Advances in Comparative Survey Research and Survey Data Harmonization

Marta Kołczyńska

RECSM Summer Methods School 2024

Course outline

Day 1

1. Cross-national surveys: overview of available data sources

2. Survey data quality and comparability: Total Survey Error framework, cross-survey differences in measurement and representation, survey quality and how to measure it

Day 2

3. Framework for survey data harmonization: representativeness and measurement

4. Representation comparability

Day 3

5. Measurement comparability

6. Measurement error in modeling

7. Wrap-up

III. Framework for survey data harmonization

1. Harmonization of representation
2. Harmonization of measurement
3. Latent trend models (if time allows)

Kołczyńska and Bürkner, 10.1093/jssam/smad024

Harmonization of measurement

Pre-requisite: items measure the same concept.

Fortunately, many survey questions are repeated across different survey project.

Trust in institutions, social trust, interest in politics, satisfaction with democracy, life satisfaction, ideologiacl self-placement on a left-right scale, self-assessed health, etc.

Satisfaction with democracy, see e.g.:Valgarðsson and Devine 2021, doi.org/10.1177/10659129211009605

Example: trust in institutions

(EB 91.5): I would like to ask you a question about how much trust you have in certain media and institutions. For each of the following media and institutions, please tell me if you tend to trust it or tend not to trust it.

Tend to trust

Tend not to trust

DK

(EVS 5): Please look at this card and tell me, for each item listed, how much confidence you have in them, is it a great deal, quite a lot, not very much or none at all?

a great deal

quite a lot

not very much

none at all

European Social Survey Round 10

CARD 9 Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly… **READ OUT**…

[Display before items B7-B12: STILL CARD 9 (IF NECESSARY: How much do you trust…) **READ OUT…]**

		No trust at all										**Complete trust**	**(Refu-sal)**	**(Don't know)**
B6	…[country]'s parliament?	00	01	02	03	04	05	06	07	08	09	10	77	88

CARD 9

No trust at all										**Complete trust**
0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**

Question wording differences (?)

English - trust (ESS, EB) vs. confidence (EVS)

The same word in:
Albania - besim
Belgium (dut) - vertrouwen
Belgium (fr) – confiance
Bulgaria – доверие
Croatia – povjerenje
Czech Republic – důvěra/ důvěřovat (noun/verb)
Denmark – tillid
Estonia – usaldate
Poland – zaufanie

Scale differences

Different response scale lengths:

2 (Eurobarometer)

4 (Afrobarometer, Arab Barometer, Asian Barometer, European Values Study, Latinobarometro, New Baltics Barometer, World Values Survey)

5 (Caucasus Barometer, ISSP, Life in Transition Survey)

7 (Americas Barometer)

10 (European Quality of Life Survey)

11 (European Social Survey)

Example: trust in institutions

Even for the same scale lengths, the labelling differs:

	EVS	CDCEE	NBB	WVS	ISSP/1991/Slovenia
1	A great deal	Totally	Complete trust	A great deal	Totally
2	Quite a lot	To a certain point	General trust	Quite a lot	A lot
3	Not very much	Little	General distrust	Not very much	Somewhat
4	None at all	Not at all	Complete distrust	None at all	Not at all

Add to this differences in country translations.

CDCEE = Consolidation of Democracy in Central and Eastern Europe; EVS = European Values Study; ISSP = International Social Survey Programme; NBB = New Baltics Barometer; WVS = World Values Survey.

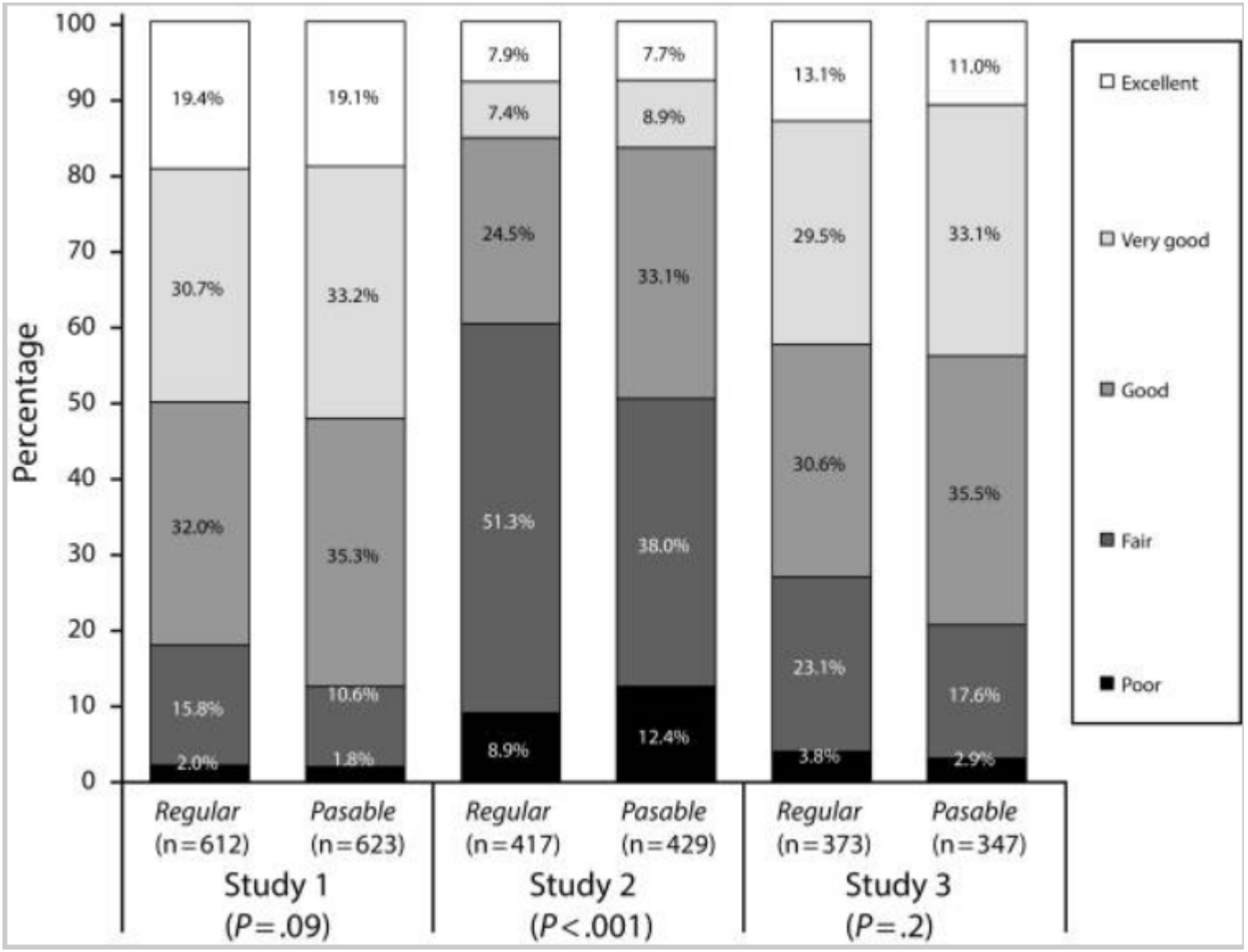
Example: Regular o pasable

Am J Public Health. 2019 December; 109(12): 1789–1796.
Published online 2019 December. doi: 10.2105/AJPH.2019.305341
PMCID: PMC6836
PMID: 31622

