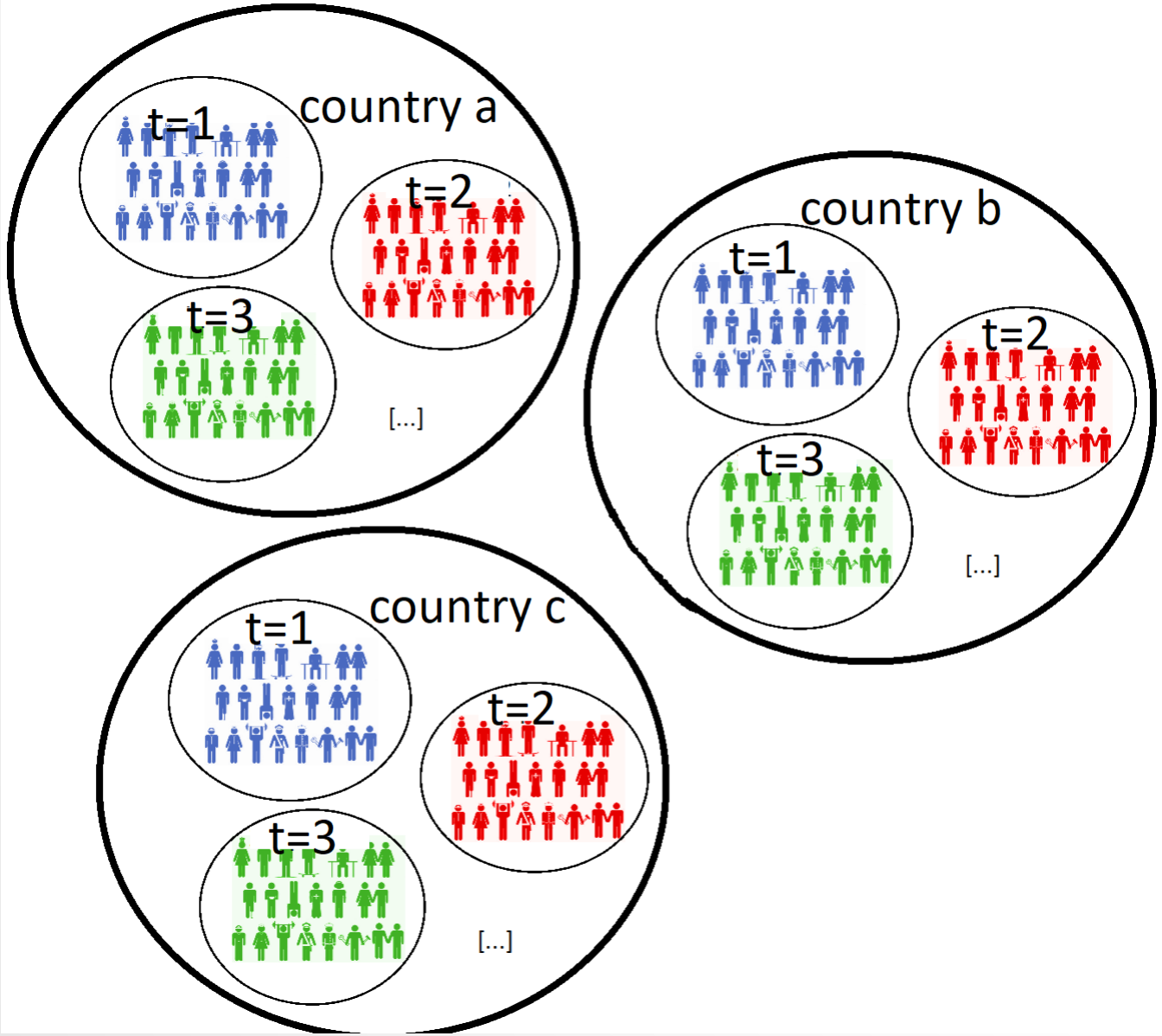
VERGLEICHENDE SOZIALFORSCHUNG MIT MEHREBENENMODELLEN IN R

Forschungspraktikum I und II
Dr. Christian Czymara
Multi-level models with pooled cross-sections

AGENDA

- Individual-level cross-national data that is collected repeatedly
- Last session: Random part of the model
- Today: Fixed part of the model
- Benefits of removing between country variance
- Tutorial: Cross-sectional and longitudinal relationships between corruption and trust in politicians

REPEATED CROSS-NATIONAL DATA


country a
t=1
t=2
t=3
[...]
country b
t=1
t=2
t=3
[...]
country c
t=1
t=2
t=3
[...]

