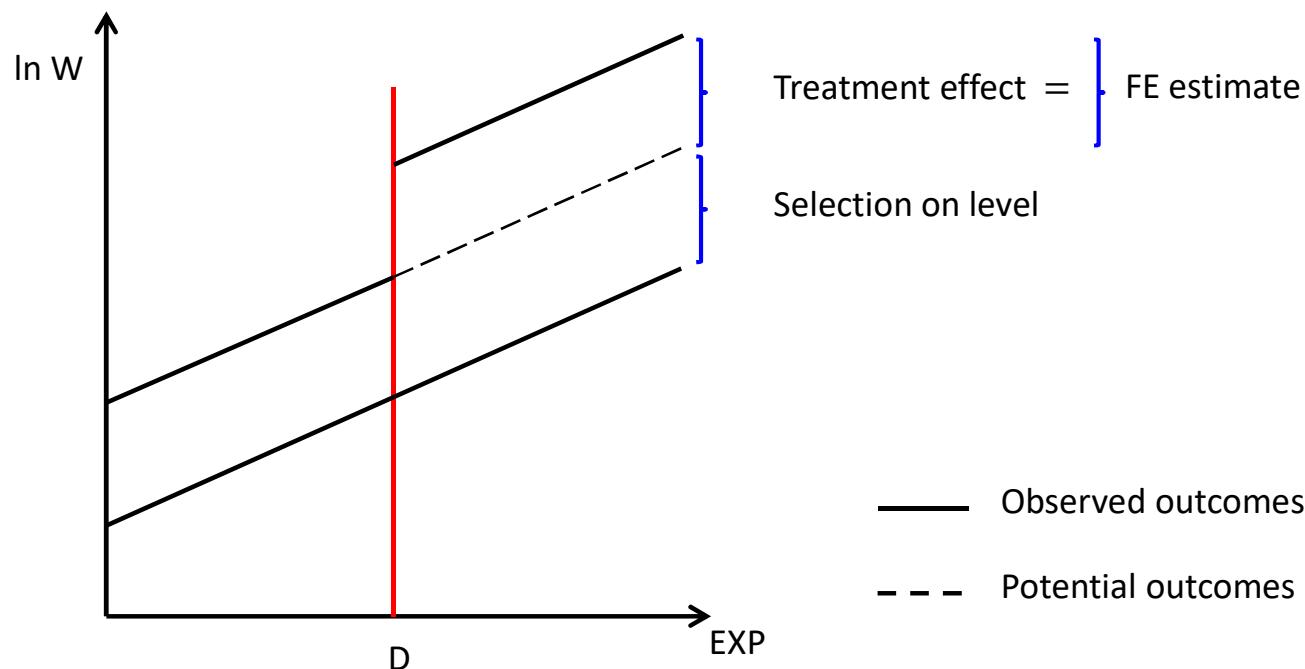
Section: The Fixed-Effects
Individual-Slopes Model

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



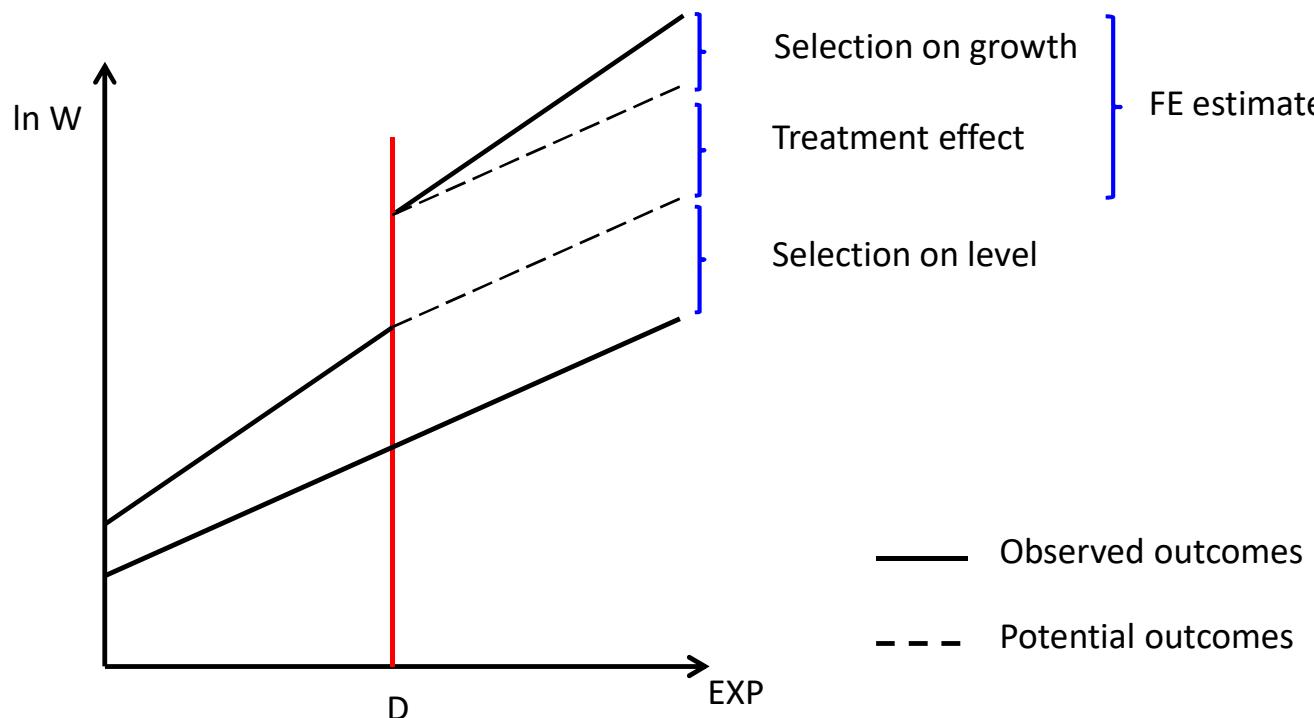
The Parallel Trends Assumption

- Another way to express the strict exogeneity assumption from a counterfactual perspective
 - Strict exogeneity implies, that the (potential) outcome trends in treatment and control group must be parallel
 - Then the FE-estimate is consistent, and captures the treatment effect



The Parallel Trends Assumption

- Parallel trends assumption violated: heterogeneous growth
 - Those on a steeper trajectory are selected into the treatment
 - E.g., younger cohorts might accumulate skills faster and self-select into treatment
 - E.g., career orientation may differ between those in treatment and control
 - This is a violation of strict exogeneity due to time-varying unobservables
 - The FE-estimate will be biased upwards (in case of diverging trends)
 - It includes the wage increase due to the steeper trend (“selection on growth”)



The FEIS Model

- Due to group specific trends we have non-parallel trends
- This can be generalized to individual specific trends

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_{1i} + \alpha_{2i}t + \xi_{it}$$

- This is the fixed-effects model with individual-specific constants and slopes (FEIS) (Wooldridge, 2010: pp. 377-381)
- α_{2i} captures individual differences in the growth of the outcome over time. $\alpha_{2i}t$ is time-varying unobserved heterogeneity (heterogeneous growth).
- In a standard FE this is part of the idiosyncratic error term

$$\varepsilon_{it} = \alpha_{2i}t + \xi_{it}$$

- As shown below, we can get rid of $\alpha_{2i}t$
- Thus, FEIS needs a weaker exogeneity assumption
 - FE: $E(\mathbf{x}'_{it}\varepsilon_{is}) = \mathbf{0}$, for all $s, t = 1, \dots, T$
 - FEIS: $E(\mathbf{x}'_{it}\xi_{is}) = \mathbf{0}$, for all $s, t = 1, \dots, T$
- In the FEIS model time-varying unobserved heterogeneity that is due to individual-specific trends is no longer a problem

How Can We Get Rid of the Individual-Specific Trends?

- Including individual level and trend dummies (LSDV)
 - This is done often intuitively in studies with country data: POLS models include country-specific intercepts *and* country-specific trends
- Second-differences estimator (SD)

$$y_{it} = \beta x_{it} + \alpha_{1i} + \alpha_{2i}t + \xi_{it}$$

$$y_{it-1} = \beta x_{it-1} + \alpha_{1i} + \alpha_{2i}(t-1) + \xi_{it-1}$$

$$y_{it-2} = \beta x_{it-2} + \alpha_{1i} + \alpha_{2i}(t-2) + \xi_{it-2}$$

- Computing first differences (FD)

$$(y_{it} - y_{it-1}) = \beta(x_{it} - x_{it-1}) + \alpha_{2i} + (\xi_{it} - \xi_{it-1})$$

$$(y_{it-1} - y_{it-2}) = \beta(x_{it-1} - x_{it-2}) + \alpha_{2i} + (\xi_{it-1} - \xi_{it-2})$$

- Now take second differences (SD)

$$(y_{it} - 2y_{it-1} + y_{it-2}) = \beta(x_{it} - 2x_{it-1} + x_{it-2}) + (\xi_{it} - 2\xi_{it-1} + \xi_{it-2})$$

- With this extension of FD we managed to get rid of α_{1i} and α_{2i} . Using POLS provides an unbiased estimate of $\hat{\beta}$, if the exogeneity assumption on ξ_{it} holds.

- SD needs at least 3 waves

- SD is efficient with $T = 3$. It is inefficient for $T > 3$

FEIS Estimation

- Detrending transformation (FEIS estimation)
 - Analogous to FE, where data are demeaned, we “detrend” the data
 - 1) Estimate for each unit the individual growth curve $y_{it} = \alpha_{1i} + \alpha_{2i}t + \zeta_{it}$ and get predicted values $\hat{y}_{it} = \hat{\alpha}_{1i} + \hat{\alpha}_{2i}t$
 - 2) Subtract predicted values from actual outcomes to get detrended outcomes $\tilde{y}_{it} = y_{it} - \hat{y}_{it}$
 - 3) Repeat steps 1) and 2) to detrend also the regressors $\tilde{x}_{it} = x_{it} - \hat{x}_{it}$
 - 4) Pool the detrended data and run a POLS regression
 - The intuition is that after detrending only variation around the trend is left. Only this around-trend variation is used to estimate the causal effect. Thus, heterogeneous growth can no longer bias estimates
- This approach can be generalized (see next slide)
 - Trends of higher order can be modeled
 - Individual slopes for non-time variables can be dealt with analogously
 - An ado `xtfeis.ado` was written by Volker Ludwig
 - `findit xtfeis`

The General FEIS Model

- The general model is

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \mathbf{z}_{it}\boldsymbol{\alpha}_i + \xi_{it}$$

where \mathbf{z}_{it} is a $(1 \times J)$ vector of variables with individual-specific slopes (including a constant)

- For detrending one needs at least $J + 1$ observations
- The detrended model is

$$\tilde{y}_{it} = \tilde{\mathbf{x}}_{it}\boldsymbol{\beta} + \tilde{\xi}_{it}$$

- Note that after detrending also \mathbf{z}_{it} has gone. FEIS provides no effect estimates for the \mathbf{z}_{it}
- To get consistent estimates, it must hold

$$E(\mathbf{x}'_{it}\xi_{is}) = \mathbf{0}, \quad \text{for all } s, t = 1, \dots, T$$

- which is a weaker form of the strict exogeneity assumption from FE estimation
 - Idiosyncratic errors are expected to be unrelated to regressors only conditional on individual slopes

FEIS Simulation

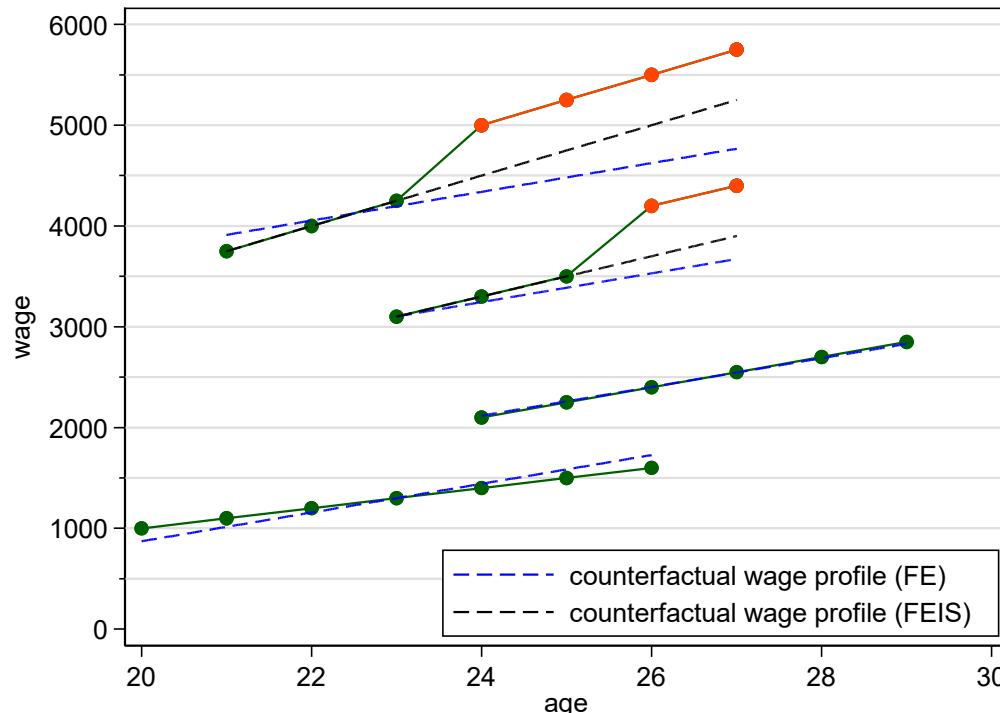
- Discrete MWP of +500 €

Those marrying are on a steeper wage profile (non-parallel trends)

- Slopes are 100, 150, 200, 250 respectively

- Consequently FE is biased upwards (blue counterfactuals)

- FEIS gets it right (black counterfactuals)



	xtreg wage marr age, fe //FE	reg wage marr ibn.id ibn.id#c.age, nocon //POLYSIS	xtfeis wage marr, slope(age) //FEIS
marr	780	500	500
age	143		
id		-1000 -1500 -1500 -1500	
id#c.age		100 150 200 250	

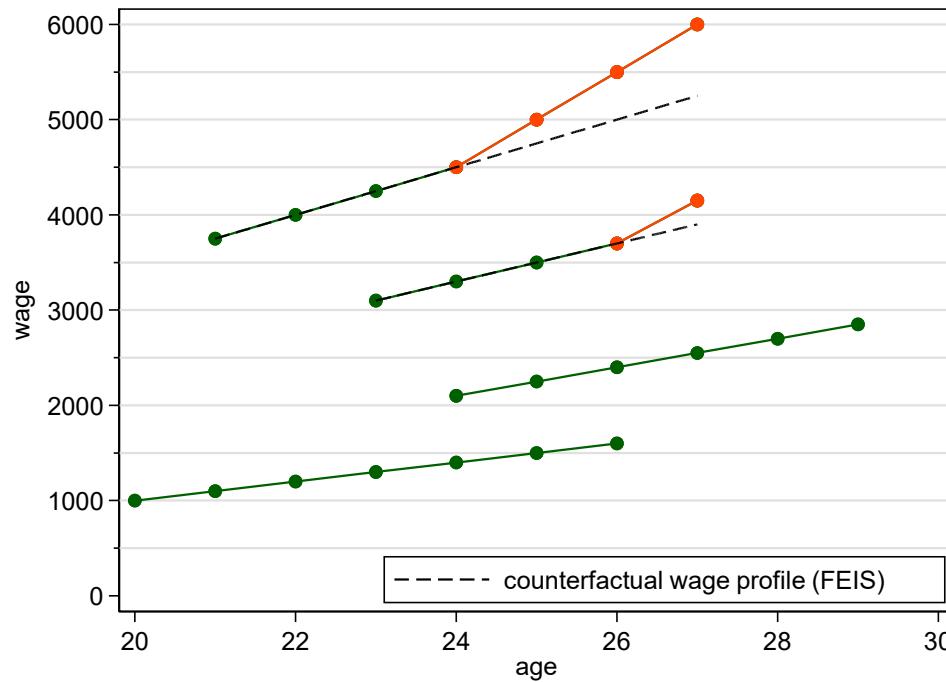
Do-File: FEIS Simulation.do

Does FEIS Overcontrol?

- Separating trends from treatment effects might be difficult
 - If FEIS is not able to separate trend and treatment effect, then FEIS will provide a biased estimate of the treatment effect
- FEIS is often used with macro panel data
 - In analogy to LSDV, FEIS can be estimated by including individual-specific trends in a POLS model (POLYSIS)
 - This has often been done with macro panel data
 - In analyzing the effect of minimum wages on employment growth on a state level, Meer/West argue that POLYSIS attenuates the real causal effect (in fact they do not use POLYSIS but DiD):
 - “We show that … research designs incorporating state-specific time trends are prone to erroneously estimated null effects” (Meer/West 2015: 2)
 - Their argument is that the state-specific time trends capture part of the treatment effect, if treatment changes the trend
 - Their argument seems intuitively plausible
 - However, our simulations below show that this is not true, if the treatment effect is correctly specified
 - Probably in their simulations the treatment effect is not specified properly

Does FEIS Overcontrol?

- **Accumulating MWP of +0 €, +250 €, +500 €, +750 €**
Those marrying are on a steeper wage trend (non-parallel trends)
- Consequently FE is biased upwards
- FEIS gets it right
- **Thus, there is no indication that FEIS does overcontrol!!**



	xtreg wage1 i.marrdur age, fe	//FE
	reg wage1 i.marrdur ibn.id ibn.id#c.age, nocon	
	xtfeis wage1 mdur?, slope(age)	//FEIS
	-----	-----
Variable	FE	POLSIS
marrdur		
1	194	0
2	542	250
3	964	500
4	1336	750
id		
age	128	
1		-1000
2		-1500
3		-1500
4		-1500
id#c.age		
1		100
2		150
3		200
4		250

Do-File: FEIS Simulation.do

Happiness Example: FEIS

```
xtfeis happy marry loghhinc, slope(age) cluster(id) //FEIS1
```

Variable	FE	FEIS1	FEIS2	FEIS3
marry	0.1675 7.3696	0.1599 6.3837	0.1616 6.5204	0.1462 5.3227
age	-0.0412 -23.9129		-0.0437 -21.5316	
loghhinc	0.1255 10.1681	0.0796 5.9522		
N	118201	118201	118201	113266

Data: Happiness2.dta

Do-File: Happiness 5 Regressions.do

legend: b/t

- We allow for heterogeneous growth (FEIS1), heterogeneous income effects (FEIS2), or both (FEIS3)
 - Note that N is smaller, because all models are estimated with $T \geq 3$ (FEIS3 $T \geq 4$)
 - Given the explorative results from above it is not surprising that the marriage effect is not much affected by allowing for heterogeneous growth
 - However, the income effect is strongly reduced by allowing for heterogeneous growth
 - Those on a less declining happiness trajectory get more likely an income increase

Testing for Heterogeneous Growth

- An explorative test on the group level
 - Interact the linear age trend with an treatment indicator (FE)

xtreg happy marry i.treat##c.age loghhinc, fe vce(cluster id)						
	happy	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
	marry	.1321504	.0233535	5.66	0.000	.0863746 .1779262
	age	-.0446538	.0024492	-18.23	0.000	-.0494546 -.0398531
	1.treat##c.age	.0076853	.0033946	2.26	0.024	.0010315 .0143391

- We see that the treatment group shows a significantly less steeper happiness decline
- However, this might simply be due to the positive treatment effect (that in this specification is only controlled by the marriage dummy)
- Thus, we include distributed FEs up to $t = -6$

xtreg happy i.ym i.treat##c.age loghhinc, fe vce(cluster id)						
	happy	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----						
					
	age	-.0446845	.002449	-18.25	0.000	-.0494849 -.0398842
	1.treat##c.age	.0099943	.0088322	1.13	0.258	-.0073178 .0273065

- Now there is no longer a significant non-parallel trend



Ludwig, Volker and Josef Brüderl (2018) Is There a Male Marital Wage Premium? New Evidence from the United States. ASR.