Impact of the Terms "*Regular*" or "*Pasable*" as Spanish Translation for "Fair" of the Self-Rated Health Question Among US Latinos: A Randomized Experiment

Sunghee Lee, PhD,[✉] Fernanda Alvarado-Leiton, MS, Elizabeth Vasquez, DrPH, and Rachel E. Davis, PhD

*Scale 1: excelente—muy buena—buena—**regular**—mala*
*Scale 2: excelente—muy buena—buena—**pasable**—mala*



Harmonization of response scales

Goal:

The same value on the latent scale should result in the same observed values.

Different observed values should reflect different latent values.

Harmonization of response scales

How to compare responses to scales of different lengths and with different response labels?

One potential solution: collect more data, i.e. survey where each respondent answers two or more question versions to create a „conversion table”.

Drawback: not very practical.

Harmonization of response scales

Most popular solutions:

- Dichotomization
- Linear rescaling
- Latent variable models

(ignoring response scale differences is not an option)

Dichotomization

☑ Convenient and quick to implement.

☑ Dichotomized DVs can be analyzed with logit / probit models.

☒ Throws away useful information.

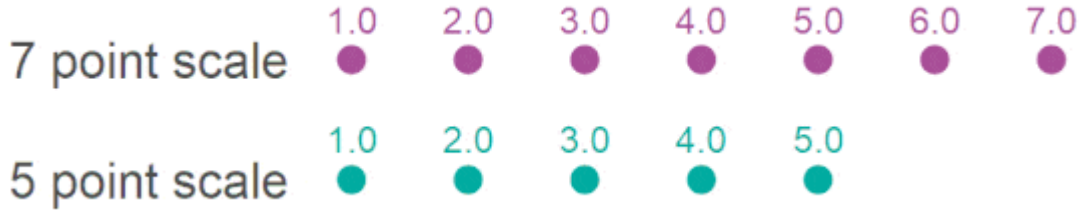
☒ For odd numer scales, it requires a decision about the mid-point:

- Consider the mid-point a positive value
- Consider the mid-point a negative value
- Split respondents in the middle category between 0 and 1 (how?)

  → Will inevitably introduce different amounts of bias across data sources

Linear rescaling

Recoding of scales to a common range, i.e. stretching or squeezing.



Singh, 2021, „(Not) by any stretch of the imagination: A cautionary tale about linear stretching.” https://doi.org/10.34879/gesisblog.2021.30

Linear rescaling

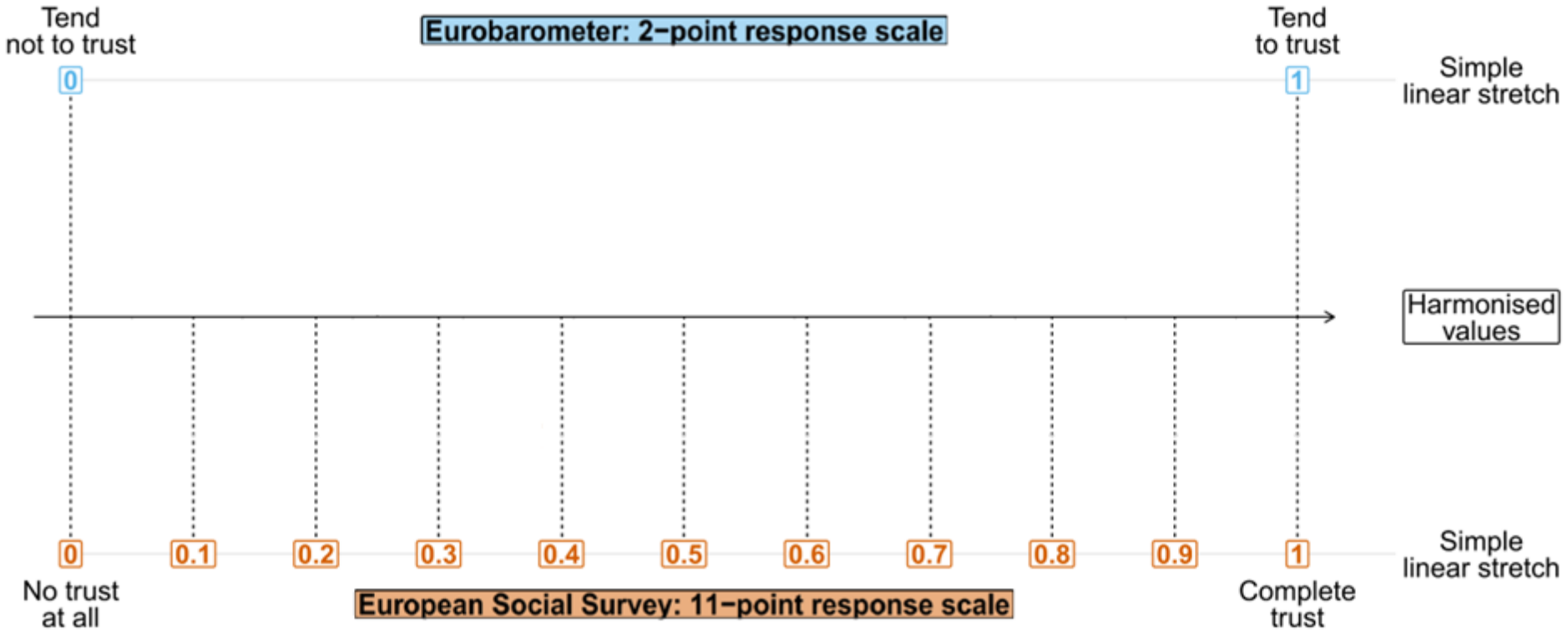


Fig. 1 Linear stretching: comparing the simple and modified approaches **(modified)**