A THREE-LEVEL RE MODEL FOR RCSD

- $y_{ijt} = \beta_0 + \beta x_{ijt} + \gamma z_{jt} + \delta z_j + u_j + u_{jt} + e_{ijt}$
- Indexes
  - i: individual-level
  - j: country-level
  - t: time-level
- x_{ijt}: individual-level variable (effect: β)
- z_{jt}: time-varying country-level variable (effect: γ)
- z_j: time-constant country-level variable (effect: δ)

RANDOM EFFECTS PARAMETERS

- $y_{ijt} = \beta_0 + \beta x_{ijt} + \gamma z_{jt} + \delta z_j + u_j + u_{jt} + e_{ijt}$
- u_j: Random effect on country level (level 3 error)

→Variance between countries

- u_{jt}: Random effect on country-year level (level 2 error)

→Variance within countries / between country-years

- e_{ijt}: Idiosyncratic error (level 1 error)

→Variance within country-years / between individuals

WITHIN AND BETWEEN COUNTRY VARIANCES

WITHIN AND BETWEEN VARIANCE

- $y_{ijt} = \beta_0 + \beta x_{ijt} + \gamma z_{jt} + \delta z_j + u_j + u_{jt} + e_{ijt}$
- This model estimates effect for time-varying country variables (variables on the CY level) as a combination of variance on the country and the country-year level
- The country level includes all time-stable differences of countries (based only on between country variance)
- The country-year level includes all time-varying differences of countries (based only on within country variance)
- Would it not be nice to account for everything time-stable?