The Male Marital Wage Premium (MWP)

- Married men earn more than unmarried men
 - “... one of the most well documented phenomena in social science”
(Waite & Gallagher 2000: 99)
- Early studies used cross-sectional data
 - Self-selection: high wage men more attractive marriage partners
- However, also recent longitudinal studies find a MWP
 - Ahituv/Lerman (2007) Demography
NLSY79, FE (fixed-effects) regression: 7.6 %
 - Pollmann-Schult (2011) European Soc. Rev.
SOEP 1985-2008, FE regression: 4.2 %
- Thus, using the best available data and methodology,
it seems marriage makes men more productive workers
 - Remark: Not the effect on labor hours is investigated here,
but the effect on productivity (gross hourly wage rate)
- However, we are not convinced
 - Self-selection may operate on wage growth (not only on level)

Explanations for a Causal MWP

- Family economics (Becker 1981)
 - Precondition: there is a traditional division of labor
 - Married men specialize on market work
They accumulate more market specific skills
 - Married women specialize on household work
Married men are released from strenuous housework
They can put more effort in their market work
- Lifestyle explanation
 - After marriage men are domesticated by their wives
- Demand side explanation
 - Paternalism of employers

specialization

work effort

domestication

employer favoritism

Arguments for a Spurious MWP

- (Self)-selection of high wage males into marriage
 - They gain more from specialization and therefore are more willing to marry
 - They are more attractive marriage partners
 - Due to their higher wage
 - Due to other unobservables correlated with wage e.g. cognitive skills, social skills, beauty
- It is not only level, but also „steepness“ of the career
 - Promising young men (steep wage-profile) are attractive partners
 - Standard FE models yield upwardly biased estimates
- To get unbiased estimates one should use FEIS
$$\ln(w_{it}) = x_{it}\beta + \gamma m_{it} + \alpha_{1i} + \alpha_{2i} exp_{it} + \alpha_{3i} exp_{it}^2 + \xi_{it}$$
where exp is labor market experience

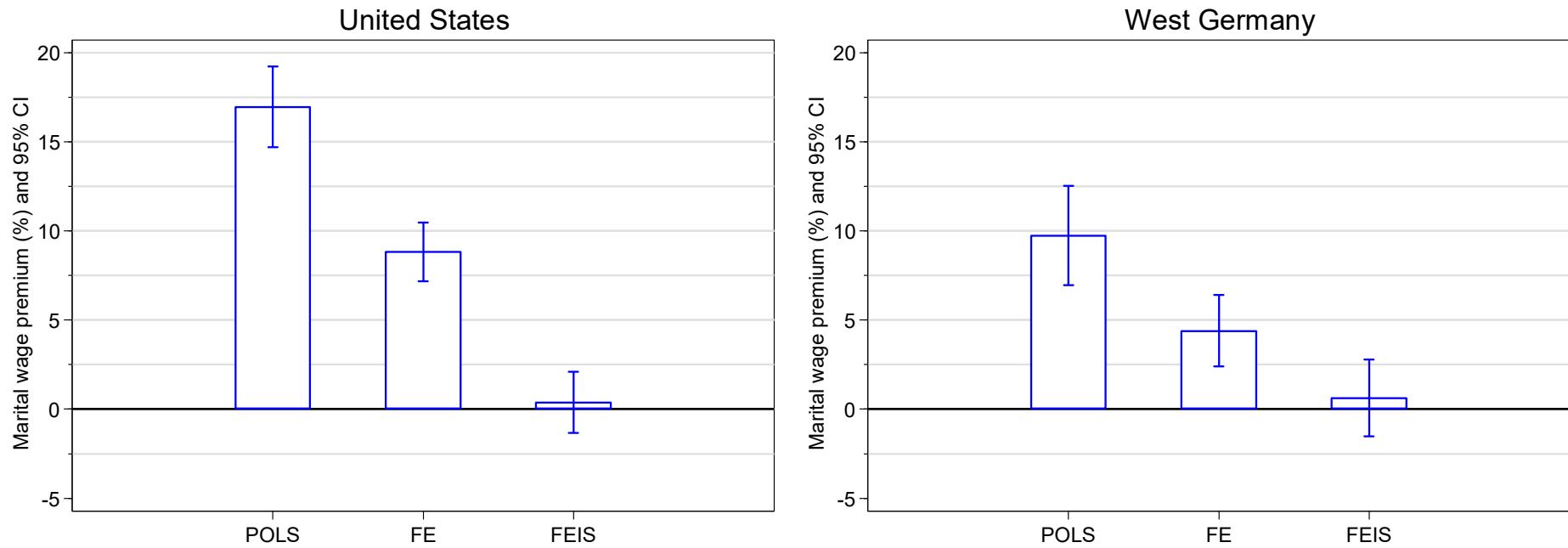
Data and Variables

- SOEP v27, 1984-2010 and NLSY79, 1979-2004
- Outcome: deflated log. hourly wage
 - NLSY: reported hourly gross earnings
 - SOEP: monthly gross earnings, divided by actual work hours * 4.36
- Treatment: marital status
 - Derived from (monthly) marriage biography, 4 states
 - Never-married, 1st marriage, separated/divorced, remarriage
- Control variables
 - Labor market experience (linear and squared)
 - Years of education
 - Currently enrolled in education
 - Firm tenure (yrs.)
 - Number of (biological) children
 - Dummies for survey year

Sample Restrictions

- SOEP
 - West German resident males
 - Cohorts 1946 to 1975
 - No self-employees, private sector workers
 - Never-married when first observed, at least 4 obs. (N=1,520)
- NLSY79
 - Males
 - No self-employees
 - Never-married when first observed, at least 4 obs. (N=4,452)

Results on the MWP



POLS grossly overestimates the MWP. Our FE results replicate the results found by Ahituv/Lerman (2007) and Pollman-Schult (2011). However, as the FEIS results show, these estimates are still too high.
In both countries there is no MWP!

Chapter VII:

Further Linear Panel Models

Section: Mixed-Coefficients Panel Models

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

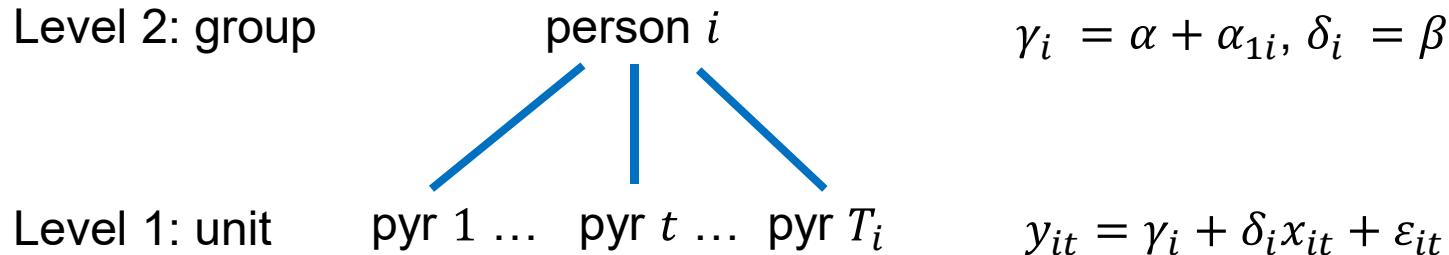


Mixed-Coefficients Panel Models

- Panel data are a special case of multi-level data
 - On the lower level (level-1) we observe person-years that are nested in persons (the higher level-2)
 - Therefore, models for multi-level data (synonym: hierarchical linear models) can also be applied to panel data
- In Stata terminology these are “mixed-effects models”
 - Some coefficients have fixed values (across persons): fixed effects
 - Other coefficients are random (across persons): random effects
 - Obviously this terminology is completely confusing. Therefore, I use the terms “fixed coefficients”, “random coefficients”, and “mixed-coefficients models”
- Further, random coefficients can be treated as
 - Random-effects
 - I.e., (exogenous) random variables drawn from a (normal) distribution
 - Fixed-effects
 - I.e., (potentially endogenous) random variables that are differenced out
- Within the multi-level tradition it is standard to assume that random coefficients are “random-effects”

The Basic Model

Level 2: group



$$\gamma_i = \alpha + \alpha_{1i}, \delta_i = \beta$$

Level 1: unit

$$y_{it} = \gamma_i + \delta_i x_{it} + \varepsilon_{it}$$

- The level-1 equation allows for person-specific coefficients

$$y_{it} = \gamma_i + \delta_i x_{it} + \varepsilon_{it}, \quad \text{where } \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$$

- Random intercept model (RI)

$$\begin{aligned} \gamma_i &= \alpha + \alpha_{1i}, \text{ where } \alpha_{1i} \sim N(0, \sigma_{\alpha_1}^2) && \text{intercept random coefficient} \\ \delta_i &= \beta && \text{slope fixed coefficient} \end{aligned}$$

- This gives $y_{it} = \alpha + \beta x_{it} + \alpha_{1i} + \varepsilon_{it}$
- Obviously, this is equivalent to the RE-model (if we treat α_{1i} as a random-effect, otherwise it would be the FE model)!
- The RI model usually is estimated by ML not GLS

- Random slope model (RS)

$$\begin{aligned} \gamma_i &= \alpha + \alpha_{1i} && \text{intercept random} \\ \delta_i &= \beta + \alpha_{2i} && \text{slope random} \end{aligned}$$

- This gives $y_{it} = \alpha + \beta x_{it} + \alpha_{2i} x_{it} + \alpha_{1i} + \varepsilon_{it}$

The Random Slope Model

- This is the random slope (in X only) model

$$y_{it} = \alpha + \mathbf{w}_{it}\boldsymbol{\varrho} + \beta x_{it} + \alpha_{2i}x_{it} + \alpha_{1i} + \varepsilon_{it}$$

- Main assumptions

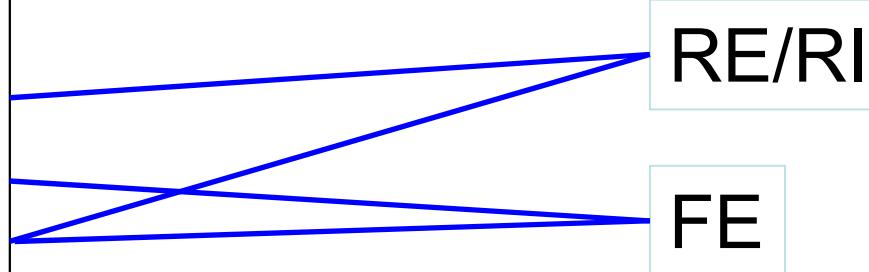
- $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$
- $\begin{bmatrix} \alpha_{1i} \\ \alpha_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\alpha_1}^2 & \tau_{12} \\ \tau_{21} & \sigma_{\alpha_2}^2 \end{bmatrix}\right)$ random coefficients follow a bivariate normal
- $E(\mathbf{w}'_{it}\varepsilon_{is}) = \mathbf{0}$, for all $s, t = 1, \dots, T$ strict exogeneity assumption
- $E(\mathbf{w}'_{it}\alpha_{1i}) = \mathbf{0}$, $E(x_{it}\alpha_{1i}) = 0$ random-effects assumption 1
- $E(\mathbf{w}'_{it}\alpha_{2i}) = \mathbf{0}$, $E(x_{it}\alpha_{2i}) = 0$ random-effects assumption 2

- Obviously, these are strong exogeneity assumptions

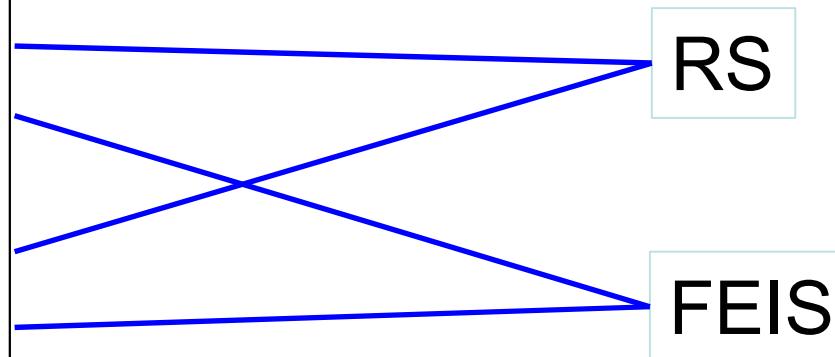
- In particular, the RS model also assumes no person-specific time-constant unobserved heterogeneity (random-effects assumption 1)
- Further, individual slopes are not allowed to correlate with the regressors (random-effects assumption 2)
- The strict exogeneity assumption is weaker compared to RE, because ε_{it} is purged by the term $\alpha_{2i}x_{it}$

Comparing Models

- Model class I
 - Random intercept
 - Random-effect
 - Fixed-effect
 - Fixed slope



- Model class II
 - Random intercept
 - Random-effect
 - Fixed-effect
 - Random slope
 - Random-effect
 - Fixed-effect



Comparing Models

- To clarify things we compare RS with FEIS
- The RS model (including time-constant z_i)

$$y_{it} = \underbrace{\alpha + w_{it}\varrho + z_i\zeta + \beta x_{it}}_{\text{fixed part}} + \underbrace{\alpha_{2i}x_{it} + \alpha_{1i}}_{\text{random part}} + \varepsilon_{it}$$

- Here the random coefficients are assumed to be “**random-effects**”
- Now the FEIS model

$$y_{it} = \underbrace{\cdot + w_{it}\varrho + \cdot + \cdot}_{\text{fixed part}} + \underbrace{\alpha_{2i}x_{it} + \alpha_{1i}}_{\text{random part}} + \varepsilon_{it}$$

- where a “.” means that the term is not estimable
- Here the random coefficients are assumed to be “**fixed-effects**”
- Exogeneity assumptions RS and FEIS
 - FEIS: $E(w'_{it}\varepsilon_{is}) = \mathbf{0}$, for all $s, t = 1, \dots, T$
 - RS: the same plus random-effects assumptions (see above)

Estimation of the RS Model (“Age” random)

. mixed happy loghhinc age id: age, vce(cluster id) cov(unstructured) stddev						
Mixed-effects regression						
Group variable: id						
Log pseudolikelihood = -219580.46						

	Robust					
happy	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
loghhinc	.1621011	.0102715	15.78	0.000	.1419693	.1822328
age	-.0297677	.001043	-28.54	0.000	-.0318119	-.0277236
_cons	6.293142	.1075543	58.51	0.000	6.082339	6.503944
-----+-----						
-----	Robust					
Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]			
-----+-----						
id: Unstructured						
sd(age)	.064252	.0015884	.0612129	.0674419		
sd(_cons)	1.947417	.037498	1.875292	2.022316		
corr(age,_cons)	-.8497485	.0068961	-.8627184	-.8356613		
-----+-----						
sd(Residual)	1.296047	.0062689	1.283818	1.308392		
-----+-----						

Interpretation of the RS Model

- This is a complex model, estimation takes quite long (a few min.)
- There is no R^2 available here (ML model!): use Pseudo- R^2 , etc.
- Fixed coefficients are interpreted as usual
- What do we learn from the random coefficients?
 - The coefficients are from a normal distribution with mean 0 and standard deviation $sd(.)$
 - The fixed age slope is -0.030
 - The random age slopes have a s.d. of 0.064, which is significant
 - “There is significant individual variation in the age slope”
 - This means that 95% of the individual slopes lie within
$$-0.03 \pm 1.96 \times 0.064 = [-0.156, +0.096]$$
 - Thus, there are also zero or even positive happiness GCs
 - We learn that the social world is complex, and that everything is possible!
Do we really want to know this?
 - Random intercept and slope correlate with -0.85
 - The higher the intercept, the lower the slope (steeper decline)

Comparing the Models

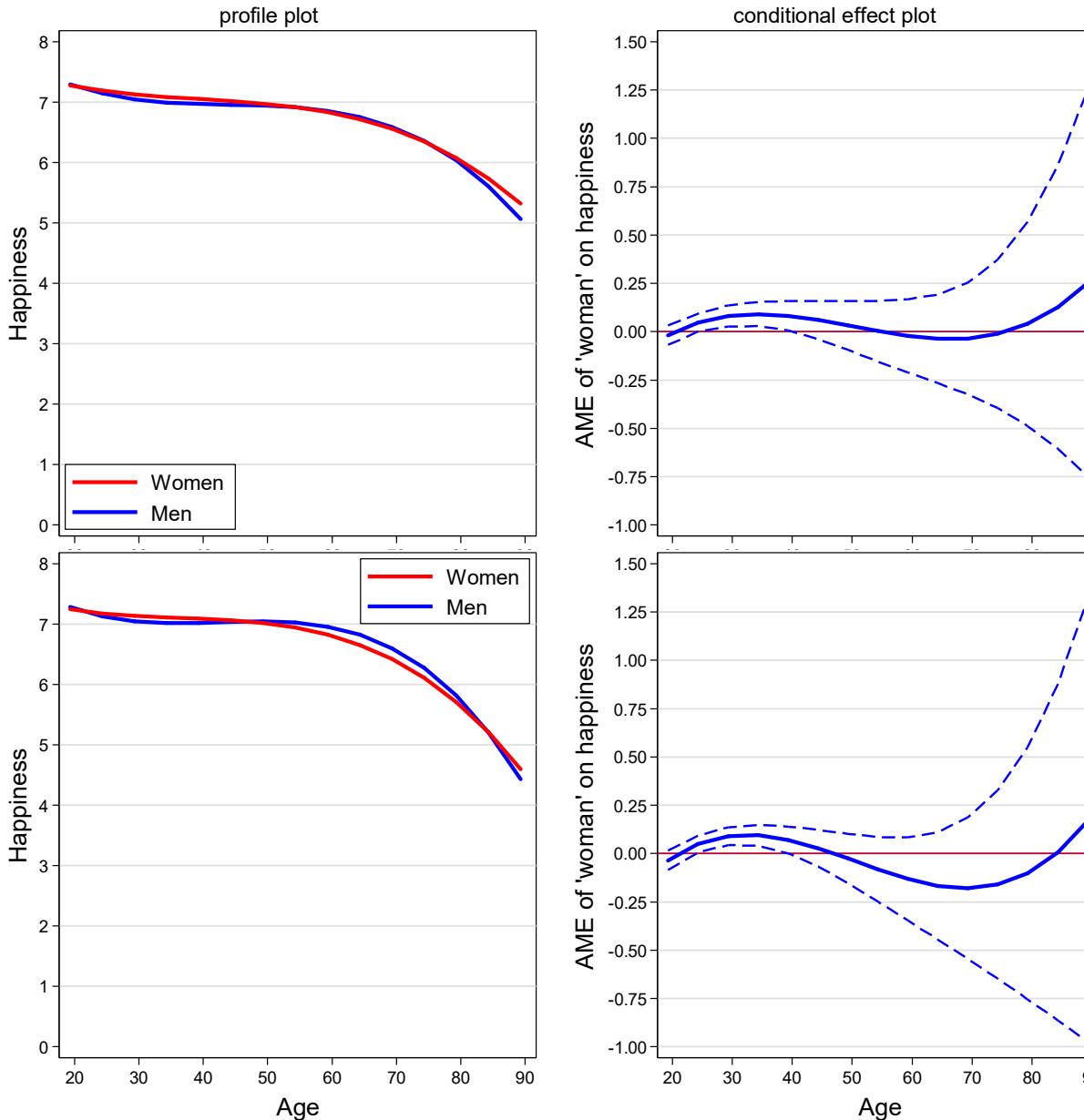
Variable	RE (GLS)	RI (ML)	RS	FE	FEIS
loghhinc	0.1692 15.9416	0.1686 15.8846	0.1621 15.7817	0.1323 10.7779	0.0824 6.1668
age	-0.0251 -24.4867	-0.0252 -24.5245	-0.0298 -28.5415	-0.0361 -23.2163	
N	121919	121919	121919	121919	118201
Data:	Happiness2.dta				legend: b/t
Do-File:	Happiness 5 Regressions.do				

- Here we focus on the income effect (“age” random-coefficient)
- RE (GLS) and RI (ML) provide very similar estimates
 - As it should be, because these are identical models but estimated with different methods
- The RS estimate of the income effect is similarly biased as RE
 - Allowing for random age slopes does not improve the estimate!
 - More formally you could do a LR-Test, compute BIC, etc.
- FEIS detrends the data for the individual happiness growth curves
 - This diminishes the income effect, because those on a less declining happiness trajectory get more likely an income increase

Group Specific Growth Curves

- Group specific GCs are often motivated within a mixed-coefficients context
 - Example: sex specific GCs (male $S = 0$, female $S = 1$) [age is A_{it}]
 - Consider a simple level-1 linear growth curve model
$$y_{it} = \gamma_i + \delta_i A_{it} + \varepsilon_{it}$$
 - Now the level-2 equations for the person-specific coefficients
$$\begin{aligned}\gamma_i &= \alpha + \vartheta S_i + \alpha_{1i} && \text{intercept group specific and random} \\ \delta_i &= \beta + \theta S_i + \alpha_{2i} && \text{age slope group specific and random}\end{aligned}$$
 - Plugging in gives:
$$y_{it} = \alpha + \vartheta S_i + \beta A_{it} + \theta(S_i \times A_{it}) + \alpha_{2i} A_{it} + \alpha_{1i} + \varepsilon_{it}$$
 - This is a RS model with sex specific GCs (RS-GC)
 - The RE-GC model that we estimated above is a special case with $\alpha_{2i} = 0$
 - Note that this model contains a so-called “cross-level interaction” $S_i \times A_{it}$
 - This is a simple kind of “latent growth curve model” (LGCM)
 - LGCMs are usually specified in the context of SEM, and allow for even more flexibility in modeling random coefficients, measurement errors, etc.
 - This easily leads to over-complex models that nobody can understand

Happiness Example: Sex Specific RS-GCs



We compare
RE-GCs (above)
RS-GCs (below)

Specification:

- cubic age
- cohort dummies
- period dummies (res.)

