

Applied Panel Data Analysis

Using Stata

(short version for teaching)

Josef Brüderl, LMU München
SoSe 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	22
– Within Estimation	31
– Within Estimation With Age/Period Effects	48
IV) The Basics of Linear Panel Regression	
– Linear Panel Models	56
– Fixed- or Random-Effects?	68
V) A Real Data Example: Marriage and Happiness	
– Preparing Panel Data	75
– Describing Panel Data	84
– The Results	87
– Interpreting Results from Panel Regressions	100

Contents II

VI) Modeling Individual Growth	
– Growth Curve Models	117
– The Age-Period-Cohort Problem	130
– Group Specific Growth Curves	140
VII) Further Linear Panel Models	
– The Fixed-Effects Individual-Slopes Model	148
VIII) Limitations of the Within Methodology	
– Limitations of Scope	163
– Violations of Strict Exogeneity	169
- Unobserved Time-Varying Confounders	
- Measurement Errors	
- Reverse Causality	
– Final Thoughts	183
References	188

Contents III

This is a shortened version of a longer script.

The long version can be found on

<https://www.ls3.soziologie.uni-muenchen.de/studium-lehre/lehrmaterialien/index.html>

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!

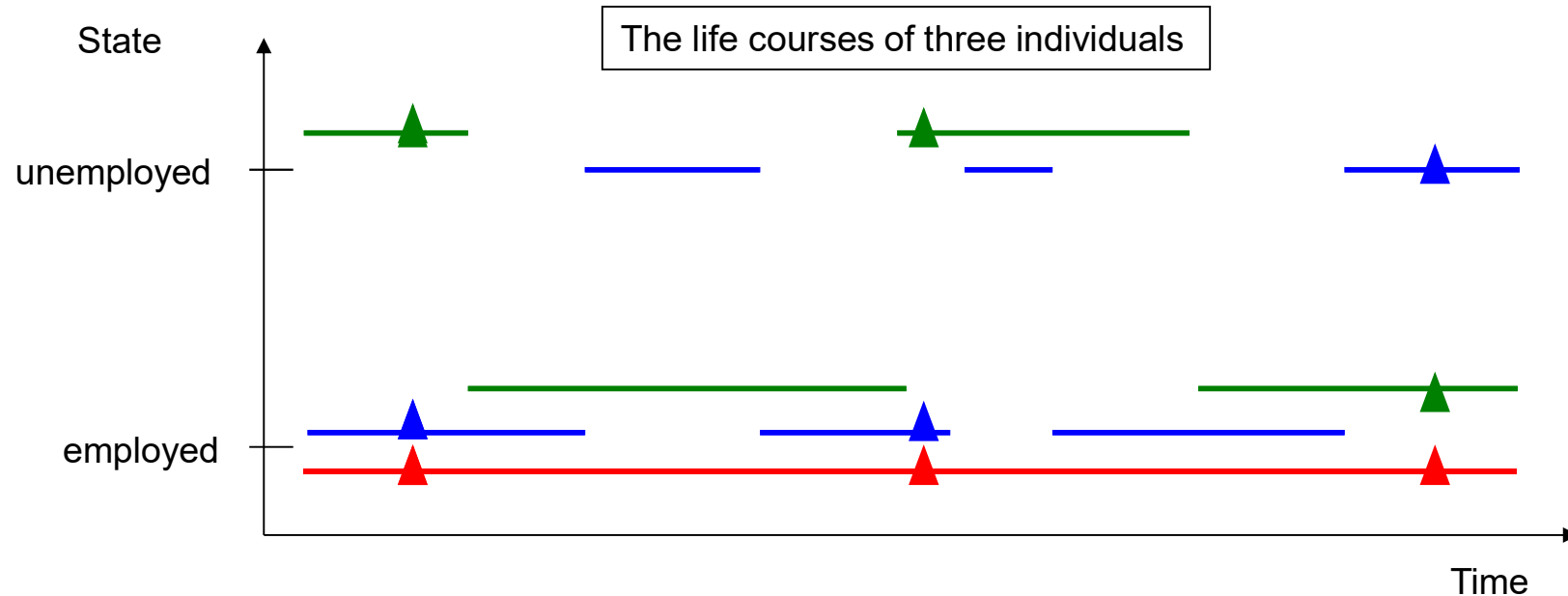
Chapter I: Panel Data

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}
\vdots	\vdots	\vdots	\vdots
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)

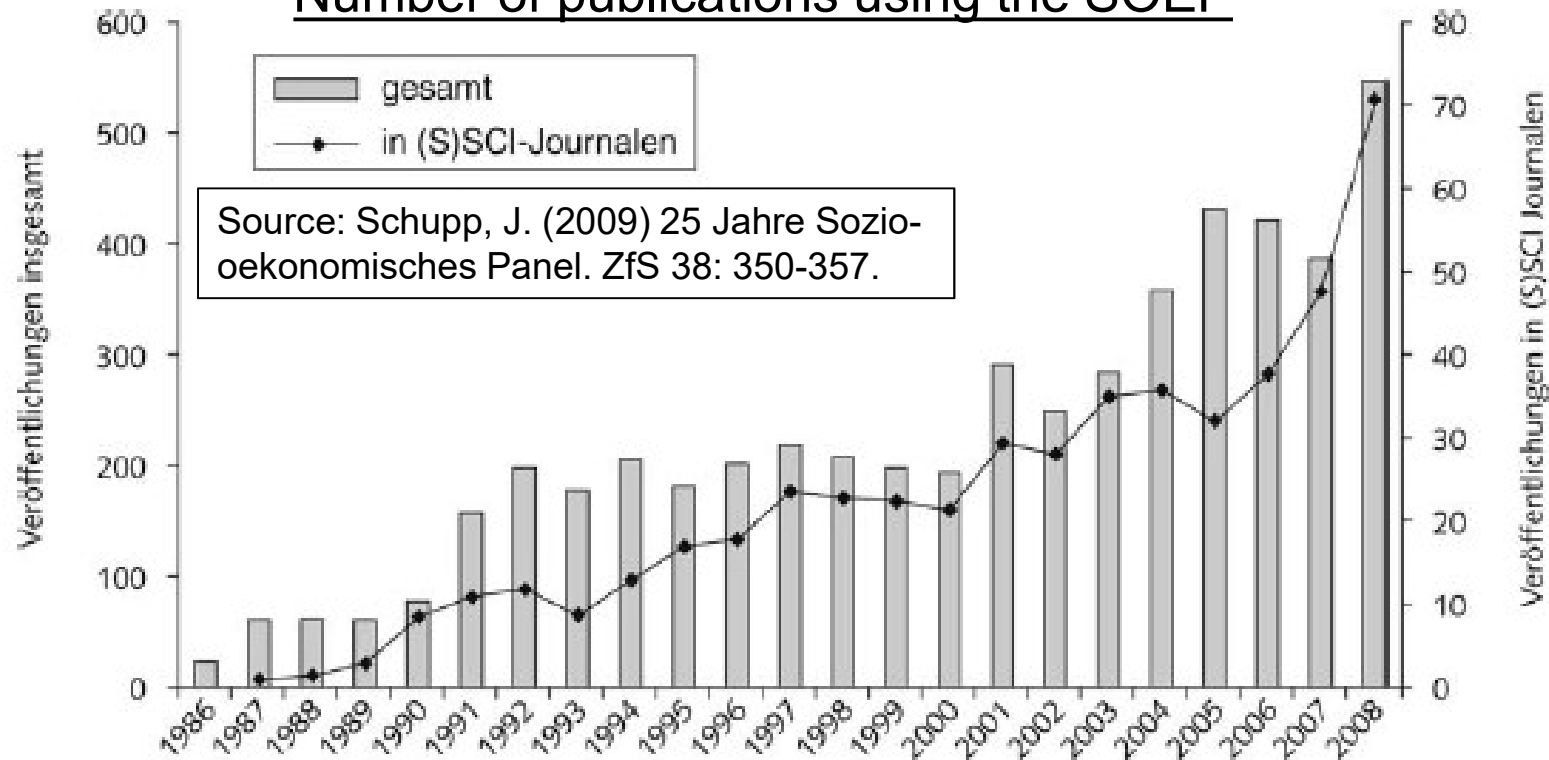
- Repeated measurements of the same variables on the same units
 - Macroeconomics, Political Science
 - Unit of analysis: countries
 - N small, T large
 - Cross-sectional time series ($\mathbf{x}\mathbf{t}$)
 - Microeconomics, Sociology
 - Unit of analysis: persons
 - N large, T small
 - (Micro) panel 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

Number of publications using the SOEP



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
 - 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
 - German internet panel (GIP)
 - GESIS Panel
- Links on German studies you can find here:
<http://www.ratswd.de/forschungsdaten/fdz>

SOEP

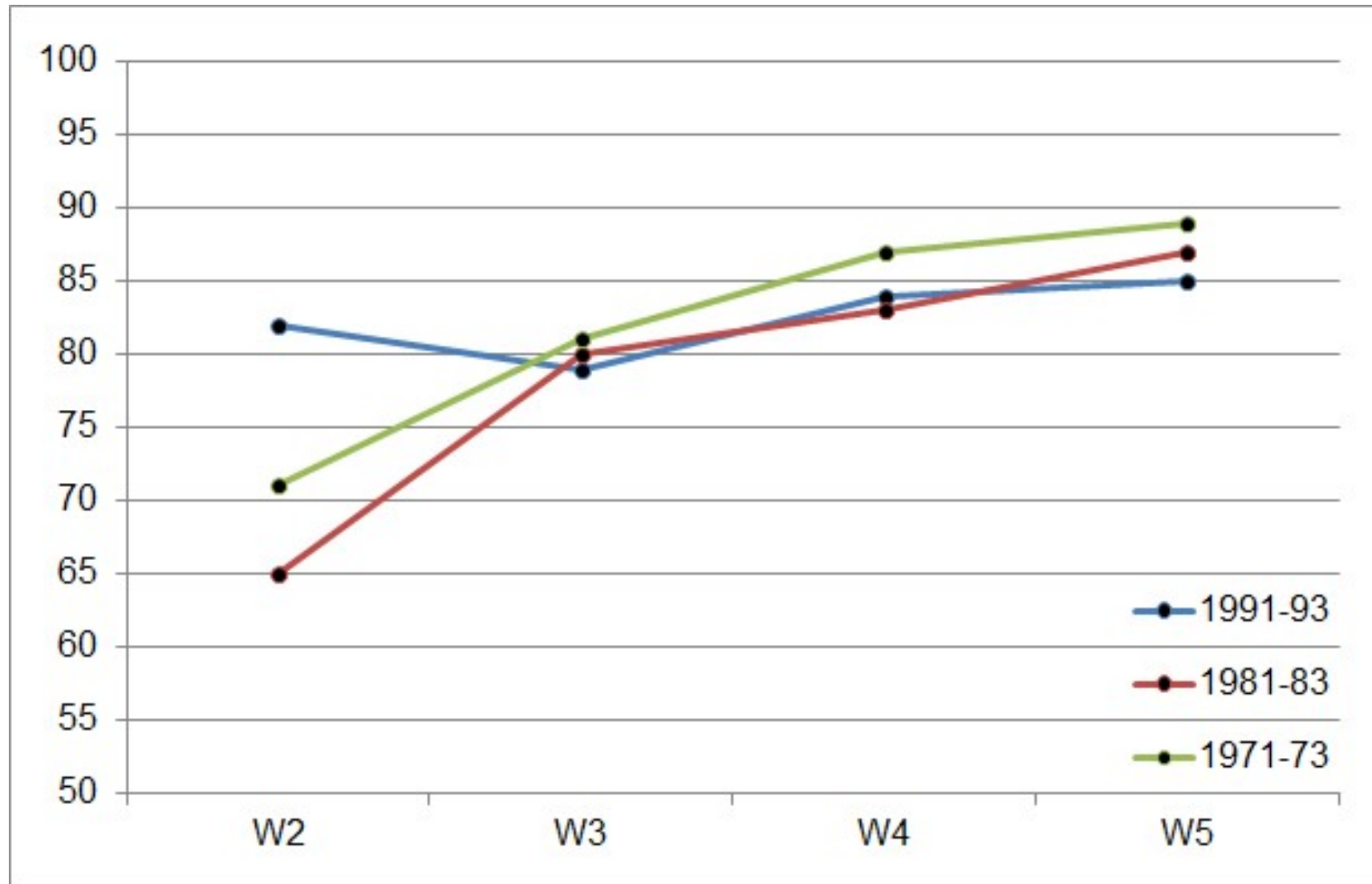
Household Panel Study

- 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>

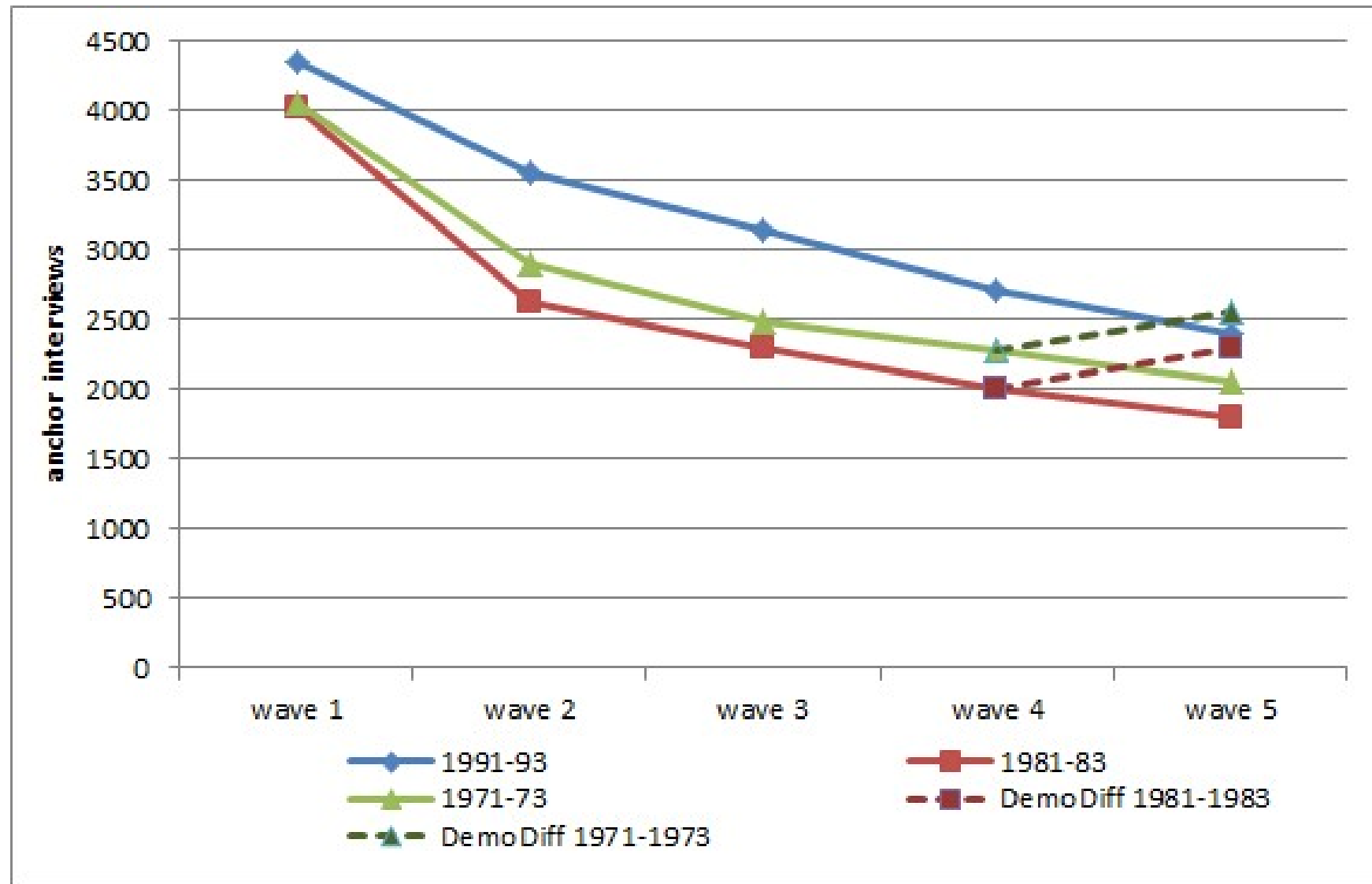
The German Family Panel (pairfam)

- 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

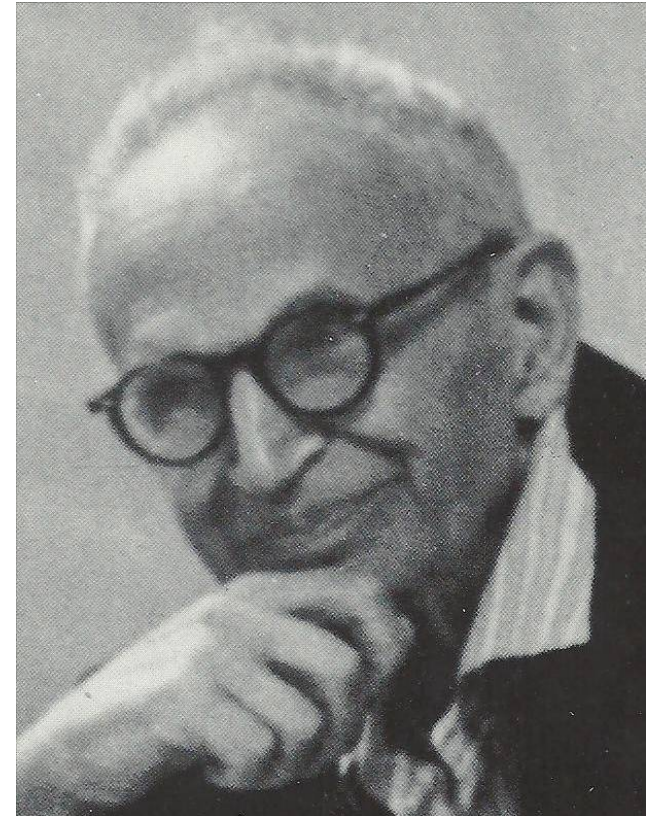
Paul Lazarsfeld on the Panel Design

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

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

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

- Estimation with panel data I

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

- 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

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

- 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

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?