Cichocki and Jabkowski, https://doi.org/10.1007/s11135-022-01523-5

Linear rescaling

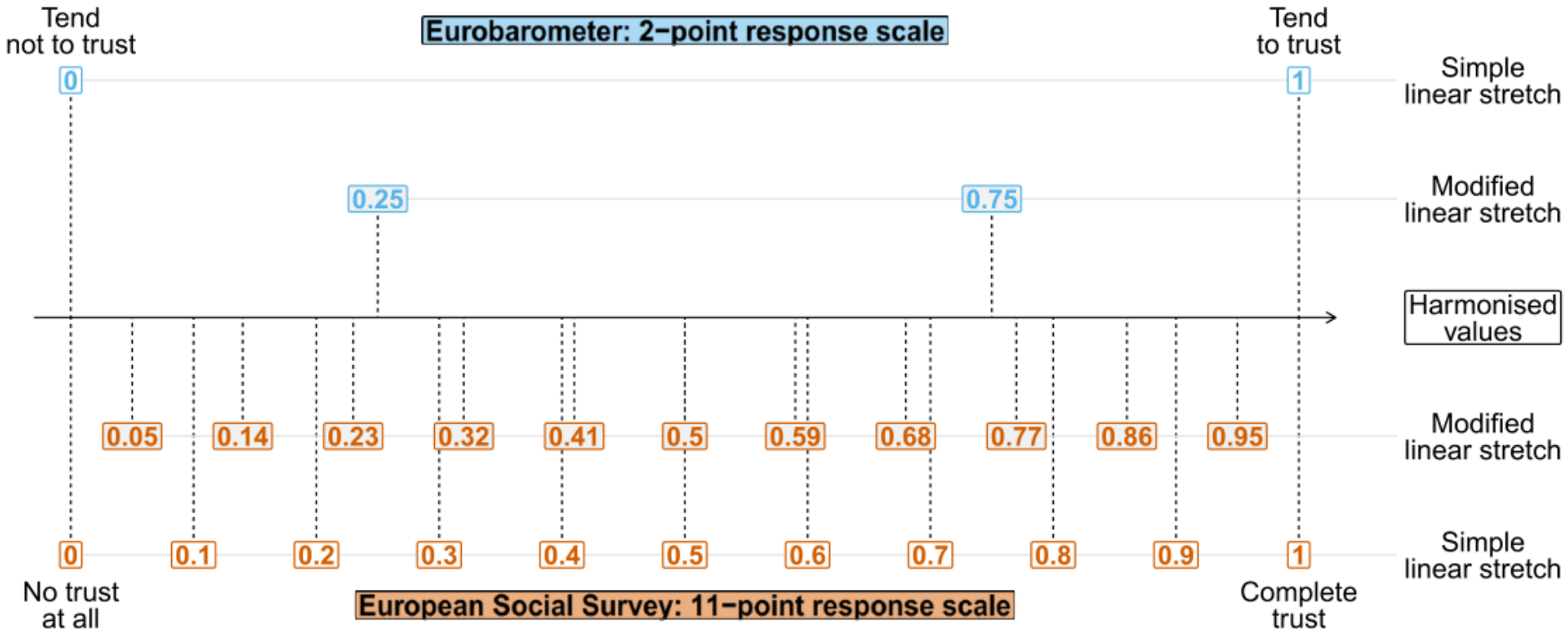
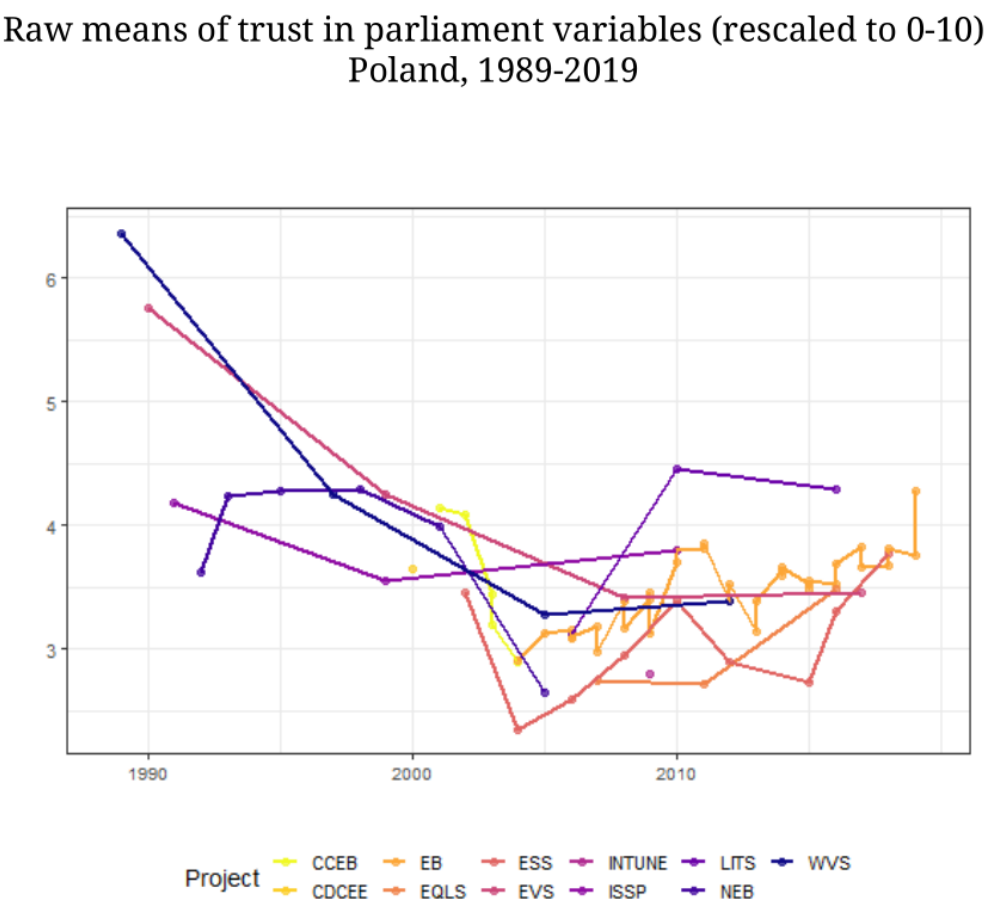