Random slopes:

- only linear age term
("this is perfectly reasonable" Rabe-Hesketh/Skrondal, 2008: 219)

Estimation of RS-GCs
takes quite long (ca. 15
minutes)

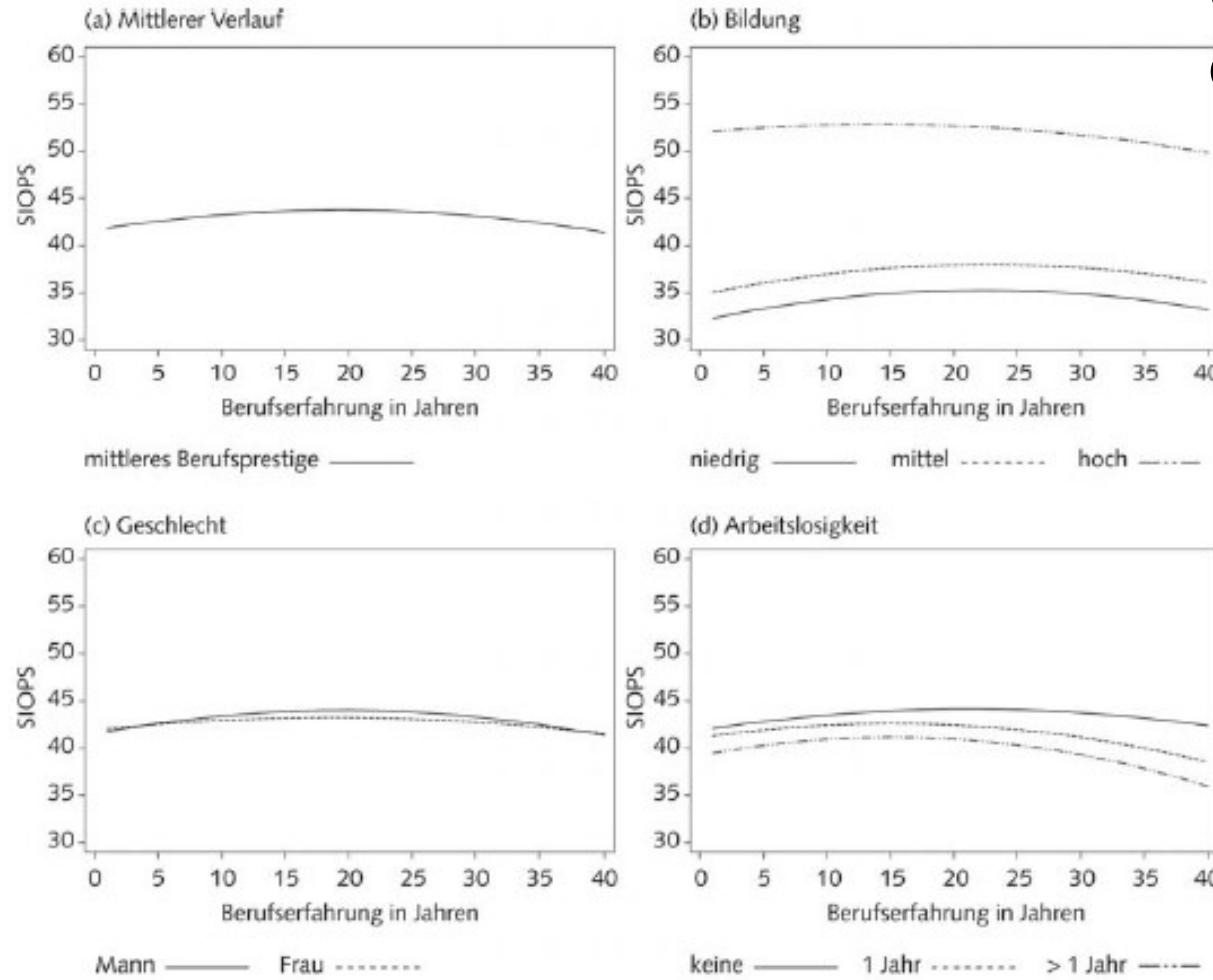
Results are very similar

Data: Happiness2.dta
Do-File: Happiness 5 Regressions.do

When To Use RS Models?

- Raudenbush (2001) strongly advocates RS models
 - He seems to be very enthusiastic about the potential of RS-GC modeling, and presents several nice examples
- However, this has to be differentiated
 - RS-GC models are in fact very helpful for describing personal growth trajectories, if it is reasonable to assume that no self-selection is operating
 - E.g., differences in vocabulary growth between boys and girls
 - However, if one wants to investigate the effects of events on outcomes, it is likely that some sort of self-selection is going on. Then one should use FE (FEIS) methodology
 - E.g., the effect of visiting a preschool on vocabulary growth
- **After all RS model estimates will be biased, if the strong random-effects assumptions do not hold**
 - Sometimes one can read statements like this: „RS models control for unobserved heterogeneity, because they include $\alpha_{2i}x_{it} + \alpha_{1i}$ “
 - Such a statement is nonsense: RS in fact rests on a weaker strict exogeneity assumption, however the random-effect assumptions are still very strong exogeneity assumptions

RS Exercise



- Stawarz (2013) estimates RS-GCs
 - Data: SOEP 1984-2009
 - Outcome: occupational prestige (SIOPS)
 - Time: labor market experience
 - GC: quadratic
 - RS: random intercept, random linear and quadratic slope terms
 - Controls: nationality, education parents, West, GDP, cohort and period grouped

Here are some results from his model. Might some of these results be biased due to self-selection?

LGCM Exercise

- Hooghe et al. (2014) analyze the impact of education and intergroup friendship on the development of ethnocentrism
 - They use three-wave panel data from Belgia (ages 16,18, 21)
 - They model group-specific GCs in a SEM framework (LGCM)
 - They include time-constant education and friendship at time 1
 - Plus indicators of education and friendship change (only on slope)

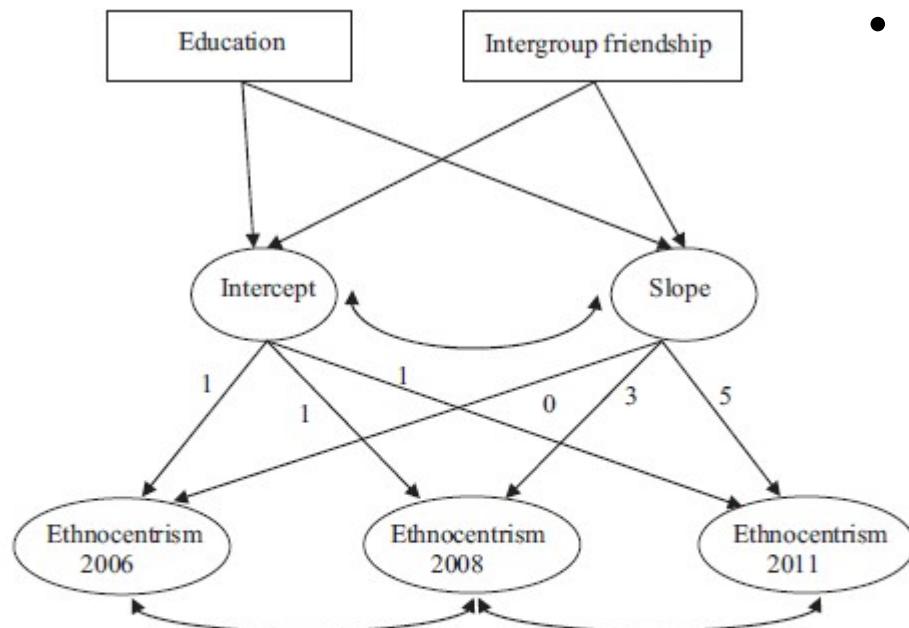
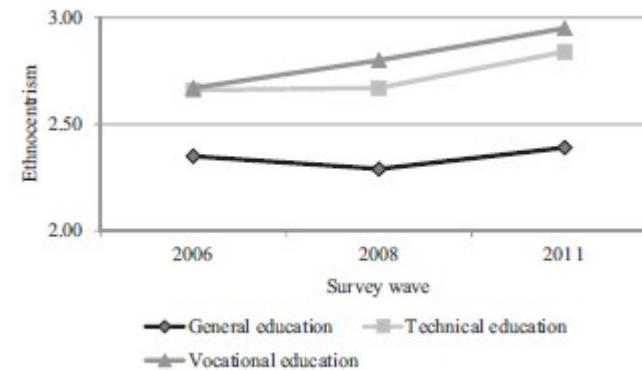


Figure 1 Two-level Latent Growth Curve Model for intra-individual variability of ethnocentrism explained by education level and intergroup friendship. Intercept loadings constrained to 1; slope loadings constrained to 0, 2 and 5

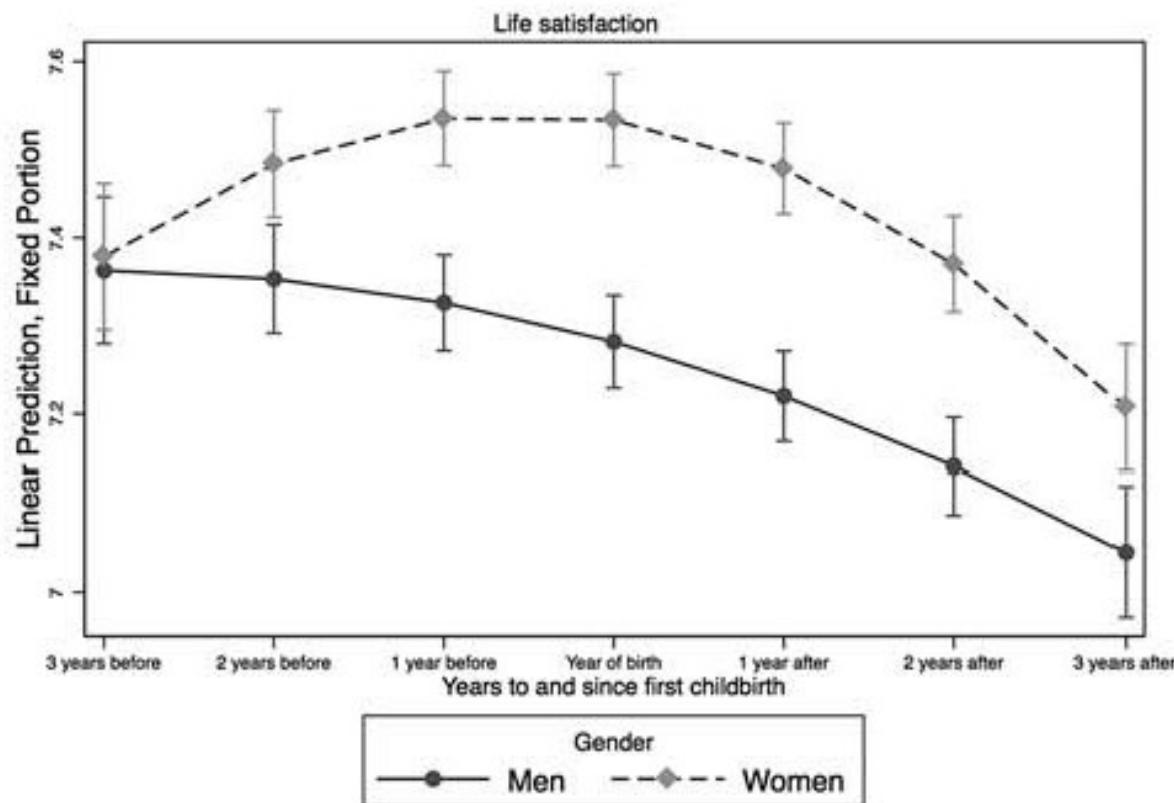
- **Findings**
 - Education lowers ethnocentrism



- Friends from different ethnic background reduce ethnocentrism (contact hypothesis)

What problems do you see?

Causal Effects via RS-GCs?



- Bernardi et al. (2017) investigate the kids-effect on happiness
- SOEP 1984-2013
- 3 years before/after first birth
- RS-GCs (linear time is RS)

What can we learn from this result on the causal effect of children on happiness?

Chapter VII:

Further Linear Panel Models

Section: The Hybrid Model

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Hybrid Model

- Often it is argued that “a major drawback” of FE is that one cannot estimate effects of time-constant variables
 - As argued above this statement is nonsense
 - Nevertheless, a hybrid model – a “compromise model” – features prominently in the literature
 - Proposed by Allison (2009) [more details in Schunck 2013]
- A hybrid model is a RE-model with decomposed variables
 - Decompose each time-varying predictor into two components
 - Between component: $\bar{x}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} x_{it}$ (person-specific mean)
 - Within component : $x_{it} - \bar{x}_i$ (demeaned variable)
 - Put both components into a RE-model

$$y_{it} = \alpha + \beta(x_{it} - \bar{x}_i) + \gamma\bar{x}_i + \delta z_i + \alpha_i + \varepsilon_{it}$$

- $\hat{\beta}$ reproduces exactly the FE estimate
- $\hat{\gamma}$ reproduces (approximately) the BE estimate
- $\hat{\delta}$ is the effect of a time-constant regressor

Mundlak Model

- Algebraically equivalent variant

$$y_{it} = \alpha + \beta x_{it} + (\gamma - \beta) \bar{x}_i + \delta z_i + \alpha_i + \varepsilon_{it}$$

- Mundlak-Model (Wooldridge 2010: pp. 332f)
- Known in the multi-level literature as “regression with contextual variable” (\bar{x}_i) (“contextual model”; Snijders/Bosker 1999: chap. 4.5)
- Advantage: it includes the untransformed x_{it} ; one needs only one new variable (the person-specific mean)
- Disadvantage: the coefficient of \bar{x}_i is the difference of within and between effect (and not interpretable)
- A test of $\beta = \gamma$ provides an alternative to the Hausman test
 - Advantage: one can test single parameters, whereas the Hausman test is an overall test
 - If $\beta = \gamma$ the Hybrid Model becomes the RE Model (and RE reproduces the FE estimate)

Hybrid Model Estimates

Variable	BE	FE	HYBRID
marry	0.3427***	0.1668***	
mmary			0.3385***
dmary			0.1668***
age	-0.0094***	-0.0413***	
mage			-0.0091***
dage			-0.0413***
loghhinc	0.4802***	0.1245***	
mhinc			0.4893***
dhinc			0.1245***
woman	0.0585**	(omitted)	0.0629**

legend: * p<0.05; ** p<0.01; *** p<0.001

. test (mmary=dmary) (mage=dage) (mhinc=dhinc)

chi2(3) = 620.65
 Prob > chi2 = **0.0000 → use the FE-model**

Data: Happiness2.dta
 Do-File: Happiness 5 Regressions.do

Interpreting the Hybrid Model

- Interpretation
 - Coefficients of the within component ($\hat{\beta}$) are FE-estimates
 - Coefficients of the between component ($\hat{\gamma}$) are “pure” between (BE) estimates (purged from the within component)
 - They provide information on how much treatment and control group differ in Y on average (purged from within-effect, i.e. “selection pure”)
 - These coefficients are (in most cases) substantively of no interest!
 - Coefficients of time-constant variables (δ) are under control of both between **and** within components
 - They are still biased, however, if z_i correlates with α_i
- Hybrid models confuse many researchers
 - Due to the presence of BE-effects that are substantively meaningless
 - They try to give the BE-effects an interpretation (see next slide)
- If one has substantive interest in the effects of time-constant variables, it is better to use group-specific growth curves!

Mis-Interpreting the Hybrid Model

- Following Curran/Bauer (2011) many researchers mis-interpret the Hybrid estimates
 - “Our statistical approach [hybrid model] distinguishes between both effects ..., differentiating the effect of levels (effects based on comparisons between regions) and changes in levels (effects based on comparisons within regions)” (Schmidt-Catran/Spies, 2016: 246)
 - Different causal effects of level (\bar{x}) and change ($x - \bar{x}$)?
 - No: According to the counterfactual approach there is only **one** causal effect. A causal effect is always the effect of a “change” in X. It is impossible that the “level” of X could show a causal effect.
 - What differs, however, is the information from which we draw causal inferences. In a between approach, we compare levels of different persons and infer from this difference the causal effect of a change in X. In a within approach, we compare the levels over time within a person and infer from this difference the causal effect of a change in X.

Hybrid RS Model

- Allowing for random slopes of the within component

$$y_{it} = \alpha + \beta(x_{it} - \bar{x}_i) + \gamma\bar{x}_i + \alpha_{2i}(x_{it} - \bar{x}_i) + \alpha_{1i} + \varepsilon_{it}$$

- Snijders/Boskers (1999: 81) favor the Mundlak RS, but as Schunck (2013) argues the hybrid RS is preferable
 - What kind of model is this?
 - Since $\hat{\beta}$ is a FE estimate this is a FE model with heterogeneous effects!
 - $\hat{\beta}$ is an estimate of the average causal effect
 - $\sigma_{\alpha_2}^2$ estimates the variation around the average effect
 - This model further weakens the assumptions (compared to FE) needed to get a consistent estimate of the average causal effect $\hat{\beta}$
 - [Unfortunately, I could not find the assumptions of the hybrid RS in the literature. So these are my own thoughts.]
 - Random-effects assumption 1 is no longer needed
 - Strict exogeneity still must hold
 - And especially random-effects assumption 2 must hold

$$E(x_{it}\alpha_{2i}) = 0$$

i.e., there must not be a correlation between X and the unobserved slope

Happiness Example: Hybrid RS Model

. mixed happy mmarry dmarry ... id: dmarry, ...	
Mixed-effects ML regression	Number of obs = 121919
Group variable: id	Number of groups = 14634
Log likelihood = -219961.49	Wald chi2(7) = 2689.66
	Prob > chi2 = 0.0000

happy Coef. Std. Err. z P> z [95% Conf. Interval]	
-----+-----	
dmarry .1453811 .0213233 6.82 0.000 .1035881 .187174	
dage -.0416513 .0010499 -39.67 0.000 -.0437091 -.0395934	
dloghhinc .1156805 .0093932 12.32 0.000 .0972701 .1340908	
woman .0639885 .0200852 3.19 0.001 .0246222 .1033549	
...	
-----+-----	
	Data: Happiness2.dta
	Do-File: Happiness 5 Regressions.do

Random-effects Parameters Estimate Std. Err. [95% Conf. Interval]	
-----+-----	
id: Unstructured	
sd(dmarry) .7687146 .0207467 .7291085 .8104721	
sd(_cons) 1.075336 .0080151 1.059741 1.091161	
corr(dmarry,_cons) .1847294 .0308048 .1237281 .2443396	
-----+-----	
sd(Residual) 1.315927 .0028854 1.310284 1.321594	
-----+-----	

Happiness Example: Hybrid RS Model

- We investigate, whether there are heterogeneous marriage effects
 - We allow for random marriage slopes
 - The effects of the contextual variables are deleted from the output
 - To save computational time we do not compute panel-robust S.E.s
- Results
 - All effect estimates are very similar to FE
 - The FE marriage effect was 0.167, here it is 0.145
 - However, now we see that the marriage effects are quite heterogeneous: s.d. is 0.77
 - This means that 95% of the marriage effects are in the interval
$$0.145 \pm 1.96 \times 0.77 = [-1.23, +1.62]$$
 - The person-specific intercepts and the marriage effects correlate with 0.18
 - Happier people gain more happiness by a marriage

Hybrid the “Default” for Panel Analysis?