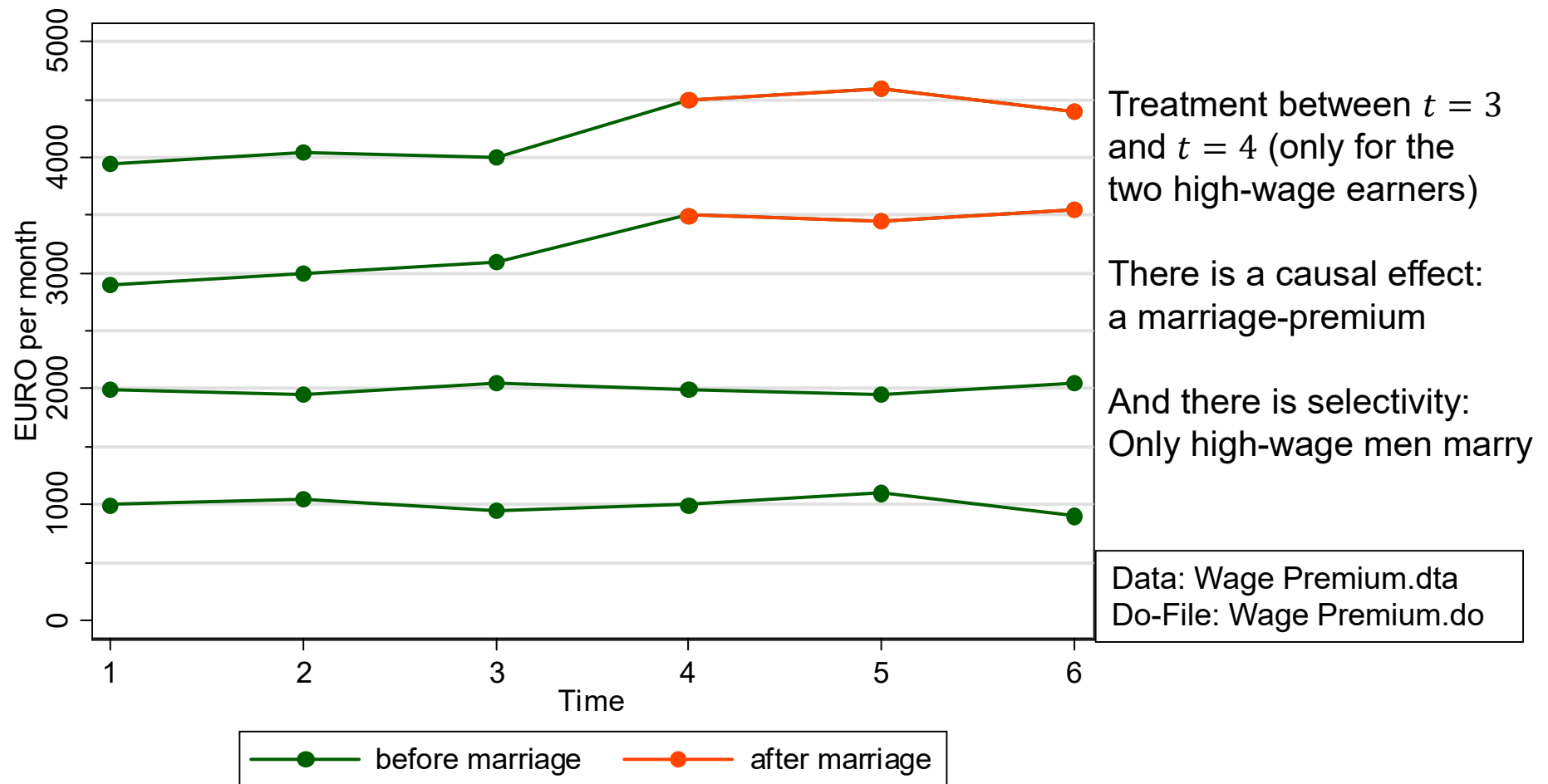
- For demonstrating how between and within estimation work we use a “toy example” with fabricated data (long-format)

```
. 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

- $\Delta_4 = 4500 - 4000 = 500$

- $\Delta_3 = 3500 - 3000 = 500$

$$\Delta_i = \frac{1}{3} \sum_{t=4}^6 y_{it} - \frac{1}{3} \sum_{t=1}^3 y_{it}$$

- 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

$$\Delta = \overline{\Delta_{i \in T}} - \overline{\Delta_{i \in C}} = \frac{500 + 500}{2} - \frac{0 + 0}{2} = 500$$

- The marriage-premium in our data is **+500 €**

- In the following we will investigate, whether different statistical regression models can recover this causal effect!

Cross-Sectional Regression

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

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

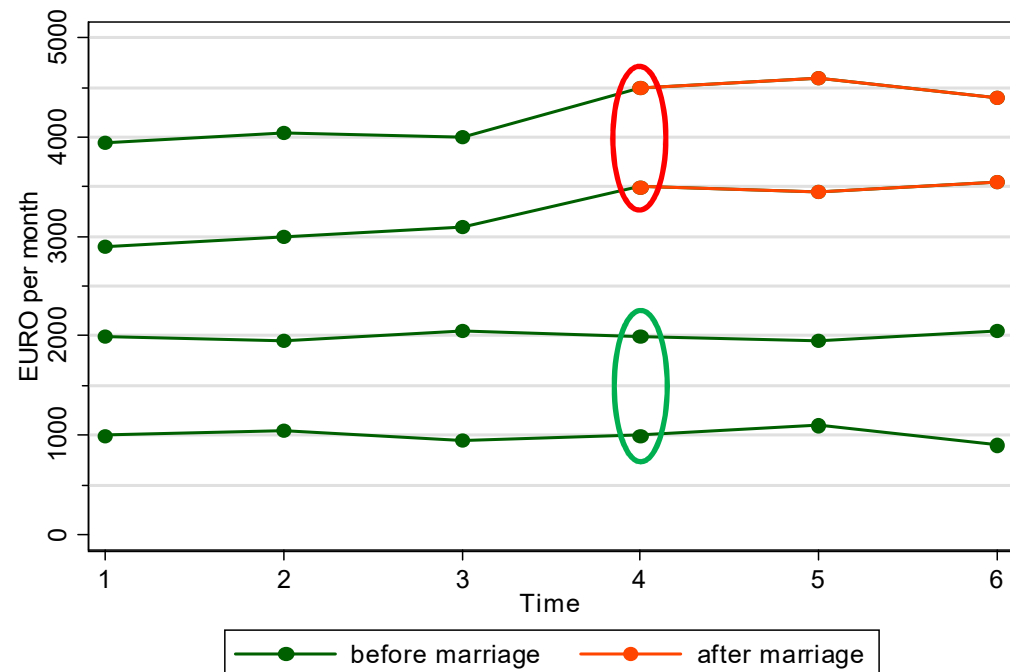
- This is a between-comparison at $t = 4$: essentially this compares average wages of married and unmarried men at $t = 4$

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

- 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

$$E(u_i|x_i) \neq 0$$

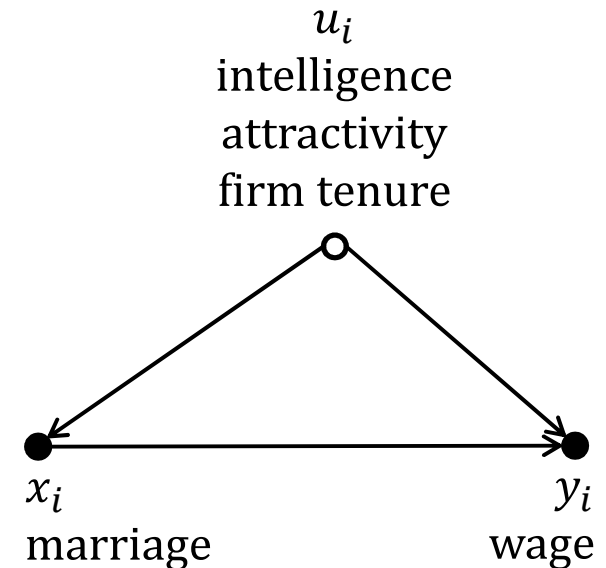
- Then it is said: the regressor is endogenous (**endogeneity**)

- **$\hat{\beta}_{OLS}$ 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**

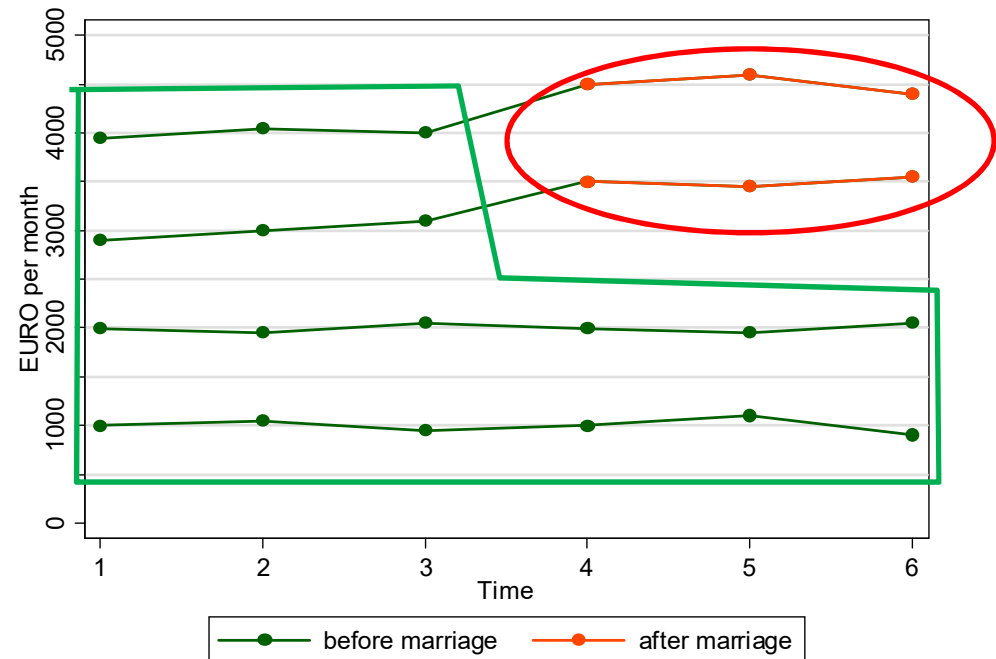
- Treatment and control groups are not built by randomization
- Instead, human beings decide according to unobservables, whether they go into treatment or not

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

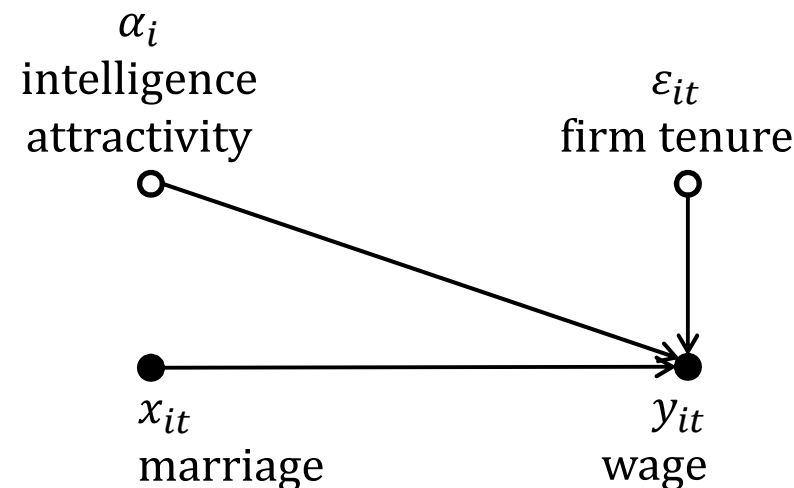
Section: Within Estimation

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

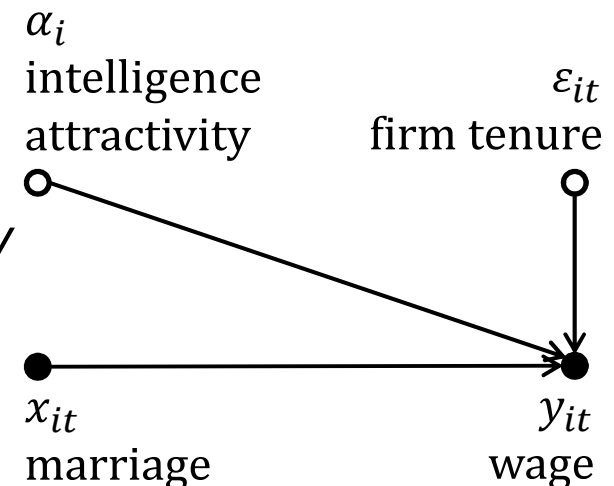
$$y_{it} = \beta x_{it} + \alpha_i + \varepsilon_{it}$$

- 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

$E(\alpha_i | x_{it}) = 0$ *random-effects assumption*
 “no (person-specific) time-constant unobserved heterogeneity”

$E(\varepsilon_{it} | x_{it}) = 0$ *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

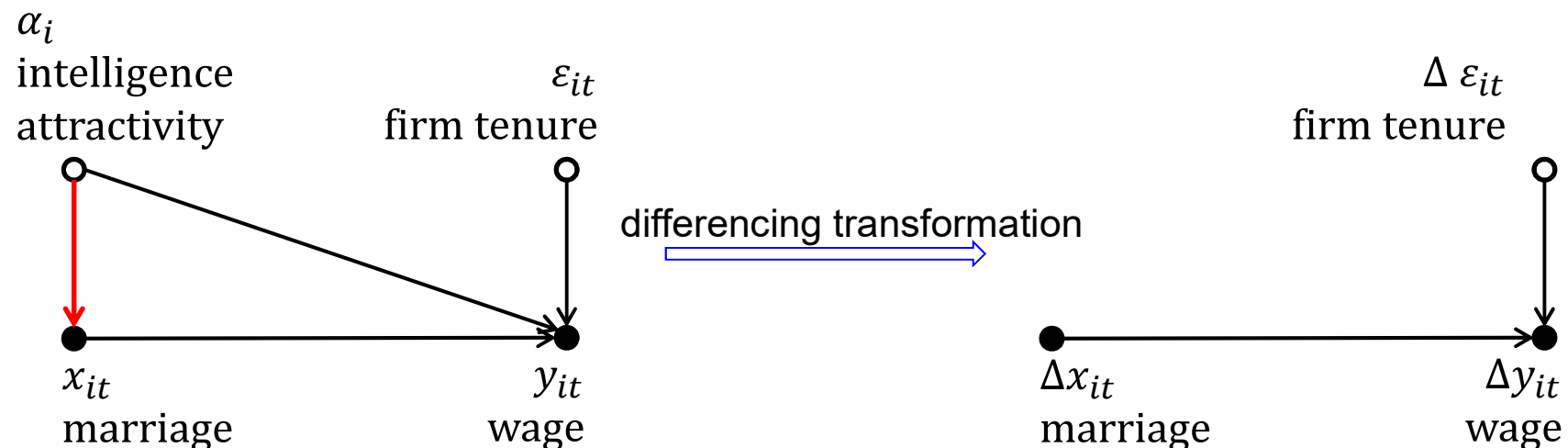
- The FD-estimator is unbiased if

$$E(\varepsilon_{it}|x_{it}) = 0 \quad \text{for } t \leq s \quad \text{sequential exogeneity assumption}$$

- 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
```

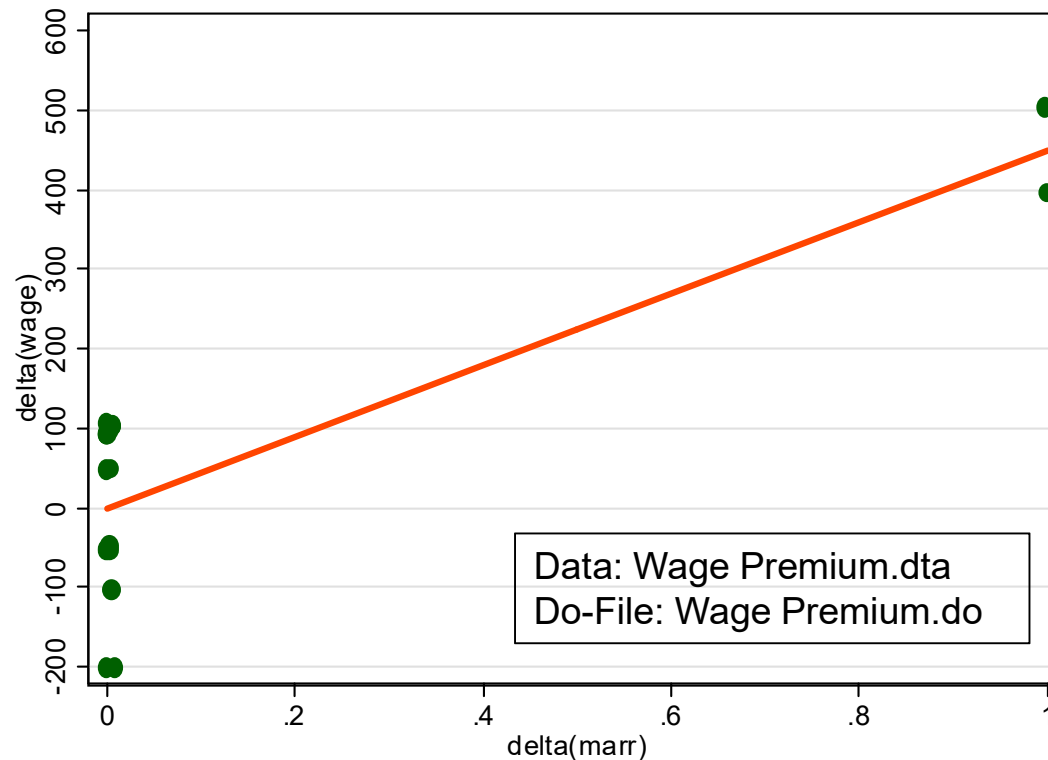
Source	SS	df	MS	Number of obs = 20		
Model	405000	1	405000	F(1, 19)	=	39.97
Residual	192500	19	10131.5789	Prob > F	=	0.0000
Total	597500	20	29875	R-squared	=	0.6778
				Adj R-squared	=	0.6609
				Root MSE	=	100.66

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

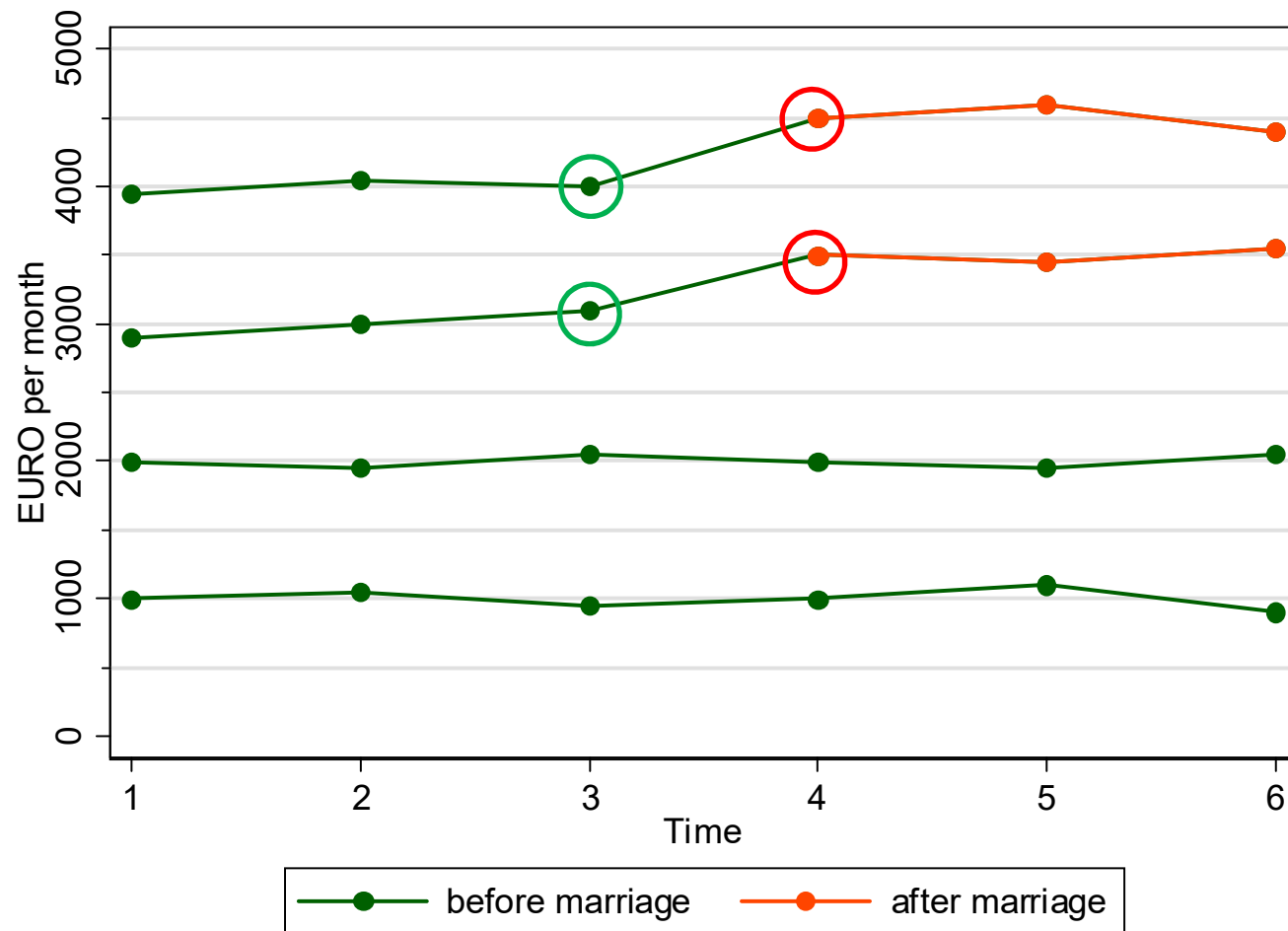
„Mechanics“ of a FD-Regression



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

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

- “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

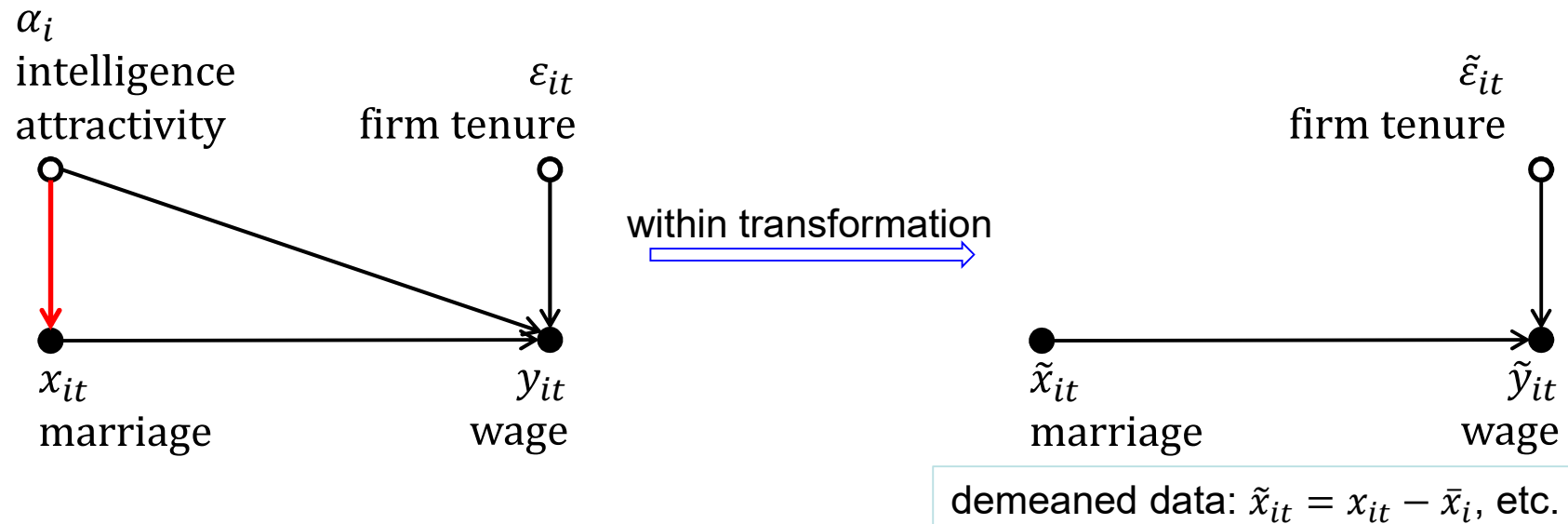
- Note: (2) is called “between regression” (BE)

Assumptions of FE Estimation

- The FE-estimator is unbiased if

$E(\varepsilon_{it}|x_{it}) = 0$ for all t and s *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
```

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

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

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

R-sq:  within  = 0.8982                      Obs per group: min =            6
      between  = 0.8351                      avg           =           6.0
      overall   = 0.4065                      max           =            6

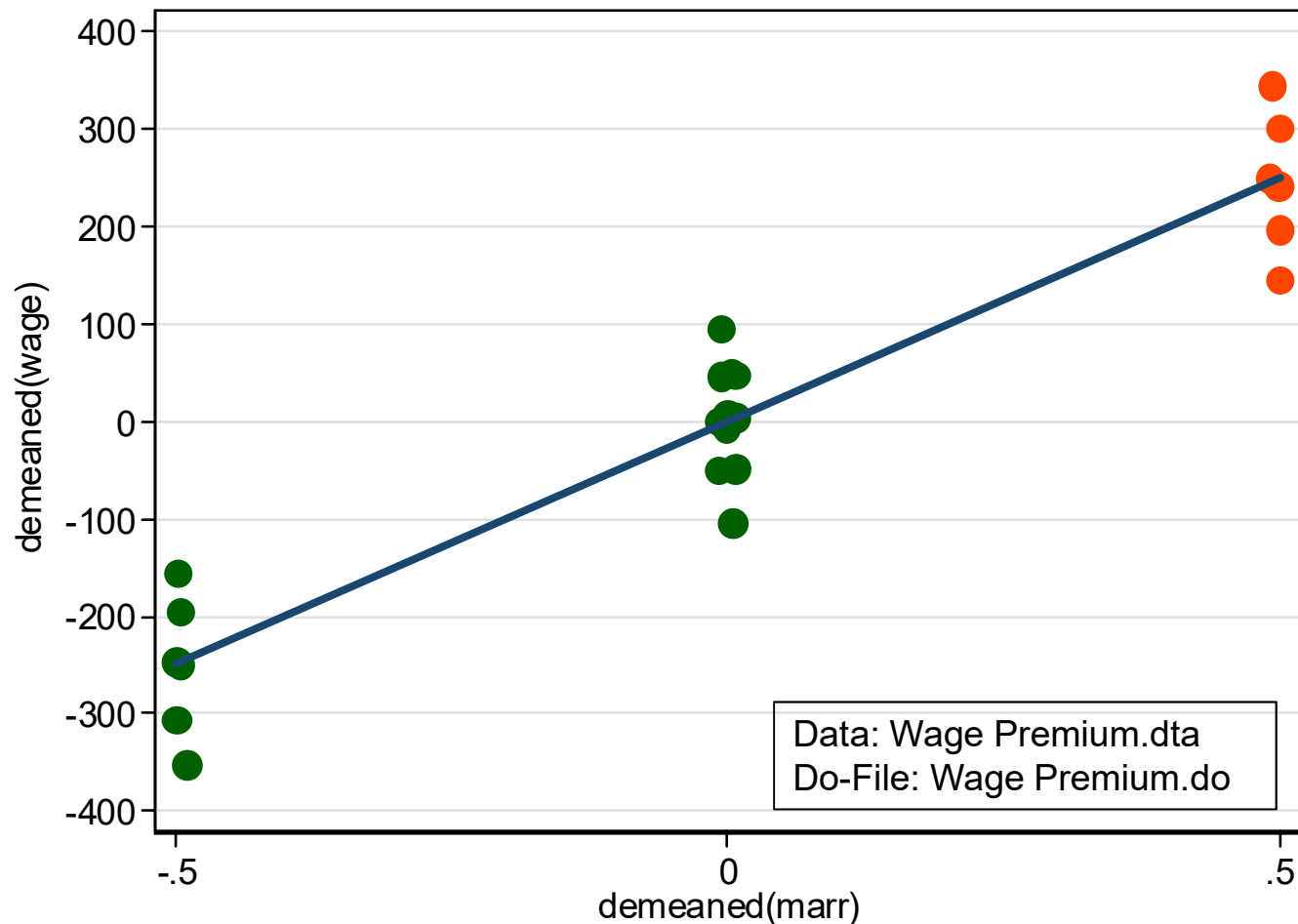
                                          F(1,19)         =       167.65
corr(u_i, Xb)  = 0.5164                      Prob > F         =       0.0000
```

```
-----+-----
      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
-----+-----
      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)

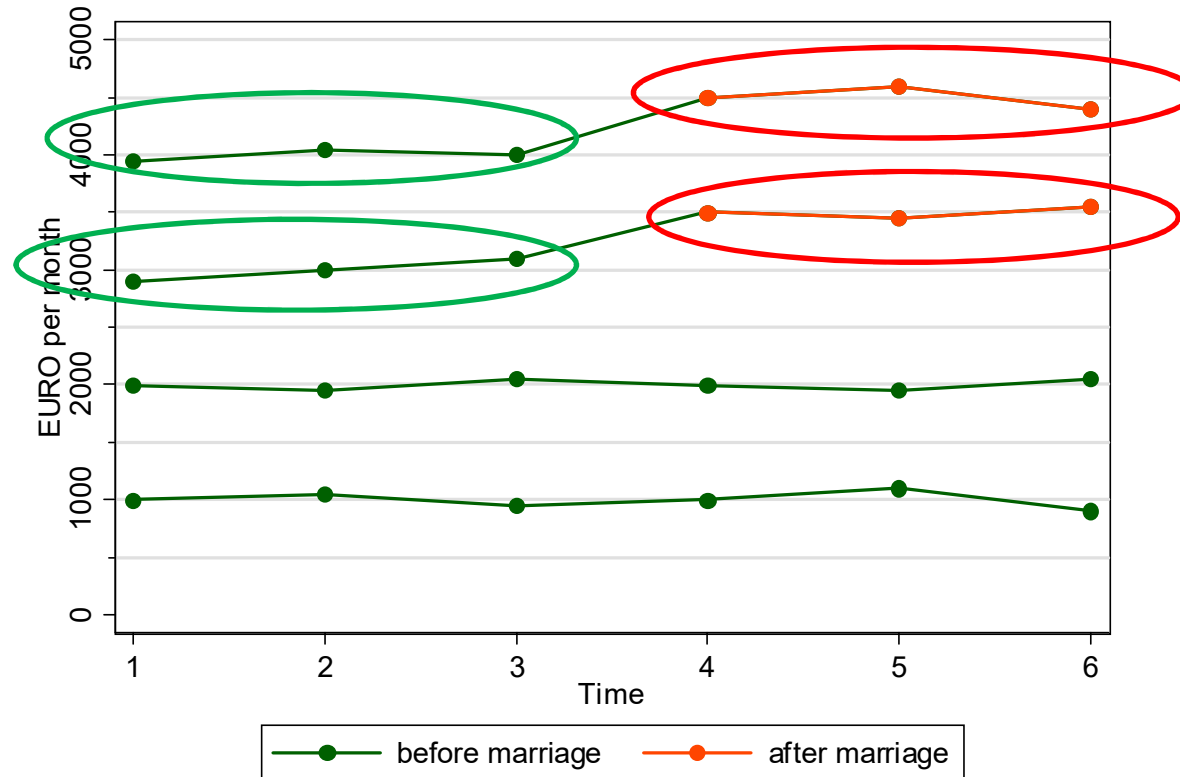
“Mechanics” of a FE-Regression



- 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

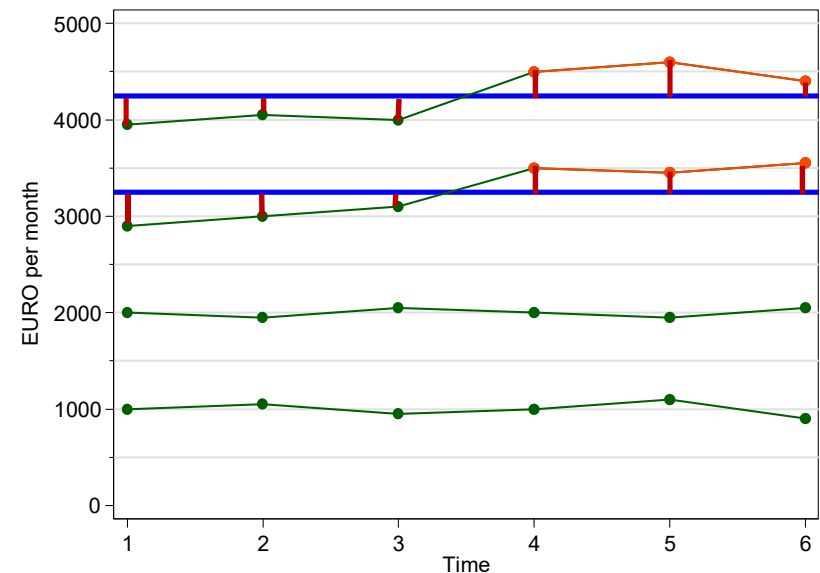
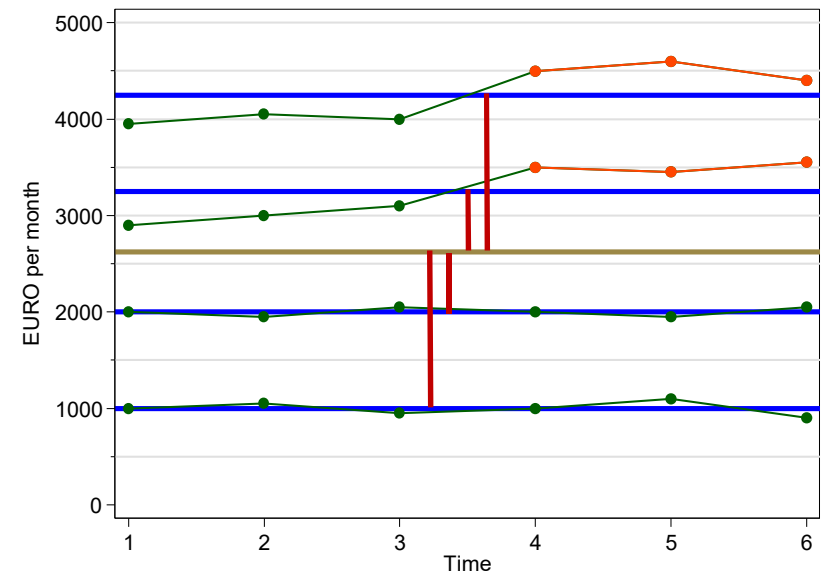
- This is the information used by a FE-regression



- 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
```

