Step-by-step or all at once: A comparison of multilevel analysis using one-step and two-step approaches

Ulrich Kohler [1]

[1] University of Potsdam
Faculty of Economic and Social Sciences

2025 Swiss Stata Conference
November 25th 2025
Hosted by the University of Bern

Contents

Aim of presentation

- Compare one-step approaches and the two-step approach for analyzing hierarchial data
- Illustrate `twostep` – a bundle of programs to ease analyzing hierarchical data with the two-step approach (Giesecke and Kohler, 2024; Kohler and Giesecke, 2021).
- Show some additional insights offered by the two-step approach.

Hierarchial data

Data with micro-level units nested within larger macro-level units.
Examples:

- Students nested in schools
- Respondents nested in countries (or cantons)
- Indivduals nested in laboratories
- Animals nested in habitats, etc.

Typical research questions

Parameter	Example
Micro level effects	Individual income on life satisfaction (controlled for country level characteristics)
Direct context effects	GDP on individual life satisfaction
Cross level interactions	GDP on the effect of individual income on life satisfaction

Contents

Notation

A linear model for hierarchial data is

$$y_{ij} = \beta_{0j} + x_{kij}\beta_{kj} + \boldsymbol{x}_{\tilde{k}ij}\boldsymbol{\beta}_{\tilde{k}j} + \epsilon_{ij} \quad (1)$$

$$\beta_{0j} = \gamma_{00} + \boldsymbol{z}_j\boldsymbol{\gamma}_0 + u_{0j} \quad (2)$$

$$\beta_{kj} = \gamma_{k0} + \boldsymbol{z}_j\boldsymbol{\gamma}_k + u_{kj} \quad (3)$$

which can be also written as

$$y_{ij} = \gamma_{00} + \boldsymbol{z}_j\boldsymbol{\gamma}_0 + \gamma_{k0}x_{kij} + \boldsymbol{z}_j x_{kij}\boldsymbol{\gamma}_k + \boldsymbol{x}_{\tilde{k}ij}\boldsymbol{\beta}_{\tilde{k}} + u_{0j} + u_{kj}x_{kij} + \epsilon_{ij} \quad (4)$$

γ_{k0} is the micro level effect, $\boldsymbol{\gamma}_0$ are the direct context effects, and $\boldsymbol{\gamma}_k$ are the cross level interactions.

Estimation approaches

Estimation approaches all assume u_{0j}, u_{kj} and ϵ_{ij} to be uncorrelated with $\boldsymbol{x}_{ij}$ and $\boldsymbol{z}_j$. Moreover u_{0j} and u_{kj} are assumed to be uncorrelated with ϵ_{ij}.

- One-step approaches
  - Cluster-robust SE (i.e. `, vce(cluster varname)`
  - Dummy-variables for macro-level units (e.g. `areg`, `xtreg, fe`)
  - Random effects (e.g. `mixed`, `xtreg, re`)
- Two-step approach: Fitting separate models for each unit of the macro level and then estimate (2) of (3) using those results.

General recommendations

- Dummy-variable approach controls unobserved macro-level heterogeneity when estimating micro-level effects.
- Cluster-robust SE should not be used to estimate CLIs.
- Dummy-variable approach cannot be used to estimate DCEs.
- Two-step approach should not be used to estimate DCEs.

See Bryan and Jenkins (2016a) or Giesecke and Kohler (2024) for details.
The random-effects approach—if specified correctly—is more efficient than the two-step approach and both approaches show better statistical performance when compared to the cluster-robust and the fixed-effects approach (Heisig et al., 2017)

Remarks on mixed

- Standard errors of the macro-level coefficients from one-step approaches with scarce macro-level data are imprecise and likely to be biased downwards (Hox 2010, 233; Raudenbush and Bryk 2002, 283).
- It is recommended to use all random coefficients (Heisig et al., 2017; Heisig and Schaeffer, 2019), but: convergance problems.
- `mixed` "borrows strengts", but:
  - Should we "borrow" observations from Luxembourg to estimate a micro level effect for Bulgaria?
  - Do we need to borrow strengths if we have enough strength within each macro-level unit?