Fig. 1 Linear stretching: comparing the simple and modified approaches

Cichocki and Jabkowski, https://doi.org/10.1007/s11135-022-01523-5

Linear rescaling: summary

- Good quick method for screening the data



Linear rescaling: summary

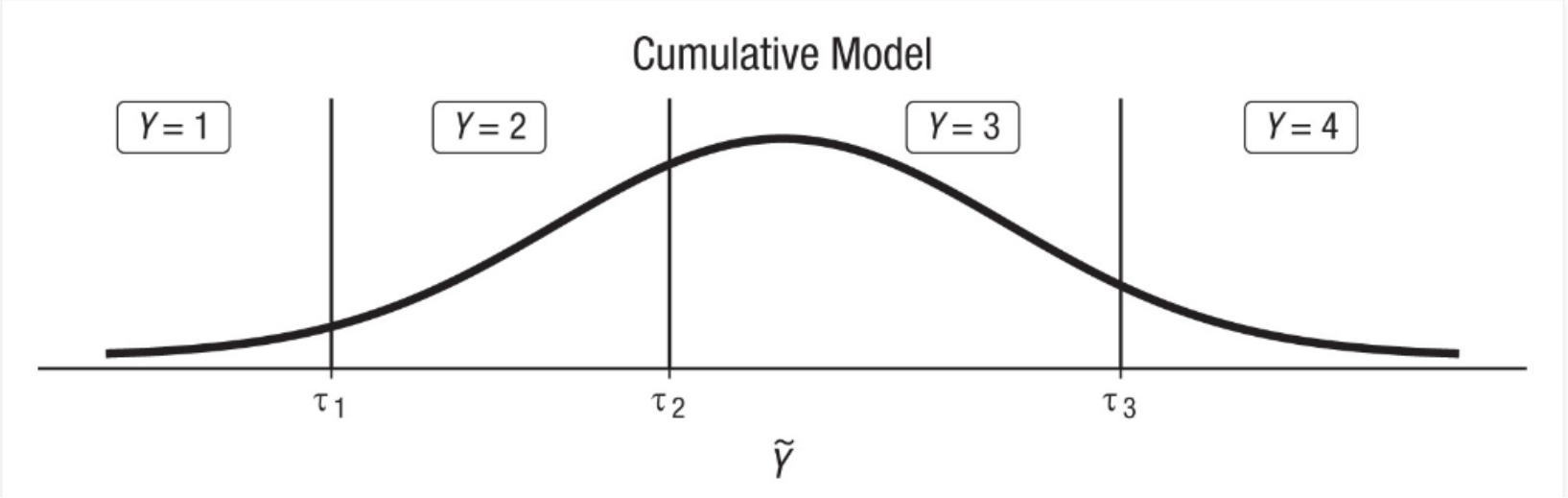
But:

- Does not account for the response mechanism
- Assumes all items have the same difficulty and response labels do not matter
- Assumes all response scales are symmetric

- The modified linear stretch restricts the range of the harmonized variable

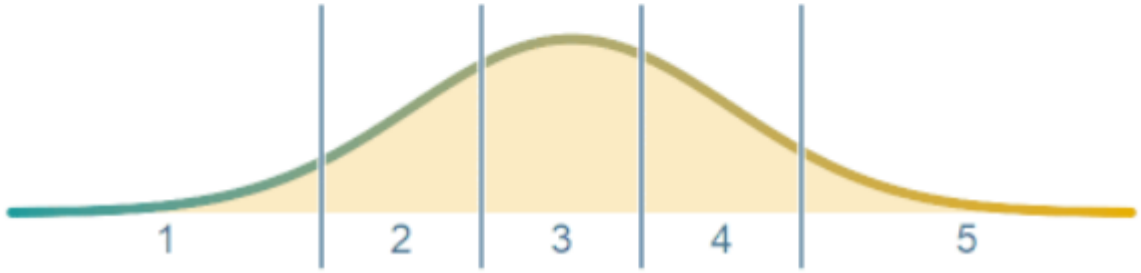
Latent variable models

Ordinal cumulative models assume that the observed ordinal variable Y, the opinion rating, originates from the categorization of a latent continuous variable $\tilde{Y}$.

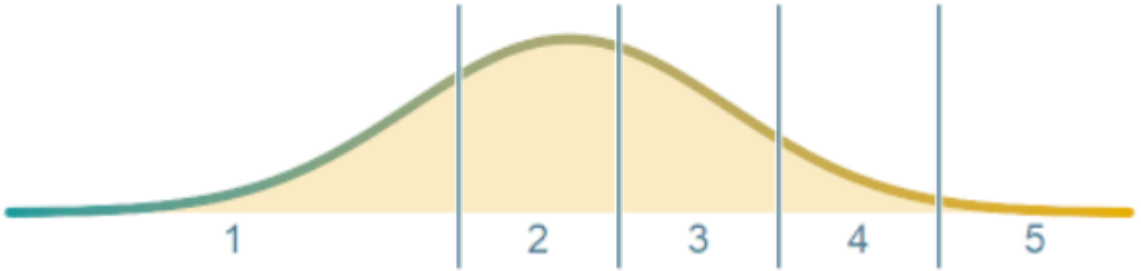


Bürkner and Vuorre 2019, https://journals.sagepub.com/doi/full/10.1177/2515245918823199

5 = trust a great deal



5 = trust completely



Singh, 2021, https://blog.gesis.org/not-by-any-stretch-of-the-imagination-a-cautionary-tale-about-linear-stretching/