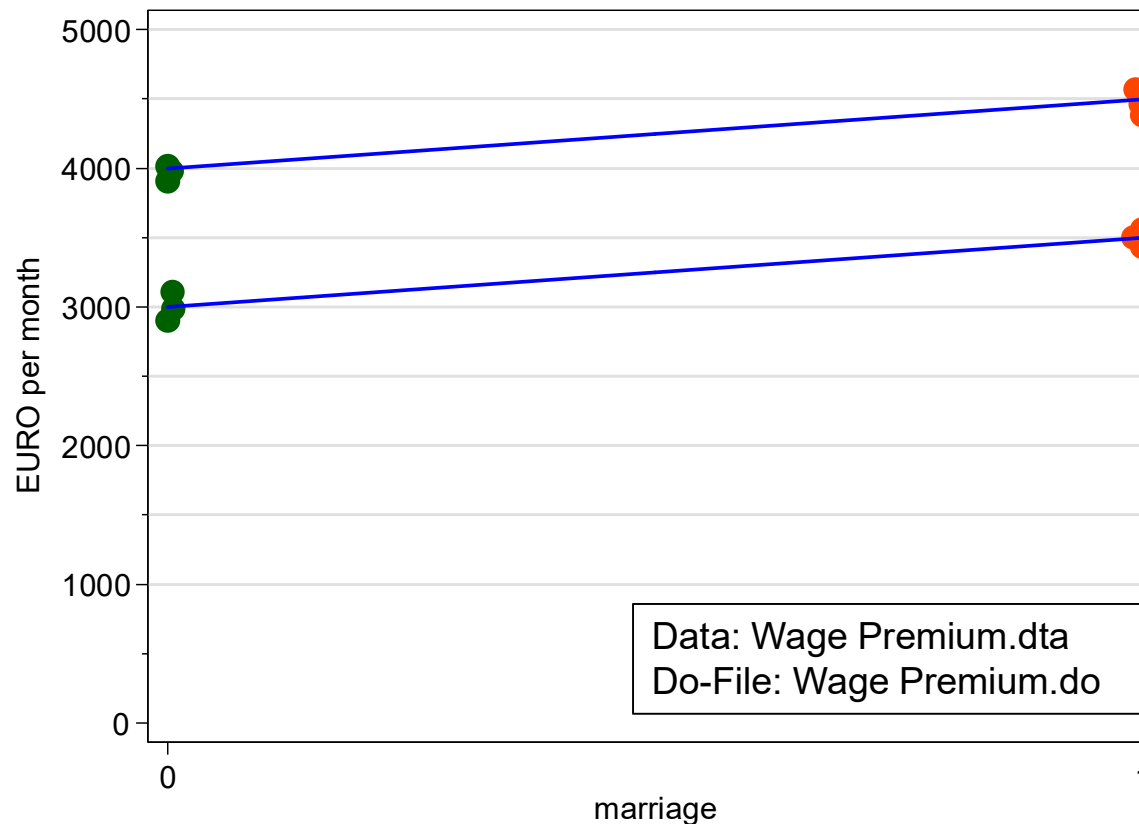
Source	SS	df	MS	Number of obs = 24			
Model	202500000	5	40500000	F(5, 19)	=	9052.94	
Residual	85000	19	4473.68421	Prob > F	=	0.0000	
Total	202585000	24	8441041.67	R-squared	=	0.9996	
				Adj R-squared	=	0.9995	
				Root MSE	=	66.886	

	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	1057.152
	2	2000	27.30593	73.24	0.000	1942.848	2057.152
	3	3000	33.4428	89.71	0.000	2930.003	3069.997
	4	4000	33.4428	119.61	0.000	3930.003	4069.997

- 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

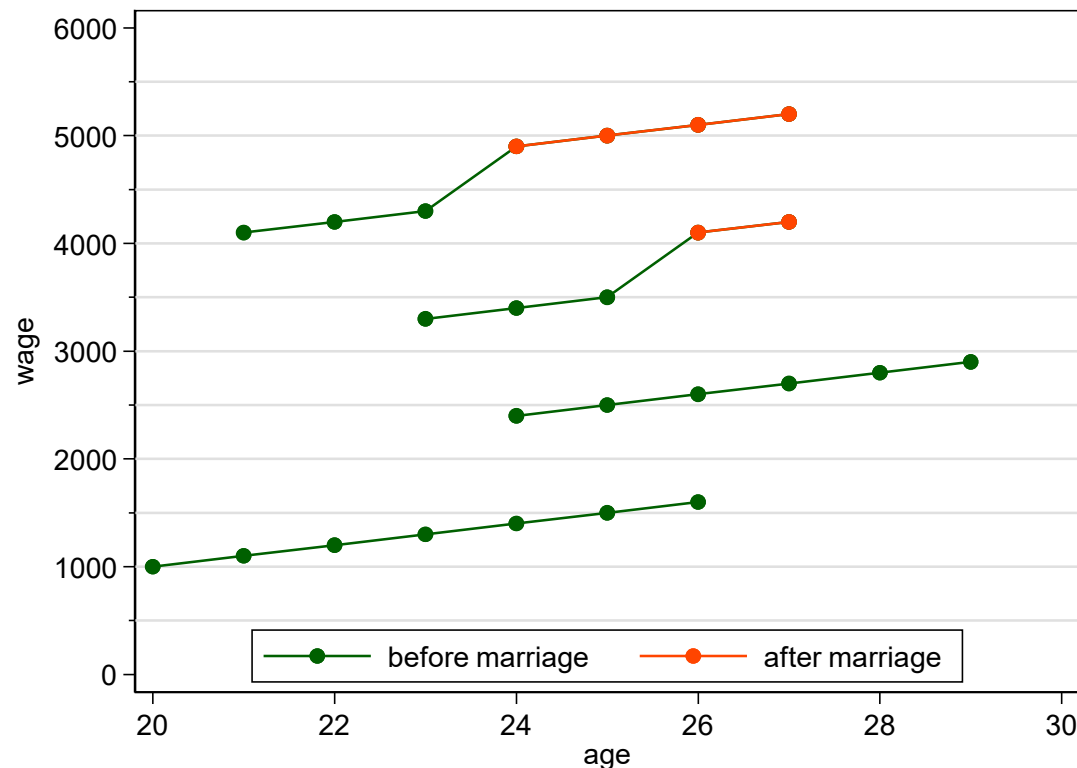
Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



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

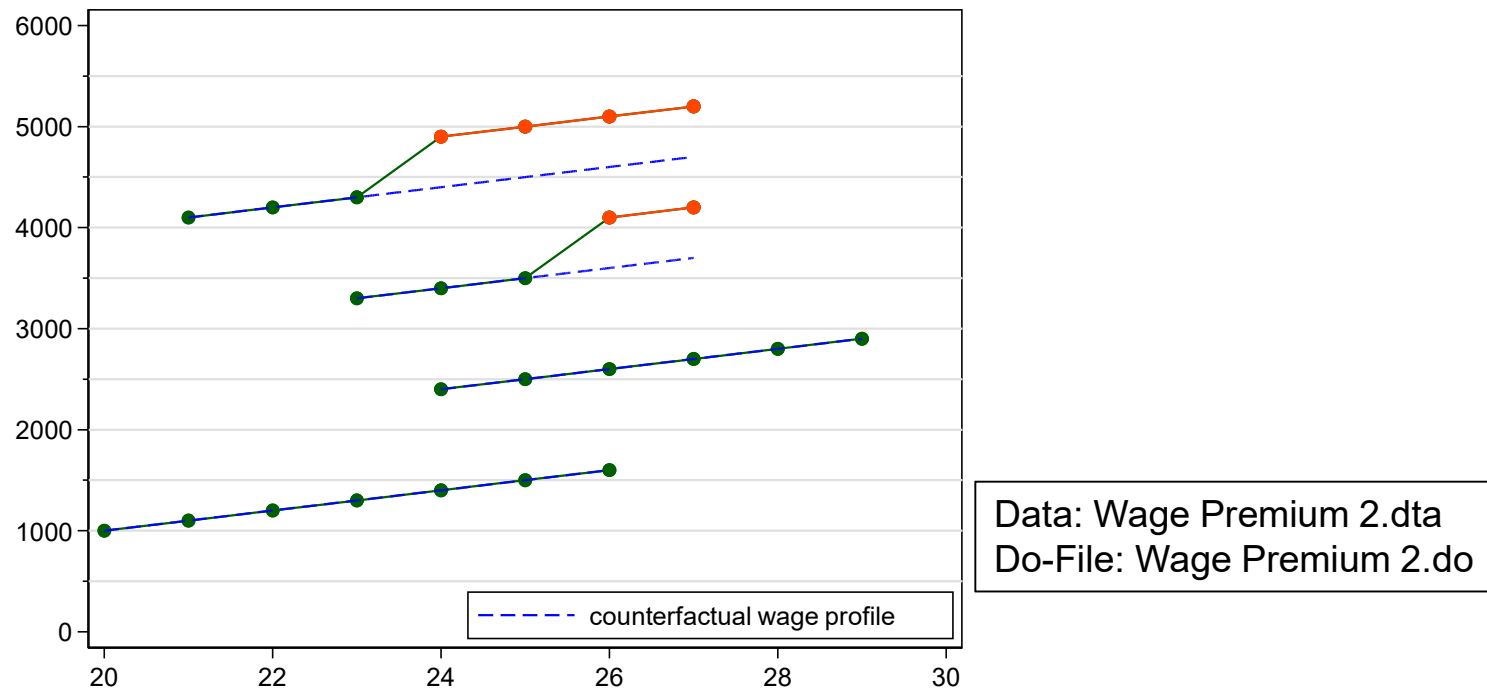
```
xtreg wage marr, fe
```

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.age`)

Do We Need a Control Group?

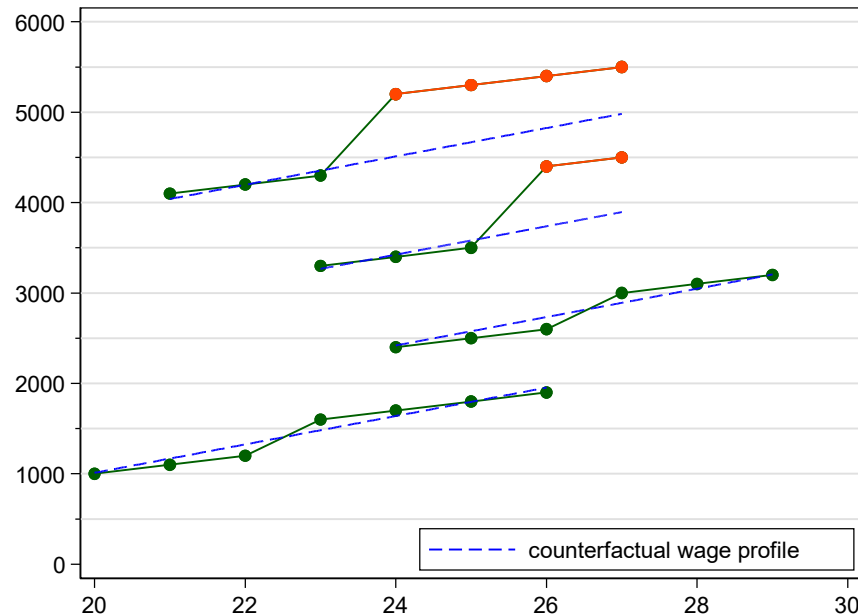
- In this example we would not even need the control group, because we have enough pre-treatment observations for estimating the wage profile



- 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 $t = 3.5$



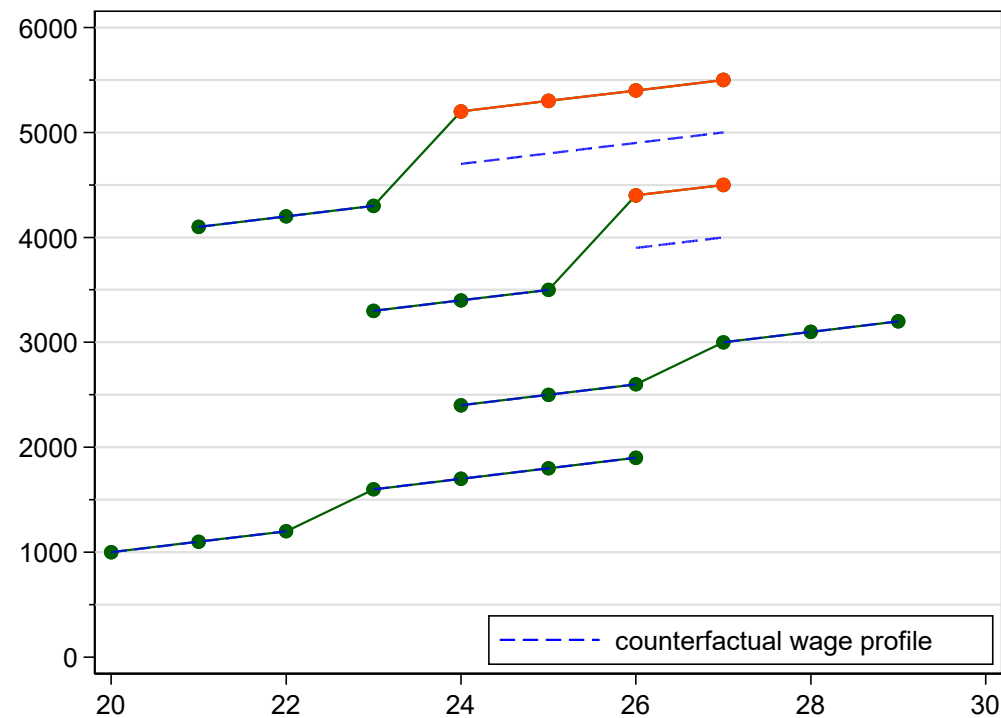
Now, a FE model without period control is biased

```
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)

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

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\delta} + \alpha_i + \bar{\varepsilon}_i$$

- “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 \mathbf{x}_{it}
 - The FE estimator is consistent even if $E(\mathbf{x}'_{it}\alpha_i) \neq \mathbf{0}$
- For consistent estimates we need
 - $E(\mathbf{x}'_{is}\varepsilon_{it}) = \mathbf{0}$, for all $s, t = 1, \dots, T$ *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

R-sq: within = 0.8982

between = 0.8351

overall = 0.4065

Obs per group: min = 6

avg = 6.0

max = 6

Random effects u_i ~ Gaussian

corr(u_i, X) = 0 (assumed)

theta = .96138358

Wald chi2(1) = 128.82

Prob > chi2 = 0.0000

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)				

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

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 $[\alpha_i]$ 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, sociologists 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

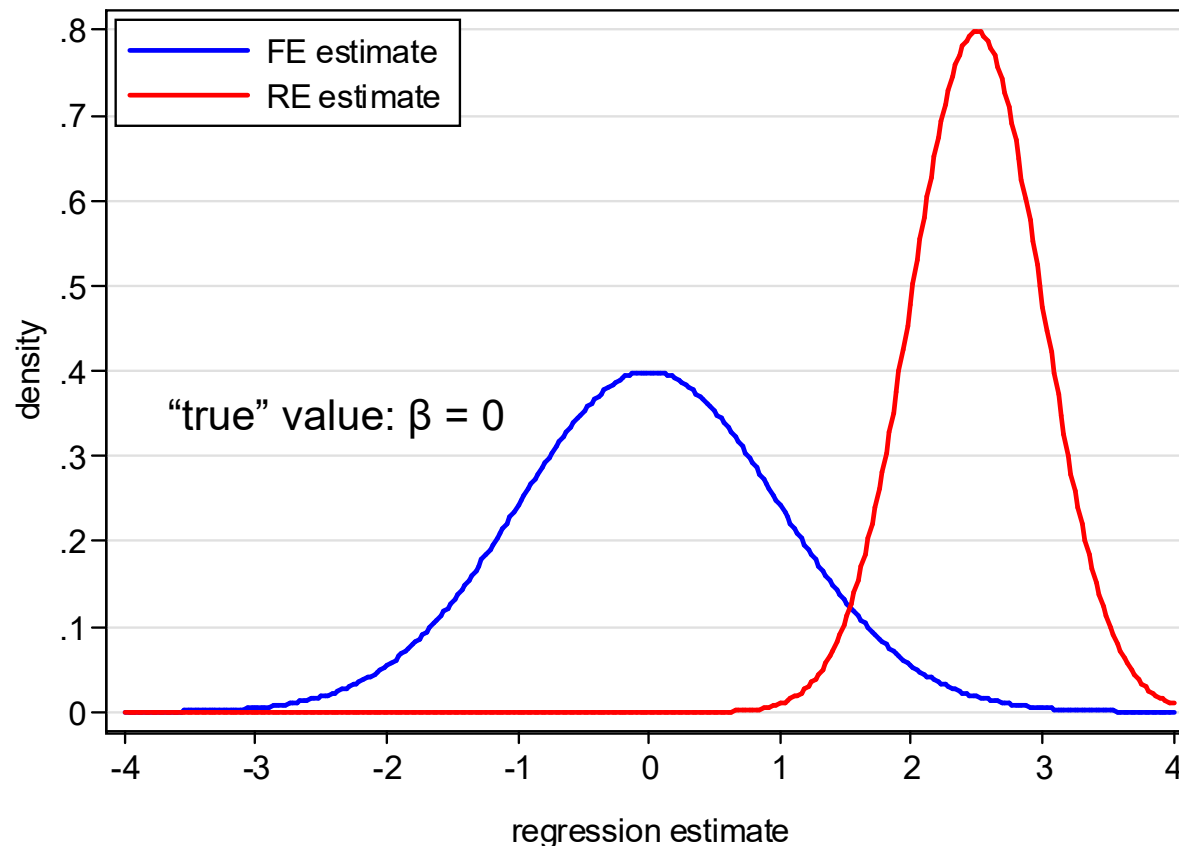
- Within estimators cannot estimate the effects of time-constant variables
 - E.g., sex, nationality, race, social origin, birth cohort, etc.
- Often it is argued that wiping out all time-constant regressors is a shortcoming of FE estimation
 - However, wiping out all time-constant regressors is **not** 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.



A thought experiment to illustrate the point

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

$$H_0 : \hat{\beta}_{RE} = \hat{\beta}_{FE}$$

$$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})' [\hat{V}(\hat{\beta}_{RE}) - \hat{V}(\hat{\beta}_{FE})]^{-1} (\hat{\beta}_{RE} - \hat{\beta}_{FE})$$

$$H \overset{a}{\sim} \chi^2(k)$$

- If you are **not** able to reject H_0 , then you can use RE

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



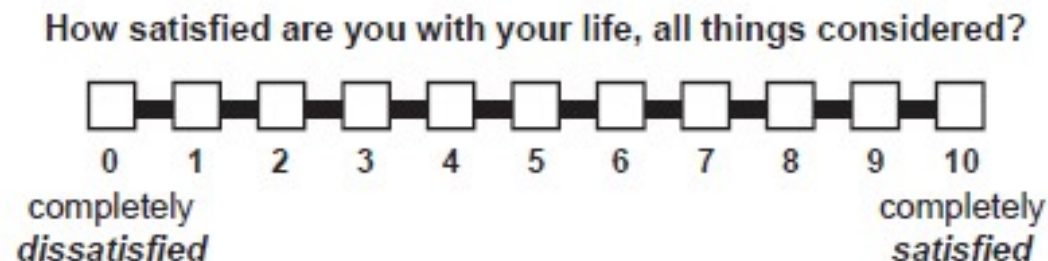
Example: Does Marriage Make Happy?

- 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

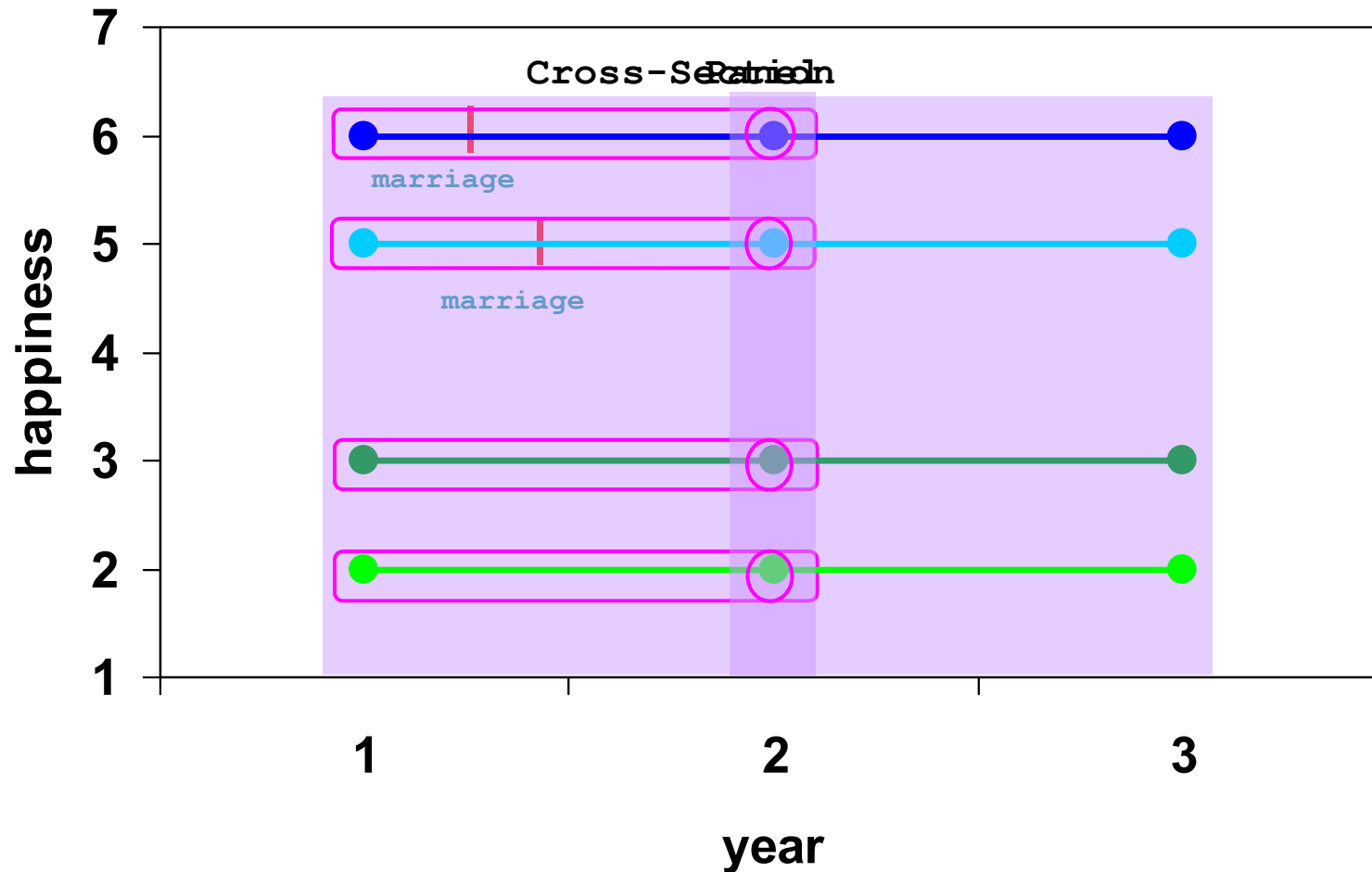
155. In conclusion, we would like to ask you about your satisfaction with your life in general.

 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:
marriage dummy

age:
age in years

loghhinc:
natural logarithm of
annual post-government
household income

yrsmarried:
years since marriage

woman:
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

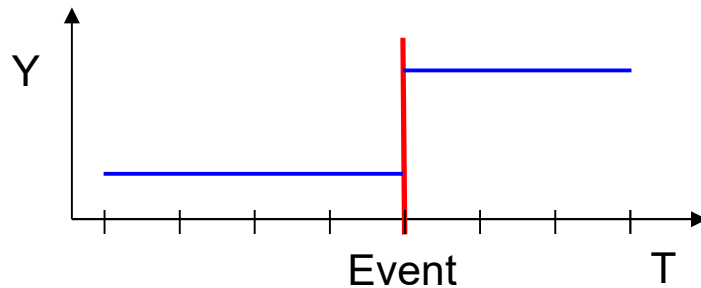
y_{it}, \dots, y_{iT} measured after event

- 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?

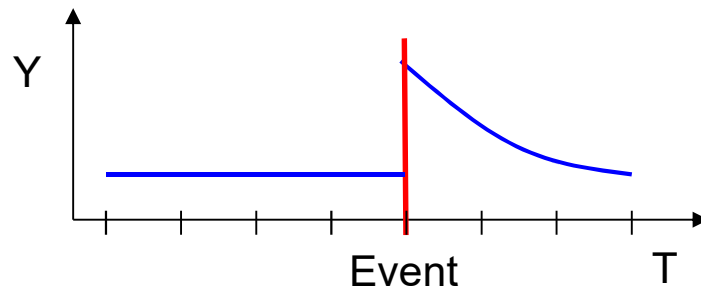
- With panel data we can investigate the time path of a causal effect
 - Termed “impact function” (IF) by Andreß et al. (2013)
 - Different impact functions can be modeled



Step impact function

Immediate and permanent impact

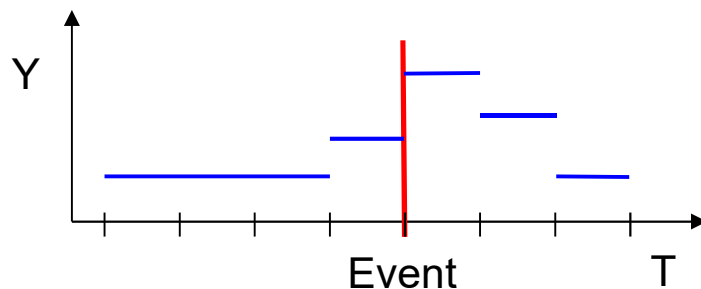
- event dummy $(0,0,0,0,\mathbf{1},\mathbf{1},\mathbf{1})$



Continuous impact function

Immediate, but transitory impact

- event dummy $(0,0,0,0,\mathbf{1},\mathbf{1},\mathbf{1})$
- linear event time $(0,0,0,0,\mathbf{0},\mathbf{1},\mathbf{2})$
- quadratic event time $(0,0,0,0,\mathbf{0},\mathbf{1},\mathbf{4})$



Dummy impact function

Arbitrary impact (including anticipation effect)

- dummy event time
 - -1 dummy $(0,0,0,\mathbf{1},0,0,0)$
 - 0 dummy $(0,0,0,0,\mathbf{1},0,0)$
 - 1 dummy $(0,0,0,0,0,\mathbf{1},0)$
 - 2 dummy $(0,0,0,0,0,0,\mathbf{1})$

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
```

```
. xtodes, 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	5.521111111111	Sample F (2000)
551	3.77	9.291111	Sample H (2006)
390	2.67	11.95111	
348	2.38	14.33	11111111111111111111	Sample A/B (1984)
347	2.37	16.7011111111	Sample G (2002)
337	2.30	19.0011	
302	2.06	21.0711111	
271	1.85	22.92111111	
250	1.71	24.631111111111111111	
235	1.61	29.61	11.....	Attrition sample A/B
222	1.52	31.13	111.....	Attrition sample A/B
141	0.96	36.96	1111.....	Attrition sample A/B
8981	61.37	100.00	(other patterns)	
14634	100.00		XXXXXXXXXXXXXXXXXXXXXXXXXXXX	

Describing Panel Data II

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