Remarks on twostep

- Model estimated at the second step is heteroscedastic (sampling variance differ varies between the macro-level units).
- May be neglible for large numbers of observations at the micro-level (Donald and Lang 2007; Wooldridge 2010, 891–892).
- Weighting matrices exist to correct the problem (Hanushek, 1974; Borjas and Sueyoshi, 1994; Lewis and Linzer, 2005; Donald and Lang, 2007) (implemented in `twostep`).
- Weighting may not work for small numbers of units at the macro-level (Bryan and Jenkins, 2016b)

Contents

Examples

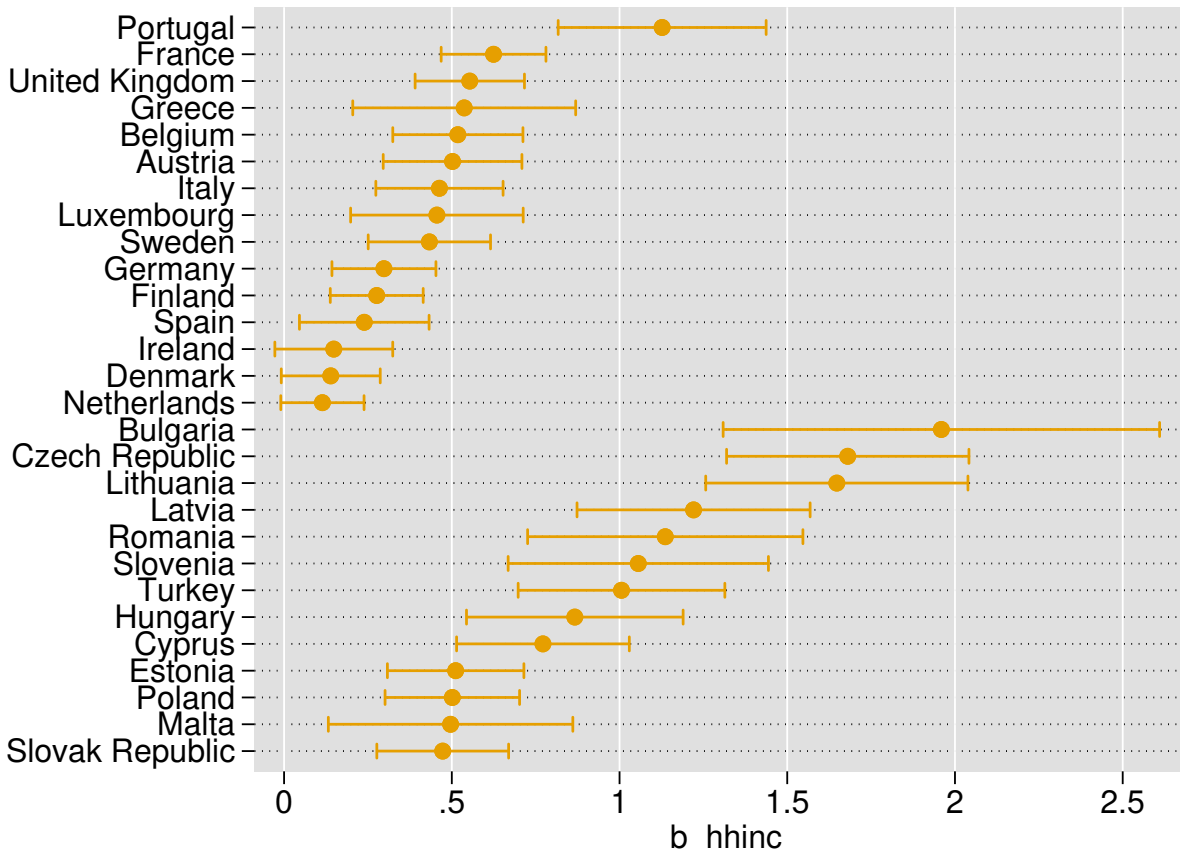
```
. use ../eqls2003_twostep, clear
. regress lsat hhinc gdppcap i.sex age, vce(cluster cntrynum)
. areg lsat hhinc i.sex age, absorb(cntrynum)
. mixed lsat hhinc gdppcap i.sex age || cntry: hhinc i.sex age, cov(unstruc
  tured)
. twostep cntry: reg lsat hhinc i.sex age || edv _b_hhinc
```

Approach	Coef.	Std.Err.	*t*
Cluster-robust	0.5763189	0.0690109	8.351
Fixed effects	0.5214747	0.0218251	23.893
Random effects	0.6787515	0.0828717	8.190
Two-step approach	0.6755608	0.0876774	7.705

One-step approaches are implicitly weighted by number of micro level observations within macro level unit. Two-step approach implicitely use equal weights for each macro level unit.