- Bell/Jones (2015) call the hybrid within-between RE (REWB). They argue that instead of FE, REWB should become the default
 - Because it provides FE estimates
 - Plus all the advantages of multilevel models
- In principle they are right. Some comments on their arguments:
 - REWB provides additionally between effects
 - With individual panel data these are (mostly) substantively meaningless
 - REWB provides additionally effects of time-constant regressors
 - In this respect, RE-GCs are much more informative
 - REWB allows for cross-level interactions
 - It is tricky to get these correctly (cf. Schunck 2013)
 - It is easier to obtain these with FE (group-specific causal effects)
 - REWB allows for heterogeneous causal effects (hybrid RS)
 - Yes. But what do we do with this information?
 - “The social world is complex and everything is possible”?
 - REWB allows for more than two levels
 - Yes, this is attractive if you want to do FE with more than two levels
- Thus I would argue: use FE as a “default” because it is less complex
 - Only if you would like to do complex modelling (heterogeneous effects, more than two levels, context matters substantively) then REWB is an attractive model

Chapter VII:

Further Linear Panel Models

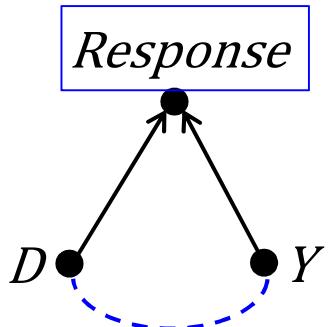
Section: Panel Regression with Missing Data

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Missing Data

- Sources of missing data in panel data (Young/Johnson 2015)
 - Unit-nonresponse in the first wave
 - Within-wave missingness (gap in the data)
 - Item-nonresponse
 - Question not included in every wave
 - Whole-wave missingness
 - Temporary unit-nonresponse (gap in the data)
 - Permanent unit-nonresponse (panel attrition)
- Nonresponse (NR) might bias regression estimates
 - If only (observable) variables affect response (MAR)
 - If both the treatment **and** the outcome affect response (NMAR)
 - Endogenous selection bias (Elwert and Winship 2014)
- Methods to deal with the problem (under MAR)
 - Regression controlling for response variables (complete case analysis)
- Methods to deal with the problem (under NMAR)
 - Selection models (are not robust; “technification”)
 - Weighting (developed for first-wave nonresponse and attrition)
 - Multiple imputation (mainly for gaps, but also attrition)



Weighting Panel Data

- With weighting one tries to compensate differential NR
 - Units that are “underrepresented” in the sample get larger weights
- With panel data there are different types of weights
 - Cross-sectional weights (CW)
 - Are constructed by adjusting a panel sample (cross-section) to population data (post-stratification weight)
 - Should compensate for unit-nonresponse in the first wave
 - Could (in principle) be constructed also for later waves
 - Longitudinal weights (LW)
 - $LW_t = 1/P_t$ where P_t is the probability of participating in wave t
 - P_t is estimated from a logit regression on response/nonresponse in wave t conditional on regressors from wave $t - 1$
 - Should compensate for permanent unit-nonresponse
 - Panel weights (PW): these are finally the weights for the pyrs
$$PW_t = CW_1 \times LW_2 \times \cdots \times LW_t$$
 - If there is a gap: $LW = 1$

Weighting Panel Data: Example

		Year 1				Year 2	
population	A	1000			80% Surv.	800	
	B	1000			90% Surv.	900	
sample	A	50% RR	50		20% RR		10
	B	100% RR	100		50% RR		50
weighted sample	A		1/0.5 = 2	100		1/0.5*1/0.2 = 10	100
	B		1/1 = 1	100		1/1*1/0.5 = 2	100

- We draw a 10% random sample from a population. Group A has a response rate (RR) of 50%, group B has 100%. Thus the resulting sample is not “representative” of the population.
- By comparing the sample group distribution with the population distribution, we see that members of group A get the cross-sectional weight 2, group B gets weight 1
- The weighted sample is then again representative of the population
- In year 2 group A has a RR of 20%, B of 50%. The respective longitudinal weights are $1/0.2=5$ and $1/0.5=2$. Multiplying the CWs from year 1 and the LWS from year 2 we arrive at the panel weights 10 resp. 2. Multiplying the sample with the panel weights brings the numbers back to the population distribution in year 1.
- We see, however, a drawback of this procedure: It is not guaranteed that this is representative of the population in year 2. If the population changes for instance due to mortality the weighted sample might deviate from the population.
- The problem could be solved if we would construct a new CW for year 2.

Example: Marriage and Happiness (pairfam)

- In this example we replicate the SOEP results with pairfam
 - We obtain very similar results
- In a next step we compare with the weighted results
 - As described above we compute the panel weights by multiplying up the cross-sectional weights and the longitudinal weights

```
gen pweight = dxpsweight if year==1 //W1: cross-sectional weight  
by sort id (year): //W2+: multiplying the lweight(t)  
replace pweight = pweight[_n-1] * lweight if year>1
```

- Some panel weights are absurdly high: mean = 1.74, max = 67.3
- Then we use areg to obtain an FE analysis with panel weights

```
areg happy marry age wave1 [pw=pweight], absorb(id) vce(cluster id)
```

 - xtreg does not accept weights that change over waves
 - Note: pairfam provides CWs only for wave1
 - The CW for wave 1 is already a combination of two weights
 - Post-stratification weights (comparing to Mikrozensus 2008)
 - Design weights (the three cohorts are not equally sized in the population)

Example: Marriage and Happiness (pairfam)

	FE	AREG
	unweight.	weighted
marry	0.272	0.289
	0.056	0.068
	4.88	4.27
age	-0.078	-0.074
	0.009	0.012
	-9.06	-6.40
wave1	-0.149	-0.139
	0.025	0.034
	-5.98	-4.09
N	27169	27169
N_clust	6572	6572

legend: b/se/t

- The unweighted FE results are very similar to the SOEP results!
 - New is the wave1-effect: probably due to panel-conditioning happiness was lower in wave 1
- Weighting doesn't change the results
 - Coefficients change only marginally
 - S.E.s are somewhat larger (some of this increase can already be seen in the unweighted areg)

Should we Use Weighted Regressions?

- Statistical theory says “yes”
- However, in practice it is unclear whether weighting really improves our analyses
 - It works only well, if our weighting models are well specified
 - In practice, they often will be mis-specified
 - Then weighted analyses might be even more biased than unweighted
 - Real designs are often much more complicated
 - New refreshment samples and grown-up children are integrated
 - This complicates the construction of weights enormously
 - Only real experts will grasp the complexity
- Recommendation
 - Do not use weighted panel analysis
 - Only if reviewers demand it

Multiple Imputation

- With multiple imputation (MI) missing values are imputed
 - MI is model based where the information available is used
- There are several challenges when using MI with panel data
 - With within-wave item nonresponse MI should use both within-wave and across-wave (panel) information
 - With whole-wave missingness we have to impute whole waves
 - Time-constant variables have to be fetched forward
 - Timing information has to be imputed (when happened the event?)
- Simulations by Young/Johnson (2015)
 - MI of within-wave missingness improves FE estimates (somewhat)
 - MI of whole-wave missingness does not improve FE estimates
- Recommendation
 - As with weighting the issues involved are extremely complex and careless usage of MI might easily increase bias
 - For instance, imputing time-constant variables or “not applicables” would mess up the whole analysis

Chapter VIII:

Non-Linear Panel Models:

Fixed-Effects Logit

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Nonlinear Panel Regression

- Linear model: modeling conditional means

$$E(y_{it}|x_{it}, \alpha_i) = \beta x_{it} + \alpha_i$$

- Person-specific errors are additive, we can difference them out

- Nonlinear model: modeling conditional densities

$$f(y_{it}|x_{it}, \alpha_i) = f(y_{it}, \beta x_{it} + \alpha_i, \gamma)$$

- Here person-specific errors can usually not be differenced out
 - Thus, with the parameters of interest (β, γ), one has to estimate the nuisance parameters ($\alpha_1, \dots, \alpha_N$), also termed incidental parameters. This creates the **incidental parameter problem**:

- As $N \rightarrow \infty$ the number of parameters to estimate goes to infinity, parameter estimates are inconsistent.
 - Solution I: RE-estimation
 - Solution II: in some special cases it is possible to condition out the nuisance parameters (conditional FE-models)

Nonlinear Panel Regression

- FE-model vs. RE-model
 - As in the linear case: if person-specific errors are present and are correlated with the regressors, then RE is biased but FE is consistent
 - A Hausman test can be used
- Interpretation of Coefficients (see Allison 2009: pp. 36)
 - With nonlinear models, the coefficients of pooled and FE models have different interpretations
 - Pooled models: population-averaged estimates
 - Marginal effects are identical for each subject
 - FE (and RE) models: subject-specific estimates
 - The marginal effect depends on α_i , therefore it differs for each subject
 - These are considered to be more accurate estimates of the underlying causal mechanism

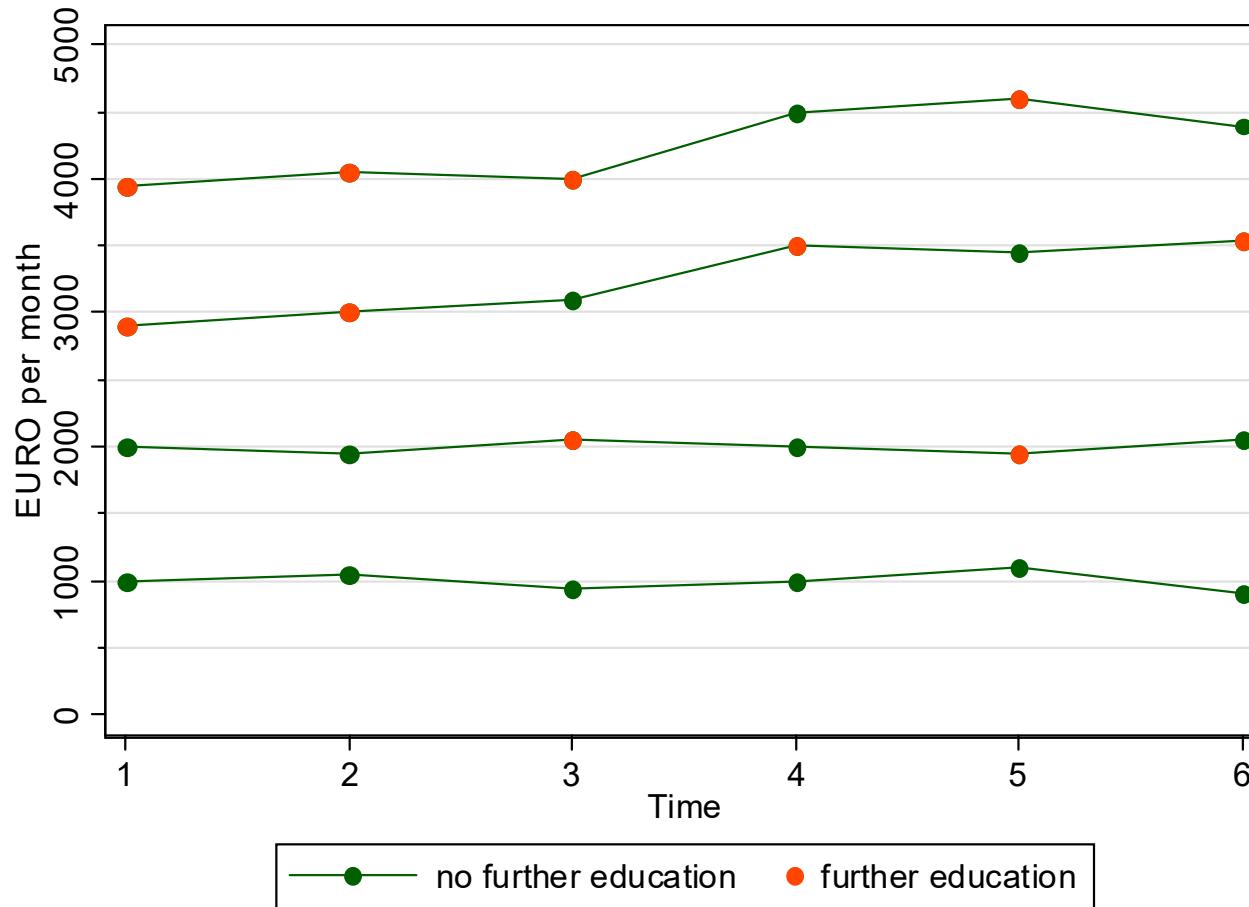
Fixed-Effects Logit

- Logistic regression model with fixed-effects

$$P(y_{it} = 1) = \frac{\exp(\beta x_{it} + \alpha_i)}{1 + \exp(\beta x_{it} + \alpha_i)}$$

- No incidental parameter problem, because the α_i can be conditioned out (conditional likelihood)
- Advantage of the FE-methodology: Estimates of β are unbiased even in the presence of time-constant unobserved heterogeneity
- Persons who have only 0s or 1s on the dependent variable are dropped. For FE-logit you need data with sufficient variance on Y , i.e., generally you will need panel data with many waves!
- Subject-specific estimates: Probability interpretation not possible! One has to use the sign interpretation (or odds)

Example: Further Education



Y: further education
X: wage
Does a wage raise
increase $P(\text{further educ})$?

The data show
a) a negative effect of a
wage raise
b) self-selection of high
wage males into further
educ

Data: Wage Premium.dta
Do-File: Wage FE Logit.do

Estimation Results

```
. replace wage = wage/1000      //wage rescaled  
. * Pooled-Logit  
. logit feduc wage
```

Data: Wage Premium.dta
Do-File: Wage FE Logit.do

feduc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
<hr/>					
wage	.9478211	.4256097	2.23	0.026	.1136414 1.782001
_cons	-2.93141	1.299289	-2.26	0.024	-5.47797 -.3848511

```
. * FE-Logit  
. xtlogit feduc wage, fe
```

note: multiple positive outcomes within groups encountered.

note: 1 group (6 obs) dropped because of all positive or all negative outcomes.

Conditional fixed-effects logistic regression Number of obs = 18
Group variable: id Number of groups = 3

feduc	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
<hr/>					
wage	-2.18767	2.32178	-0.94	0.346	-6.738276 2.362935

Linear Probability Model as Alternative

- Much easier to interpret is the LPM
 - We can interpret the coefficient estimates

```
. xtreg feduc wage, fe

Fixed-effects (within) regression                         Number of obs     =      24
Group variable: id                                     Number of groups  =       4

R-sq:  within  = 0.0606                                Obs per group: min =       6
          between = 0.9007                               avg =      6.0
          overall = 0.2455                               max =       6

corr(u_i, Xb)  = -0.9844                             F(1, 19)           =     1.23
                                                       Prob > F        = 0.2820

-----
          feduc |      Coef.    Std. Err.      t    P>|t| [95% Conf. Interval]
-----+
          wage |  -.5389222   .4866634    -1.11   0.282    -1.55752   .4796759
          _cons |  1.831338   1.280712     1.43   0.169    -.8492242  4.511899
-----+
```



Chapter IX:

EHA with Repeated Events

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Event History Analysis with Repeated Events

- Multiple episodes in the dataset
 - Problem: dependent episodes, biased S.E.
 - Potential: within estimation
- Analysis options
 - Analyzing time to first event only
 - Discards much information
 - Analyzing episodes separately
 - Does not make use of the within information
 - Pooled estimation
 - Biased S.E. and sub-optimal use of the within information
 - Random-effects models
 - Sub-optimal use of the within information
 - Fixed-effects models
 - Uses the within information (biased S.E., but vce() option)

Continuous-Time: Pooled- and RE-Cox

- Pooled-Cox
 - Pooling the data and applying Cox
 - Using panel-robust S.E.s (`vce(cluster id)`)
- Proportional hazards „shared frailty“ model (RE-Cox)

$$r_{ij}(t) = r_0(t) \exp(\beta x_{it}(t) + \alpha_i)$$

- i person index, j episode index
 - α_i person-specific error term
 - RE-model (`shared(id)`)
- α_i and $x_{ij}(t)$ are correlated (under normal circumstances)
 - Pooled estimates and RE-estimates will be biased

Continuous-Time: FE-Cox Regression

- FE-estimator via adding person dummies
 - Does not work (incidental parameter problem)
- FE-Cox: absorb α_i in the base rate (Allison 2009: chap. 5)

$$r_{ij}(t) = r_{0i}(t) \exp(\beta x_{it}(t))$$

- Stratified Cox regression: $r_{0i}(t)$ is not estimated (partial likelihood), thus there is no incidental parameter problem
- `stcox ... strata(id)`

Discrete-Time: FE-Logit

- Alternative: using discrete-time analysis
- Logistic regression model with fixed effects

$$P_{ij}(t) = \frac{\exp(c(t) + \beta x_{it}(t) + \alpha_i)}{1 + \exp(c(t) + \beta x_{it}(t) + \alpha_i)}$$

- $c(t)$ models duration dependence (often t and $\ln(t)$)
- Estimation via conditional likelihood

Modeling Effects of Time/History

- How to define analysis time?
 - Clock ticks from the first onset of the risk (elapsed time approach)
 - Clock is reset after each event (gap time approach)
 - In most situations the gap time approach will be more appropriate
- Modeling effects of history
 - Effect of „number of previous events“ (occurrence dependence, OD)
 - Positive (spurious) OD very likely, if observation window of fixed length
 - FE-estimates reduce the bias
 - Mechanics: is the hazard rate of episodes with a higher serial number really higher? (within persons!)
 - Similar arguments for effect of “cumulative duration” or “% of life time spent in the state X”

Chapter IX:

EHA with Repeated Events

Example: Duration of Unemployment

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

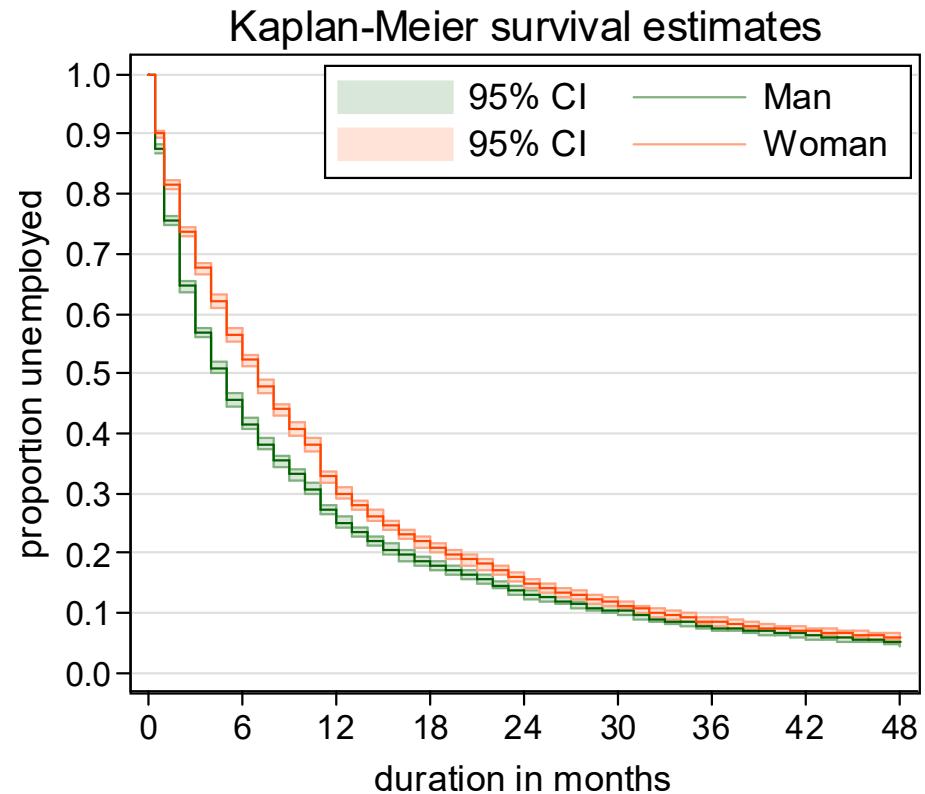
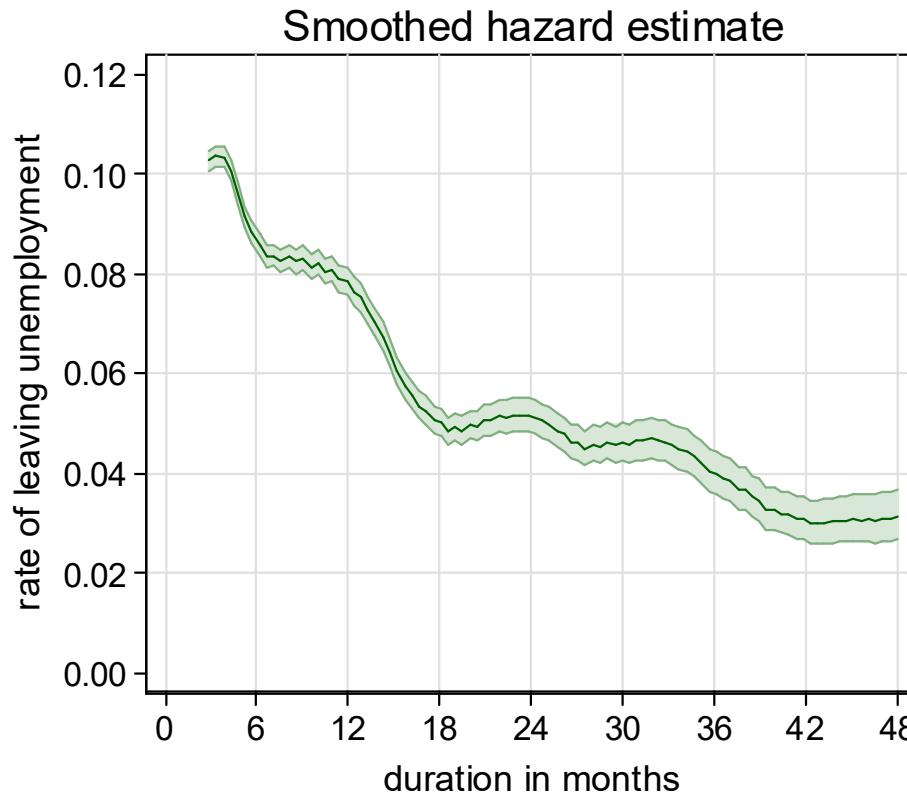


Example: Duration of Unemployment

- Data: SOEP 1984-2009 (v26)
 - All unemployment episodes from ARTKALEN 1983-2008
 - Left-censored episodes are dropped
 - Event: leaving unemployment (all destinations)
 - Gap time approach: clock starts at 0 with every new episode
 - The data set:
 - 10,949 respondents provide
 - 20,826 unemployment episodes (18,245 failures)
 - 230,412 month splits
 - Control variables:
 - Time-constant: gender, living in West 1989, migration background
 - Time-varying: age, education, period (spring, winter, seam, 2003-06)
- Research question
 - What is the effect of previous unemployment on the rate of leaving unemployment?
 - A negative OD is hypothesized (“scarring effect”)

Leaving Unemployment

Data: Unemployment.dta
Do-File: Unemployment 2 Regressions.do



The hazard rate peaks at three months, and then shows a monotonic decline (besides of “heaping” effects).

Men leave unemployment faster (within the first two years).

FE-Cox Results

```
stcox nprev age educ spring winter seam y0306, strata(id) efron vce(cluster id)
```

Stratified Cox regr. -- Efron method for ties

```
No. of subjects      =      17655          Number of obs     =   200448  
No. of failures     =      15616  
Time at risk        =    200448          Wald chi2(7)     =     415.87  
Log pseudolikelihood = -8143.5372      Prob > chi2      =     0.0000
```

(Std. Err. adjusted for 9171 clusters in id)

		Robust				
_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
nprev	.9068937	.0155766	-5.69	0.000	.8768722	.937943
age	.9919157	.0064574	-1.25	0.212	.9793398	1.004653
educ	1.023851	.0259907	0.93	0.353	.9741568	1.076081
spring	1.226209	.0426469	5.86	0.000	1.145408	1.31271
winter	.6272161	.0281982	-10.38	0.000	.5743137	.6849915
seam	1.226957	.0612777	4.10	0.000	1.112546	1.353134
y0306	.8353965	.0347889	-4.32	0.000	.7699199	.9064414

Stratified by id

We see negative OD (“scarring”)

Comparing Models for Repeated Events

Variable	Pooled Cox	Pooled Logit	FE-Cox	FE-Logit	
nprev	1.01	1.01	0.91***	0.81***	# previous unempl. epis.
age	0.98***	0.97***	0.99	0.99	age in years
female	0.83***	0.81***			female
west	1.11***	1.13***			living in West 1989
migrant	0.79***	0.77***			migration background
educ	1.07***	1.08***	1.02	1.11***	education in years
spring	1.31***	1.36***	1.23***	1.33***	February, March, April
winter	0.69***	0.67***	0.63***	0.64***	November, January
seam	1.68***	1.90***	1.23***	1.83***	December
y0306	0.82***	0.80***	0.84***	0.70***	period 2003-2006
t		0.99***		1.00	
lnt		0.90***		2.19***	
_cons		0.15***			
N	200448	200448	200448	181233	Splits
N_sub	17655	17655	17655		Episodes
N_fail	15616	15616	15616		Failures
N_g	9171	9171	9171	7613	Persons

legend: * p<0.05; ** p<0.01; *** p<0.001

All models with panel-robust S.E.s (FE-Logit with bootstrap)

FE-Cox keeps only persons, where a within comparison is possible (not seen in the output)

a) at least two failures,

b) one failure plus at least one censor: censored duration \geq failure duration

Contrary, the FE-Logit includes also persons with only one failure episode

It drops all persons with only 0s (or only 1s)

A Caveat on Estimating OD Effects

- A standard EHA model (pooled or RE-analysis) will provide upwardly biased OD estimates
 - In a fixed observation interval, subjects with more episodes will have (*ceteris paribus*) shorter durations (higher rates)
 - OD will be spuriously positive
- A FE-approach is not prone to this bias
 - Within subjects one looks, whether durations increase with the serial number of the episode (if yes: negative OD)
 - However, there might be a bias in the opposite direction
 - Chamberlain (1985) argued that censoring is no longer independent
 - Allison (1996) showed with simulations that there is a potential for a negative OD bias (also for “cumulative duration dependence”)
 - Especially severe if few episodes per subject (and/or many censored episodes)
 - Take negative OD found in FE-models also with caution!

Chapter X:

Limitations of the Within Methodology

Section: Limitations of Scope

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Limitations of the Within Methodology

- The main point with FE is that it discards potentially “**contaminated**” between variation (Allison 2009)
 - It uses only within variation to estimate the causal effect of an event
 - Thus, to identify the causal effect FE needs only the assumption that the *within variation is exogenous*
- Several implications follow:
 - 1) FE can only be applied, if there is within variation.
 - 2) For descriptive purposes it makes sense to use also the between variation.
 - 3) Because the FE estimator uses only the within variation, it is only generalizable to those units that show within variation, i.e. it is an ATT.
 - 4) If the within variation is not exogenous the FE estimator will yield biased estimates.

Is There Enough Within Variation?

- FE identifies the causal effect by using within variation only
 - Therefore, one has to ask for every variable of interest
 - whether there are enough changes? (categorical variables)
 - whether there is enough within variation? (metric variables)
- Example: education
 - When investigating the effects of human capital on wages, it is difficult to use within methodology, because education does not change during the labor market career of most people

Descriptive Questions

- In Social Sciences descriptive questions are also important
 - Some even argue that we first should have good descriptions of the social world, before we start with analyzing causal effects
- For descriptive questions one has to use between variation
 - For describing group differences use POLS
 - Not RE, because RE is „biased“ towards FE
 - In fact, panel data are sub-optimal for this purpose.
Trend data would be better for describing group differences!
- Example
 - What is the earnings difference between East and West Germans?
 - This is a descriptive question. Investigate it with POLS
 - What is the causal effect on earnings of moving from West Germany to East Germany?
 - This is a causal question. Investigate it with FE

Within Methods Provide Only ATTs

- When estimating causal effects, one always has to ask, what kind of treatment effect do I identify?
 - An „average treatment effect“ (ATE) that generalizes to the whole population
 - Or an “average treatment effects on the treated” (ATT) that generalizes only to those, who potentially can experience the treatment
- Within estimators identify an ATT
 - They use only within variation of those, who experienced an treatment
 - Therefore, we can only generalize to those, who potentially can/will experience the treatment
- For causal analysis ATTs are preferable
 - The „Not Treated“ could have a completely different treatment effect (effect heterogeneity) (ATC). Perhaps this is even the reason why they did not choose treatment. Therefore, the ATE is „biased“ by the ATC
 - Contrary, the ATT is what one wants: „what is the effect of a treatment, on those who will experience that treatment“?
 - In experiments we „force“ some subjects into treatment, who under „natural“ conditions never would have taken the treatment. Thus, an experiment identifies an ATE
 - However, policy measures should be based on ATEs, because with a policy measure one usually wants to „force“ some subjects into treatment, who under „natural“ conditions never would have taken the treatment

Within Methods Provide Only ATTs

- Example: marriage premium
 - Estimating the wage difference between married and unmarried people (by POLS) generalizes to the whole population. However, this is not a treatment effect, it is only a descriptive group difference
 - The marriage premium estimated with FE generalizes only to those, who eventually will marry. It does not apply to „hard-core-singles“
 - The latter may even stay single, because for them the treatment effect is completely different (i.e., a negative marriage premium)
 - In an experiment some „hard-core-singles“ would be „forced“ to marry. Therefore, the experiment will provide an ATE
 - Imagine we found a positive ATT: It is not guaranteed that policy measures introduced to increase marriage would rise the wage level
- Example: sex change
 - Sometimes one observes enough true sex changes. Applying FE to such data provides an ATT for transsexuals only

Chapter X:

Limitations of the Within Methodology

Section: Violations of Strict Exogeneity

- Unobserved Time-Varying Confounders
- Measurement Errors
- Reverse Causality
- Panel Attrition
- Panel Conditioning

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

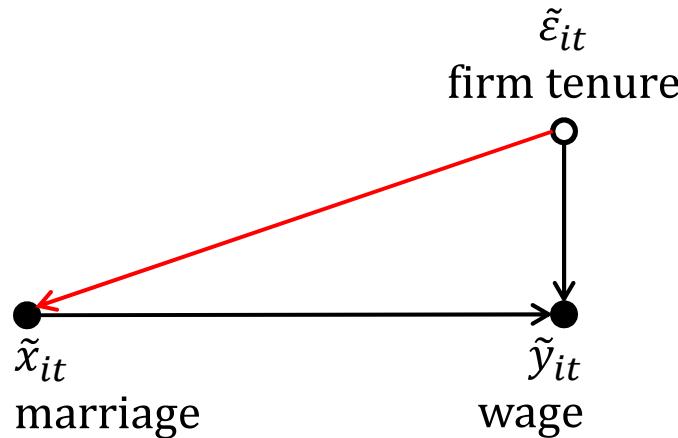


Endogeneity

- FE estimates are biased under endogeneity:
$$E(\mathbf{x}'_{is} \varepsilon_{it}) \neq \mathbf{0}, \text{ for some } s, t = 1, \dots, T$$
- Note that endogeneity can have several sources
 - Unobserved time-varying confounders
 - Errors in reporting X (measurement errors)
 - Y affects also X (reverse causality)
 - Endogenous selection bias (collider bias, attrition)
- Endogeneity thus is the consequence of substantive mis-specification or methodological deficiencies
 - When arguing for exogeneity, one has to make it plausible, that **none** of the four sources is at work

Unobserved Time-Varying Confounders

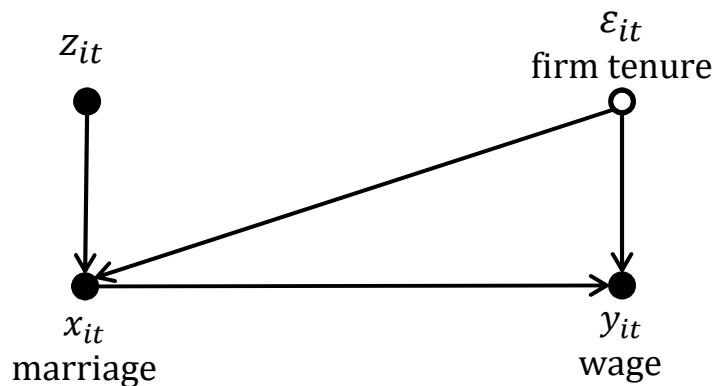
- Unobservables that affect both X and Y



- So far we assumed that the red arrow does not exist
- However, there might be plausible arguments for such an arrow
 - For instance, with longer firm tenure the job is more secure what could affect contemporaneous or future chances for marriage
 - Or past (non) marriage could affect firm tenure.
- Time-varying unobservables can produce non-parallel trends

What to Do? Instrumental Variables

- Instrumental variables (IV) estimation (`xtivreg`)
 - FD-IV and FE-IV are available.
 - If “sequential exogeneity” is maintainable,
then FD-IV can use the lagged regressors as a valid instrument

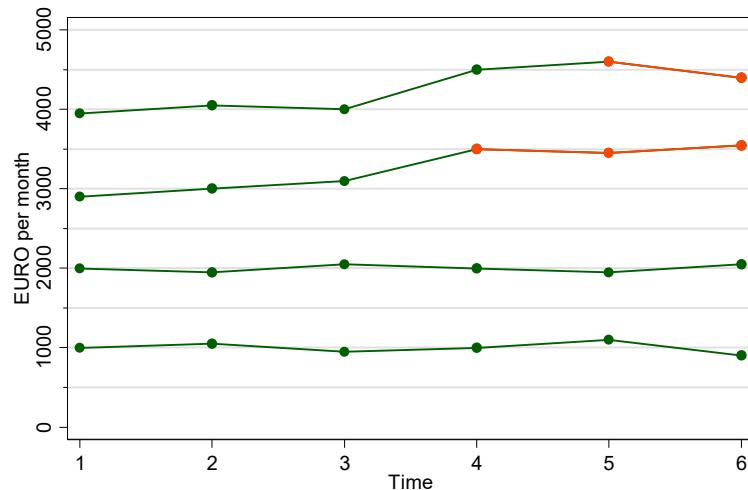


- Use IV methods only, when you have good arguments for the validity of the instruments used

Measurement Errors

- In a bivariate regression, measurement errors generally produce an “attenuation bias”
 - With more X-variables the direction of the bias is unknown
- Within estimators increase the problem
 - In fact, compared with pooled-OLS the bias due to measurement errors is amplified by using FD- or FE-estimators, because taking the difference of two unreliable measures generally produces an even more unreliable measure
 - On the other hand, pooled-OLS suffers from bias due to unobserved heterogeneity
 - Simulation studies show that generally the latter bias dominates
 - The suggestion is, therefore, to use within estimators: unobserved heterogeneity is a "more important" problem than measurement error.

Example for Measurement Errors



The highest earner reported marriage one year too late.
Now FE is biased, but POLS is still biased much more!

```
. regress wage marr1
```

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
marr1	1610.526	556.2303	2.90	0.008	456.9754 2764.077
_cons	2289.474	253.8832	9.02	0.000	1762.952 2815.995

```
. xtreg wage marr1, fe
```

wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
-----+-----					
marr1	441.1765	72.57272	6.08	0.000	289.28 593.0729
_cons	2533.088	29.16108	86.87	0.000	2472.053 2594.123

Reverse Causality

Reverse causality:

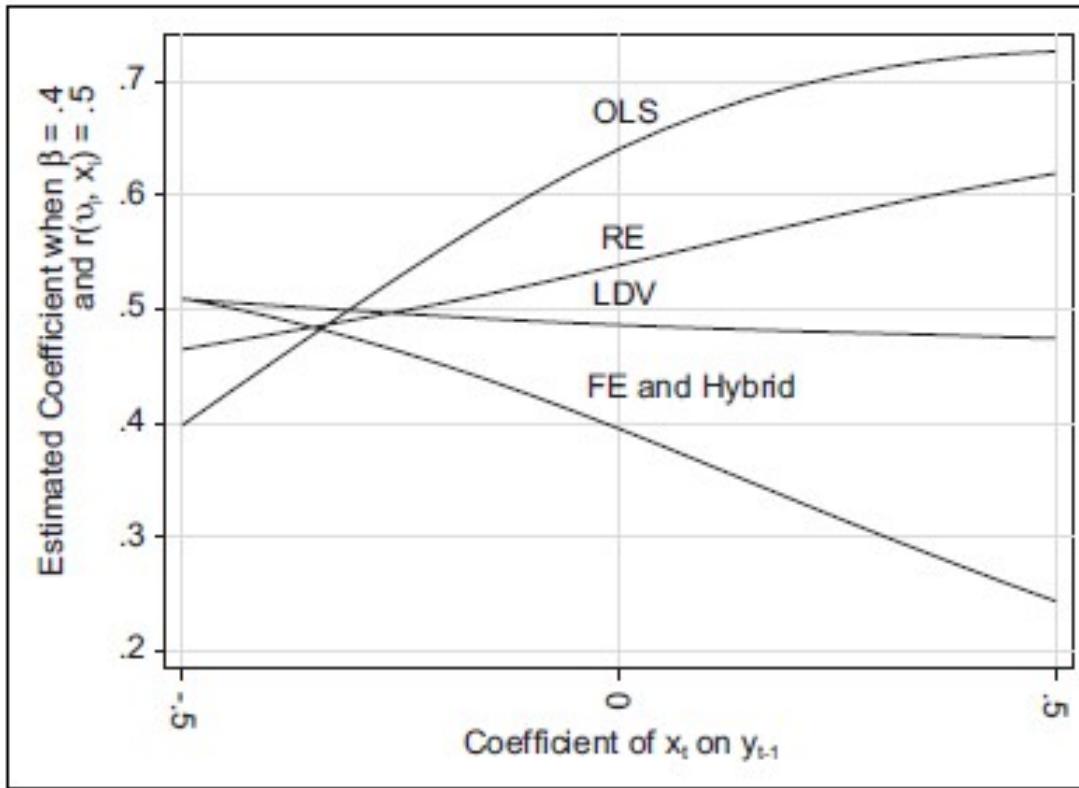
$$Y \rightarrow X$$

Reciprocal causality (simultaneity):

$$Y \rightleftarrows X$$

- Under reciprocal causality strict exogeneity is violated and therefore FE estimates are biased
 - **Any statistical model is biased under reciprocal causality!**
 - Some people believe that SEM will help (see next section)
 - See simulations on next slide
- Example: especially psychic concepts may show simultaneity
 - E.g., self-esteem and depression
- Many researchers see reciprocal causality almost everywhere
 - Fortunately, this is often only because of imprecise reasoning
 - There are two special cases, where the problem is less severe
 - Feedback mechanisms
 - Not reverse causality, but selection

Simulations from Vaisey/Miles (2014)



- Model
 - Person-specific heterogeneity
 - Reciprocal causality
- **None of the estimators works!**
 - If there is no reverse causality ($\gamma = 0$) FE works