```
. xttrans marry , freq
```

		marry (t+1)		
marry (t)		0	1	Total
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?

```
. xtsum marry age loghhinc woman
```

Variable		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

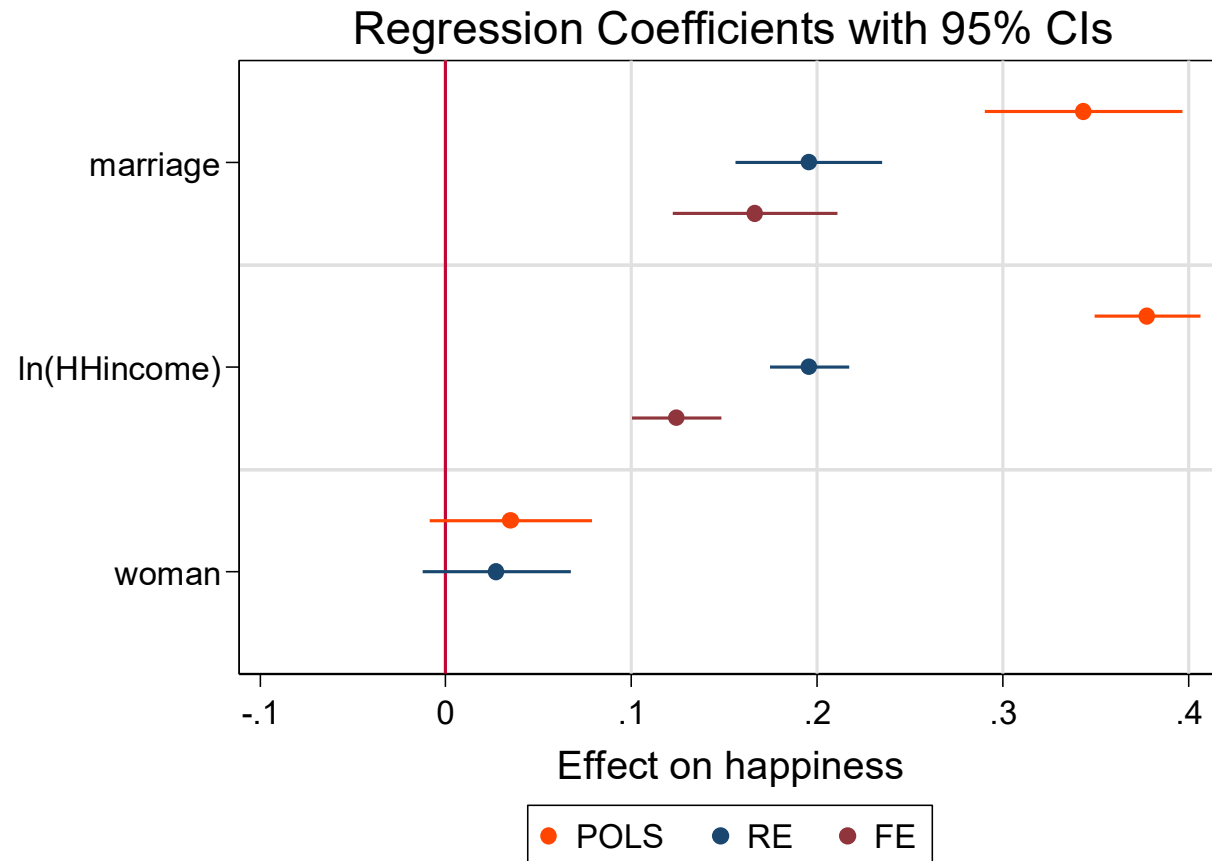
FD: Due to gaps in the data we loose some groups (clusters) and observations.

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	-0.139	0.136	0.186
	0.0139	0.0144	0.0186	0.0167
age			-0.033	-0.039
			0.0015	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
Group variable: id

Number of obs = **121919**
Number of groups = **14634**

R-sq: within = **0.0162**
between = 0.0221
overall = 0.0152

Obs per group: min = 2
avg = 8.3
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)				

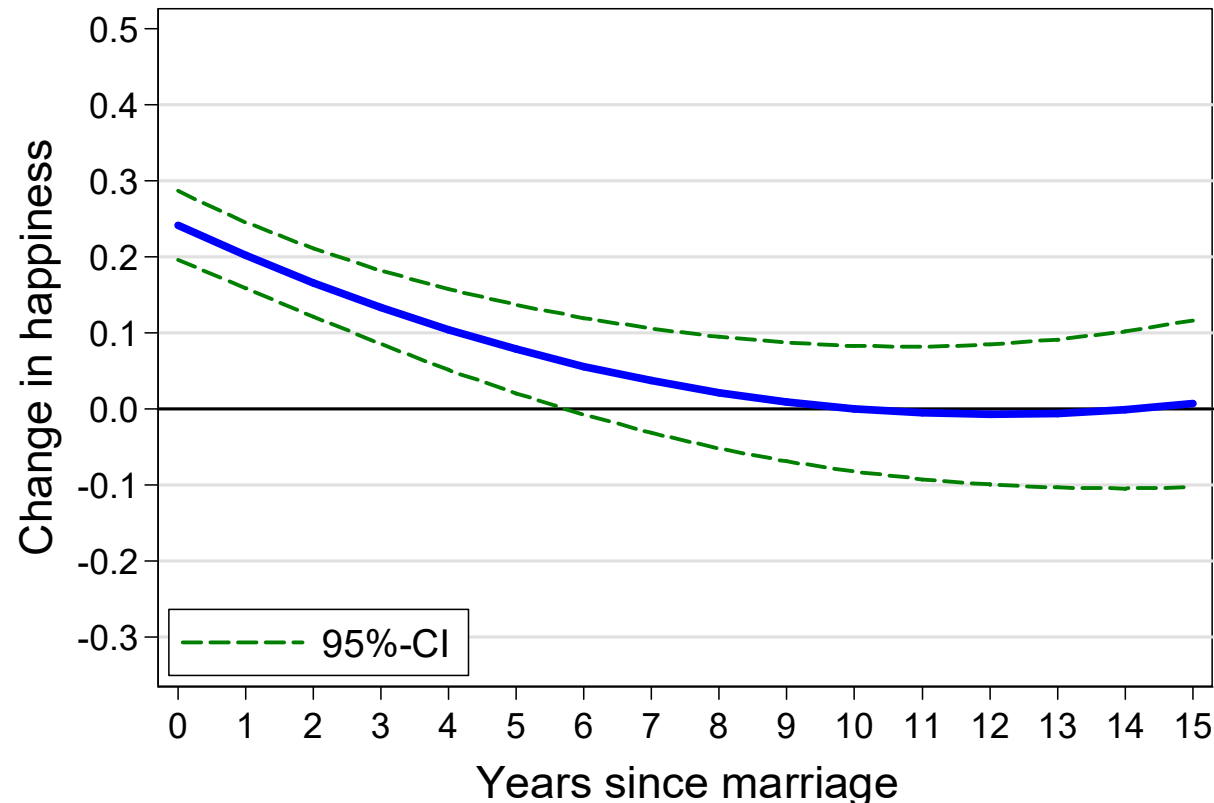
Data: Happiness2.dta
Do-File: Happiness 3 Regressions.do

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)

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

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

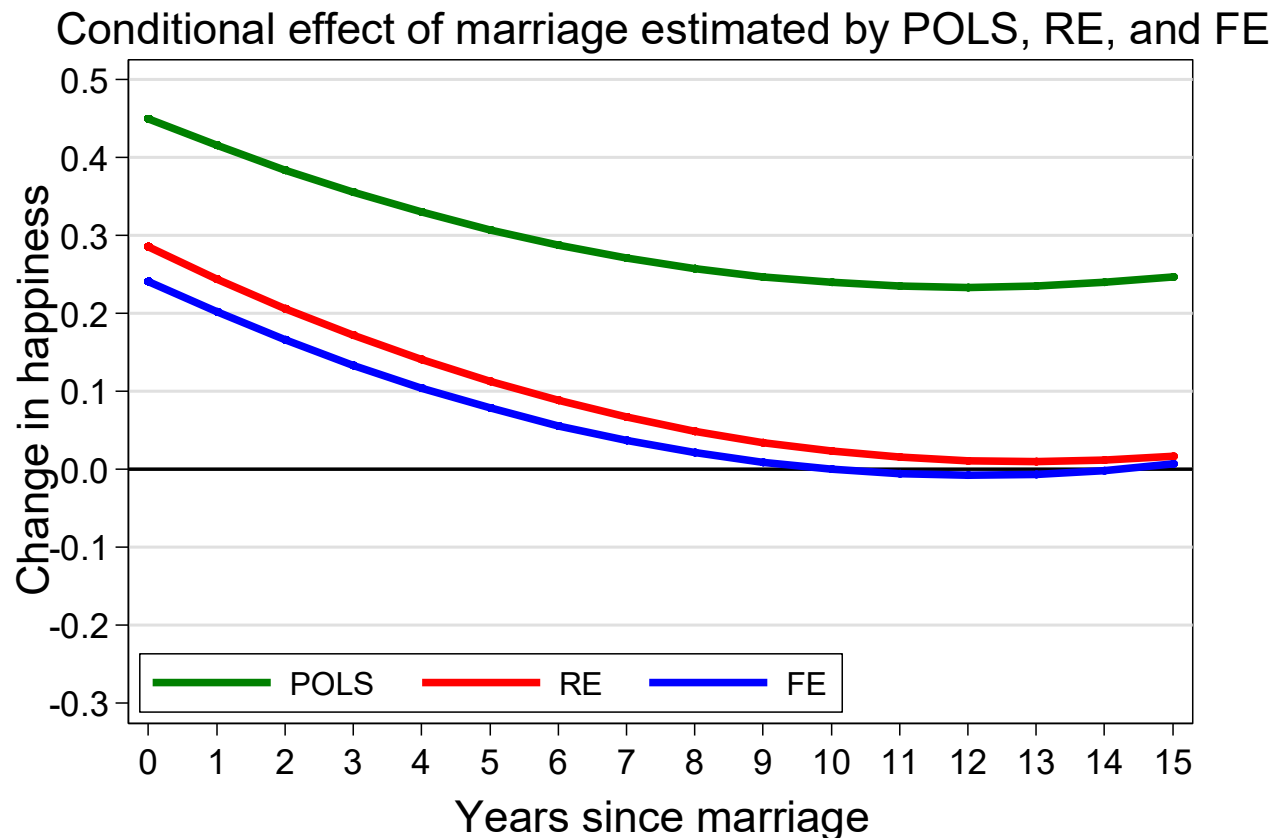


What is the reference point?

- 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

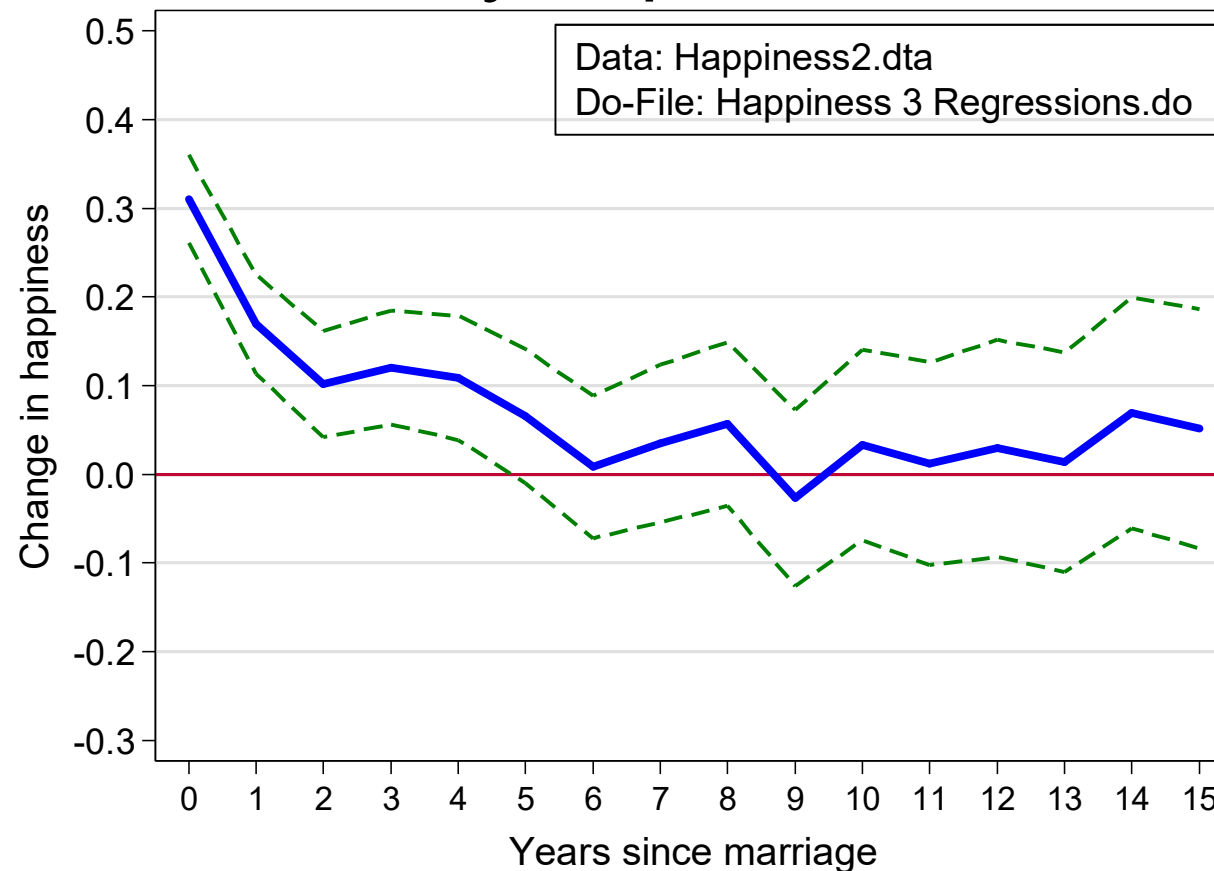
- Comparing the conditional marriage effect over models
 - We see that POLS heavily over-estimates the marriage effect
 - Due to self-selection
 - RE slightly over-estimates the marriage effect
 - RE works quite well with these data (probably due to the long panels)



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 (y_m)
 - 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 ($i . y_m$) [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

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“

Deutsche Post 

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

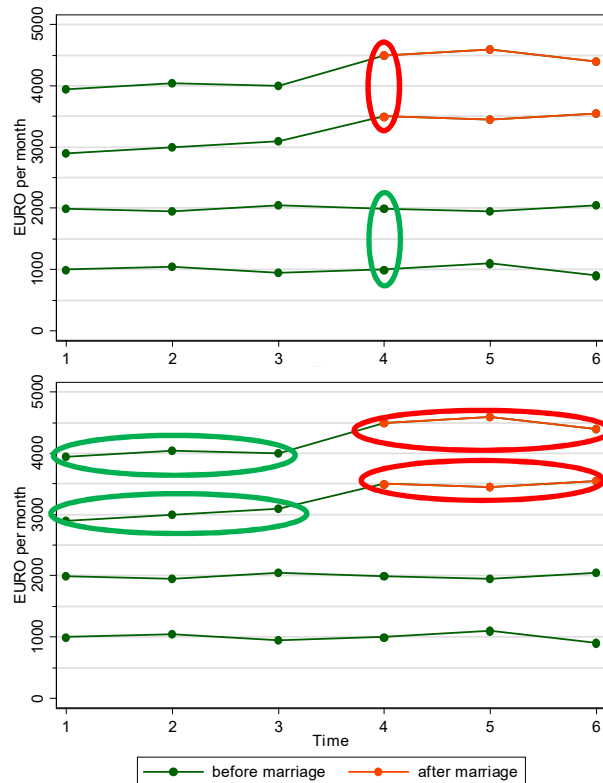
Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



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)

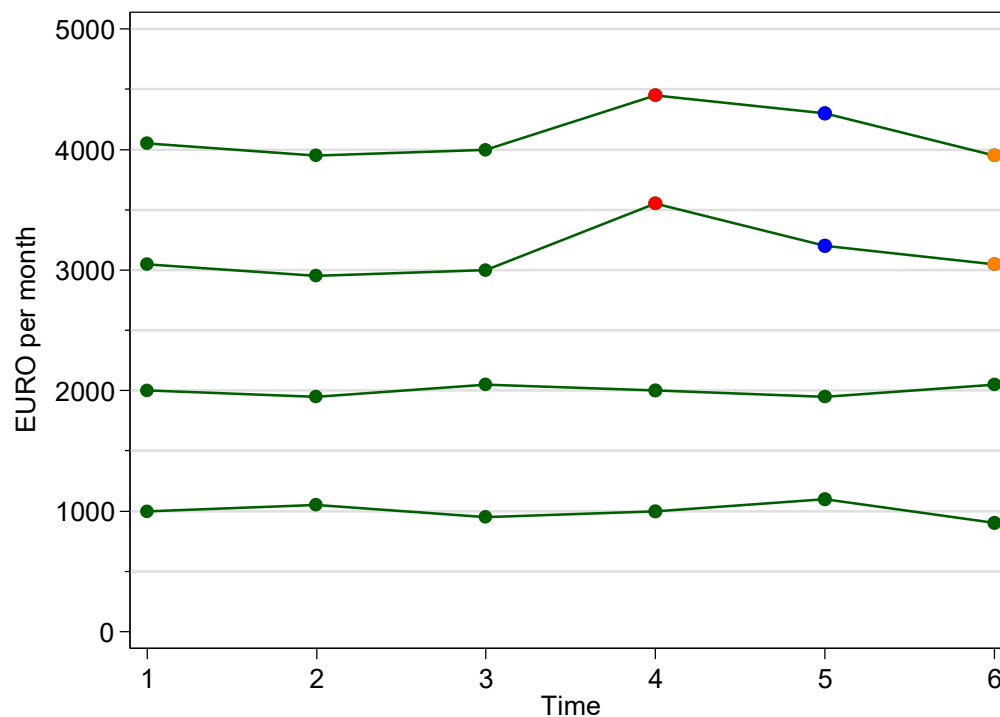
The descriptive interpretation is always correct.