Presenting results with twostep

```
. twostep cntry: regress lsat hhinc i.sex age || dot _b_hhinc eu15
```



Contents

Examples

```
. regress lsat hhinc gdppcap i.sex age, vce(cluster cntrynum)
. areg lsat hhinc gdppcap i.sex age, absorb(cntrynum)
. mixed lsat hhinc gdppcap i.sex age || cntry: hhinc i.sex age, cov(unstruc
  tured)
. twostep cntry: reg lsat hhinc i.sex age || edv _b_cons gdppcap
. center hhinc age
. twostep cntry: reg lsat c_hhinc i.sex c_age || edv _b_cons gdppcap
```

Approach	Coef.	Std.Err.	*t*
Cluster-robust	0.0772236	0.0166585	4.636
Fixed effects	0	0	.
Random effects	0.0371283	0.0093513	3.970
Two-step approach I	0.0521872	0.0175978	2.966
Two-step approach II	0.0571071	0.0111544	2.966

Twostep approach depend on the “meaning of zero” in micro-level covariates.

Presenting results with twostep

```
. twostep cntry: regress lsat hhinc i.sex age || cprplot _b_cons gdppcap
```

Contents

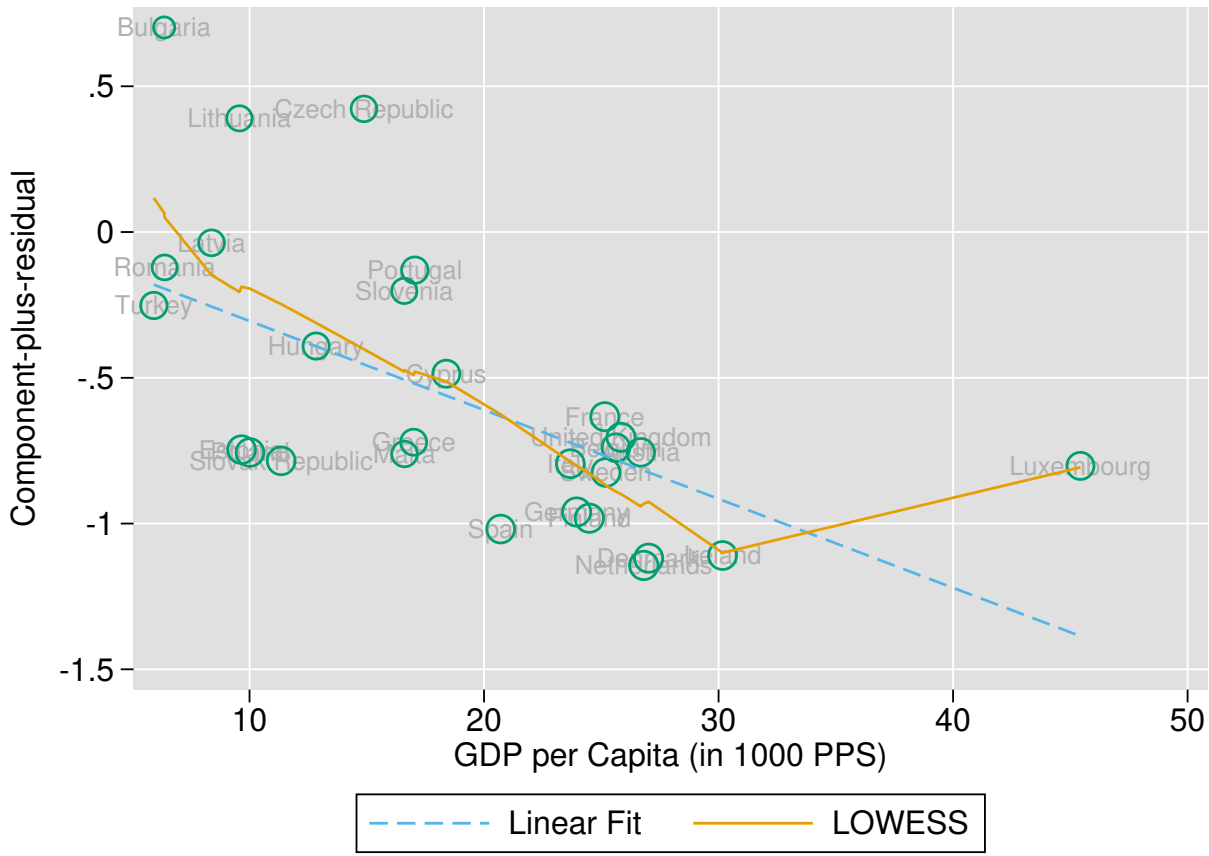
Examples

```
. regress lsat c.hhinc##c.gdppcap i.sex age, vce(cluster cntrynum)
. areg lsat c.hhinc##c.gdppcap i.sex age, absorb(cntrynum)
. mixed lsat c.hhinc##c.gdppcap i.sex age || cntry: hhinc i.sex age, cov(un
  structured)
. twostep cntry: reg lsat c.hhinc i.sex age || edv _b_hhinc gdppcap
```