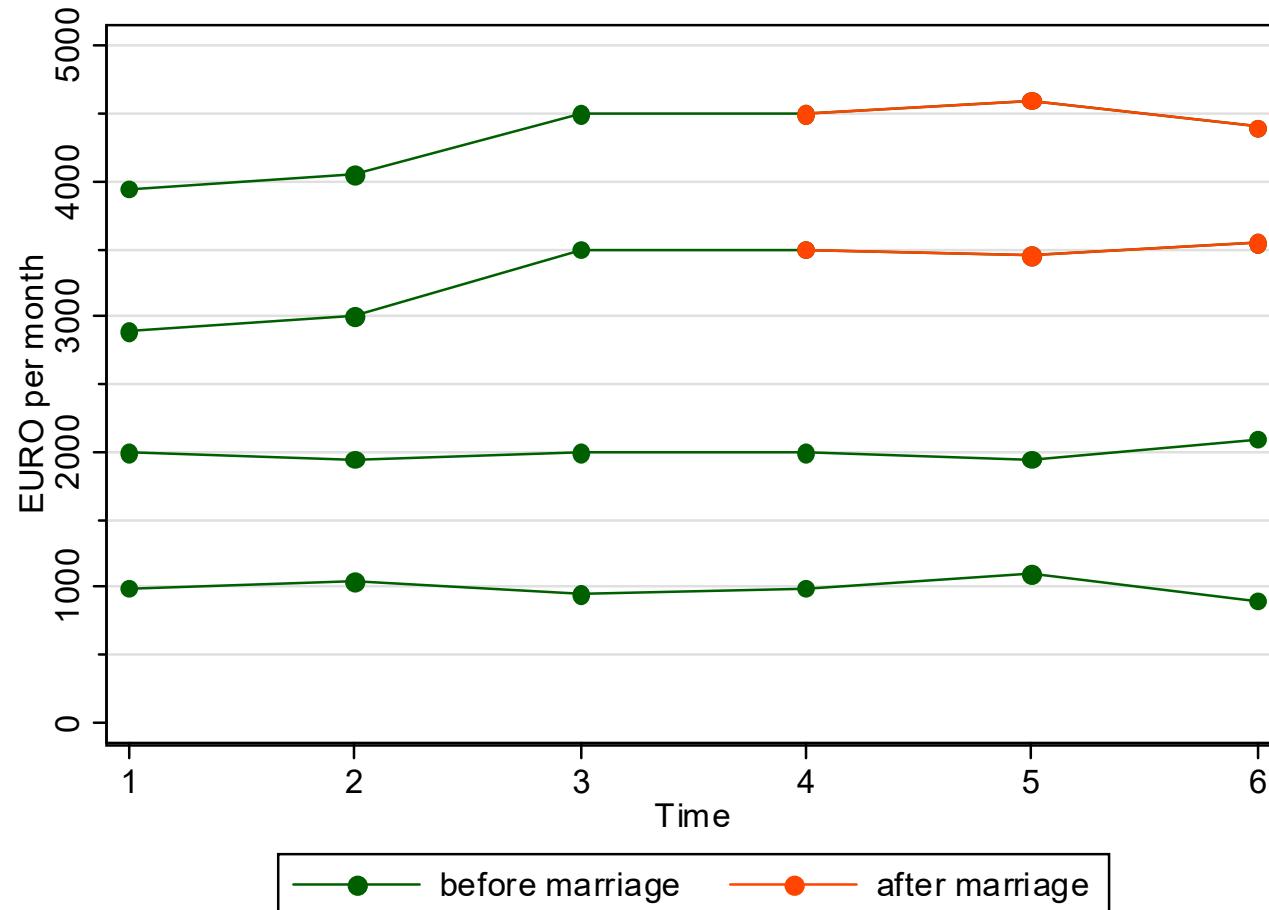
Special Case: Feedback Mechanisms

- An important case are feedback mechanisms
 - The covariates might react to shocks on the outcome
$$\varepsilon_t \rightarrow y_t \rightarrow x_{t+1} \rightarrow y_{t+1} \rightarrow \dots$$
 - Then $x_{i,t+1}$ will correlate with ε_{it}
 - This violates strict exogeneity but not sequential exogeneity!
 - **In this situation FD is not biased!!**
- Example: Does a new team coach increase performance?
 - Causal effect: a new team coach improves performance a little
 - A new coach is hired, if performance of a soccer team gets worse
 - FE will estimate a negative effect of the new coach on performance
 - FD will estimate the true (small) positive effect
- Robustness check
 - Compare FE and FD estimates. If they differ there might be feedback
 - Problem: feedback is statistically not distinguishable from anticipation
 - We need substantive knowledge to decide what is going on

Special Case: „Spurious“ Causality

- We have only reverse causality: $X \not\rightarrow Y$ $Y \rightarrow X$
 - The treatment reacts to shocks on the outcome
$$\varepsilon_t \rightarrow y_t \rightarrow x_{t+1}$$
- Example: male marital wage premium (MWP)
 - No causal effect of marriage on wages
 - But causality runs the other way: Men get a job promotion plus more wage, and then marry
 - FE will estimate a positive MWP
 - FD will estimate 0 MWP
- Thus, FD is a solution in this situation

Example for Spurious Causality



Modified Data:

The wage jump at $t = 3$ triggers marriage, i.e., causality runs the other way.

There is no additional causal effect of marriage!

There is still selectivity.

Data: Wage Premium.dta
Do-File: Wage Premium.do

Example for Spurious Causality

```
. xtreg wage2 marr, fe
-----  
          wage2 |      Coef.    Std. Err.      t    P>|t|    [95% Conf. Interval]  
-----+-----  
        marr |     350    88.27456     3.96    0.001    165.2392    534.7608  
       _cons |    2575    38.22401    67.37    0.000    2494.996    2655.004  
-----+-----  
  
. regress D.(wage2 marr), noconstant
-----  
          D.wage2 |      Coef.    Std. Err.      t    P>|t|    [95% Conf. Interval]  
-----+-----  
        D1.marr |      0    128.7593     0.00    1.000   -269.4962    269.4962  
-----
```

- Now the FE estimator is biased
- The FD-estimator provides the correct answer!
 - Substantive knowledge tells us that this is probably feedback not anticipation

Data: Wage Premium.dta
Do-File: Wage Premium.do

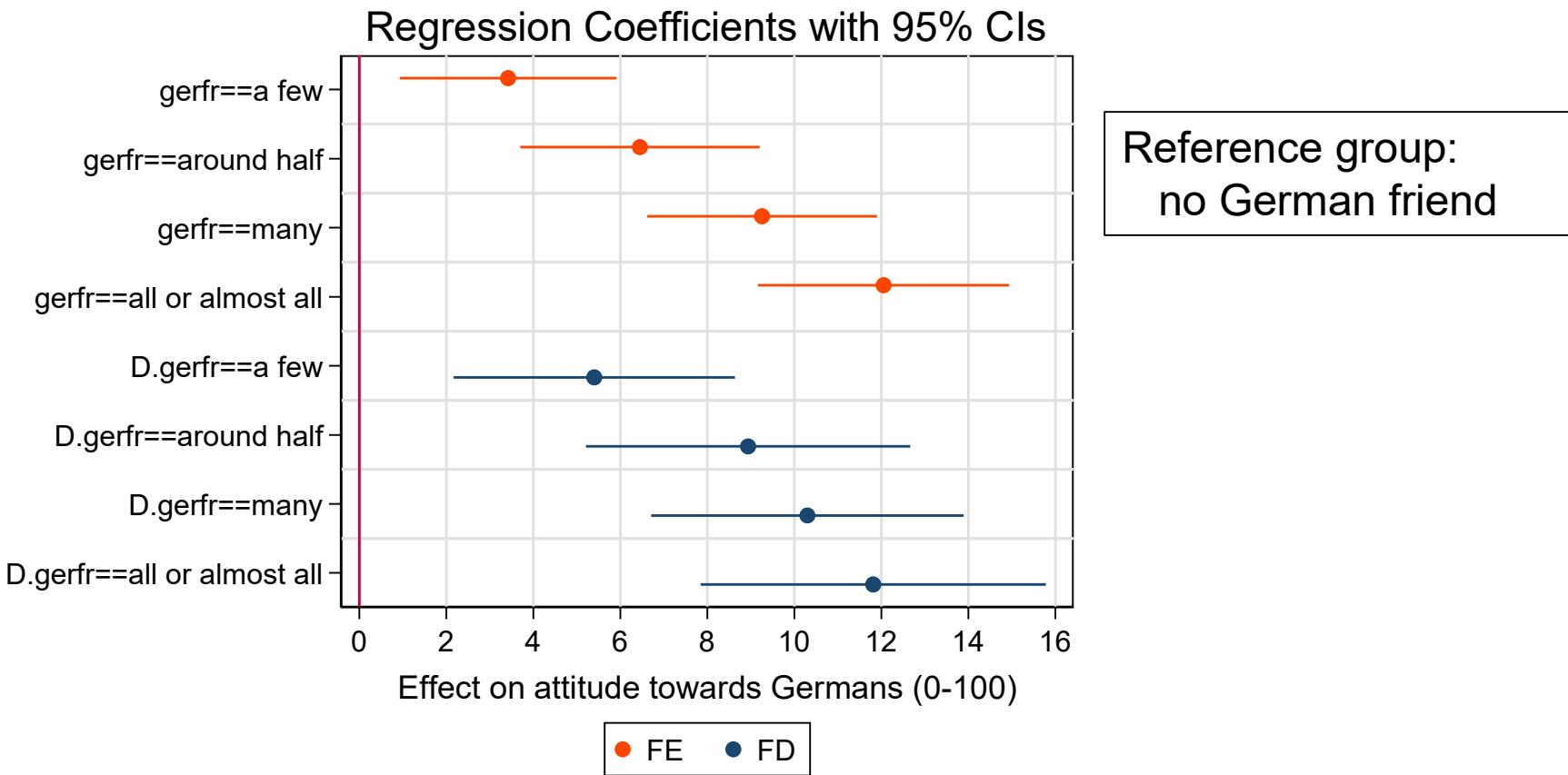
Not Reverse Causality, but Selection

- Researchers often construct a reciprocal causality problem where there really is none
 - The reasoning goes like this
 - The treatment causally affects the outcome, and high-outcome persons are more likely to get treatment
 - However, the second part of the argument is not reverse causality, but it is selection!
 - As long as selection is modelled correctly (via α_i (FE), or $\alpha_{1i} + \alpha_{2i}t$ (FEIS)), fixed-effects methods work fine

Example: Interethnic Friendships

- According to the contact hypothesis interethnic friendships reduce the amount of interethnic stereotypes
 - Many argue that one has a reciprocal causality problem, because those with less stereotypes will have more interethnic friends
 - But the second part of the argument clearly is a selection argument
 - FE will work fine here
 - At most, one could argue that there is a feedback mechanism: if your stereotype is reduced (for whatever reason), you add new interethnic friends to your network
 - In this situation FE will not work, but FD will work fine
- Example: data from CILS4EU-DE
 - Waves: 1, 3, 4, 5, 7. Sample: only migrants
 - Outcome: attitude towards Germans (scale 0-100)
 - Treatment: share of Germans in friendship network (5 categories)
 - Control: age linearly

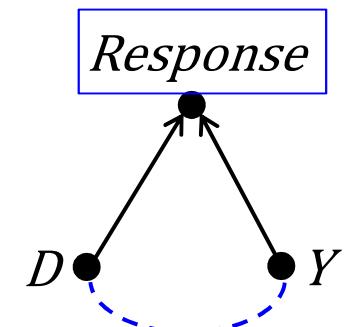
Example: Interethnic Friendships



- FE shows very strong effects: increasing the share of German friends increases positive attitude towards Germans
- This result holds with FD
 - Thus for this example we would conclude that the FE result is not an artifact of reverse causality

Panel Attrition

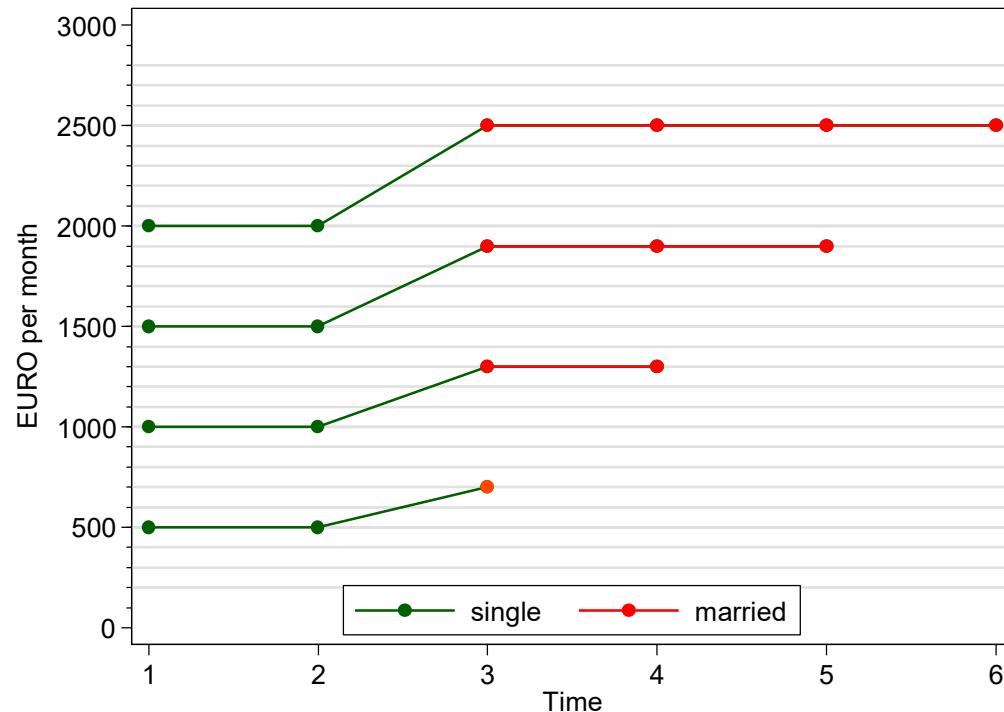
- Panel attrition is seen as a major threat to panel analyses
 - This assertion has to be qualified: attrition is not always a problem!
 - Case I) Attrition produced by (unobservable) regressors only
 - FE estimation offers some advantages here! Attrition correlated with α_i does not bias FE estimates (cf. Wooldridge 2010: 837ff)!
 - I.e., attrition associated with person-specific characteristics (observed or unobserved!) does not bias FE estimates
 - FEIS is even more robust (allowing for attrition correlated with het. trends)
 - Case II) The treatment **and** the outcome affect response (NMAR) (endogenous selection bias, Elwert and Winship 2014)
 - This produces endogeneity and the FE estimates will be biased also
- What can we do in case II?
 - There are special attrition selection models, there is weighting and imputation
 - But this is technification!
 - However, there is an obvious robustness check
 - Restrict analysis to only those observed continuously. If results remain, then it is unlikely that attrition biased the results



Attrition Biases Impact Functions

- If there is effect heterogeneity, attrition might bias results
 - If effect strength is connected with the probability of attrition then those observed for a long time under treatment are a select sample
 - This will bias the impact function
- Example: Step impact function
 - Those observed longer after treatment get a higher weight in FE estimation
- Example: Modeling the marriage effect with a dummy impact function [see next slide]
 - Assume a constant marriage impact
 - But those, with a larger marriage impact stay longer married
 - Then it will appear, as if the marriage effect increases with marriage duration
 - At higher marriage ages only those with the larger effects are left

Example: Dummy Impact Function



FE regress	
1. year marr	350***
2. year marr	383***
3. year marr	408***
4. year marr	428***
N	18
N_g	4

Data: Attrition Bias.dta
Do-File: Attrition Bias.do

- Here we have four men with different, but constant impacts
 - 200, 300, 400, 500 resp.
 - If all four were observed for four waves after marriage, we would estimate a constant impact of 350 € (no problem)
 - However, in this example we observe those with the smaller effects shorter (perhaps because they are frustrated due to their lower MWP)
 - Therefore, we get the erroneous impression that impact increases with time since event

Panel Conditioning

- Panel conditioning might bias panel analyses
 - If respondents react to measurement this might produce
 - Response effects (measurement errors)
 - Or even behavioral change
- The methodological literature is ambivalent on whether panel conditioning exists
 - Response effects can be found occasionally
 - Behavioral change has not been found
- Example: happiness (Kratz/Brüderl 2019)
 - Over the first three waves, SOEP respondents adapt their happiness responses
 - Such a response effect can easily be controlled by wave dummies
 - 1.wave: +0.28*** 2.wave: +0.10*** 3.wave: +0.04***
 - Works only in a multi-cohort, multi-subsample panel like the SOEP

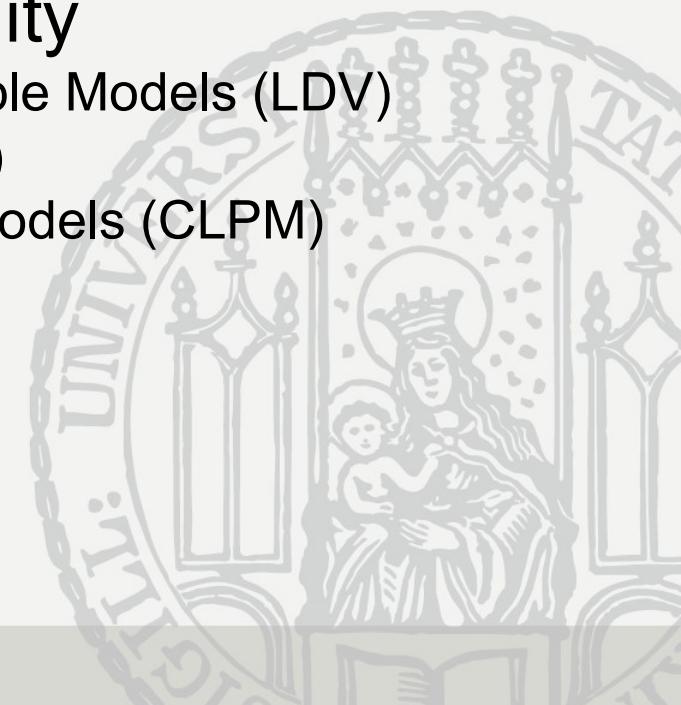
Chapter X:

Limitations of the Within Methodology

Section: Three Problems of Causality

- State Dependence: Lagged Dependent Variable Models (LDV)
- Lagged Effects: Lagged Effects Models (LEM)
- Direction of Causality: Cross-Lagged Panel Models (CLPM)

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis





State Dependence: Lagged Dependent Variable Models (LDV)



Correlation Over Time

- Panel variables usually are correlated over time.
This can be due to two sources:
 - Spurious state dependence:
time-constant unobservables affect the outcomes at all times
$$y_{it} = \alpha_i + \varepsilon_{it}$$
 - It follows that $\text{Corr}(y_{it}, y_{i,t-1}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$
 - True state dependence:
past outcomes affect “causally” future outcomes
(some kind of “reproduction” / “tendency for stability”)
 - Naturally modeled with a lagged dependent variable (LDV)
$$y_{it} = \rho y_{i,t-1} + \alpha_i + \varepsilon_{it}$$
 - $|\rho| < 1$ (stationary outcome process)
 - ρ captures the “amount of reproduction” / “autoregression”
 - If $\alpha_i = 0$, it follows that $\text{Corr}(y_{it}, y_{i,t-1}) = \rho$
 - [The general formula can be found in Cameron/Trivedi, 2005: 763]

LDV Models

- Desire to model true state dependence leads to LDV models

$$y_{it} = \rho y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$$

- Often termed “dynamic panel models”.
 - Panel models without LDV are termed “static”. These are, however, also dynamic, because they also model time
- LDV models have “intuitive appeal” and have been the standard models for panel analysis since Lazarsfeld’s time
- Sometimes LDV models are also motivated, with the argument that the LDV captures unobserved heterogeneity
 - But this is a sub-optimal approach, given that we have panel data

Estimating LDV Models

- Introducing the LDV complicates estimation enormously!
 - POLS / RE estimation is inevitably biased
 - $y_{i,t-1}$ is necessarily correlated with α_i , because α_i affects the outcome at all times. Thus estimates of ρ are inevitably biased.
 - If $y_{i,t-1}$ is correlated with x_{it} (this is the case, if $\beta \neq 0$) not only the estimate of ρ but also of β will be biased (the bias of ρ is transferred)! [proofs can be found in Angrist/Pischke, 2009: chap. 5.4]
 - FE estimation of LDV models does also not work!
 - Strict exogeneity is necessarily violated (“Nickell bias”) (Wooldridge 2010: 290)
 - Thus, one has to use instrumental variable (IV) methods
 - Arellano/Bond (AB) suggest using FD and $y_{i,t-2}$ as IV (if the “sequential exogeneity” assumption holds, lags are valid IVs).
 - `xtabond` in Stata
 - However, if ε_{it} is serially correlated (this will be the case with real data almost always), the AB estimator is biased also

Why Is the LDV Model so Popular?