Harmonization of response scales

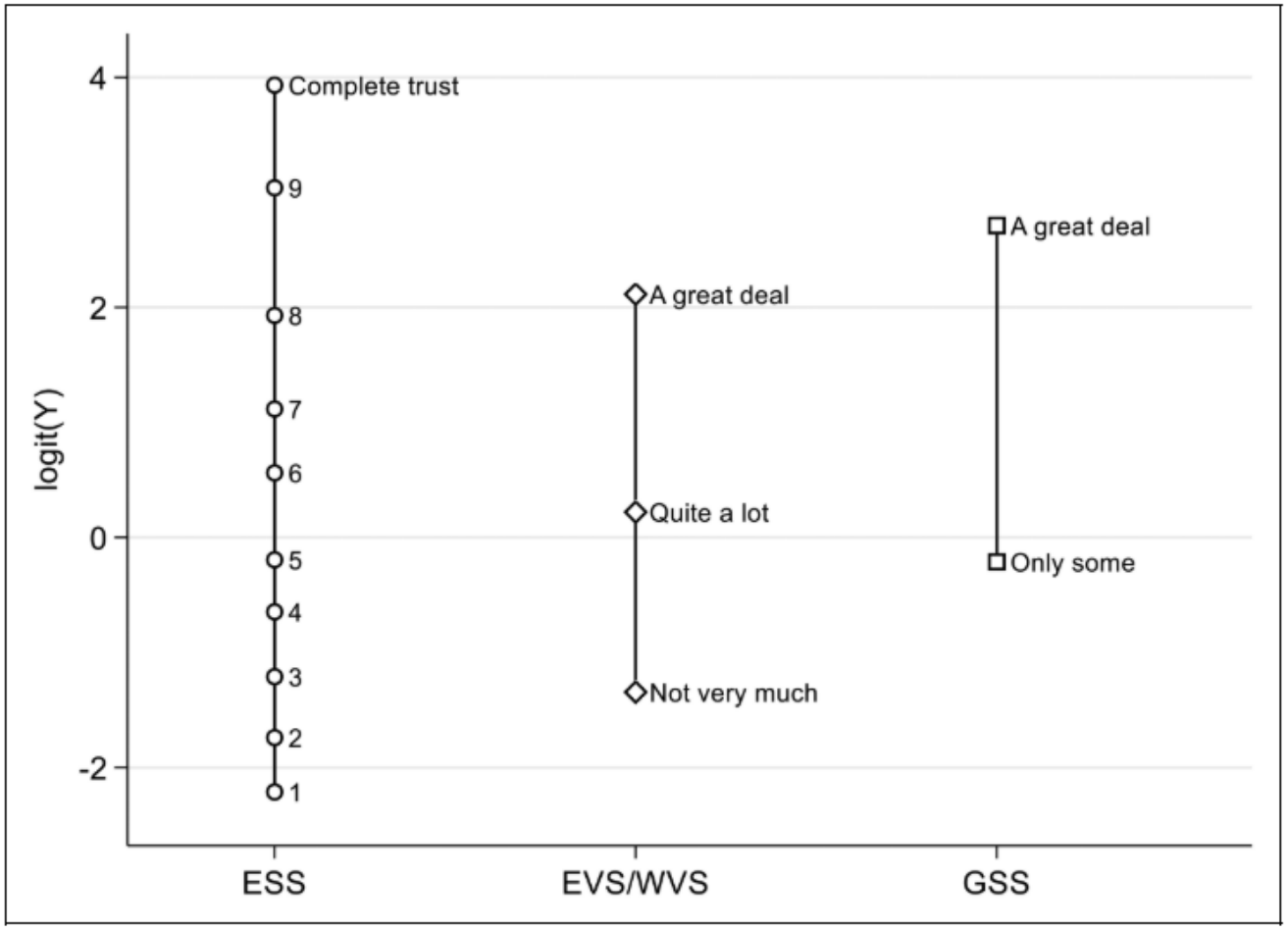
Latent variable models also make it possible to account for the different scale lengths and differences in threshold sets.

Markus Gangl. 2023. A Generalized Ordered Logit Model to Accommodate Multiple Rating Scales. Sociological Methods & Research, https://doi.org/10.1177/00491241231186

+ ado in Stata

Kołczyńska, Bürkner, Kennedy, and Vehtari, 2024. Modeling Public Opinion Over Time and Space: Trust in State Institutions in Europe, 1989-2019. Survey Research Methods 18(1). https://doi.org/10.18148/srm/2024.v18i1.8119

+ code in R / brms (Stan)

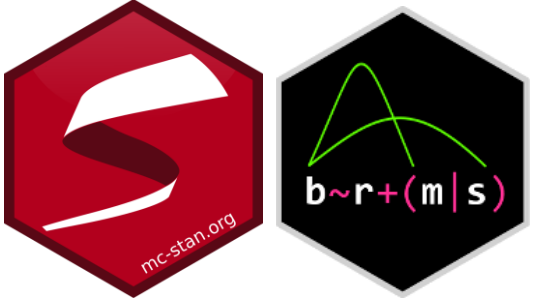


Lowest category: **0** **None at all** **Hardly any**

Figure 1. Estimated cutpoint locations for the ESS, EVS/WVS, and GSS response formats to express trust in the national parliament. *Note:* Inverted cutpoint estimates $-\alpha_{j_s}$ from model specification M1 in Table 1 (multiscale ordered logit model, full ESS–EVS–GSS–WVS sample). ESS = European Social Survey; GSS = General Social Survey; EVS/WVS = European and World Values Survey.

Markus Gangl. 2023. A Generalized Ordered Logit Model to Accommodate Multiple Rating Scales. Sociological Methods & Research, https://doi.org/10.1177/00491241231186

Exercise



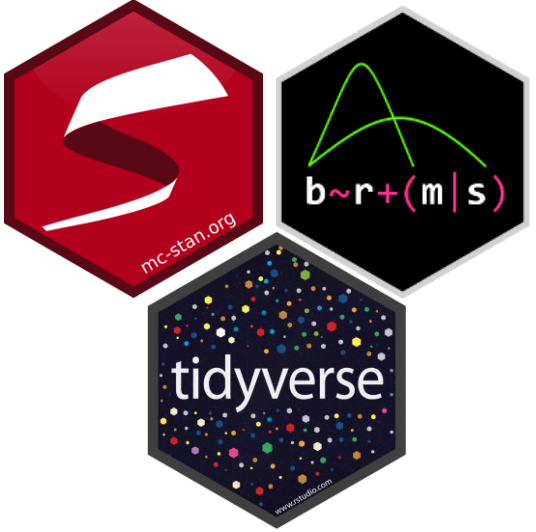
We'll use the brms package created and maintained by Paul-Christian Bürkner.

brms stands for **B**ayesian **r**egression **m**odels using **S**tan, and provides a user-friendly interface to Stan.

https://paul-buerkner.github.io/brms; https://mc-stan.org

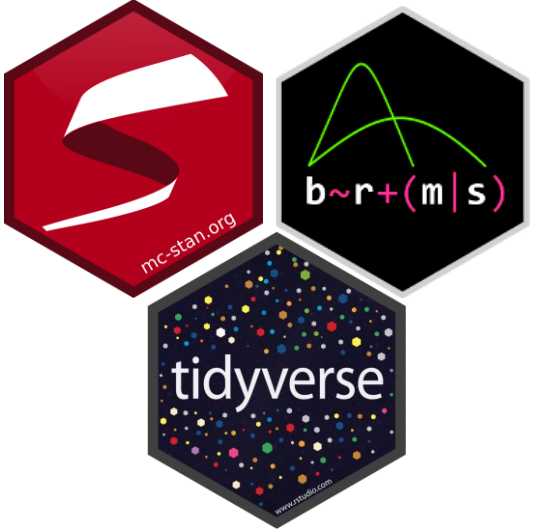
Exercise

ex4_brms.R



Exercise

ex4_brms2.R



IV. Using survey estimates

Uncertainty

Regression models assume that predictors are measured without error.