Approach	Coef.	Std.Err.	*t*
Cluster-robust	-0.0346065	0.0100692	-3.437
Fixed effects	-0.0231143	0.0025919	-8.918
Random effects	-0.0269340	0.0071865	-3.748
Two-step approach	-0.0305056	0.0078769	-3.873

Presenting results with twostep

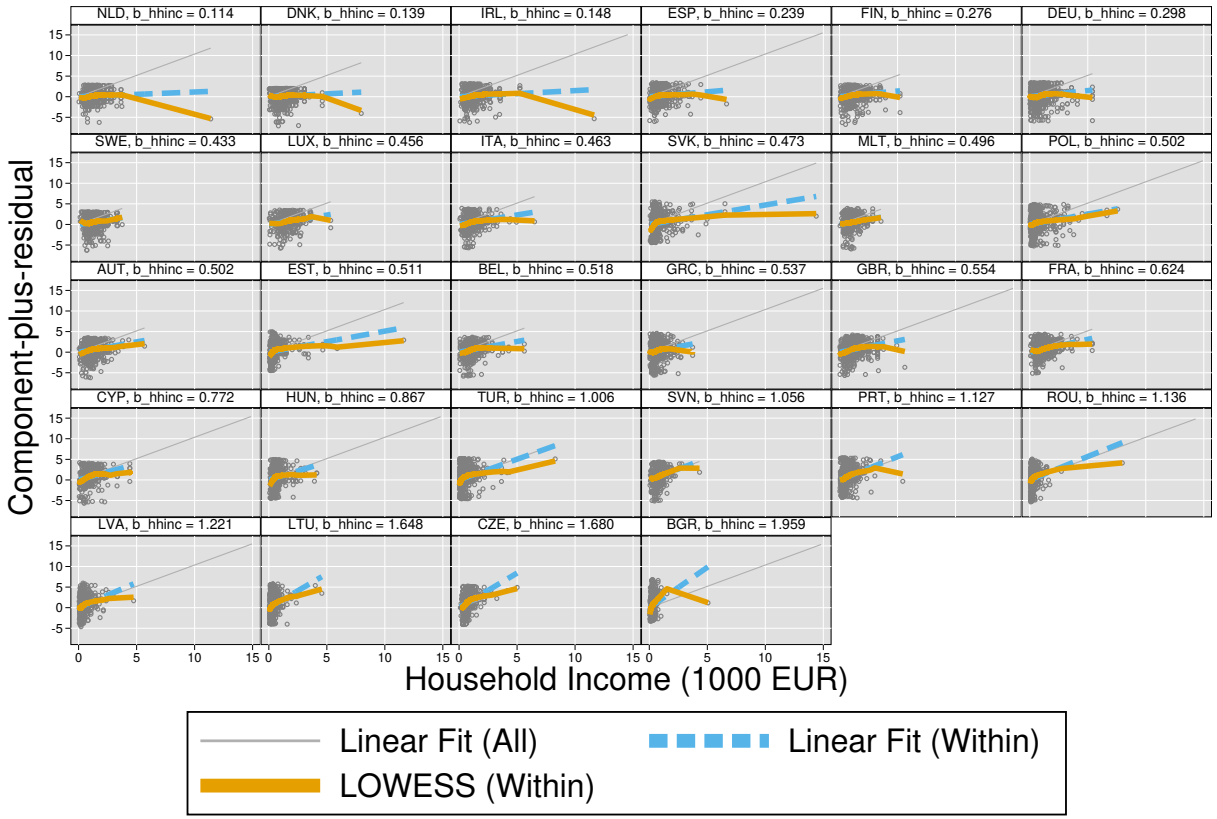
```
. twostep cntry: regress lsat hhinc i.sex age || cprplot _b_hhinc gdppcap
```



Contents

Linear model within each country appropriate?

```
. twostep cntryiso: microcpr lsat hhinc i.sex age || _b_hhinc, regopts(lwid
  th(*5)) lowessopts(lwidth(*5))
```