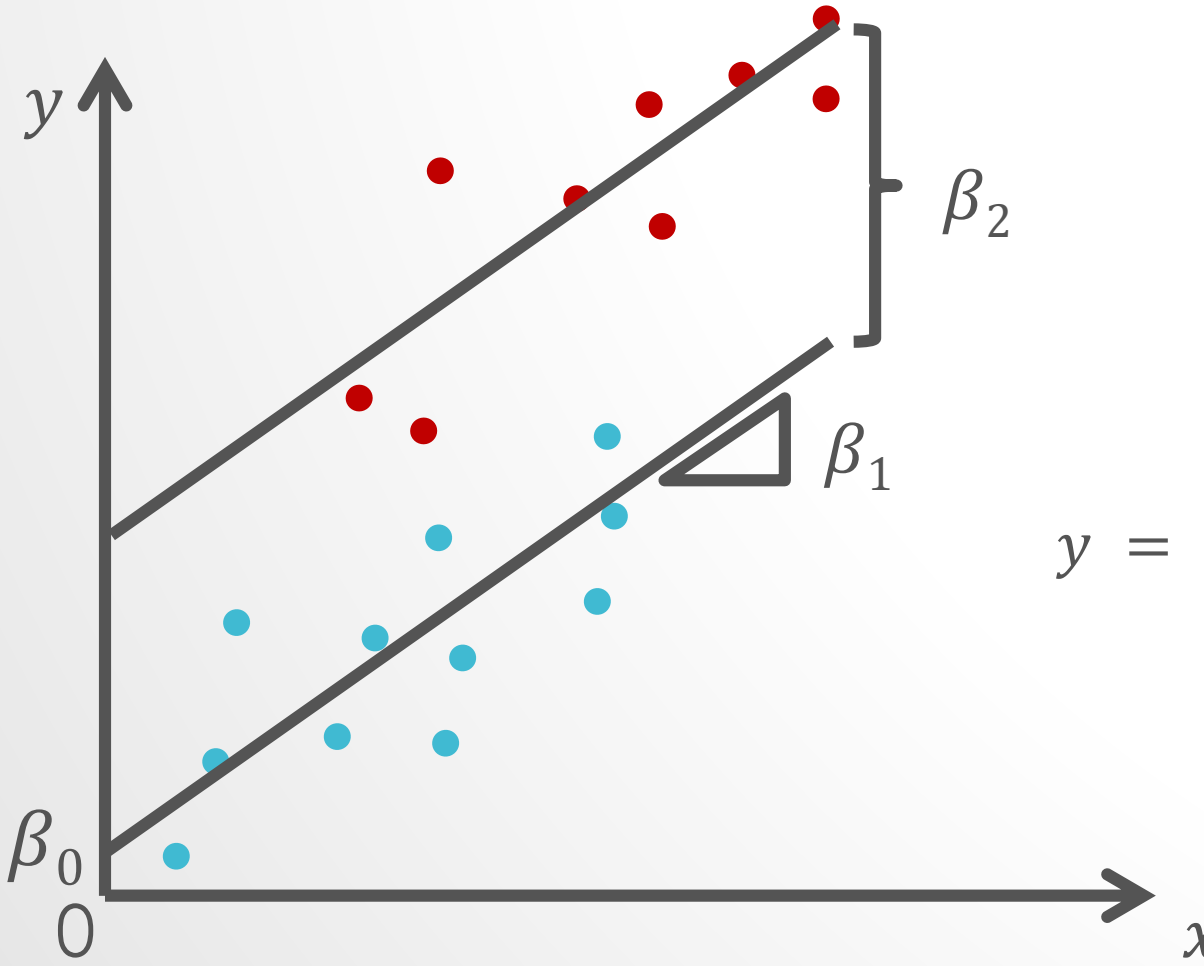
ACCOUNTING FOR TIME-STABLE DIFFERENCES

- If we eliminate all time-stable variance, we would only examine change over time
- This would account for *all historical differences* that exist between countries
- For example: People in richer countries might be open toward immigrants than those in poorer countries
- But that could also be caused by many things that relate to wealth (like a better social system)
- Does an *increase* in national wealth actually lead to more openness?
- Put differently, accounting for time-stable differences tackles model misspecification due to any time constant-confounder (so-called unobserved heterogeneity)

RECAP: COUNTRY FIXED EFFECTS

- A dummy variable for a country captures *everything* that is specific about this country
- One way to account for differences between: Statistical control (add dummy variables for each country)
- In this case, the residuals within each country are statistically independent by design (remember statistical control)
- Model estimates one effect for each country
- For example: "*Attitudes are 1.4 units more positive in the UK compared to Denmark*"
- These effects are fixed values
- So-called *Country Fixed Effects Model*

RECAP: COUNTRY FIXED EFFECTS



$$y = \beta_0 + \beta_1 x + \beta_2 country + e$$

RECAP: COUNTRY FIXED EFFECTS

- Including dummy accounts for *all differences* in y between countries
- Identification of effects variables country-level impossible
- *Country dummies "control away" all variance between countries*
- Not possible to test which *particular* differences may be important
- For example: Let us say we find a significant difference in welfare state support between Sweden und Spain
  - Is this difference due to differences in national wealth?
  - Or difference in welfare state regime?
  - …
- Methodologically, there is *no variance between countries left* that independent variables on the country level could explain
- However, testing individual level associations (or cross-level interactions) still possible

COUNTRY FIXED EFFECTS

- With cross-sectional data, there is indeed no variance on the country level left when country fixed effects are added
- With *repeated* multi-national data, however, we only eliminate variance between countries but not within countries (between country-years)
- This is actually a huge benefit
- Model: $y_{ijt} = \beta_0 + \beta x_{ijt} + \gamma z_{jt} + \beta_c countries_j + \delta z_j + u_j + u_{jt} + e_{ijt}$
- $countries_j$: country dummies
- When time trend and country dummies are added, such a model is also called *two-way fixed effects* model

FIXED EFFECTS

- $y_{ijt} = \beta_0 + \beta x_{ijt} + \gamma z_{jt} + \beta_c countries_j$ ~~$+ \delta z_j + u_j$~~ $+ u_{jt} + e_{ijt}$
- Effects of main interest here are those of variables on the country-level that vary over time (z_{jt})
- Effects of time-constant country-level variables (z_j) not estimable (but automatically controlled for)
- Effects of individual-level variables are also estimated (x_{ijt}) but cannot take advantage of the longitudinal approach (because they are only pooled cross-sections)

EXAMPLE: BRADY & FINNIGAN 2014

- Research question: *Does immigration undermine public support for social policy?*
- Data: International Social Survey Program (ISSP), waves 1996 and 2006 + OECD data
- 17 European countries observed at two time points
- Cross-sectional & longitudinal perspective
- Example: Brady & Finnigan (2014). Does immigration undermine public support for social policy?. American Sociological Review, 79 (1), 17-42.