In OLS models with one predictor, if the predictor is measured with error, the estimated effect is downwardly biased (attenuation bias).

In OLS models with more predictors and non-linear models, the situation becomes less predictable.

- Note: lagged variable models: $y_{t1} \sim y_{t0} + x1_{t0} + x2_{t0} + \ldots$

Solution: error-in-measurement models

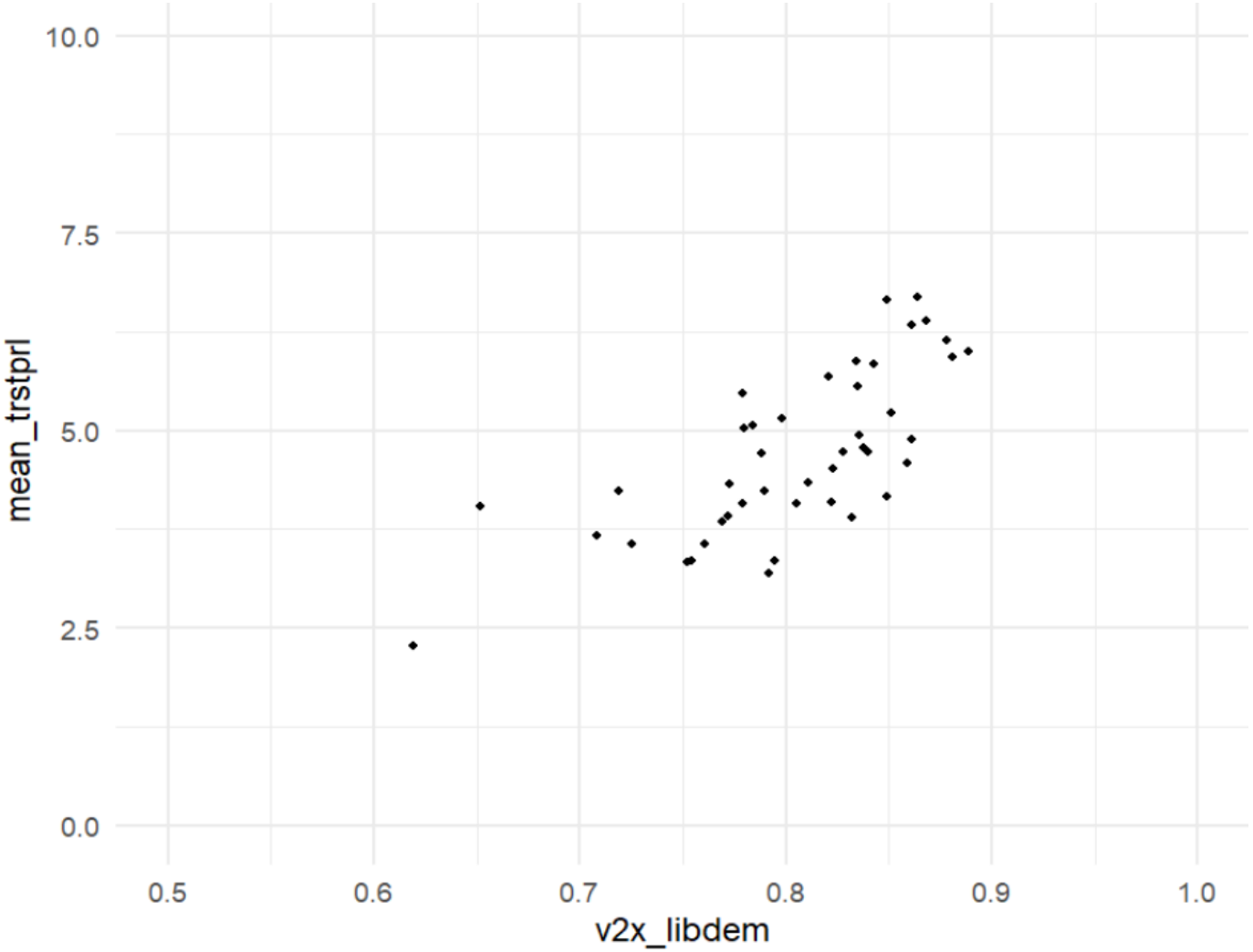
Error in measurement: ESS & V-Dem

- Varieties of Democracy indicators of liberal democracy
- Based on expert surveys, aggregated via Bayesian item response theory models
- Provide a point estimate and an uncertainty interval