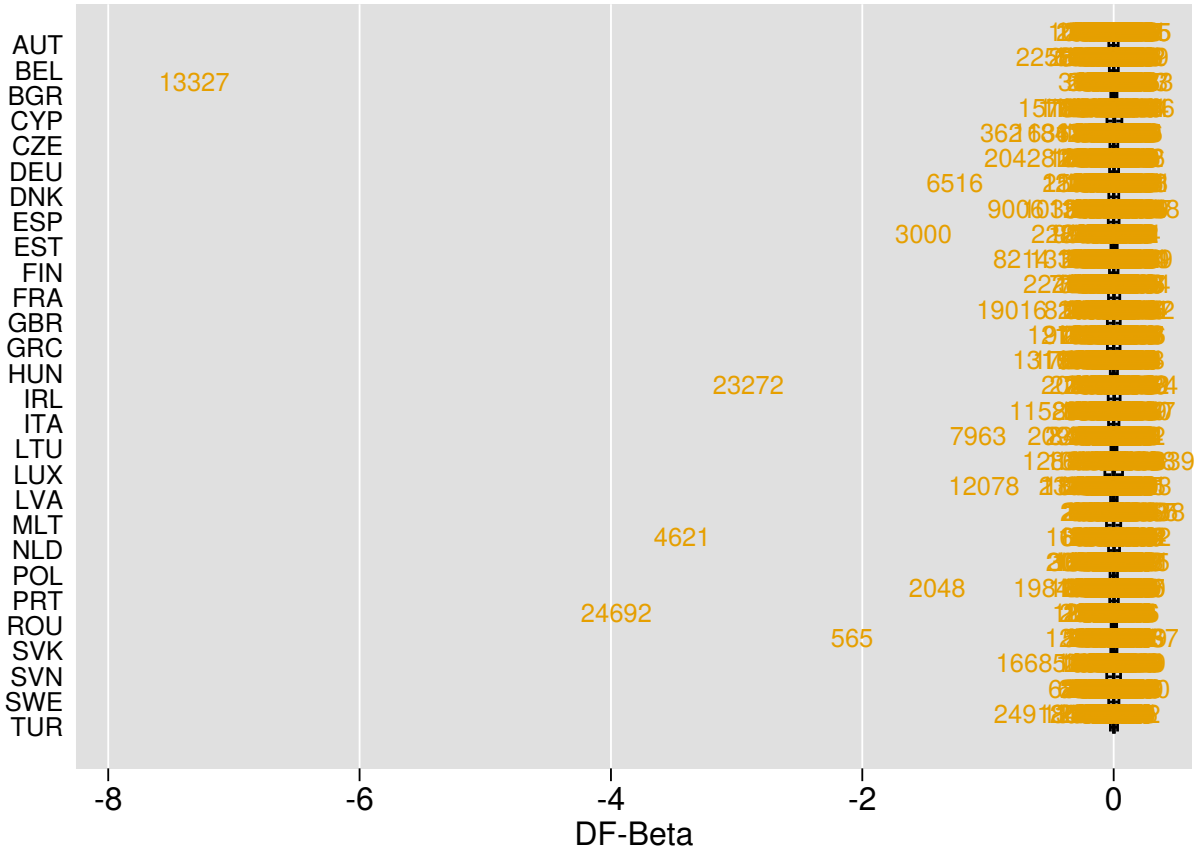


Plots ordered by field rank of b_hhinc

Several countries look suspicious. Also the dispersion of household income in Bulgaria is very low (Influential data points?).

Influential data-points in one of the micro models?

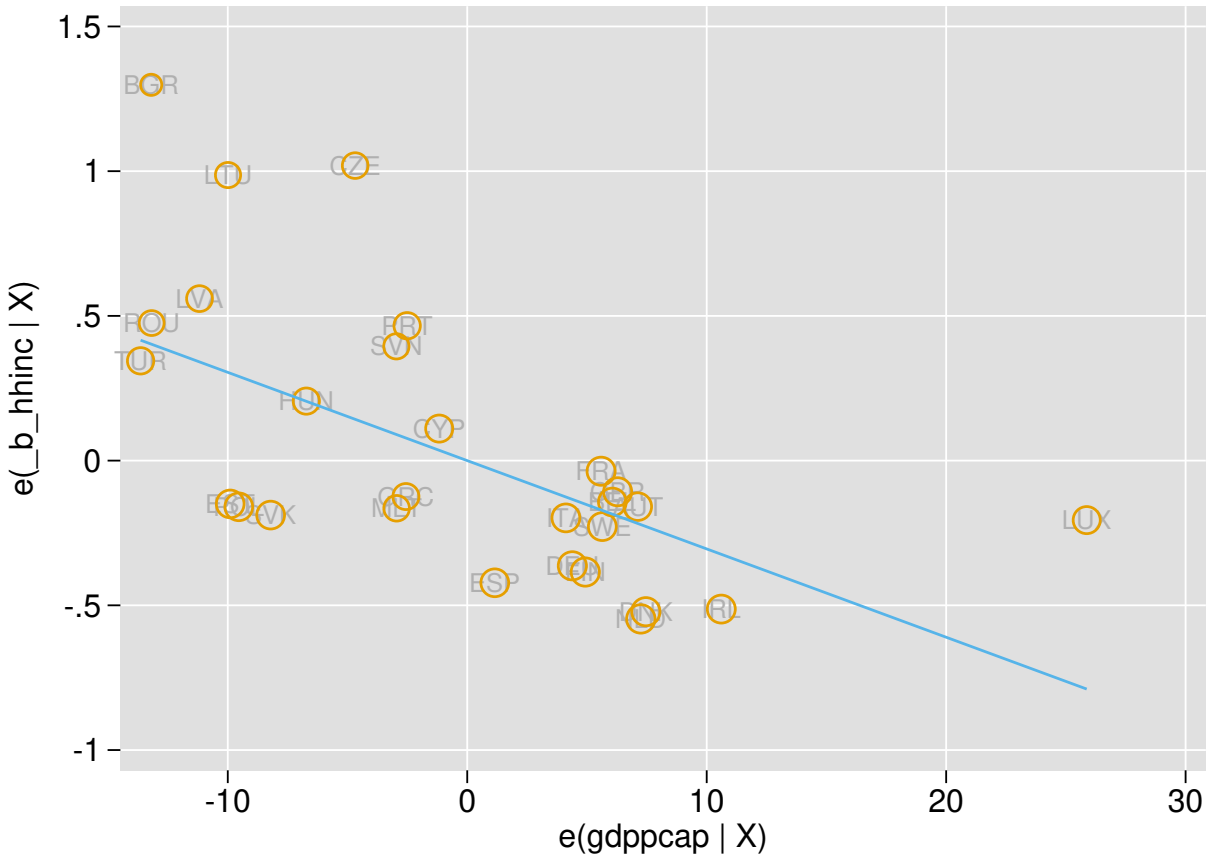
```
. twostep cntryiso: microdfb lsat hhinc i.sex age || _b_hhinc, marker(ms(i)
   mlabel(id) mlabpos(0)) overopts(label(labsize(*.8)))
```