- The “classical” argumentation can for instance be found in Finkel (1995) “Causal Analysis with Panel Data”
 - [N.B.: Finkel (1995) terms the LDV model „static-score model“]
 - The FD model, „contains one highly restrictive assumption: that the lagged dependent variable Y_{t-1} does not have an influence on either Y_t or ΔY . As we will see, this assumption is likely to be incorrect.” (Finkel, 1995: 6)
 - I.e., Finkel argues that there is always true state dependence
 - “If it is assumed that the error term in the static-score model is uncorrelated with both X_t and Y_{t-1} , then the coefficient can be estimated consistently through ordinary least squares (OLS) regression.” (Finkel, 1995: 11)
 - He sees no special problem with the exogeneity assumption, because he does not start from an error decomposition formulation of the model. Therefore, he oversees that the error term necessarily correlates with Y_{t-1} , if there are person-specific time-constant unobservables

Are LDV Models Useful?

- Consistent estimation of LDV models seems impossible
 - Simple estimation methods (POLS, FE) do not work
 - Even the specially developed AB method does not work
- → **Do not use LDV models!**
- Further, there is no real need to use LDV models
 - Often it is argued in an intuitive fashion “There is certainly some kind of reproduction going on”, but without explicating the concrete mechanisms behind it
 - But state dependence in most cases will be “spurious”, produced by stable unobservables that affect the outcome at all time points

$$y_{it} = \alpha_i + \varepsilon_{it}$$

- Example: stability of wages
 - Caused by stable, unobserved traits: productivity, personality traits, etc.
 - Not by some mysterious “wage reproduction mechanism”

Happiness Example: Introducing a LDV

Variable	FE	POLS-LDV	FE-LDV	AB-LDV
happy				
L1.		0.50***	0.10***	0.11***
marry	0.17***	0.17***	0.15***	0.13***
age	-0.04***	-0.02***	-0.03***	-0.04***
loghhinc	0.12***	0.18***	0.11***	0.05**
woman	(omitted)	0.02	(omitted)	(omitted)
N	121919	104671	104671	88751
N_clust	14634	14511	14511	

legend: * p<0.05; ** p<0.01; *** p<0.001

- Lagged happiness is introduced in the regressions by L.happy
- POLS-LDV gets the marriage effect quite well!
 - However, the age and income effects are strongly biased
- With FE-LDV only the age effect is besides the point
- AB-LDV on the other side estimates a particularly low income effect

LDVs are Colliders

- The problem in DAG notation (Morgan/Winship 2007: 71)

- In all three scenarios confounding via D and Y_{t-1}

- I) SD: true, no spurious

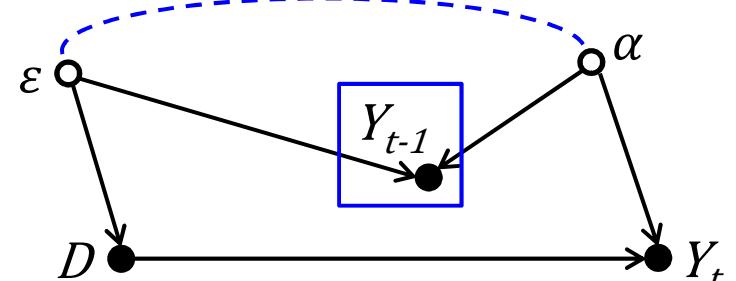
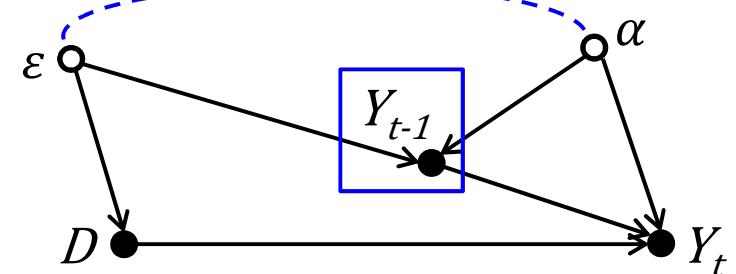
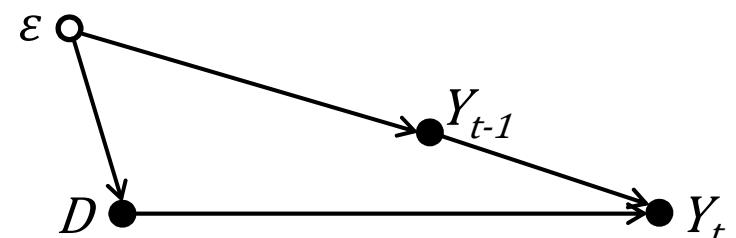
- Here the LDV blocks the confounding path. The LDV model works!
- Here the LDV in fact “captures unobserved heterogeneity”

- II) SD: true, spurious

- But in a more realistic scenario we also have person-specific confounders (spurious SD). Now the LDV is a collider
- If we do not control, we have confounding
- If we control, a new confounding path opens up
- We have an unsolvable dilemma!

- III) SD: no true, spurious

- Here we have no problem
- Only when controlling for the LDV!



The Simulation Setup

- Monte Carlo simulations from Brüderl/Ludwig (2015)
 - Data generating process ($N = 200$, $T = 10$)

$$y_{it} = \rho y_{i,t-1} + \beta x_{it} + \alpha_{1i} + \alpha_{2i}t + \varepsilon_{it}$$
 - X is a binary treatment indicator (50% treated (=1) after $t = 5$)
 - Scenarios:
 - Heterogeneity w.r.t. α_{1i} : purely random heterogeneity
 - Selection w.r.t. α_{1i} : unobserved heterogeneity (levels only)
 - Selection w.r.t. α_{2i} : unobserved heterogeneity (levels and growth)
 - TSD: true state dependence ($\rho \neq 0$)

Scenario	$\hat{\beta}_{\text{POLS}}$	$\hat{\beta}_{\text{RE}}$	$\hat{\beta}_{\text{FE}}$	$\hat{\beta}_{\text{FEIS}}$	$\hat{\beta}_{\text{RS}}$	$\hat{\beta}_{\text{POLS-LDV}}$	$\hat{\beta}_{\text{AB-LDV}}$
Heterogeneity w.r.t. α_{1i}	con.	con.	con.	con.	con.	inc.	con.
Selection w.r.t. α_{1i}	inc.	inc.	con.	con.	inc.	inc.	inc.
Selection w.r.t. α_{2i}	inc.	inc.	inc.	con.	inc.	inc.	inc.
TSD, het. w.r.t. α_{1i}	inc.	inc.	inc.	inc.	inc.	inc.	con.
TSD, sel. w.r.t. α_{1i} or α_{2i}	inc.	inc.	inc.	inc.	inc.	inc.	inc.

Source: Brüderl/Ludwig (2015)

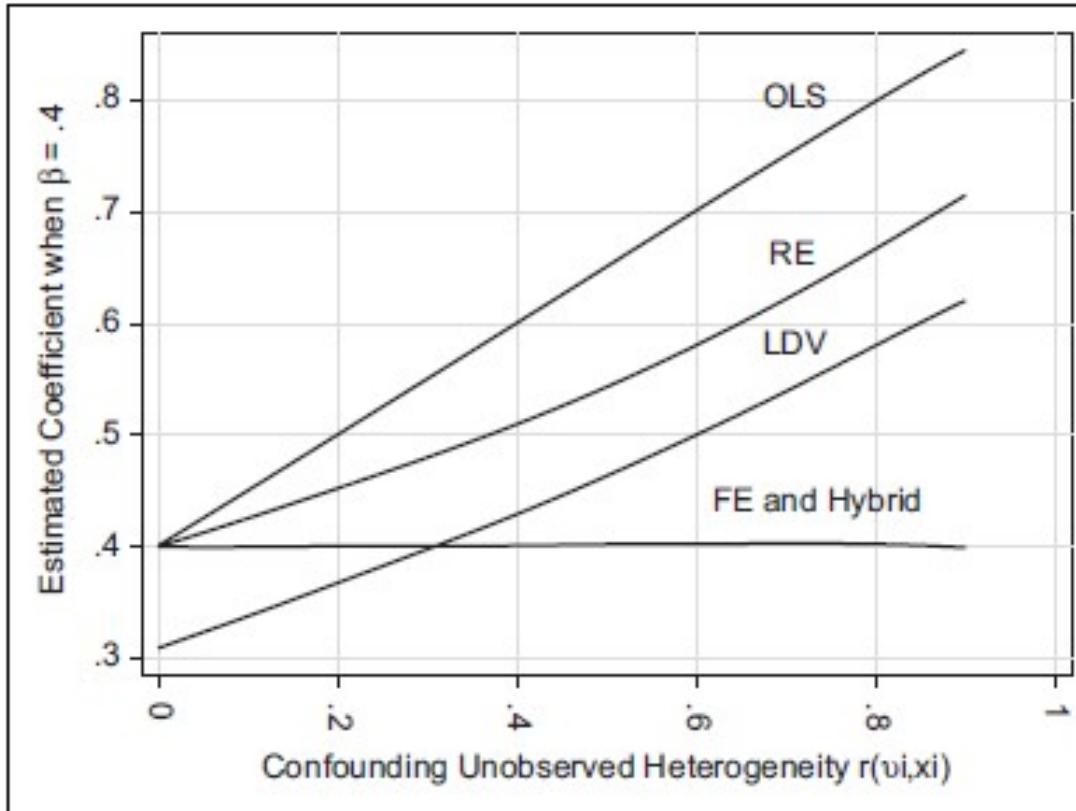
Simulation Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	het. α_1	het. α_1	sel. α_1	sel. α_2	sel. α_2	het. α_1	sel. α_1	sel. α_1
	$\beta = 1$	$\beta = 0$	$\beta = 1$					
	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0.5$	$\rho = 0.5$	$\rho = 0.5, \gamma = 0.2$
POLS	1.005 (0.139)	-0.006 (0.139)	1.896 (0.140)	2.495 (0.890)	1.257 (0.892)	1.557 (0.269)	3.314 (0.272)	3.297 (0.280)
RE	1.004 (0.077)	-0.001 (0.077)	1.139 (0.077)	1.482 (0.598)	0.649 (0.599)	1.506 (0.127)	1.864 (0.130)	1.860 (0.147)
RS	1.005 (0.078)	-0.001 (0.078)	1.131 (0.078)	0.917 (0.126)	0.904 (0.126)	1.303 (0.103)	1.239 (0.104)	1.202 (0.111)
FE	1.004 (0.080)	-0.000 (0.080)	1.002 (0.080)	1.391 (0.572)	0.594 (0.574)	1.502 (0.125)	1.771 (0.127)	1.760 (0.145)
FEIS	1.002 (0.128)	0.003 (0.128)	1.003 (0.128)	0.994 (0.128)	1.001 (0.128)	1.051 (0.141)	0.853 (0.141)	0.847 (0.153)
POLS-LDV	0.644 (0.082)	-0.003 (0.077)	1.053 (0.081)	0.093 (0.065)	0.040 (0.063)	0.585 (0.059)	0.775 (0.065)	0.713 (0.066)
AB-LDV	1.007 (0.147)	0.003 (0.149)	1.002 (0.128)	0.993 (0.205)	0.908 (0.206)	0.993 (0.146)	1.011 (0.127)	0.884 (0.133)

Simulation Results

- (1) and (2): purely random heterogeneity
 - Here all models are consistent; RE and RS are efficient
 - However, POLS-LDV is even in this innocuous situation biased, only if there is no treatment effect (2) does this model work!
- (3): selection on level
 - Here only FE and FEIS work well (FE more efficient)
 - AB-LDV works only because there is no serial correlation in ε_{it}
- (4) and (5): selection on level and growth
 - Here only FEIS works well
 - Note that also RS is substantially biased, even though it allows for individual-specific slopes (that however must not correlate with X)
- (6) - (8): true state dependence
 - Here only AB-LDV works, but only in the unrealistic scenarios (6, random heterogeneity) and (7, no serial correlation in ε_{it})
 - In the realistic scenario (8, selection and serial correlation) it is also biased. Thus under true state dependence no estimator works.

Simulations from Vaisey/Miles (2014)



- Model (scenarios (1) and (3)) $y_{it} = 0.4x_{it} + \alpha_i + \varepsilon_{it}$
 - Person-specific heterogeneity $r(x_{it}, \alpha_i) \in [0, 0.9]$
- Only FE works!
 - LDV doesn't even work, if $r(x_{it}, \alpha_i) = 0$



Lagged Effects: Lagged Effects Models (LEM)



Lagged Effects Model

- So far we assumed instantaneous effects

$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

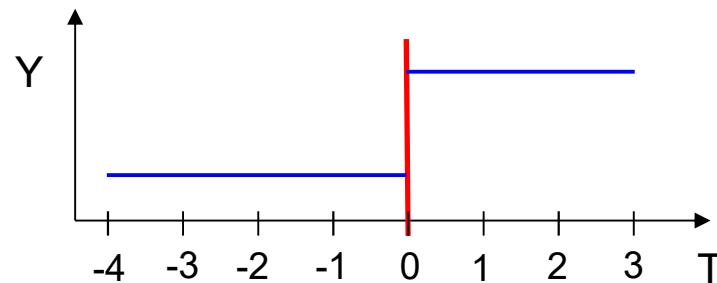
- Often researchers assume lagged effects (LEM)

$$y_{it} = \beta x_{i,t-1} + \alpha_i + \varepsilon_{it}$$

- In the first place, LEMs simply model lagged effects, i.e. effects that show up only after a certain time (here one period)
 - LEM estimated by FD (LFD), estimated by FE (LFE)
-
- However, the use of LEMs is often motivated by the intention to rule out reverse causality
 - The intuition is that $\Delta y_{it} \rightarrow x_{it}$ is more likely than $\Delta y_{it} \rightarrow x_{i,t-1}$

LEM Biases Within Estimation

- LEM works, if the process is truly lagged by one period
- LEM produces biased estimates, however, if the causal effect shows up immediately
 - Consider the following toy data with a binary treatment at $t = 0$



- There is an immediate causal effect of $y_0 - y_{-1}$
 - FD (and FE) would correctly estimate $\beta = y_0 - y_{-1}$
 - However, LFD will estimate $0 = y_1 - y_0$
 - LFE will also be biased downwards,
because y_0 is considered pre-treatment
- LEM bias: if the specified lag (data lag) is longer than the true lag (process lag) then we see a downward bias

Simulating the LEM Bias

- Vaisey/Miles (2014) simulate the LEM bias more precisely

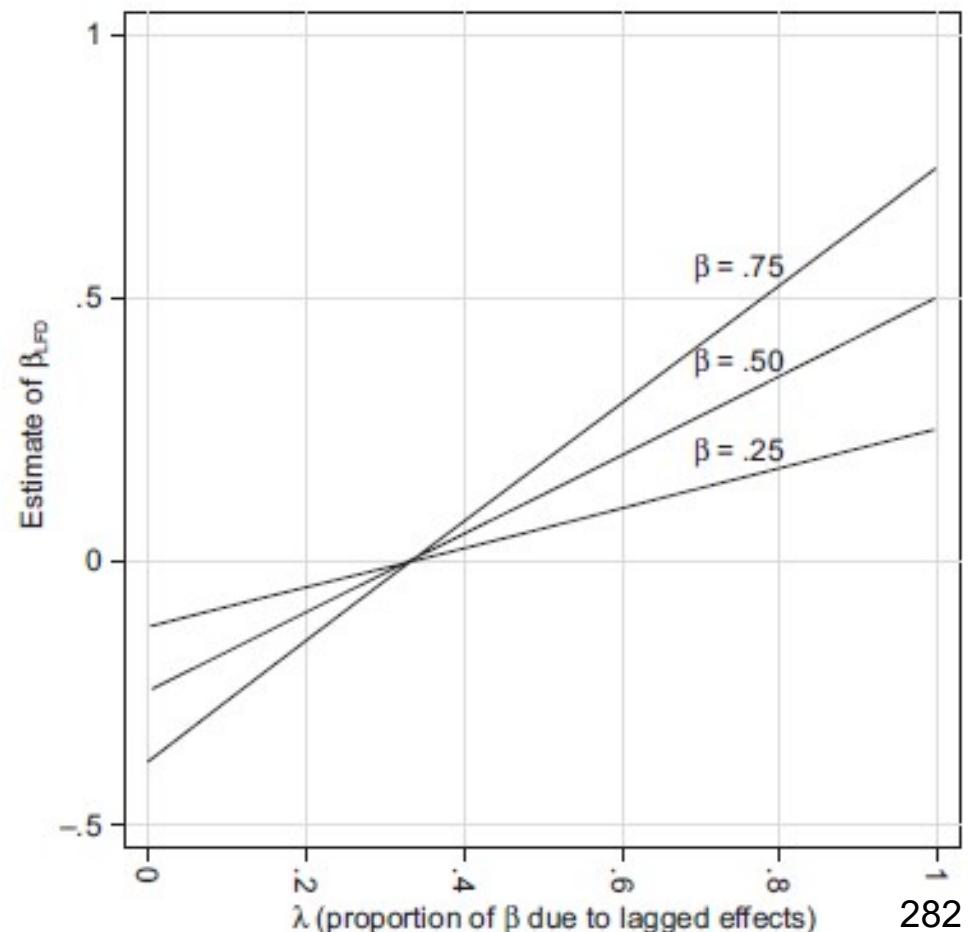
$$y_{it} = (1 - \lambda)\beta x_{it} + \lambda \beta x_{i,t-1} + \alpha_i + \varepsilon_{it}$$

- If $\lambda = 0$ instantaneous effects only, if $\lambda = 1$ lagged effects only

If there are lagged effects only ($\lambda = 1$), then LEM (first difference estimator) provides correct estimates

If there are instantaneous effects only ($\lambda = 0$), then LEM provides estimates with the opposite sign: $-\frac{1}{2}\beta$

There is a symmetrical bias, if there are lagged effects and we specify a non-lagged model



LEM Lessons

- LEM-Bias is very likely
 - In most social/psychological processes the causal impact shows up within a few hours/days
 - Then a LEM will produce artificial negative effects (provided that the true effects are positive)
- Thus, it is no good idea to use lagged effects per default
 - It is also no good idea to use non-lagged effects per default
 - Always theorize about the lag structure of the causal effect
- It is not a good idea to use LEM to avoid reverse causality
 - LEMs do not cure reverse causality (Reed 2015)
 - If there is feedback one would have to model “lead effects”
 - Instead they likely produce a LEM bias what produces completely wrong (opposite sign) results



Direction of Causality: Cross-Lagged Panel Models (CLPM)



Direction of Causality

- It is often asserted that from panel data we can learn about the direction of causality
 - LEMs are used to establish the direction of causality

$$\begin{aligned}y_{it} &= \beta x_{i,t-1} + \alpha_i + \varepsilon_{it} \\x_{it} &= \varphi y_{i,t-1} + v_i + \xi_{it}\end{aligned}$$

$$X \rightarrow Y \text{ if } \beta > \varphi \quad \text{or} \quad X \leftarrow Y \text{ if } \beta < \varphi$$

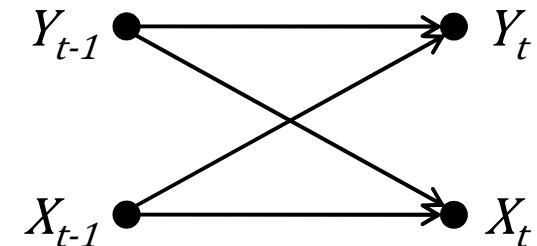
- However, due to LEM bias this does not work. It is very likely that the LEM will produce completely wrong (opposite sign) results
- **The assertion that with panel data we can decide about the direction of causality is wrong**

The Cross-Lagged Panel Model (CLPM)

- The CLPM has been seen as the „crown“ of all panel analysis methods (Duncan 1969)
- The CLPM combines both LEMs and LDVs

$$y_{it} = \rho y_{i,t-1} + \beta x_{i,t-1} + \alpha_i + \varepsilon_{it}$$

$$x_{it} = \varphi y_{i,t-1} + \delta x_{i,t-1} + \nu_i + \xi_{it}$$



- It seems that a CLPM models
 - True state dependence
 - Reciprocal causality
 - Lagged effects
- Researchers even want to identify causal dominance
 - $X \rightarrow Y$ if $\beta > \varphi$ or $X \leftarrow Y$ if $\beta < \varphi$

RI-CLPM ?