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.

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



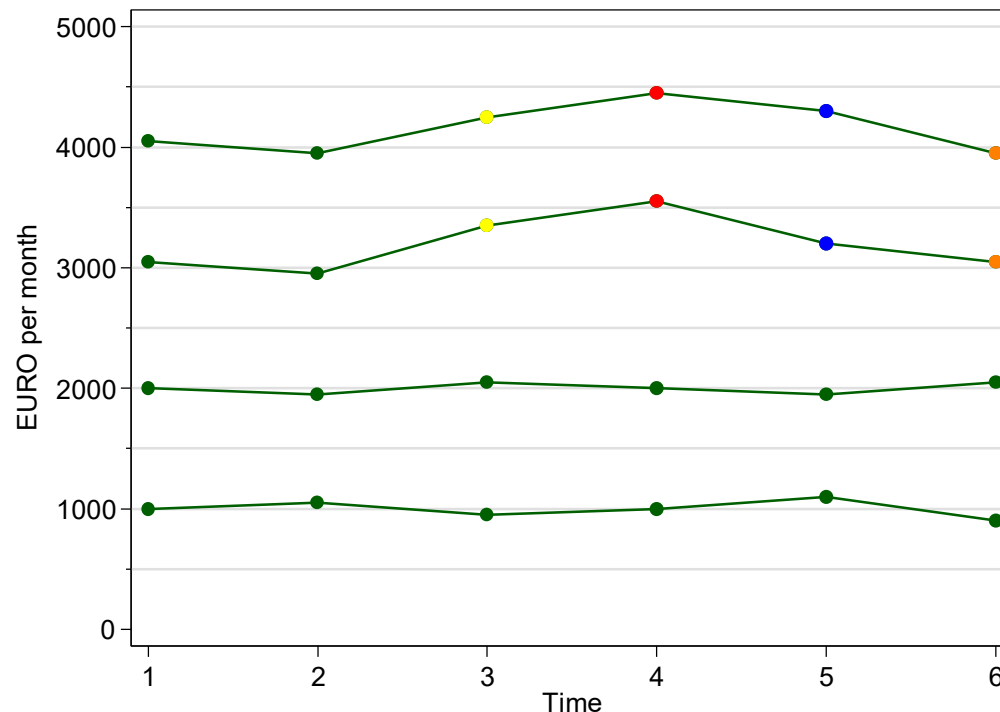
Data: Impact Function.dta
Do-File: Impact Function.do

• Example

- 0.y 500
- 1.y 250
- 2.y 0
- The effect of 0.y is the average wage of the red points minus the average wage of the green points (only the treated)
- The effect of 1.y is ... blue points ...
- The effect of 2.y 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

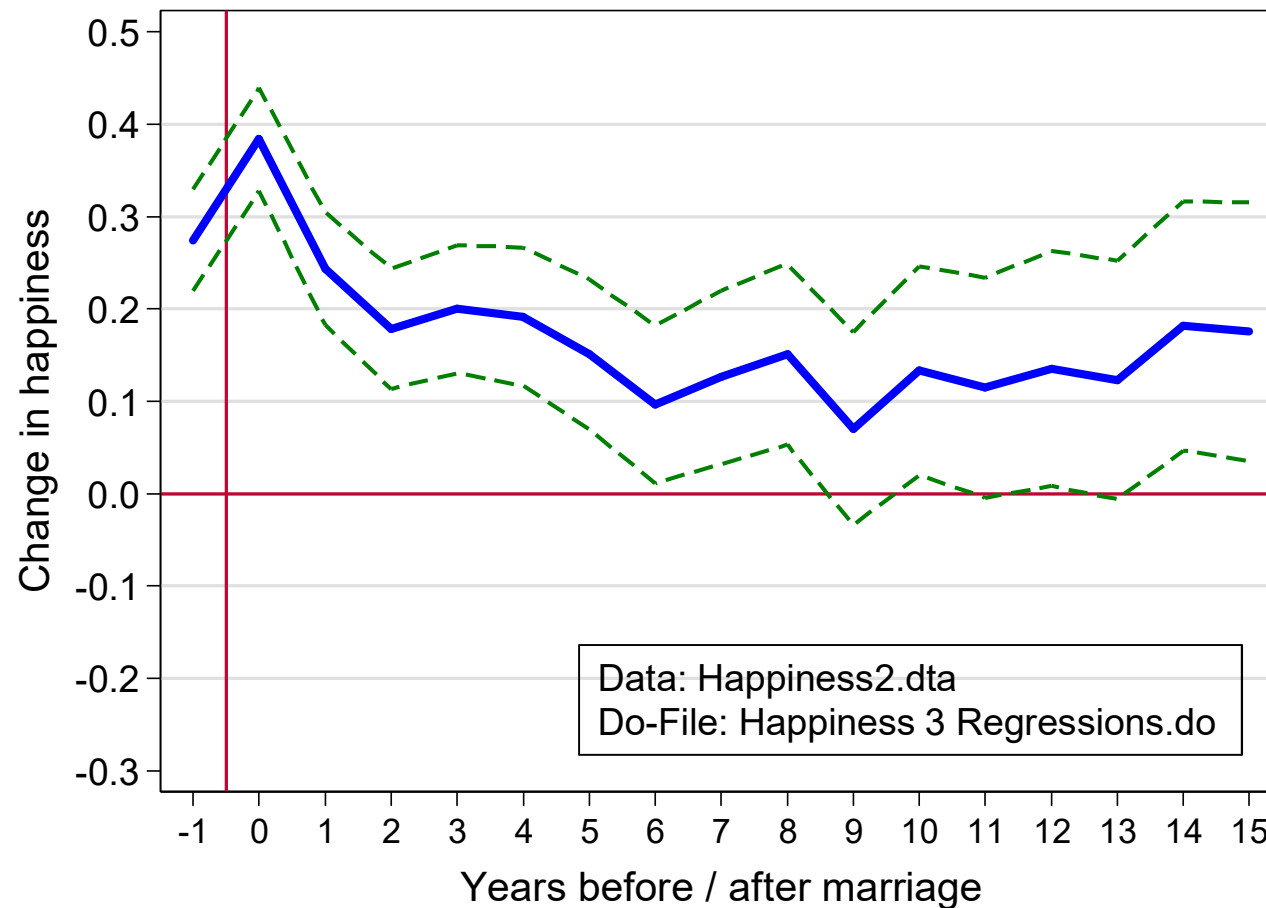
0 .ym	400
1 .ym	150
2 .ym	-100

- Remedy: we add a -1-dummy

-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 $t = -1$

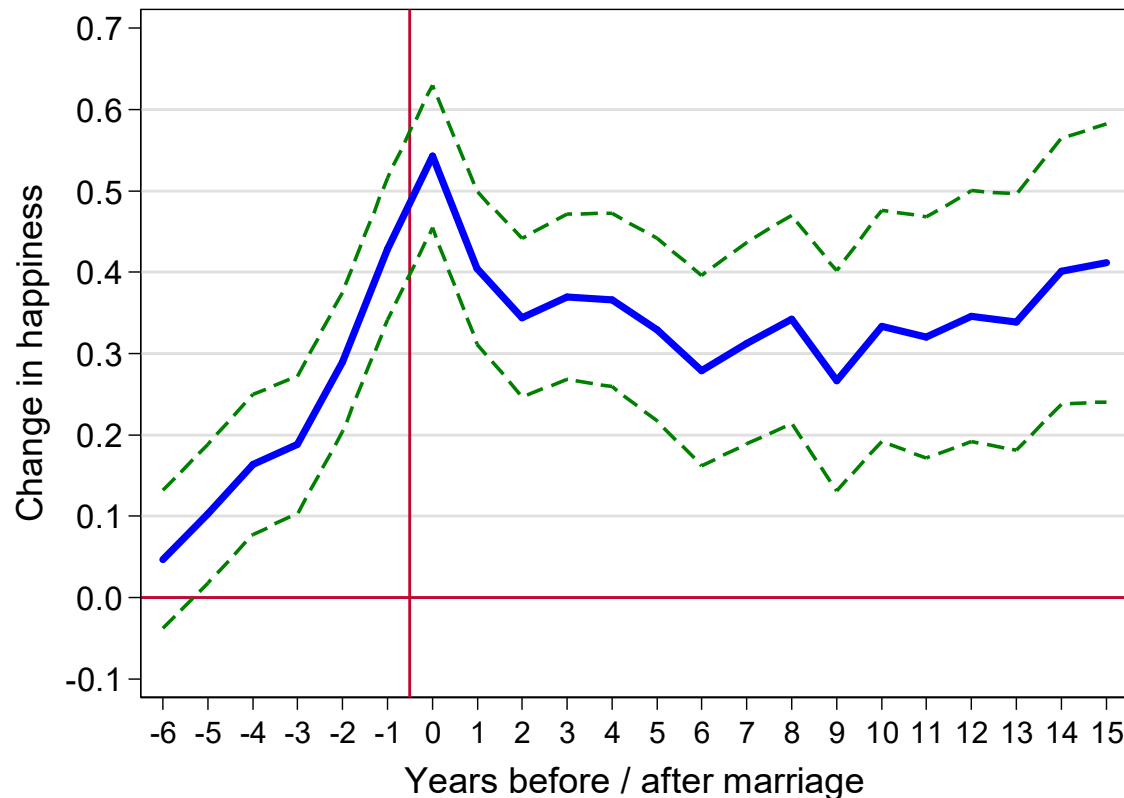


We observe a strong anticipation effect

- 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!

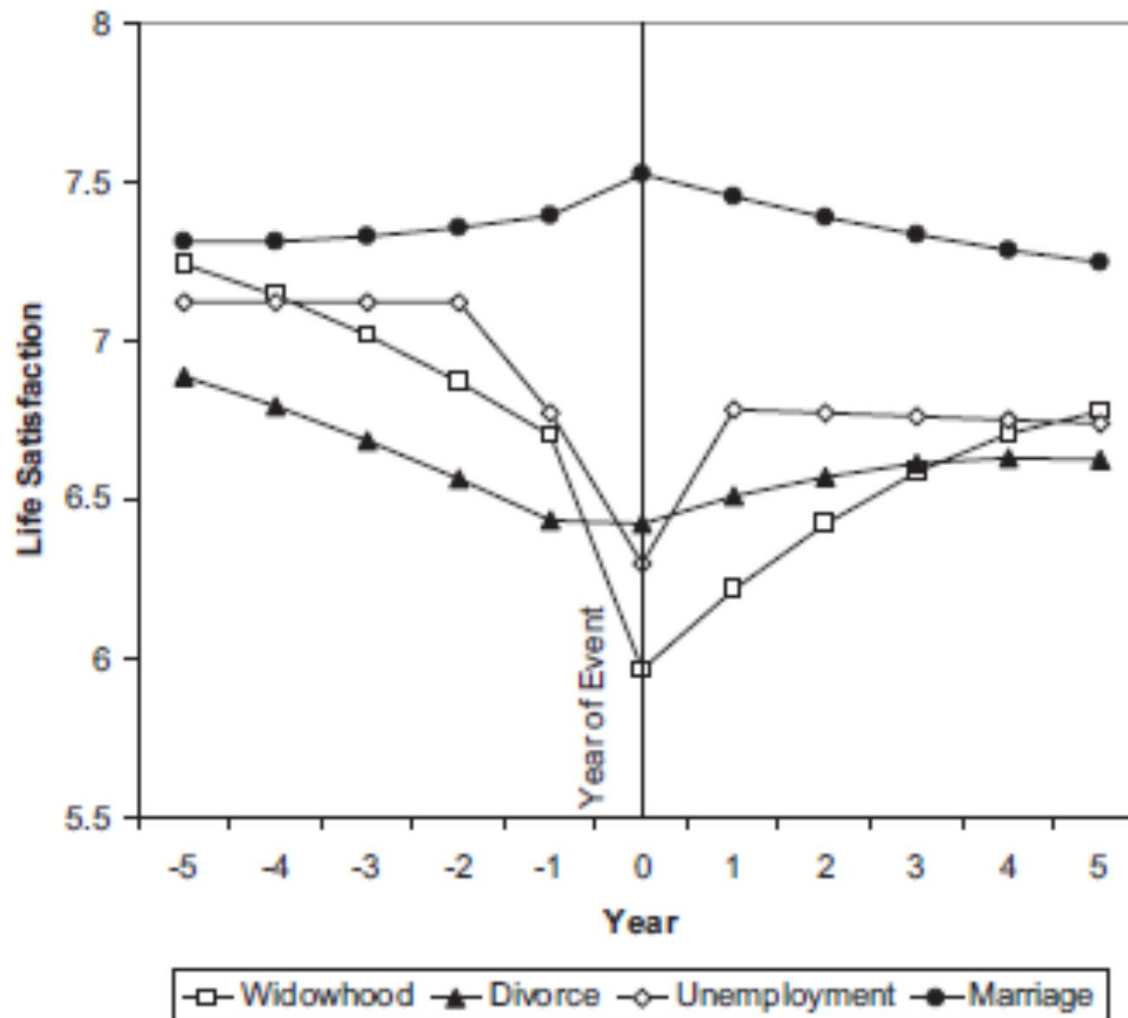
But this is nonsense, and due to the fact that we have omitted an important variable (being in a partnership).

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



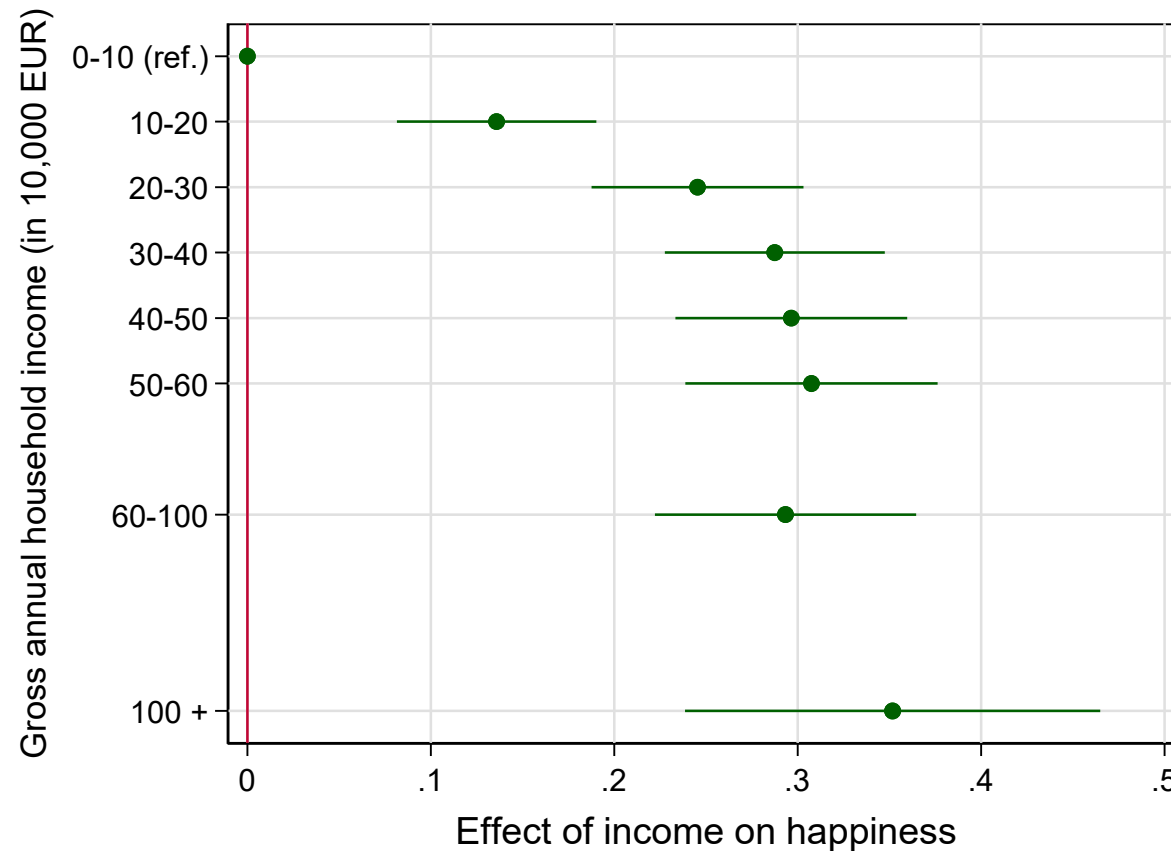
Diener, E. et al. (2006) Beyond the Hedonic Treadmill. *American Psychologist* 61 (4): 305-314.

- 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

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 paradoxon”
 - 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”

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

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

$\beta_1 = \beta$ (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

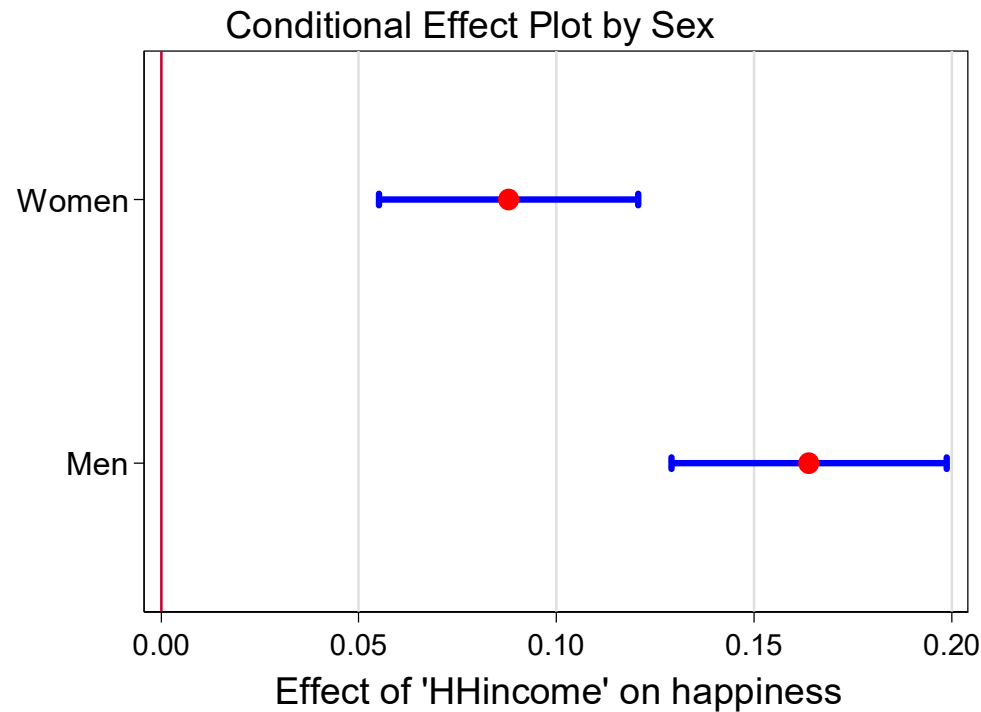
```
. xtreg happy marry i.woman#c.loghhinc age, fe vce(cluster id)
note: 1.woman omitted because of collinearity
```

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
happy							
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

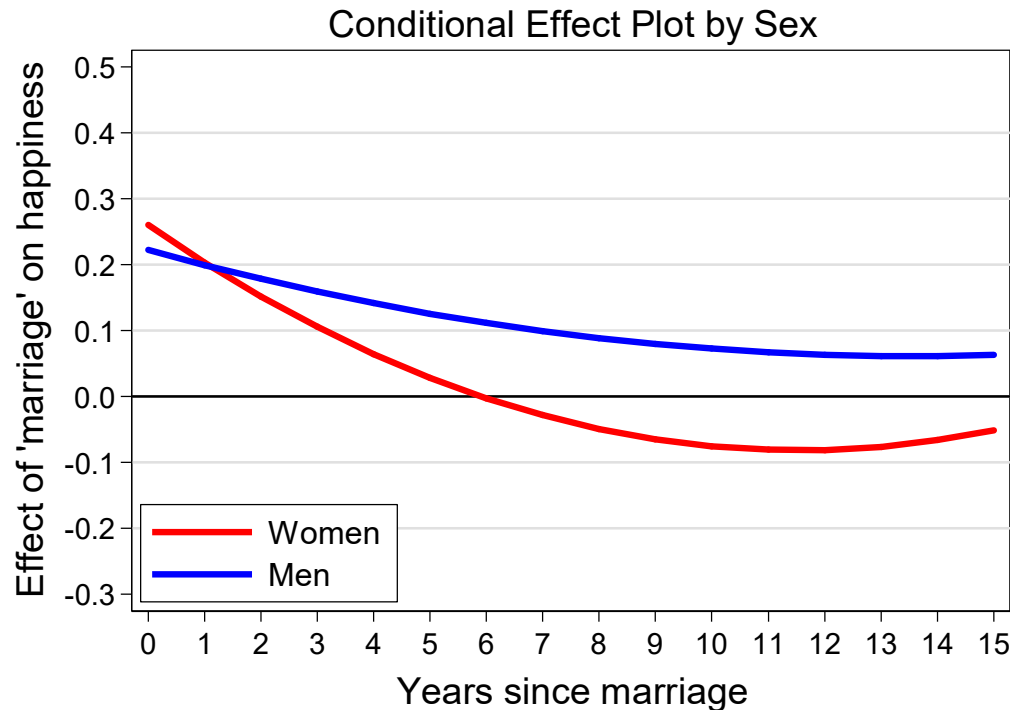


The nested effects specification is often more informative

- 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
```

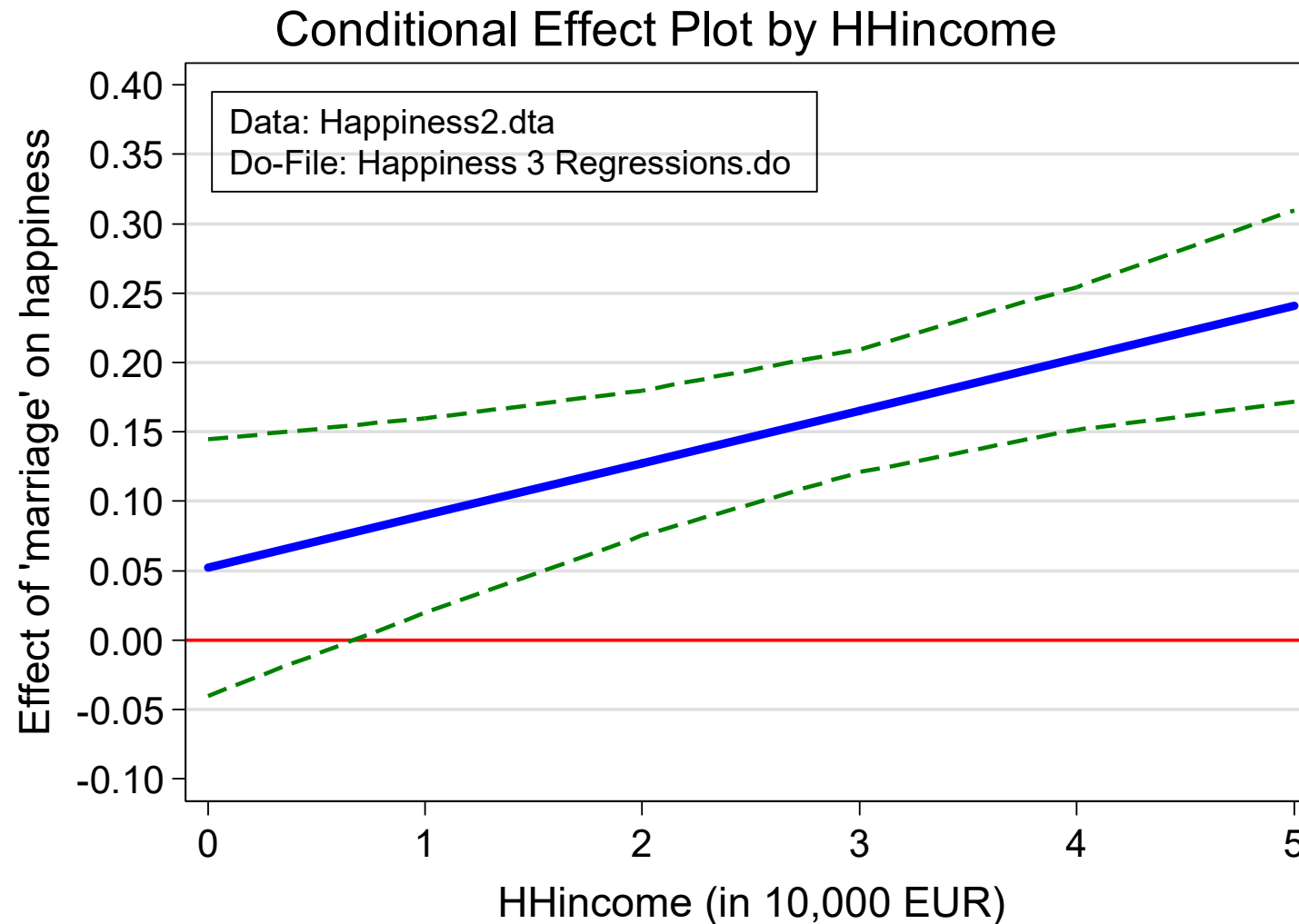
Interaction of Two Time-Varying Variables