Observation 13327 in Bulgaria is very suspicious! Would removing this observation change the CLI? All influential data points operates towards zero (Take Logs?).

Influential data points on the macro-level?

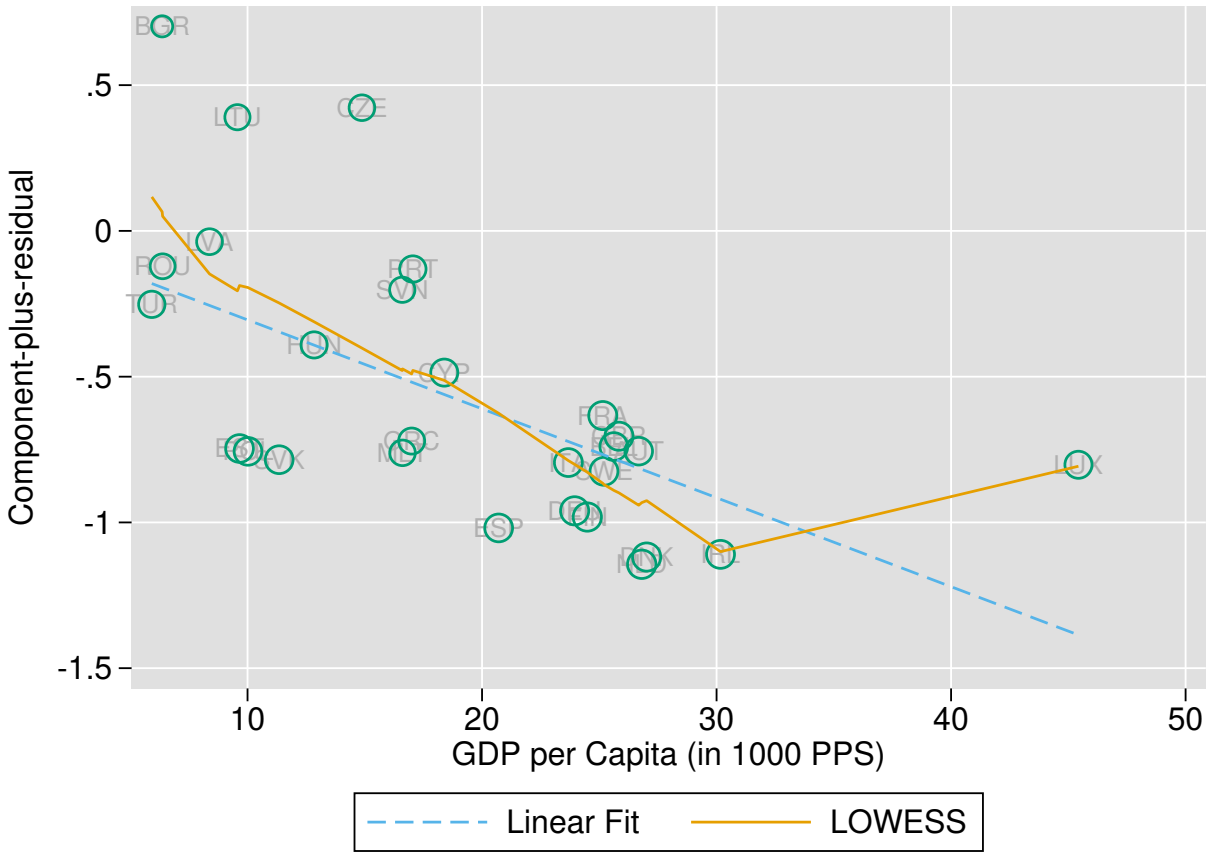
```
. twostep cntryiso: reg lsat hhinc i.sex age || avplot _b_hhinc gdppcap
```