- In psychology estimates from the CLPM are believed to be “the most appropriate measures for studying causality in longitudinal correlational data” (Hamaker et al. 2015: 102)
- However, as is well-known in econometrics, CLPM estimates are biased
 - If there are person-specific confounders (“trait-like, time invariant”)
 - LDV Bias, LEM Bias
- Hamaker et al. (2015) demonstrate this for psychologists
 - They suggest to allow for fixed-effects in the CLPM
 - Confusingly they term these “random intercepts” (RI-CLPM)
 - But what about the estimation problems (LDV, LEM Bias)?
 - As is standard in the SEM literature no assumptions are discussed

ML-SEM ?

- Allison et al. (2017) modify the CLPM
 - Introduce fixed-effects
 - Focus on one equation (do not specify functional form and lag structure of the other equation)
 - Use ML-SEM estimation

$$y_{it} = \rho y_{i,t-1} + \beta x_{i,t-1} + \alpha_i + \varepsilon_{it}, \quad \text{for } t > 1 \quad (1)$$

$$x_{it} = f(y_{i1}, \alpha_i, \varepsilon_{is}), \quad \text{for } s < t \quad (2)$$

- Assumptions
 - Multivariate normality of Y, X
 - No serial correlation of errors: $\text{Cor}(\varepsilon_{it}, \varepsilon_{is}) = 0$
 - Sequential exogeneity: $\text{Cor}(x_{it}, \varepsilon_{is}) = 0$, whenever $s \geq t$
- They ignore a potential LEM-Bias
 - It is “unimportant” whether X is lagged or not

Example: Makes Singlehood Depressive?

- The German Family Panel (pairfam)
 - Annual panel survey
 - 14 waves 2008 – 2021
 - One hour CAPI/CASI interviews with anchor
 - Multi-actor design: partner, parents, children 8-15
 - First wave 2008/09
 - Random sample from the population registers
 - 3 birth cohorts: 1971-73, 1981-83, 1991-93
 - About 4000 respondents (anchors) for each cohort (N=12,402)
- Pairfam V8 (2017)
 - Only original sample
 - Waves 2 – 8 used
 - Balanced panel (valid answers on all variables from wave 2 – 8)
 - N = 3,330
 - NT = 23,310

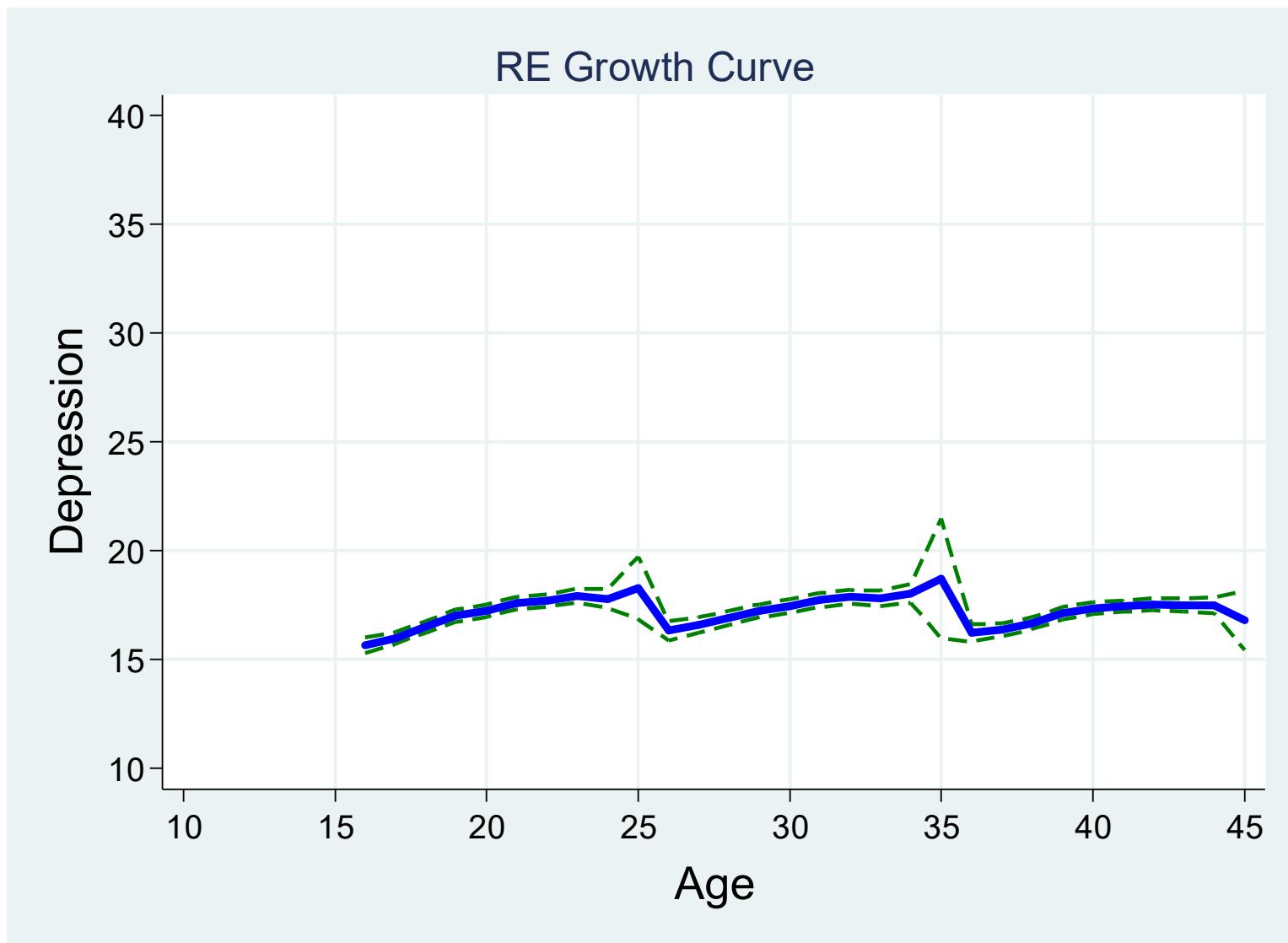
Depression

- State-trait-depression scale (STDS)
 - 10 items, additive index, range 10-40, standardized in regressions

Im Folgenden finden Sie eine Reihe von Feststellungen, mit denen man sich selbst beschreiben kann. Bitte lesen Sie jede Feststellung durch und wählen Sie aus den vier Antworten diejenige aus, die angibt, wie Sie sich im Allgemeinen fühlen.

- | | |
|---------|---------------------------------|
| per2i1 | Meine Stimmung ist schwermüdig. |
| per2i2 | Ich bin glücklich. |
| per2i3 | Ich bin deprimiert. |
| per2i4 | Ich bin traurig. |
| per2i5 | Ich bin verzweifelt. |
| per2i6 | Ich bin in gedrückter Stimmung. |
| per2i7 | Ich fühle mich gut. |
| per2i8 | Ich fühle mich sicher. |
| per2i9 | Ich bin ruhig und gelassen. |
| per2i10 | Das Leben macht mir Spaß. |
-
- | | |
|------------------|----------------------------|
| Fast nie | <input type="checkbox"/> 1 |
| Manchmal | <input type="checkbox"/> 2 |
| Oft | <input type="checkbox"/> 3 |
| Fast immer | <input type="checkbox"/> 4 |

Depression and Age



Results From Different Models

	POLS	FE	FE Estimated with ML-SEM	FD	ML-FE# Estimated with ML-SEM “Single” predetermined	ML- SEM#
Single	.26***	.13***	.14***	.15***	.20***	---
Single _{t-1}	---	---	---	---	---	-.16***
Age/10	.06***	.44***	.39***	.45***	.36***	.13***
Woman	.20***	---	---	---	---	.12***
Depr _{t-1}	---	---	---	---	---	.26***

* p<0.05; ** p<0.01; *** p<0.001. Constants are invariant across time periods.

Models did not converge

- There is selection of depressive persons into singlehood
 - The effect of single halves from POLS to FE
- The age effect in POLS is suppressed by a cohort effect
- There might be a feedback effect: a depressive attack could lead to separation of a partnership
 - ML-FE “causal” effect of “single” even increases?
 - ML-SEM now the effect is even negative: LEM-Bias?
 - FD no indication that we might have a problem with feedback

Chapter XI:

Final Thoughts

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



Technification

- Traditionally, it has been believed that more complex methodology equals better research (“technification”)
 - Structural equation modeling (SEM)
 - SEM supposedly can take regard of both simultaneity and measurement errors
 - Problems of these methods
 - They are not robust
 - Research fields that use these methods are full of contradictory results
 - **These methods have produced a big mess in social research!**
- More and more social researchers oppose:
 - “Those who worship at the altar of complex methods are prone to the error of thinking that technical sophistication can substitute for knowledge of the subject matter, careful theorizing, and appropriate research design” (Firebaugh 2008: 207f)
 - “We cannot rely on statistical wizardry to overcome faulty data and research design” (Firebaugh 2008: 208)

What to Do? Simple Methods

- **Keep it simple!**
 - Use FD
 - If endogeneity is due to non-parallel trends then use FEIS
 - Synthetic control group method (SCM)
 - DID-matching on panels (life courses) before treatment occurs
 - Feedback (reverse causality) no longer a problem
- **Be critical!**
 - If you have doubts on the validity of the strict exogeneity assumption, be careful when interpreting FE results
 - If you find no good IV or simple method, be conservative and conclude that even the panel data at hand do not help in identifying the causal effect

What to Do? Better Research Design

- Invest in “**shoe leather**”
 - Invest in better research design
 - Try to collect better data (panel data) that include (i) the time-varying unobservables, and (ii) more precise measures
 - Try to collect better data by (natural) experiments
- One cannot rule out reverse causality by panel data and statistics alone
 - The point is that even with panel data our measurement of event timing is imprecise, because we measure only the contemporaneous values. Thus, we do not know, whether first X or Y changed!
 - To improve, one would need better timing information. For discrete events we could ask timing retrospectively. For continuous treatments we would need daily (or so) waves!

What To Do? Precise Thinking

- Clear causal questions
 - Treatment and outcome should be separate concepts
 - Panel data are good for identifying the “effects of events”
- Do not construct a problem, where there really is none
 - Avoid intuitive reasoning and always provide precise mechanisms behind causal effects/associations
 - Is there really true state dependence?
 - Do we really have lagged effects?
 - Is there really reverse causality?

And: Always Discuss Your Assumptions!

- An essential lesson from the “new wave of causal analysis” is that you always should discuss the plausibility of the assumptions that you need to identify the causal effect
- Example: marriage and happiness
 - Measurement errors in X
 - We see no obvious problems with “hard facts” as used here
 - Time-varying unobservables
 - We did certainly not control for some obvious candidates:
E.g., birth of a (first) child
 - Feedback mechanisms
 - It doesn’t seem plausible that a positive happiness shock triggers marriage. It is more plausible that we see an anticipation effect.
 - Attrition
 - Robustness check: are results stable with long panels?

References

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



References I

- Abendroth, A.K. et al. (2014) The Parity Penalty in Life Course Perspective. *ASR* 79: 993-1014.
- Ahituv, A. and R. Lerman (2007) How Do Marital Status, Work Effort, and Wage Rates Interact? *Demography*, 44: 623-647.
- Allison, P.D. (2009) Fixed Effects Regression Models. Sage.
- Allison, P.D., R. Williams and E. Moral-Benito (2017) Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects. *Socius* 3: 1-17.
- Andreß, H.J., K. Golsch, and A. Schmidt (2013) Applied Panel Data Analysis for Economic and Social Surveys. Springer.
- Angrist, J.D. and J.S. Pischke (2009) Mostly Harmless Econometrics. Princeton UP.
- Bauer, G. (2015) Graphical Display of Regression Results. Pp. 205-224, in: H. Best and C. Wolf (eds.) Regression Analysis and Causal Inference. Sage.
- Bell, A. and K. Jones (2015) Explaining Fixed Effects. *Pol. Sc. Res. and Methods* 3: 133-153.
- Bernardi, L. et al. (2017) Multidimensionality of Well-Being and Spillover Effects Across Life Domains. *Research in Human Development*: 14: 26–51.
- Best, H. und T. Rüttenauer (2018) How Selective Migration Shapes Environmental Inequality in Germany. *ESR* 34: 52-63.
- Blossfeld, H.P., K. Golsch, and G. Rohwer (2007) Event History Analysis with Stata. Erlbaum.
- Brüderl, J. (2010) Kausalanalyse mit Paneldaten. S. 963-994 in: C. Wolf and H. Best (eds.) Handbuch der sozialwissenschaftlichen Datenanalyse. VS Verlag.
- Brüderl, J. and V. Ludwig (2011) Does a Smoking Ban Reduce Smoking? Evidence from Germany. *Schmollers Jahrbuch* 131: 419-429.
- Brüderl, J. and V. Ludwig (2015) Fixed-Effects Panel Regression. Pp. 327-357, in: H. Best and C. Wolf (eds.) Regression Analysis and Causal Inference. Sage.
- Cameron, A., and P. Trivedi (2005) Microeconometrics. Cambridge UP.
- Chowdhury, G. (1991) A Comparison of Covariance Estimators for Complete and Incomplete Panel Data Models. *Oxford Bulletin of Ec. and Statistics* 53: 83-93.
- Curran, P. and D. Bauer (2011) The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change. *Annual Review of Psychology* 62: 583-619.
- Dougherty, C. (2006) The Marriage Earnings Premium as a Distributed Fixed Effect. *J. of Human Resources* 41: 433-443.

References II

- Duncan, O.D. (1969) Some linear models for two-wave, two-variable panel analysis. *Psychological Bulletin* 72: 177-182.
- Elwert, F. and C. Winship (2014) Endogenous Selection Bias. *Ann. Rev. Soc.* 40: 31-53.
- Finkel, S. (1995) Causal Analysis with Panel Data. Sage.
- Firebaugh, G. (2008) Seven Rules for Social Research. Princeton UP.
- Firebaugh, G., C. Warner, and M. Massoglia (2013) Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis. Pp. 113-132, in: S. Morgan (ed.) *Handbook of Causal Analysis for Social Research*. Springer.
- Frijters, P. and T. Beatton (2012) The Mystery of the U-Shaped Relationship Between Happiness and Age. *J. of Ec. Behavior & Organization* 82: 525-542.
- Gerstorf, D., N. Ram, G. Mayraz, M. Hidajat, U. Lindenberger, G. Wagner, and J. Schupp (2010) Late-Life Decline in Well-Being Across Adulthood. *Psychology and Aging* 25: 477-485.
- Giesselmann, M. and M. Windzio (2012) Regressionsmodelle zur Analyse von Paneldaten. Springer VS.
- Giesselmann, M. and M. Windzio (2014) Paneldaten in der Soziologie. KZfSS 66: 95-113.
- Goodman-Bacon, A. (2018) Difference-in-Differences with Variation in Treatment Timing. Unpublished manuscript.
- Hamaker, E.L., R.M. Kuiper and R.P. Grasman (2015) A Critique of the Cross-Lagged Panel Model. *Psychological Methods* 20:102-116.
- Halaby, C. (2004) Panel Models in Sociological Research. *Annual Rev. Sociol.* 30: 507-544.
- Hooghe, M., C. Meeusen and E. Qunitelier (2013) The Impact of Education and Intergroup Friendship on the Development of Ethnocentrism. *ESR* 29: 1109-1121.
- Jann, B. (2014) Plotting Regression Coefficients and Other Estimates. *Stata J.* 14: 708-737.
- Kohler, U. et al. (2012) Verarmungsrisiken nach kritischen Lebensereignissen in Deutschland und den USA. KZfSS 64: 223–245.
- Kratz, F. and J. Brüderl (2019) Subjective Well-Being Over the Life Course. Unpublished manuscript.
- Kroh, M. (2013) Unequal Political Voice and Family Background. Unpublished manuscript.

References III

- Legewie, J. (2012) Die Schätzung von kausalen Effekten: Überlegungen zu Methoden der Kausalanalyse anhand von Kontexteffekten in der Schule. *KZfSS* 64: 123-153.
- Ludwig, V., and J. Brüderl (2011) Is There a Male Marital Wage Premium? Resolving an Enduring Puzzle with Panel Data from Germany and the U.S. Unpublished manuscript.
- Lynn, P. (2009) (Ed.) *Methodology of Longitudinal Surveys*. Wiley.
- Meer, J. and J. West (2015) Effects of the Minimum Wage on Employment Dynamics. Unpublished WP.
- Morgan, S. and C. Winship (2007) *Counterfactuals and Causal Inference*. Cambridge UP.
- Mouw, T. (2006) Estimating the Causal Effect of Social Capital. *Ann. Rev. Soc.* 32: 79-102.
- Nickell, S.J. (1981) Biases in Dynamic Models with Fixed Effects. *Econometrica* 49: 1417-1426.
- Pollmann-Schult, M. (2011) Marriage and Earnings? *ESR* 27: 147-163.
- Rabe-Hesketh, S. and A. Skrondal (2008) *Multilevel and Longitudinal Modeling Using Stata*.
- Raudenbush, S. (2001) Comparing Personal Trajectories and Drawing Causal Inferences from Longitudinal Data. *Ann. Rev. Of Psychology* 52: 501-525.
- Reed, W.R. (2015) On the Practice of Lagging Variables to Avoid Simultaneity. *Oxf. Bull. of Ec. and Stat.* 77: 897-905.
- Rippeyoung, P. and M. Noonan (2012) Is Breastfeeding Truly Cost Free? *ASR* 77: 244-267.
- Schmidt-Catran, Alexander W. and Dennis C. Spies. 2016. "Immigration and Welfare Support in Germany." *American Sociological Review* 81(2):242-61
- Schunck, R. (2013) Within and between estimates in RE models. *Stata J.* 13: 65-76.
- Snijders, T., and R. Bosker (1999) *Multilevel Analysis*. Sage.
- Sobel, M. (2012) Does Marriage Boost Men's Wages? Identification of Treatment Effects in Fixed Effects Regression Models for Panel Data. *JASA* 107: 521-529.

References IV

- Stawarz, N. (2013) Inter- und Intragenerationale Mobilität: Eine simultane Analyse unter Verwendung von Wachstumskurven. ZfS 42: 385-404.
- Stutzer, A. and B.S. Frey (2005) Does Marriage Make People Happy, or Do Happy People Get Married? J. of Socio-Economics 35: 326-347.
- Vaisey, S. and A. Miles (2017) What You Can – and Can't – Do With Three Wave Panel Data. SMR 46: 44-67
- Waite, L. and M. Gallagher (2000) The Case for Marriage: Why Married People Are Happier, Healthier, and Better Off Financially. New York.
- Wolbring, T. et al. (2013) Needs, Comparisons, and Adaptation. ESR 29: 86-104.
- Wooldridge, J. (2010) Econometric Analysis of Cross Section and Panel Data. MIT Press.
- Wunder, C., A. Wiencierz, J. Schwarze, and H. Küchenhoff (2013) Well-Being over the Life Span. Review of Economics and Statistics 95: 154-167.
- Young, R. and D.R. Johnson (2015) Handling Missing Values in Longitudinal Panel Data With Multiple Imputation: JMF 77: 277-294.