BRADY & FINNIGAN 2014: VARIABLES

- Dependent variable: Six welfare attitudes (e. g.: "*Do you think it should or should not be the government's responsibility to reduce income differences between rich and poor*") → Individual level
- Core explanatory variable: percent foreign born of the total population
- Plus individual- (e. g., education) and country-level (e. g., social welfare expenditures) controls

BRADY & FINNIGAN 2014: MODEL

- $\ln\left(\frac{p_{ijt}}{1-p_{ijt}}\right) = y_{ijt} = \beta_0 + \beta_x x_{ijt} + \beta_z z_{jt} + \beta_c C_j + \beta_w W_t$
- x_{ijt}: individual level variables
- z_{jt}: country-(year-)level variables
- C_j: country dummies
- W_t: time dummy (2006)

BRADY & FINNIGAN 2014: RESULTS

Table 4. Two-Way FE Models of Welfare State Attitudes on Percent Foreign Born and Individual- and Country-Level Controls in 13 Affluent Democracies in 1996 and 2006: Odds Ratios and *Z*-Scores

	Jobs	Unemp.	Income	Retirement	Housing	Healthcare
Percent Foreign Born	.945***	1.001	.947***	1.091**	.996	.895***
	(–4.992)	(.061)	(–4.611)	(2.590)	(–.223)	(–3.555)
Percent Foreign Born	.944***	1.000	.944***	1.079*	.966	.876***
	(–5.130)	(–.010)	(–4.774)	(2.151)	(–1.918)	(–4.049)
Social Welfare Expenditures	.969	.992	.972	.968	.882***	.924*
	(–1.776)	(–.400)	(–1.579)	(–.935)	(–6.293)	(–2.319)
Percent Foreign Born	.930***	1.002	.921***	1.067	1.000	.892**
	(–4.206)	(.110)	(–4.665)	(1.800)	(–.008)	(–3.036)
Employment Rate	1.012	.999	1.020*	1.042	.996	1.003
	(1.258)	(–.097)	(2.093)	(1.839)	(–.336)	(.170)
N	31,272	31,017	30,971	32,061	31,116	32,028

Note: All individual-level controls from Table S6 in the online supplement included but not shown.
*$p < .05$; **$p < .01$; ***$p < .001$ (two-tailed tests).

DEMEANING

RECAP: GROUP MEAN CENTERING

- Center at the country mean
- O… ossible for individual-level variables (for country v… ntry mean is simply its value)