Luxembourg affects the CLI more than other countries. Is this appropriate?

Linear model on macro-level appropriate?

```
. twostep cntryiso: reg lsat hhinc i.sex age || cprplot _b_hhinc gdppcap
```



Looks reasonable—except the usual suspicious.

What if . . . ?

. . . we remove the suspicious data?

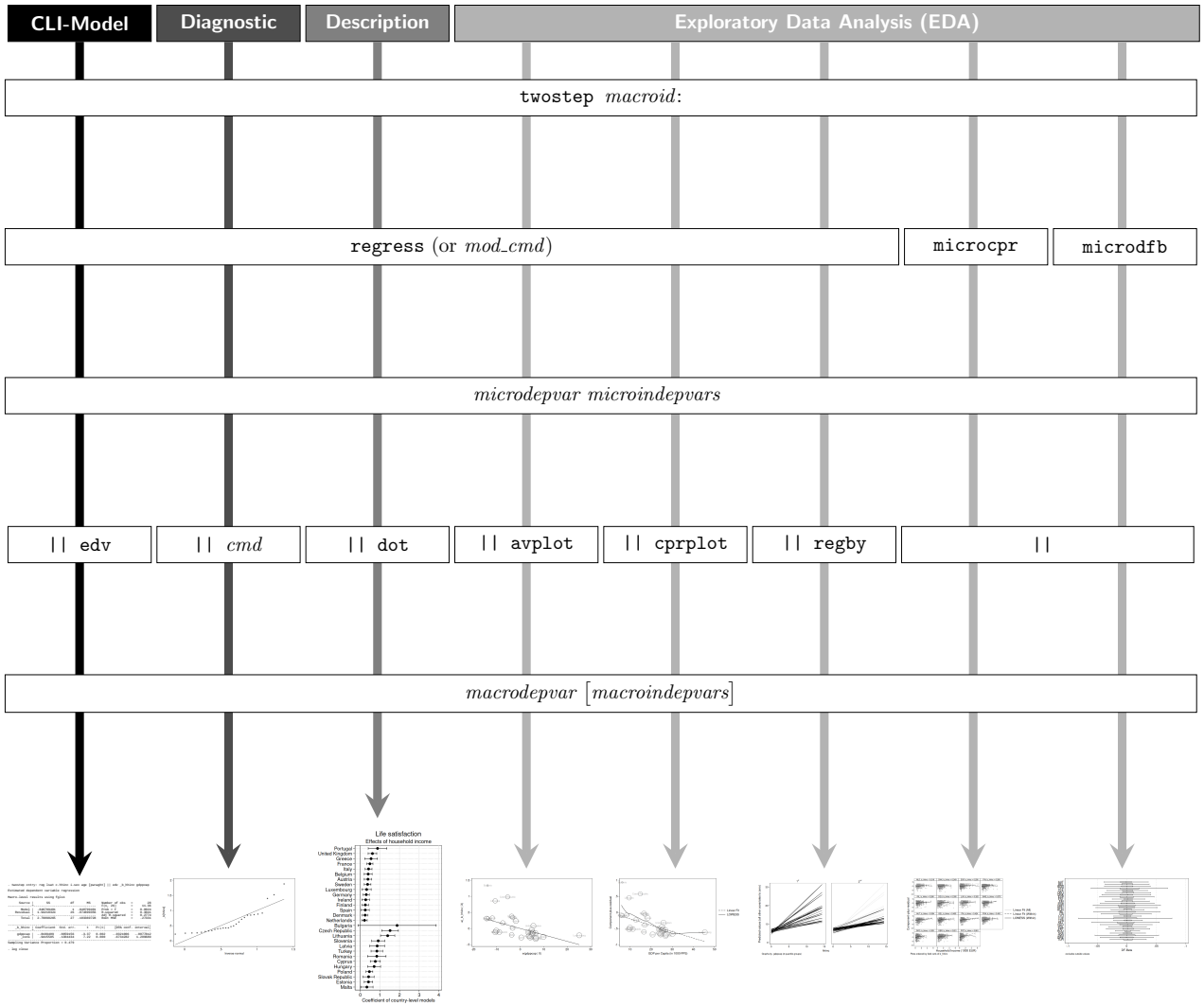
```
. mixed lsat c.hhinc##c.gdppcap i.sex age || cntry: hhinc i.sex age, cov(un
  structured)
. mixed lsat c.hhinc##c.gdppcap i.sex age || cntry: hhinc i.sex age if id!=
  13327, cov(unstructured)
. mixed lsat c.hhinc##c.gdppcap i.sex age || cntry: hhinc i.sex age if cntr
  yiso != "LUX", cov(unstructured)
. mixed lsat c.hhinc##c.gdppcap i.sex age || cntry: hhinc i.sex age if cntr
  yiso != "LUX" & id != 13327, cov(unstructured)
```