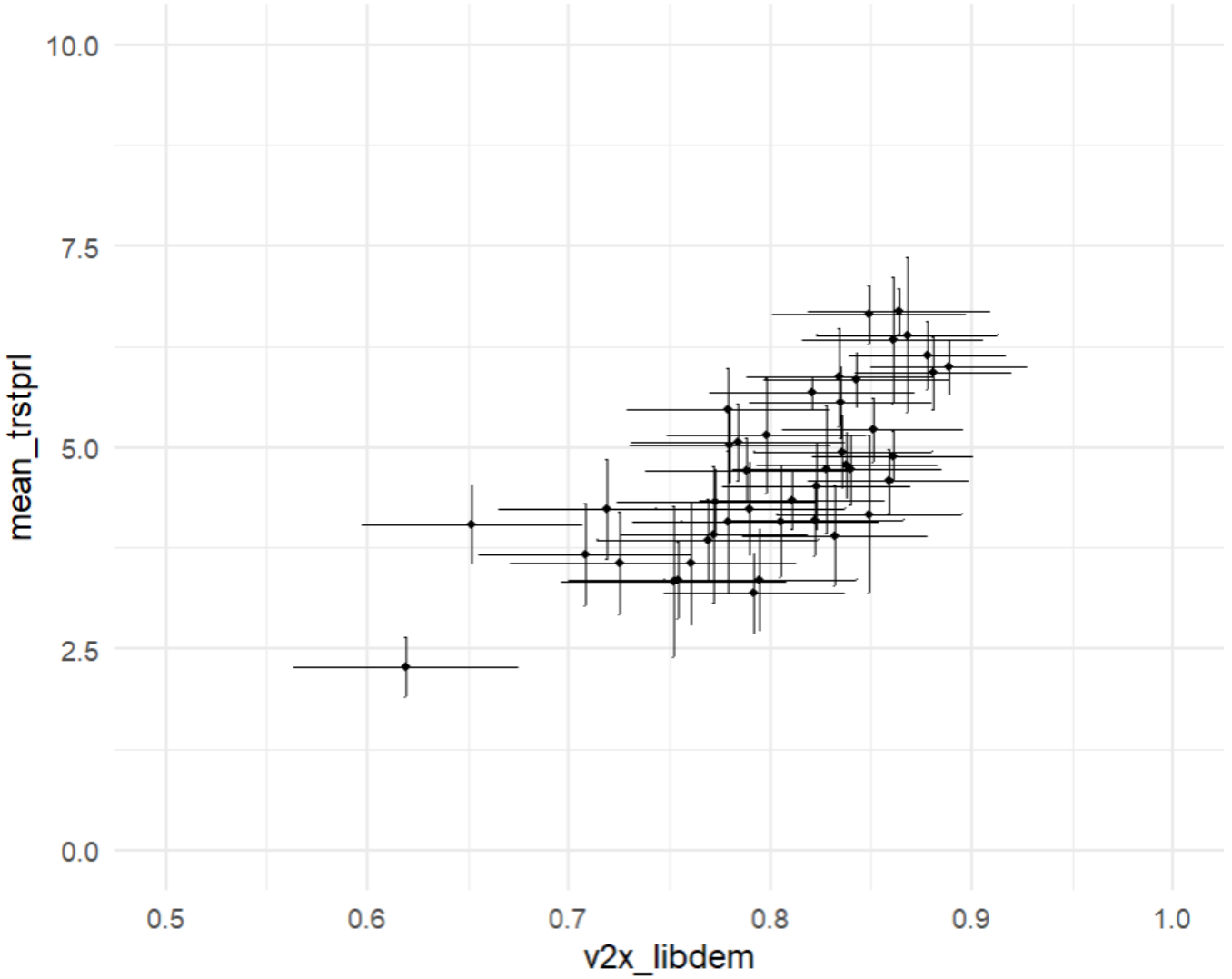
- European Social Survey
- Means of trust in parliament by country-year
- Standard error of the mean

Data source: https://www.v-dem.net/ & https://www.europeansocialsurvey.org/

Error in measurement

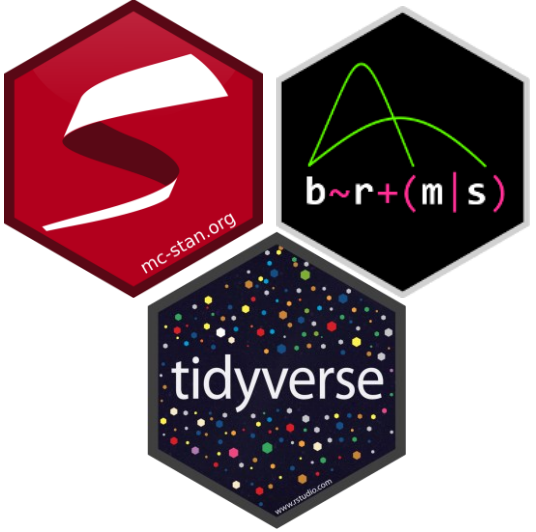


Error in measurement



Exercise

ex6_error_in_measurement.R



American Political Science Review, Page 1 of 18
doi:10.1017/S0003055419000558

In the Mood for Democracy? Democratic Support as Thermostatic Opinion

CHRISTOPHER CLAASSEN *University of Glasgow*

(…) Using new panel measures of democratic mood varying over 135 countries and up to 30 years, this article finds little evidence for such a positive feedback effect of democracy on support. Instead, it demonstrates a negative thermostatic effect: increases in democracy depress democratic mood, while decreases cheer it. (…)

AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Does Public Support Help Democracy Survive?

Christopher Claassen University of Glasgow

(…) The article then demonstrates a positive effect of support on subsequent democratic change, while adjusting for the possible confounding effects of prior levels of democracy and unobservable time-invariant factors. (…)

APSR, doi.org/10.1017/S0003055419000558; AJPS, doi.org/10.1111/ajps.12452

1. Measuring Democratic Mood

Data for 137 countries,
from 14 survey projects
and 1391 national surveys.

1.1. Survey Questions Used to Measure Democratic Mood

1. Three statements items

   1.1. Which of these three statements is closest to your own opinion? (AfroBarometer)
   - Democracy is preferable to any other kind of government
   - Under some circumstances, an authoritarian government can be preferable to a democratic one
   - For someone like me, it does not matter what kind of government we have.

2. "Churchill" items

   2.1. Democracy may have its problems, but it is better than any other form of government. To what extent do you agree or disagree? (ArabBarometer)

3. Strong leader items

   3.1. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Elections and Parliament are abolished so that the president can decide everything. (AfroBarometer)