- 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)
```

		Coef.	Robust 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



Chapter VI: Modeling Individual Growth

Section: Growth Curve Models

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

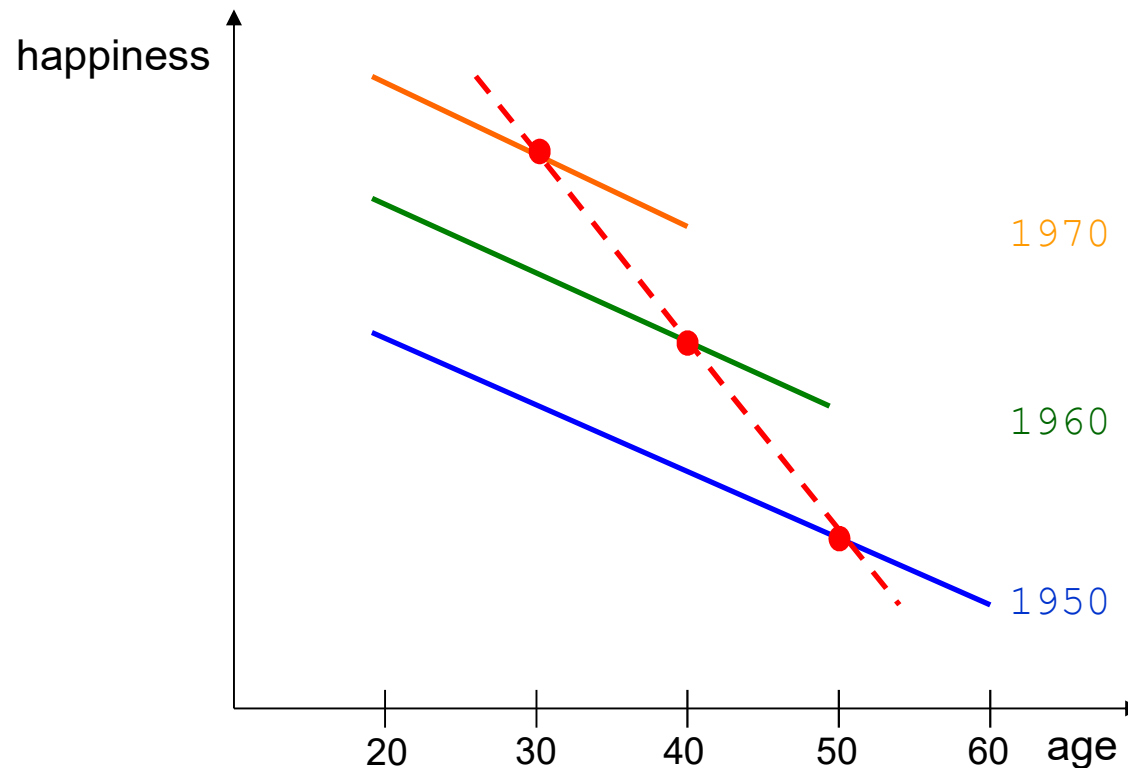


Modeling Individual Growth

- 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

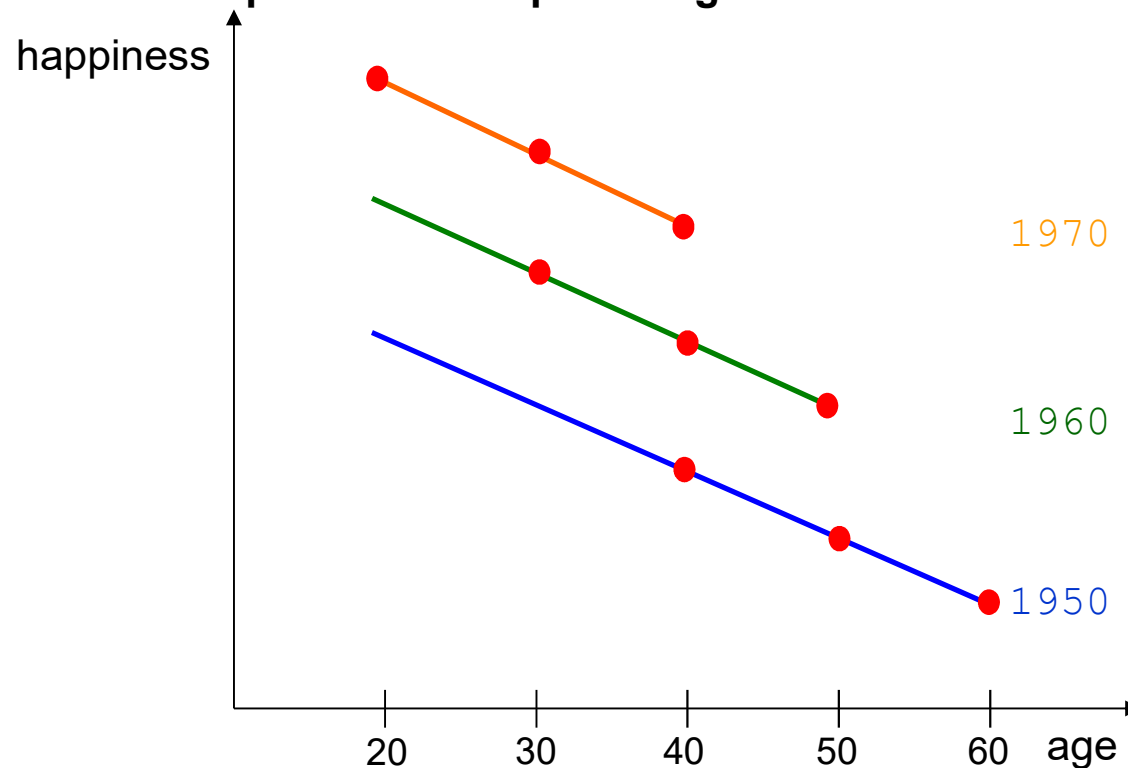
Cross-Sectional Data

- 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?

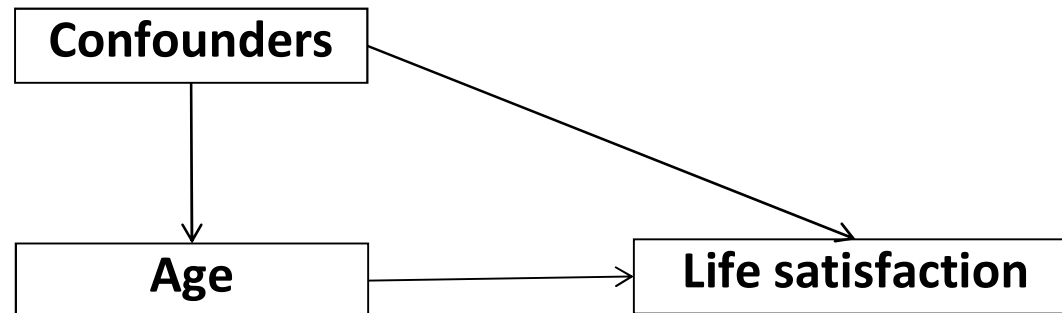
- The naive strategy is to estimate the following equation by POLS:

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

- 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



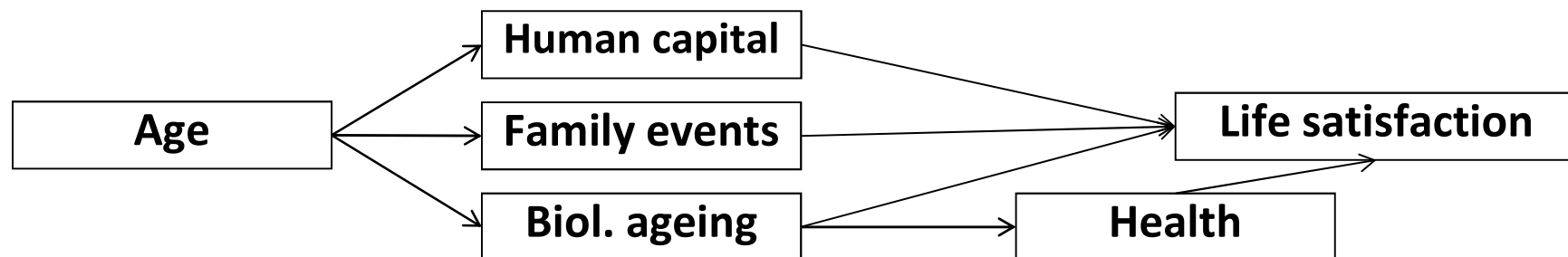
- How can it be that something affects age? After all, age is not manipulable
 - But in a sample some mechanisms might affect the sample composition at different ages (composition effects)
- Observed confounders
 - Cohort effects might distort the growth curve estimates
 - Older cohorts (born shortly 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



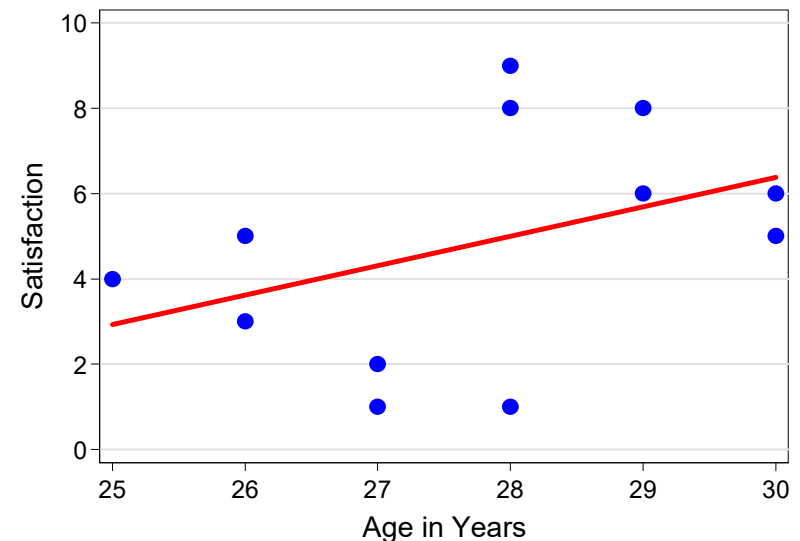
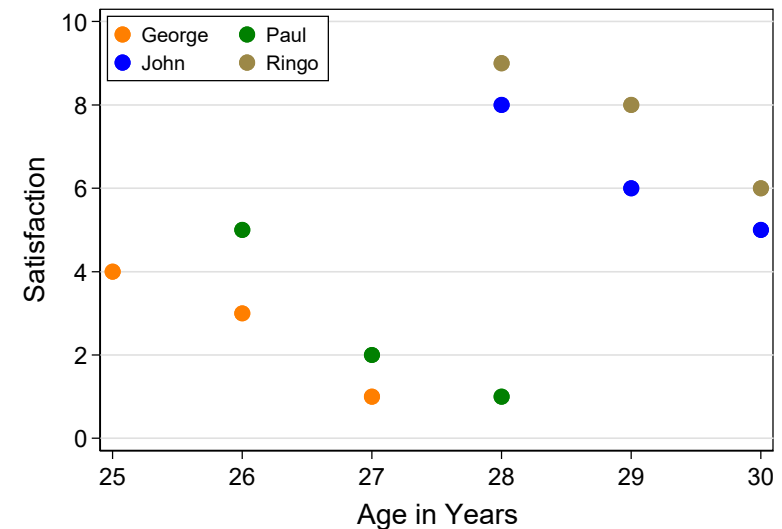
- **Methodological lesson II**
 - Do not control for intervening mechanisms
 - First, get the gross causal effect of age by controlling for confounders only
 - In a second step, try to get closer on the intervening mechanisms by adding them to the model

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

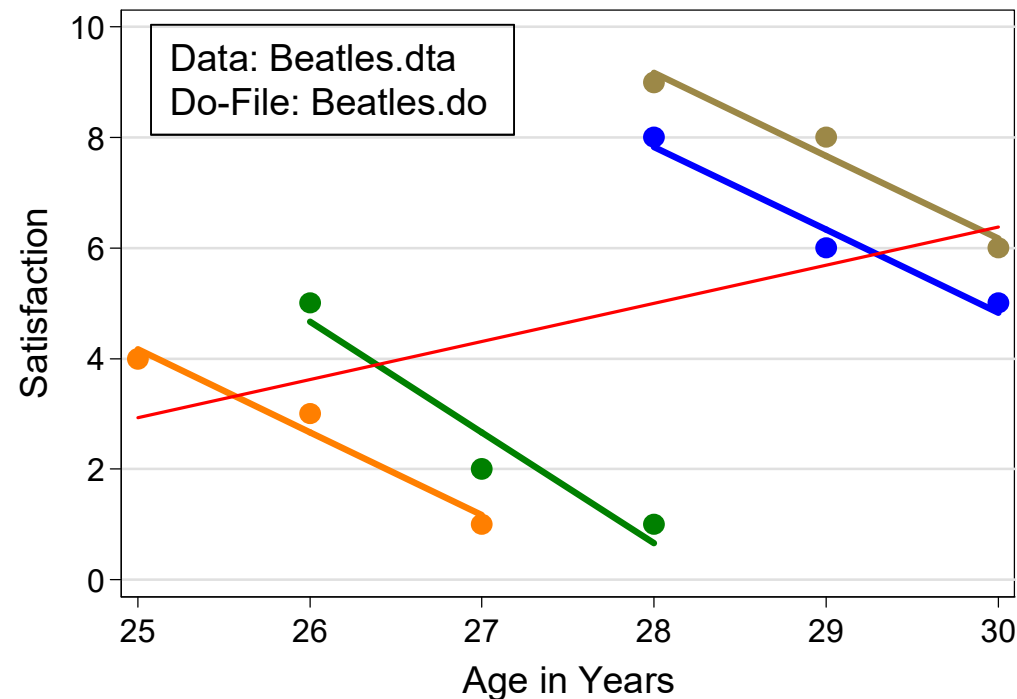
Source: Kohler/Kreuter (2009) Data Analysis Using Stata. Stata Press. Pp. 242 ff.



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

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

	happy	Coef.	Robust 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))
```

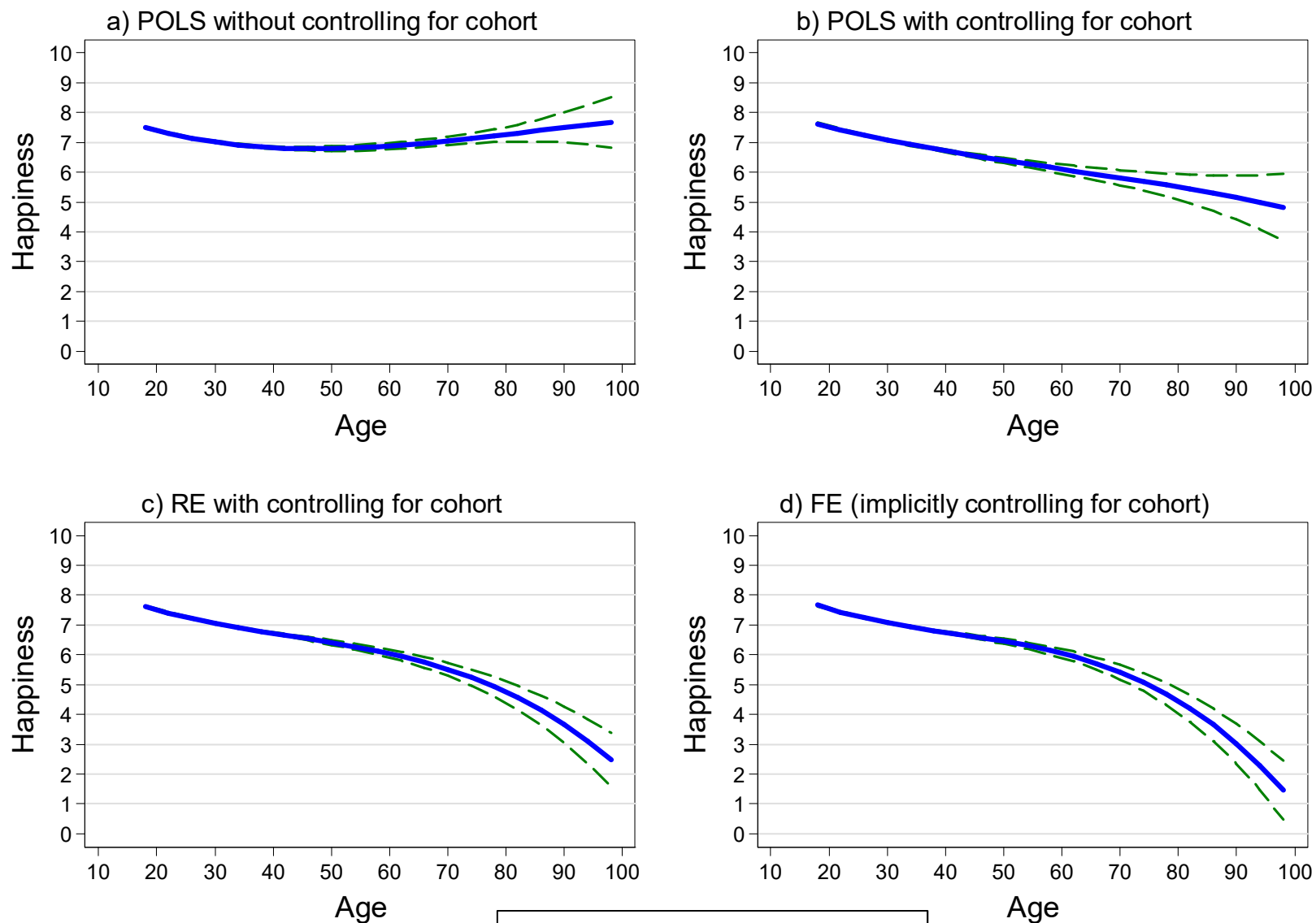
```
. marginsplot, recast(line) recastci(rline)
```

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

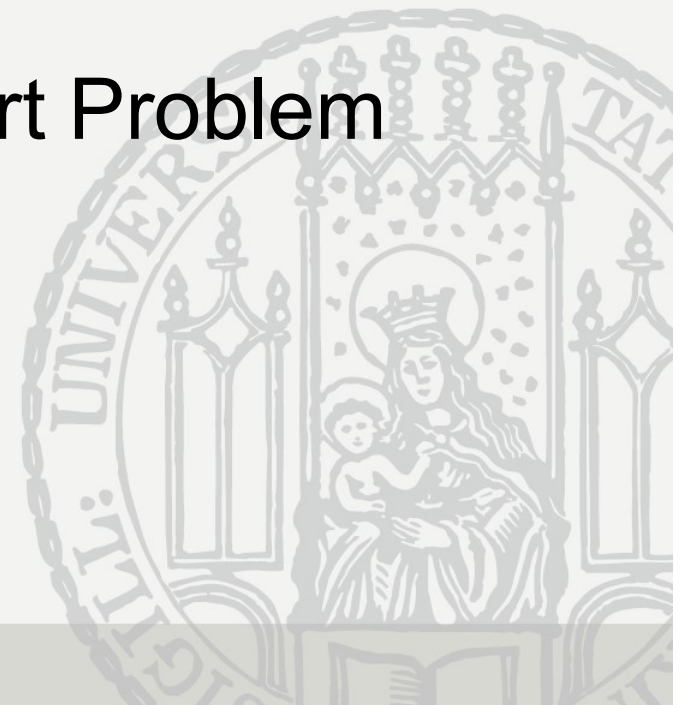
Happiness Example: Comparing Different Specifications

- 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

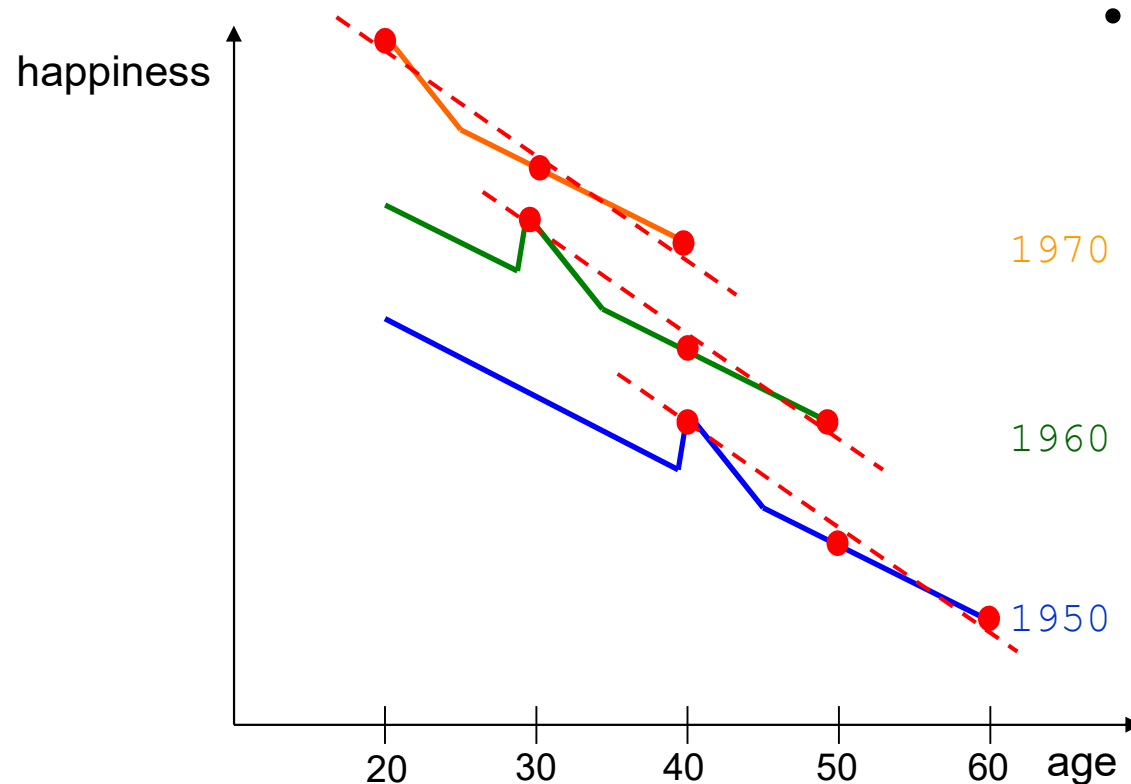
$$y_{it} = \beta x_{it} + \gamma A_{it} + \delta C_i + \mu(C_i + A_{it}) + \alpha_i + \varepsilon_{it}$$

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

- 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 Biasing 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:

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

note: 2009.year omitted because of collinearity

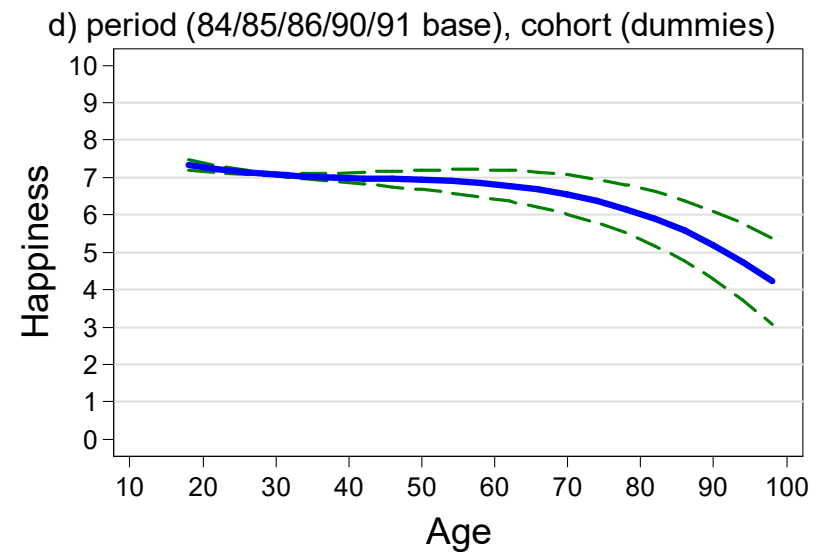
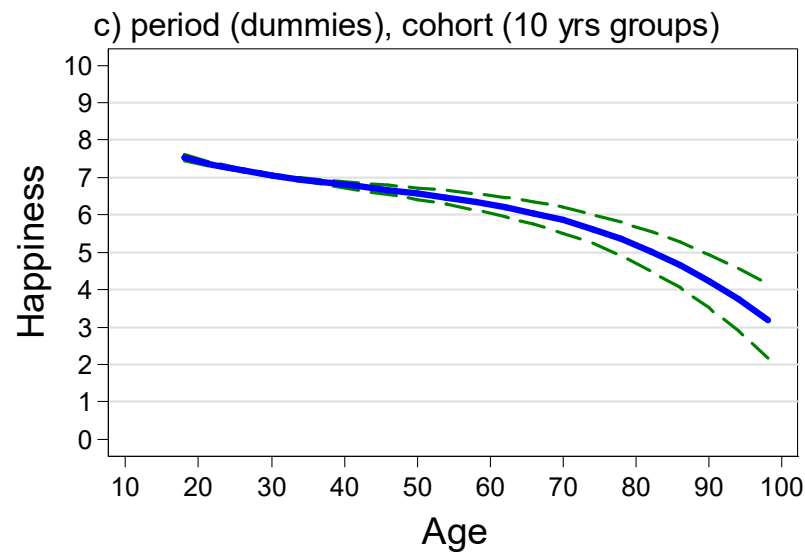
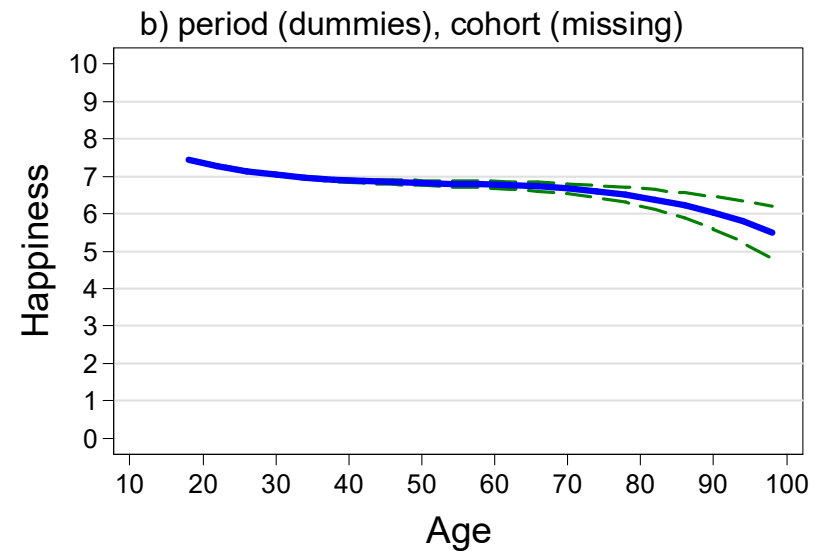
	happy	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
		-----+-----					
	age	-0.1373	0.0129	-10.64	0.000	-0.1626	-0.1120
	c.age#c.age	0.0026	0.0003	7.76	0.000	0.0019	0.0032
	c.age#c.age#c.age	-0.0000	0.0000	-8.01	0.000	-0.0000	-0.0000
		-----+-----					
	year						
	1984	0.0000	(base)				
	1985	-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

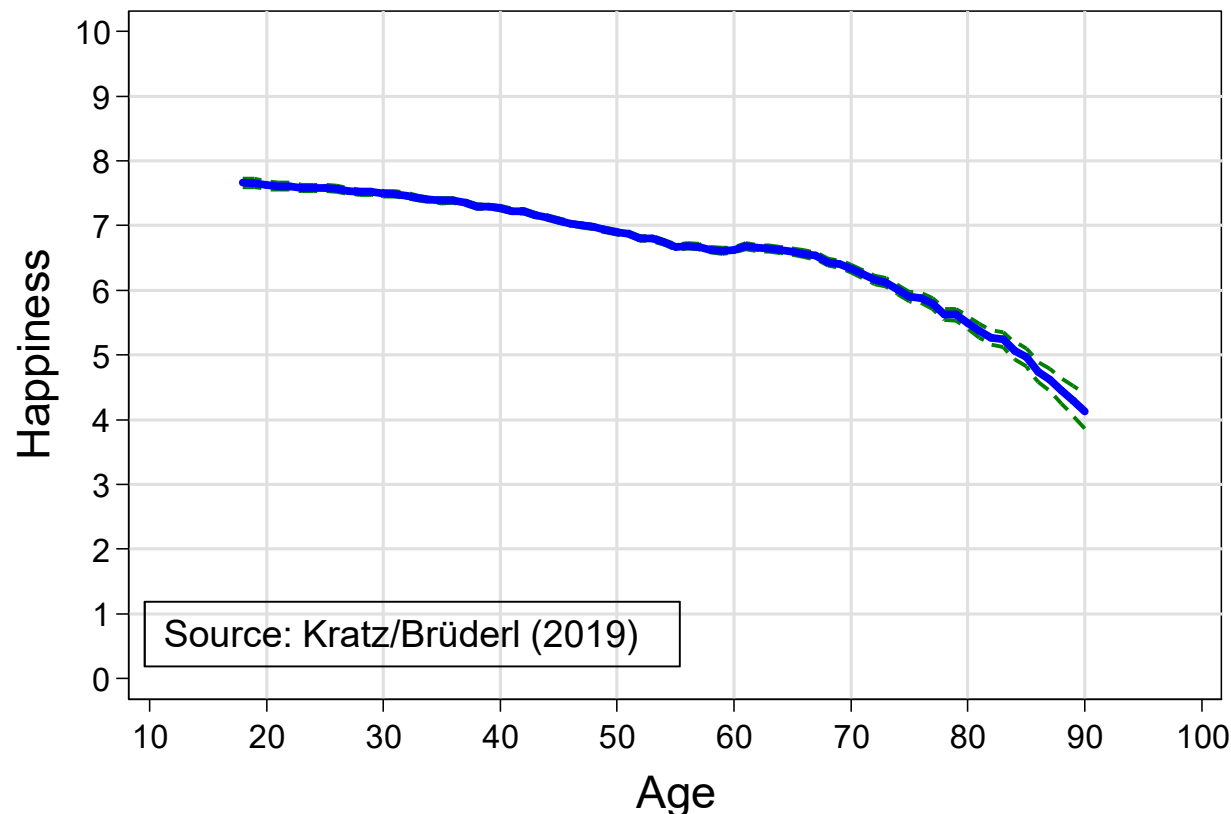


The APC Problem: A Better Solution

- 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

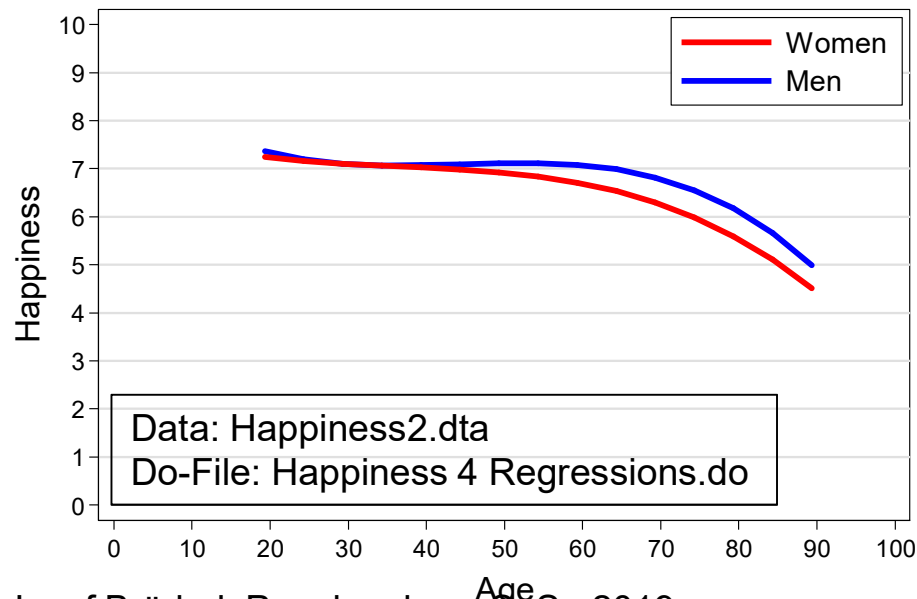
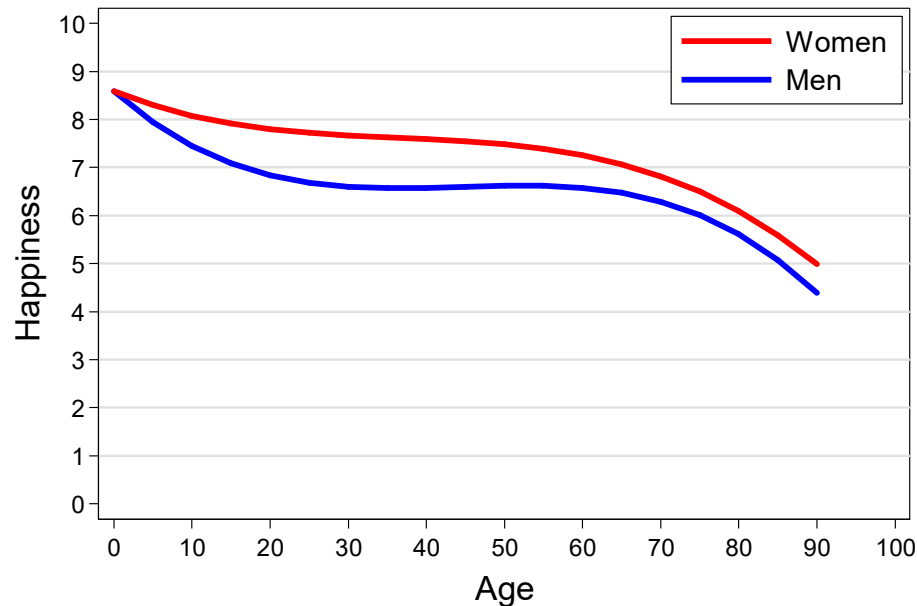
Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



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



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
Group variable: id

Number of obs = 121919
Number of groups = 14634

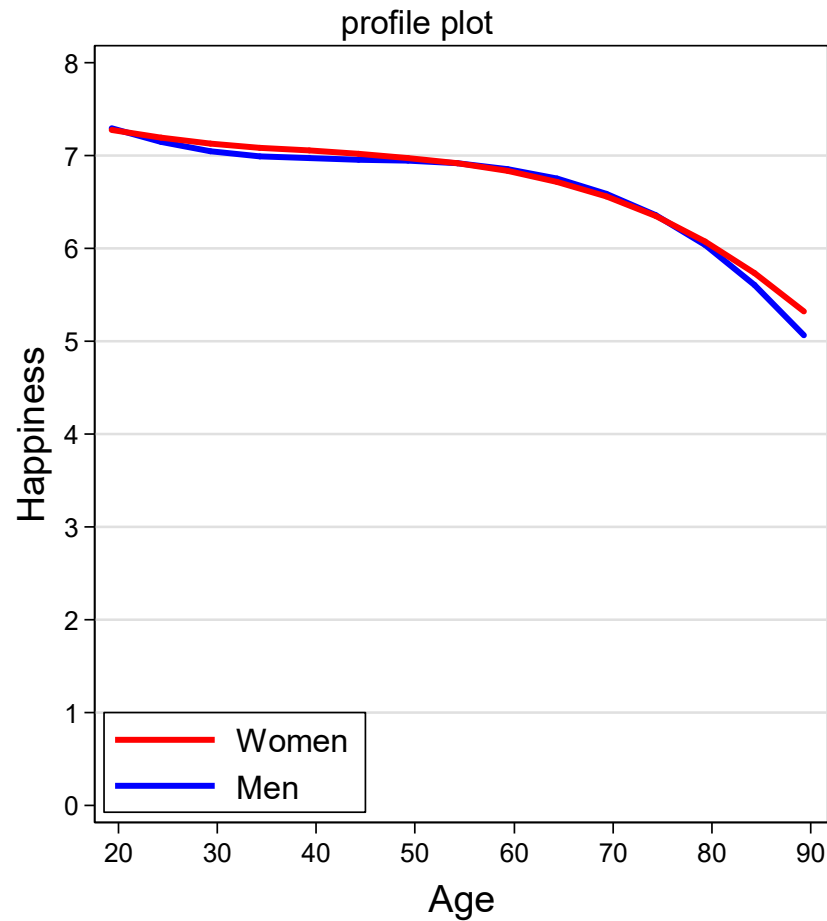
R-sq: within = 0.0199

(Std. Err. adjusted for 14634 clusters in id)

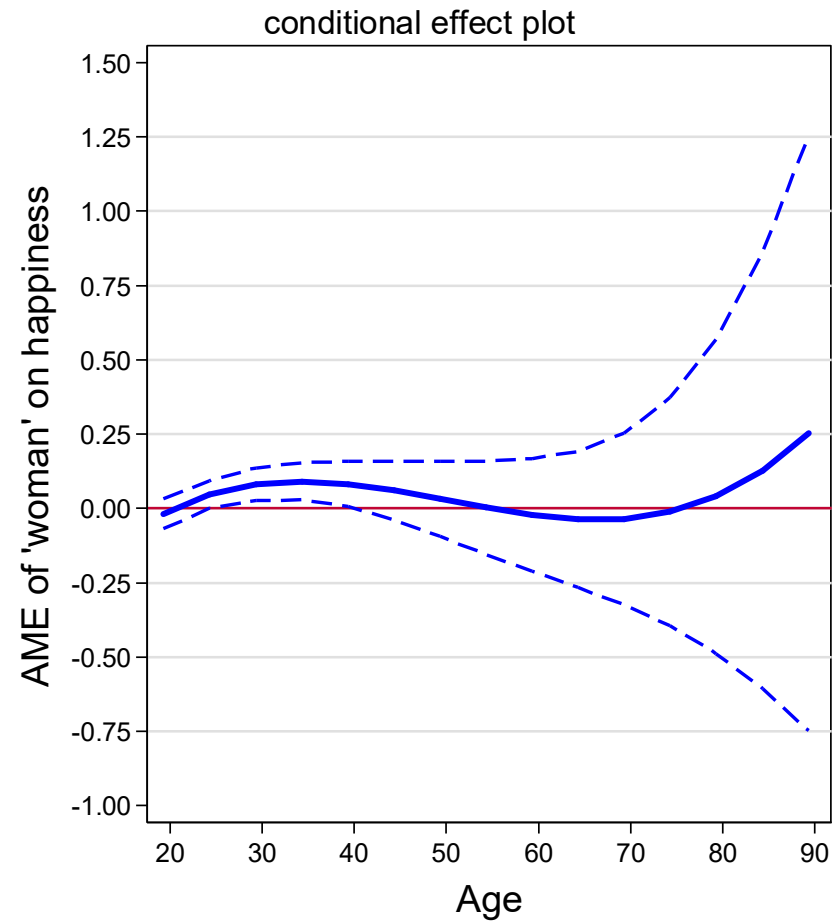
		happy	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]
1.woman			.0812636	.0280571	2.90	0.004	.0262727 .1362545
woman#c.agec							
	0		-.0143229	.0062536	-2.29	0.022	-.0265797 -.0020661
	1		-.010067	.0062785	-1.60	0.109	-.0223725 .0022386
woman#c.agec#c.agec							
	0		.0008397	.0001548	5.43	0.000	.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
```

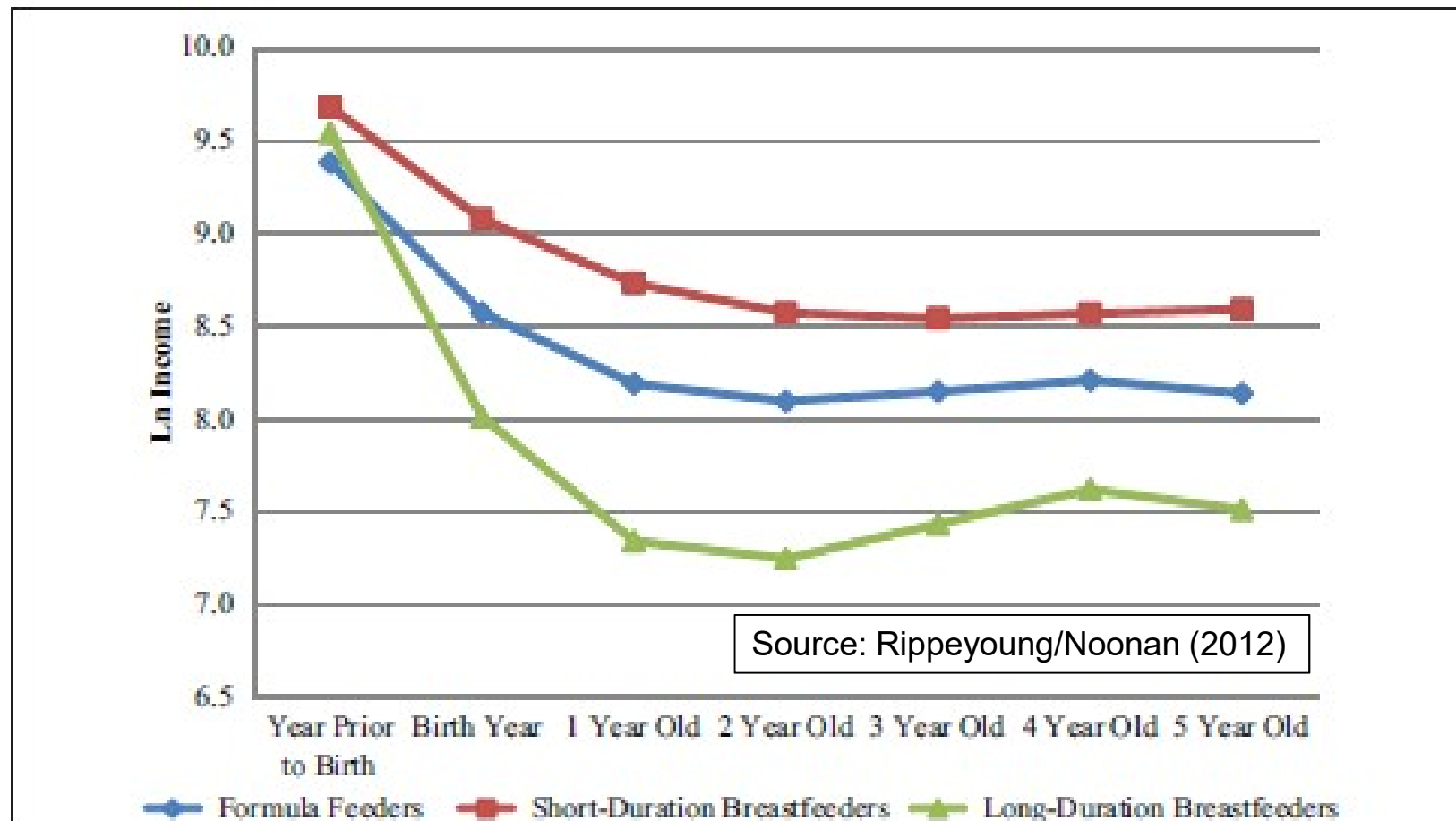