Model	Coef.	Std.Err.	*t*
All data	-0.0269340	0.0071865	-3.748
Without 13327	-0.0319394	0.0095615	-3.340
Without LUX	-0.0384881	0.0082620	-4.658
Without LUX and 13327	-0.0462515	0.0113670	-4.658

All changes are larger than the difference between random coefficient and random intercept.

Contents

Twostep syntax: Graphical overview



Acknowledgements

- This presentation was build with `texdoc` (Jann, 2016).
- The graphs in this presentation use Benn Jann's `imesh`-style and `cbline` color palette (Jann, 2018a,b,c).
- The command `center` (slide 16) was written by Jann (2004).
- `twostep` is also accessible on GiTLab:

`https://gitup.uni-potsdam.de/ukohler/twostep`

Bibliograpy I

Borjas, G. and G. Sueyoshi. 1994. A two-stage estimator for probit models with structural group effects. *Journal of Econometrics* 64: 165–182.

Bryan, M. L. and S. Jenkins. 2016a. Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review* 32(1): 3–22.

—. 2016b. Multilevel Modelling of Country Effects: A Cautionary Tale. Supplementary Material.

Donald, S. G. and K. Lang. 2007. Inference with difference-in-differences and other panel data. *Review of Economics and Statistics* 89: 221–233.

Giesecke, J. and U. Kohler. 2024. Twostep analysis of hierarchical data. *Stata Journal* 24(2): 213–249.

Hanushek, E. 1974. Efficient Estimators for Regressing Regression Coefficients. *The American Statistician* 28(1): 66–67.

Bibliograpy II

Heisig, J. P. and M. Schaeffer. 2019. Why You Should Always Include a Random Slope for the Lower-Level Variable Involved in a Cross-Level Interaction. *European Sociological Review* 35(2): 258–279.
URL https://doi.org/10.1093/esr/jcy053

Heisig, J. P., M. Schaeffer, and J. Giesecke. 2017. The Costs of Simplicity: Why Multilevel Models May Benefit from Accounting for Cross-Cluster Differences in the Effects of Controls. *American Sociological Review* 82(4): 796–827.

Hox, J. J. 2010. *Applied Multilevel Analysis: Techniques and Applications*. 2d ed. London: Routledge.

Jann, B. 2004. CENTER: Stata module to center (or standardize) variables. Statistical Software Components, Boston College Department of Economics.
URL https://ideas.repec.org/c/boc/bocode/s444102.html

Bibliograpy III

—. 2016. Creating LaTeX documents from within Stata using texdoc. *Stata Journal* 16(2): 245–263(19).
URL http://www.stata-journal.com/article.html?article=pr0062

—. 2018a. Color palettes for Stata graphics. *Stata Journal* 18(4): 765–785.
URL https://ideas.repec.org/a/tsj/stataj/y18y2018i4p765-785.html

—. 2018b. Customizing Stata graphs made easy (part 1). *Stata Journal* 18(3): 491–502.

—. 2018c. Customizing Stata graphs made easy (part 2). *Stata Journal* 18(4): 786–802.

Bibliograpy IV

Kohler, U. and J. Giesecke. 2021. TWOSTEP: Stata module to perform twostep multilevel analysis. Statistical Software Components, Boston College Department of Economics. URL `https://ideas.repec.org/c/boc/bocode/s459021.html`

Lewis, F. B. and D. A. Linzer. 2005. Estimating Regression Model in Which the Dependent Variable is Based on Estimates. *Political Analysis* 13: 345–364.

Raudenbush, S. W. and A. S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2d ed. Thousand Oaks: Sage.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2d ed. Massachusetts Institute of Technology.