→ *Now: Possible for individual and country-year level variables*

- In R:
  - ```
    data %<>%
      group_by(cntry) %>%
      mutate(var_g_cent = var - mean(var))
    ```
- Not a simple reparameterization of a model but a completely *different model* (cf. Hox 2010: 68 f.)
    ```

RECAP: WHEN TO USE GROUP MEAN CENTERING

- Usually, individual-level effects include a mixture of effects *within* and *between* countries
- Group mean centering *removes all between-group variance* (highly used with panel data, not so much with cross-national data, for a comparison see Ziller 2018)
- Only variance *within* countries (b → Now: between country-years individuals) is analyzed, all differences *between* countries are automatically controlled for
- This is equivalent to using country fixed effects (adding dummy variables for all countries but one)
- A
  le
  → *Now: Not useful for country-variables that are time constant*
- H
  e
  → *Now: Estimation of time-varying country variables still possible (and cross-level interactions, too)*

GROUP MEAN CENTERING ON COUNTRY LEVEL

- With repeated cross-sections, country-level effects include a mixture of effects *within* (between country-years, level 2) and *between* countries (level 3)
- Group mean centering *removes all between-group variance*
- Only variance *within* countries (between country-years) is analyzed, all differences *between* countries are automatically controlled for
- *This is equivalent to using country fixed effects (adding dummy variables for all countries but one)*
- Accordingly, group mean centering can not be used for country level (level 3) variables
- Highly useful when interest is in relationship between variables on the country-year level (i. e. time-varying country characteristics)

GROUP MEAN-CENTERING AS DEMEANING

- In the context of panel data, group mean centering is highly used
- Usually referred to as *demeaning* or *fixed effects transformation*
- Just like country FEs, it eliminates all variance between countries, leaving only variance within countries
- Results and interpretation of both approaches are thus equivalent
- Demeaning variables allows to estimate effects within and between countries simultaneously within the same model (see Fairbrother 2014)

REMOVING BETWEEN COUNTRY VARIANCE

- For cross-sectional comparative data (two levels), removing all variance between countries (level 2) did not make sense when interest in country-level variables because estimation of their effects impossible (no variance left to explain)
- With repeated multi-national data (three levels), remove time-stable variance between countries makes sense when we are interested in time-varying country variables
- Because time-varying country variables are not on the country-level (level 3) but on the country-year-level (level 2), where variance remains in the model
- Interpretation of coefficients is otherwise identical (*"a change in x leads to a change in y"…*)

EXAMPLE: CZYMARA (2021)

- Research question: Does demographic changes explain attitudes toward refugees in Europe?
- Data: European Social Survey, waves 7 and 8 + Eurostat data (and other)
- 22 countries in four years (~50,000 individuals)
- Cross-sectional & longitudinal perspective
- Example: Czymara (2021). Attitudes toward refugees in contemporary Europe: A longitudinal perspective on cross-national differences. Social Forces, 99 (3), 1306-1333.

CZYMARA 2021: VARIABLES

- Outcome: Anti-refugee attitude ("*Some people come to this country and apply for refugee status on the grounds that they fear persecution in their own country. Using this card, please say how much you agree or disagree that: 'the government should be generous in judging people's applications for refugee status'*.") → Individual level
- Explanatory variable: Share non-EU foreigners (and asylum applications)
- Plus individual- (e. g. education) and country-level (e. g. GDP/c) controls

CZYMARA 2021: MODEL

- Three-level hierarchical linear model (see last session)
- Demeaned variables → within effects

→(Mean variables → between effects; ignored here, but see Fairbrother 2014)

CZYMARA 2021: RESULTS

Table 3 Random effects models

	M0	M1	M2
Country variables			
Asylum appl. (WE)		−0.02	−0.02
		(0.01)	(0.01)
Asylum appl. (BE)		0.05	0.05
		(0.03)	(0.03)
Foreigner share (WE)		0.29**	0.29**
		(0.10)	(0.09)
Foreigner share (BE)		0.05	0.05
		(0.03)	(0.03)
GDP/c (WE)		−0.00	−0.00
		(0.01)	(0.01)
GDP/c (BE)		−0.01	−0.01
		(0.01)	(0.01)

DEMEANING USING THE `PARAMETERS` PACKAGE

- For manual demeaning…
  - First, calculate the country-specific mean ($\bar{x}$)
  - Then subtract this mean from the observed value ($x - \bar{x}$)
- The `demean()` function of the `parameters` package in R does this job for you
- First, create a new data frame that contains the mean (by default ending with `* between`) and the demeaned variables (by default ending with `*_within`):

```
demeaned_vars <- demean(data, select = variables,
group = "cntry")
```

- Then assign the variables you need to your existing data, e.g.:

```
data$var_within <- demeaned_vars$ar_within
```

SUMMING UP

SUMMARY

- Various cross-national data sets with individual-level information (!) are pooled → more complex data hierarchy
- Statistical dependencies must be modelled, otherwise underestimation of standard errors
- Practical and often correct solution:
  - Level 3: countries
  - Level 2: country-years
  - Level 1: individuals
- Effects of time-varying country characteristics can be estimated based purely on variation within countries over time (fixed effects estimation)

→Automatically control all possible time-constant characteristics

LITERATURE

- Fairbrother (2014). Two multilevel modeling techniques for analyzing comparative longitudinal survey datasets. Political Science Research and Methods, 2 (1), 119-140.