```
margins , at(agec=(-10(5)60)) dydx(woman)  
marginsplot, recast(line) recastci(rline)
```

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

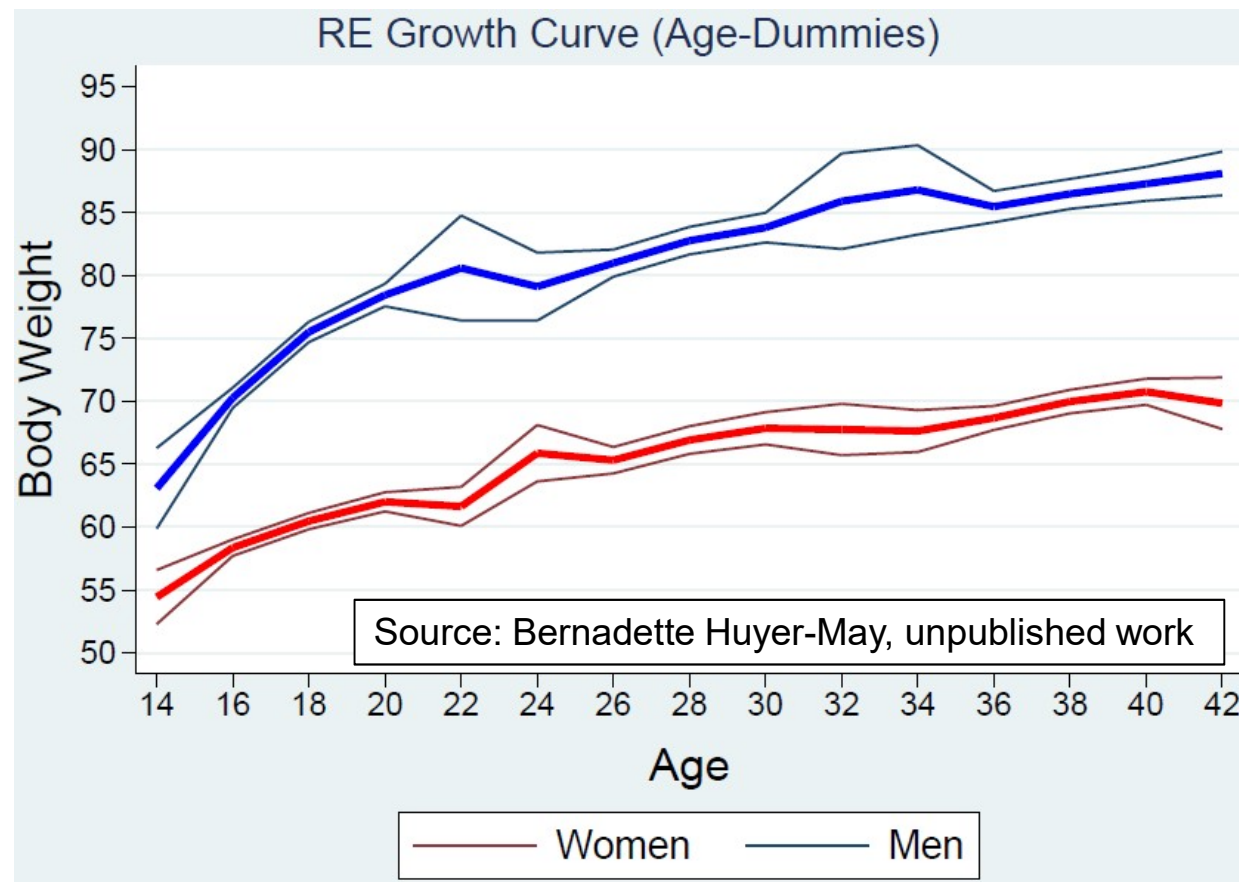
Example: The Cost of Breastfeeding

- 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

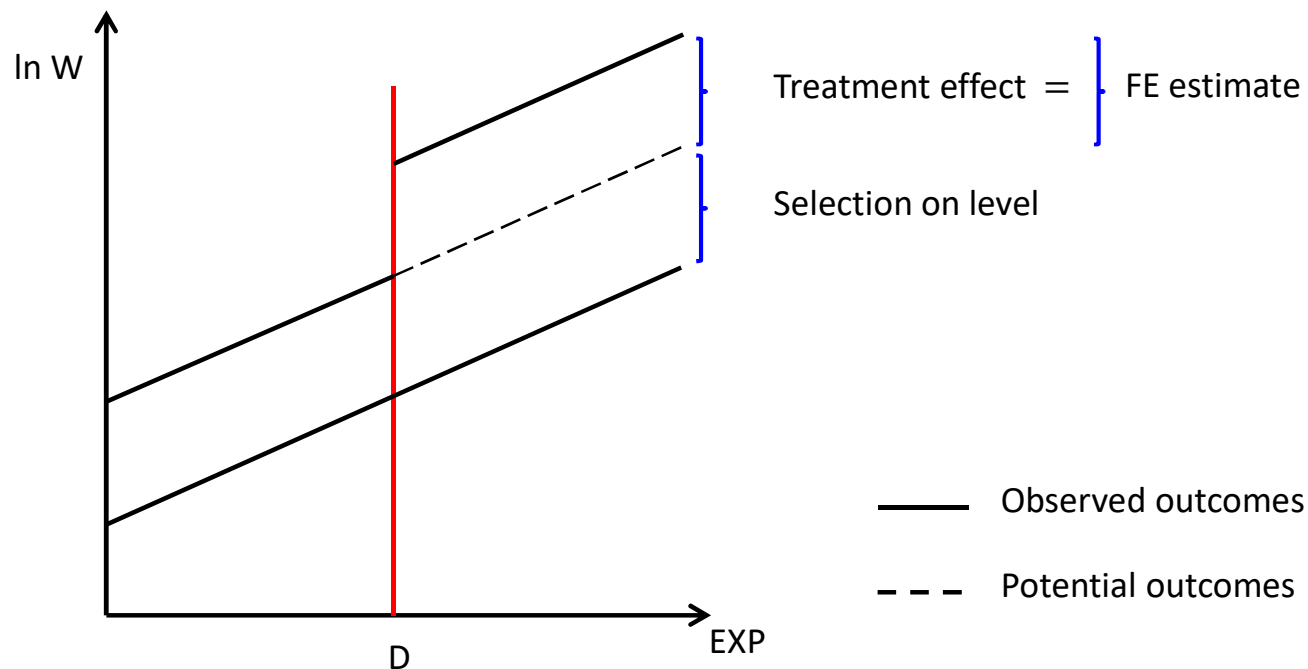
Section: The Fixed-Effects Individual-Slopes Model

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis



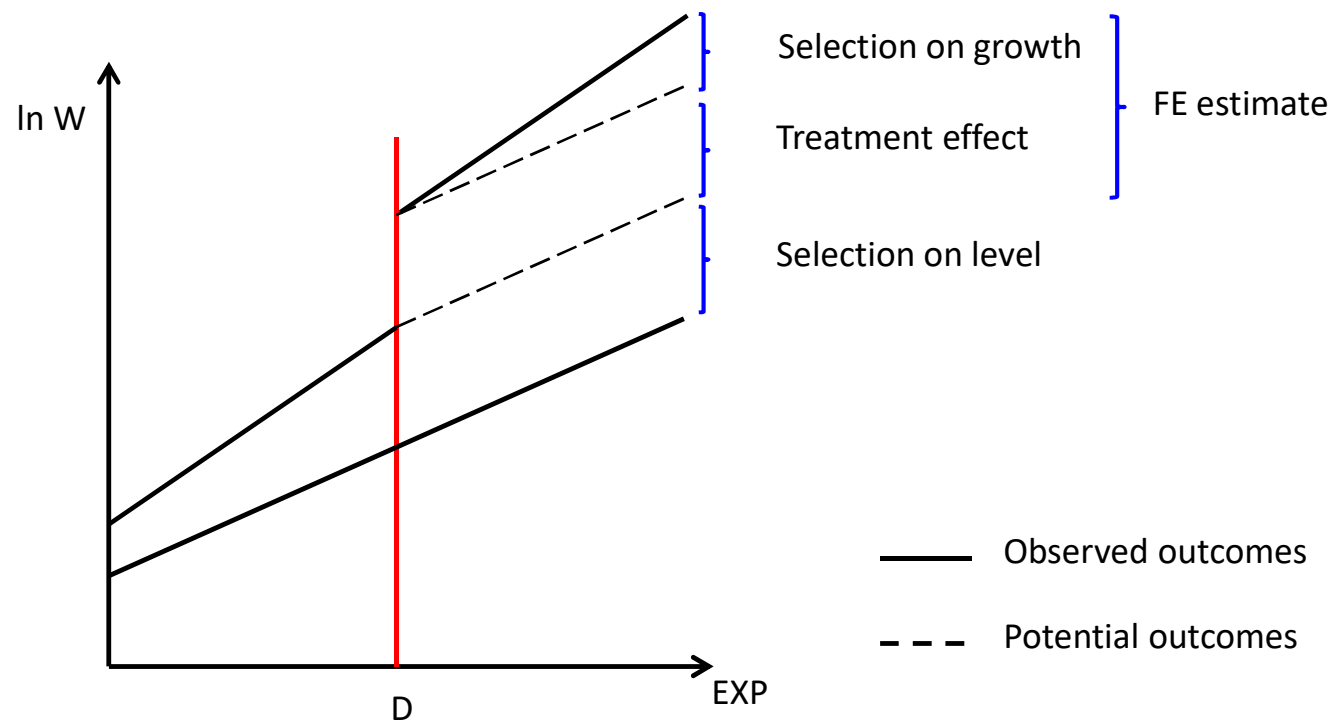
The Parallel Trends Assumption

- 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

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

- 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 //POLSIS
xtfeis wage marr, slope(age) //FEIS
```

Variable		FE	POLSIS	FEIS
marr		780	500	500
age		143		
id	1		-1000	
	2		-1500	
	3		-1500	
	4		-1500	
id#c.age	1		100	
	2		150	
	3		200	
	4		250	

Do-File: FEIS Simulation.do



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

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 specialization
 - Married women specialize on household work
Married men are released from strenuous housework work effort
They can put more effort in their market work
- Lifestyle explanation
 - After marriage men are domesticated by their wives domestication
- Demand side explanation
 - Paternalism of employers 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
$$\ln(w_{it}) = \mathbf{x}_{it}\boldsymbol{\beta} + \gamma m_{it} + \alpha_{1i} + \alpha_{2i} \exp_{it} + \alpha_{3i} \exp_{it}^2 + \xi_{it}$$
where exp is labor market experience

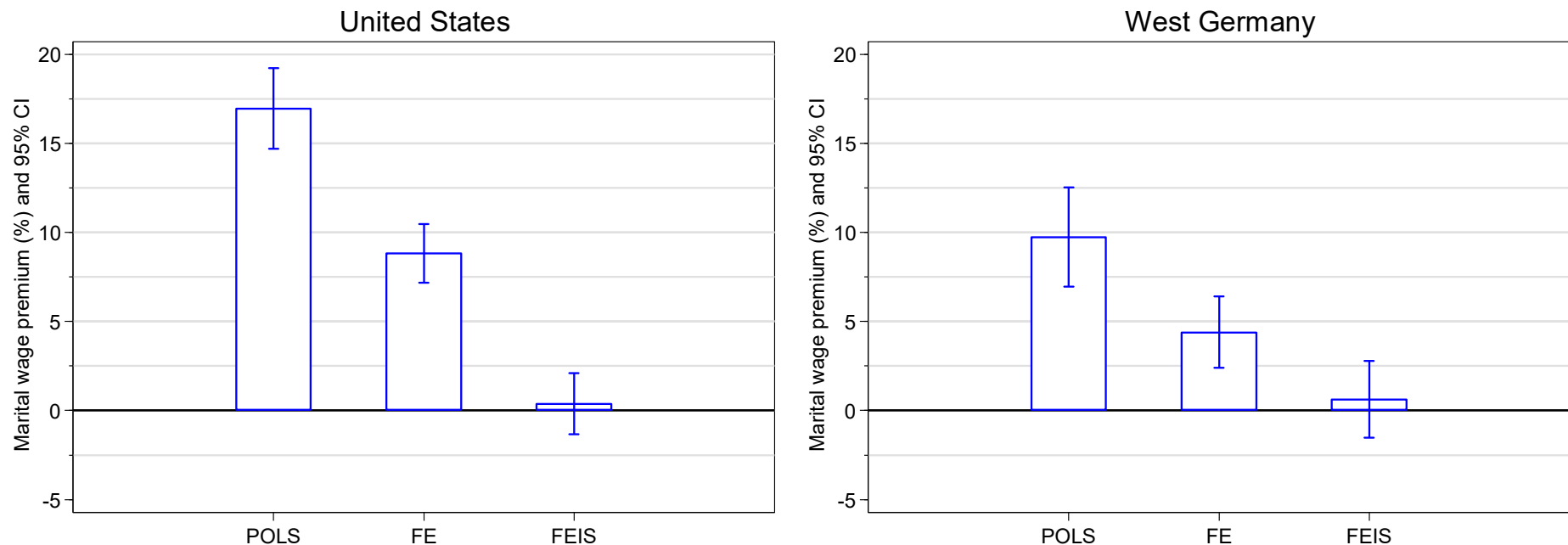
Data and Variables

- 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 VIII:

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

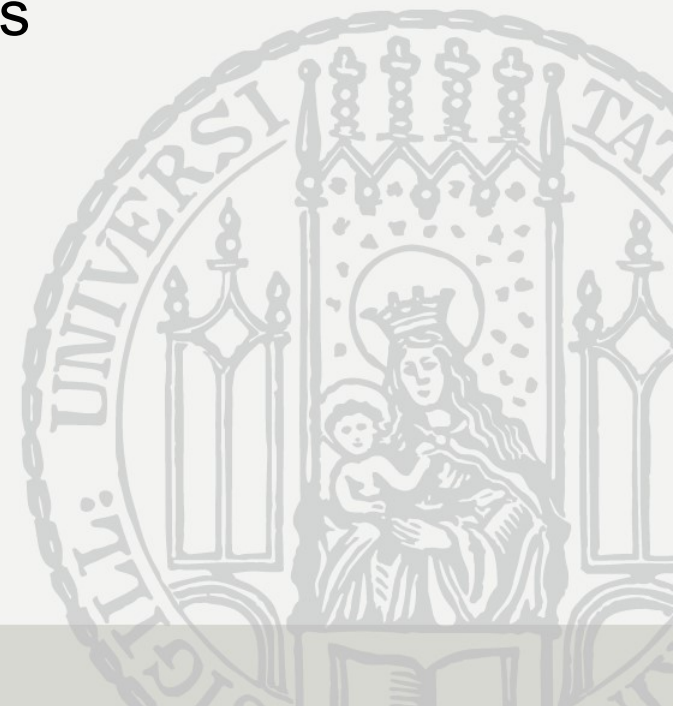
Chapter VIII:

Limitations of the Within Methodology

Section: Violations of Strict Exogeneity

- Unobserved Time-Varying Confounders
- Measurement Errors
- Reverse Causality

Josef Brüderl & Volker Ludwig
Applied Panel Data Analysis

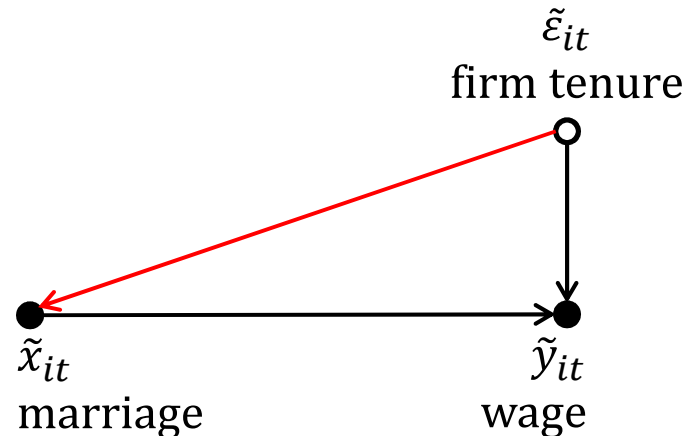


Endogeneity

- FE estimates are biased under endogeneity:
 $E(\mathbf{x}'_{is}\varepsilon_{it}) \neq \mathbf{0}$, 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

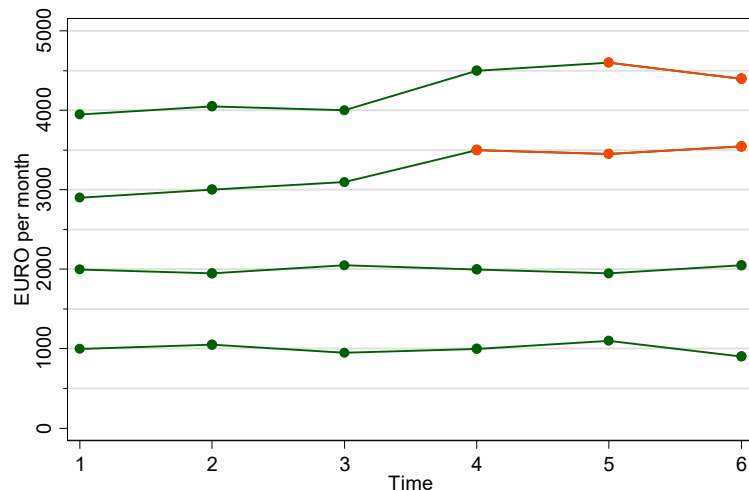


- So far we assumed that the red arrow does not exist
- However, there might be plausible arguments for such an arrow
 - For instance, with longer firm tenure the job is more secure what could affect contemporaneous or future chances for marriage
 - Or past (non) marriage could affect firm tenure.
-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

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:

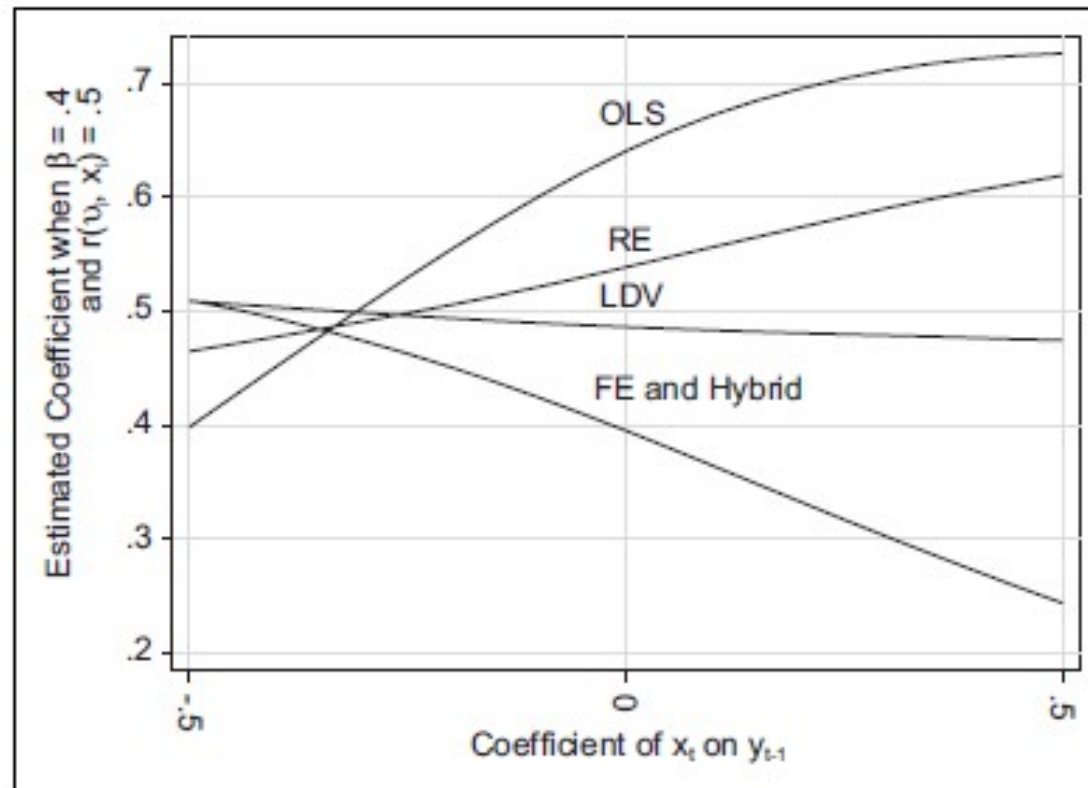
$$Y \rightarrow X$$

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

$$y_{it} = 0.4x_{it} + \alpha_i + \varepsilon_{it}$$

$$r(x_{it}, \alpha_i) = 0.5$$

$$x_{it} = \alpha + \gamma y_{it-1}$$

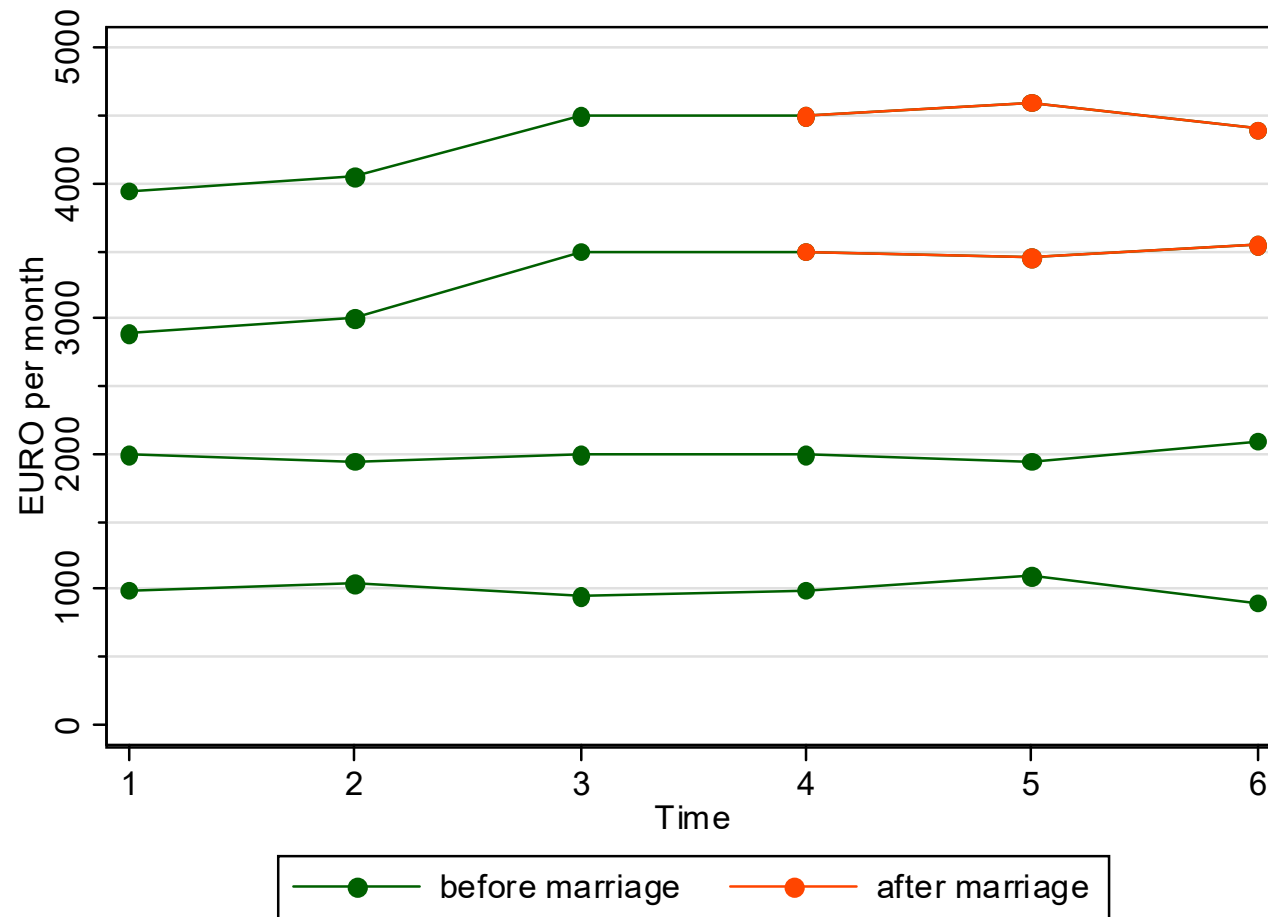
Special Case: Feedback Mechanisms

- 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: $X \nrightarrow Y$ $Y \rightarrow X$
 - 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

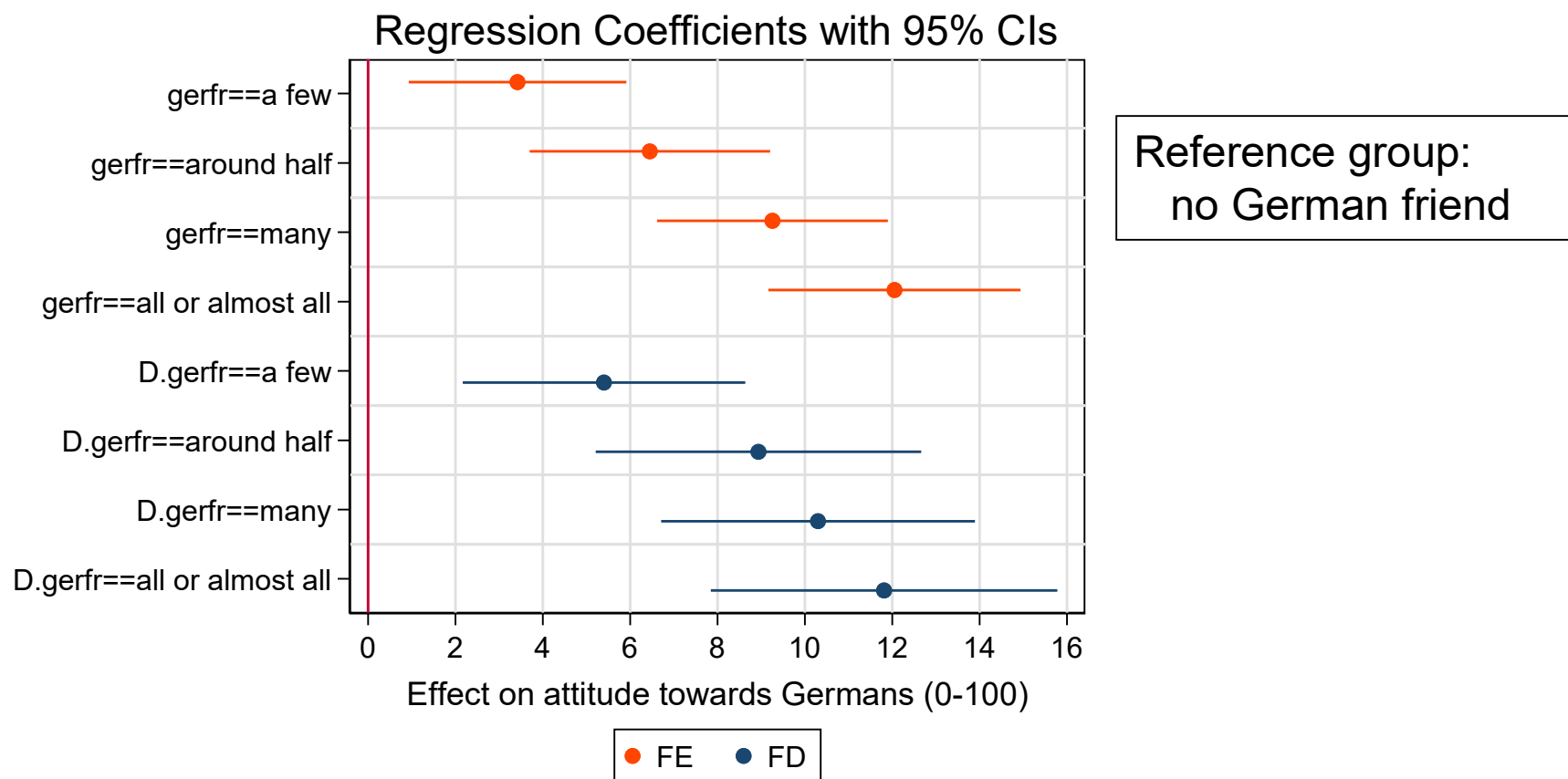
Not Reverse Causality, but Selection

- 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

Chapter VIII: 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

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. Beaton (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.