4. Military rule items

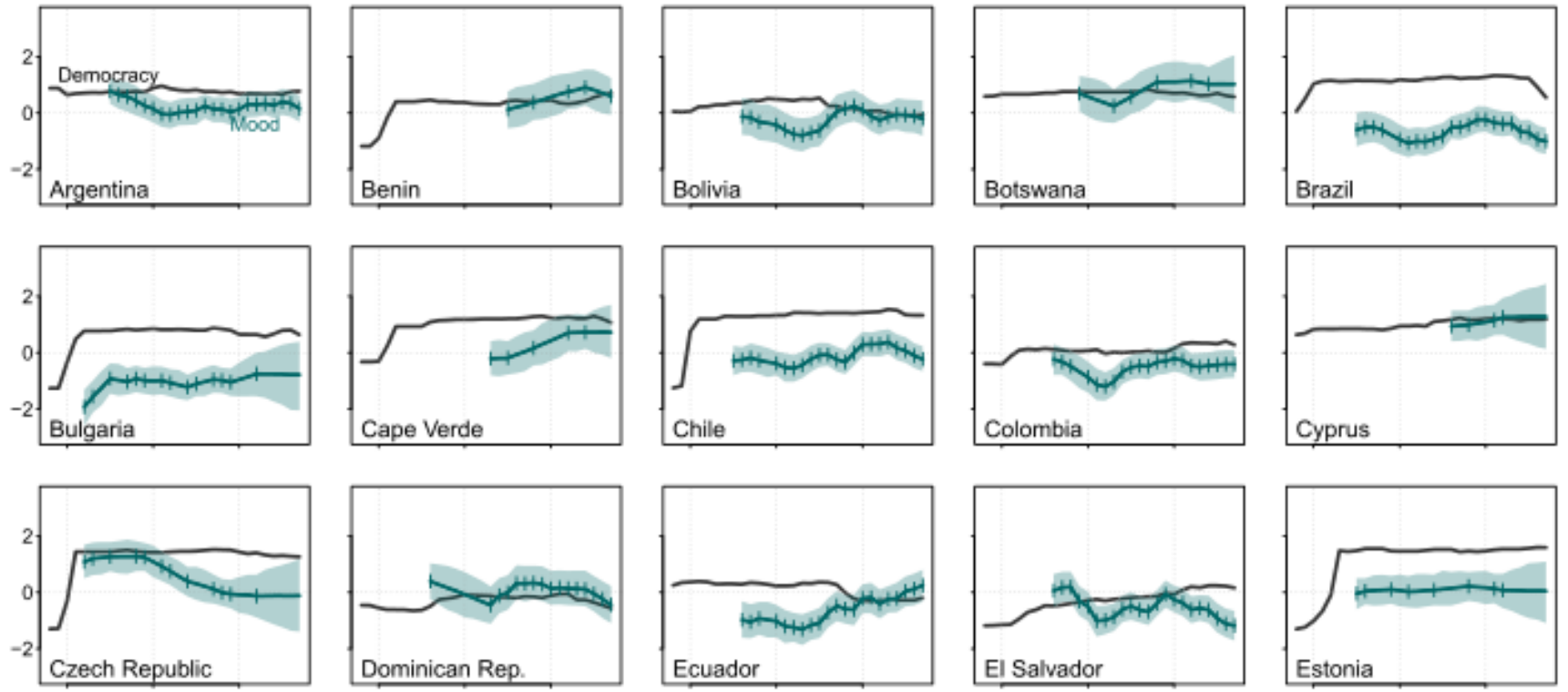
   4.1. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? The army comes in to govern the country (AfroBarometer)

5. One party rule items

   5.1. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office (AfroBarometer)

Source: Appendix to: doi.org/10.1017/S0003055419000558

FIGURE 2. The Dynamics of Democracy and Mood: New Democracies



Each plot shows V-Dem Liberal democracy (plotted using a black line) from 1988 to 2017 and estimates of democratic mood (plotted using a blue line). The shaded regions around the mood estimates indicate 90% uncertainty intervals. Vertical bars on the mood estimates indicate years in which survey data were available. Liberal democracy and mood are unit-normal standardized.

Note: V-Dem indicators also come with uncertainty estimates, which were ignored in the plot here.

APSR, doi.org/10.1017/S0003055419000558

Democracy, Public Support, and Measurement Uncertainty

Published online by Cambridge University Press: **05 May 2022**

YUEHONG 'CASSANDRA' TAI, YUE HU and FREDERICK SOLT

From the abstract:

Do democratic regimes depend on public support to avoid backsliding? Does public support, in turn, respond thermostatically to changes in democracy? Two prominent recent studies (Claassen 2020a, 2020b) reinvigorated the classic hypothesis on the positive relationship between public support for democracy and regime survival—and challenged its reciprocal counterpart—by using a latent variable approach to measure mass democratic support from cross-national survey data. Both studies, however, used only the point estimates of democratic support; we show that incorporating the concomitant measurement uncertainty into these analyses reveals that there is no support for either study's conclusion. (…)

APSR, doi.org/10.1017/S0003055422000429

Including Measurement Uncertainty in Time-Series, Cross-Sectional Analyses: The Case of Mood and Democracy

15 Pages • Posted: 20 Sep 2021 • Last revised: 5 Dec 2022

Christopher Claassen
University of Glasgow

Date Written: November 29, 2022

(…) Unresolved, however, is the issue of how to incorporate the uncertainty of measurement into subsequent analyses. One proposal is to use the "method of composition," in which multiple samples from the posterior distribution of the latent variable are analyzed, rather than a single point estimate. Using Monte Carlo studies, this paper shows, however, that the method of composition performs poorly in a dynamic context when the latent variable is endogenous to other covariates. In such circumstances (which are likely to be widespread) a unified model that measures latent variables and estimates structural connections in a single step is more accurate and is recommended.

SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3924934

Conclusions

More surveys – more opportunities, but also more work.

Importance of survey quality and comparability assessment – validation, checking documentation and using additional helper tools.

Both measurement and representation errors need to be assessed and addressed.

Many of the issues of *ex post* survey data harmonization apply to any comparative analysis, but are often ignored.

Conclusions

There is an art component to this science.

All models are wrong, but ...

Post scriptum on reproducibility

- Survey data harmonization projects use a lot of data from different sources, require a lot of data processing, and many decisions (some of them seem small and inconsequential, but may not be)

- All decisions and analytic choices need to be clearly documented (and, ideally, justified)

- Data processing and analyses need to be reproducable (by others, which includes your future self) from the reading in of the different data sources to the final analysis outcomes

Resources (1)

- A General Primer for Data Harmonization (2024), Cheng et al., https://www.nature.com/articles/s41597-024-02956-3
- Handbook on Survey Data Harmonization in the Social Sciences (2023): https://onlinelibrary.wiley.com/doi/book/10.1002/9781119712206
- Cross-cultural survey guidelines, chapter on harmonization (2016): https://ccsg.isr.umich.edu/chapters/data-harmonization/
- Maelstrom Research guidelines for rigorous retrospective data harmonization (2016): https://academic.oup.com/ije/article/46/1/103/2617181
  - Project website: https://www.maelstrom-research.org/
- The Comparative Panel File: Harmonized Household Panel Surveys from Seven Countries: https://academic.oup.com/esr/article/37/3/505/6168670

Resources (2)

- Ranjit Singh's blog at GESIS: https://blog.gesis.org/author/dr-ranjit-k-singh/
- European Social Survey Webinar: Harmonising survey data across different survey modes by Ranjit Singh: https://www.youtube.com/watch?v=EhllEyCgL5s
- MRP tutorials, etc.: http://joshuamccrain.com/index.php/mrp-in-r/
- World Association for Public Opinion Research (WAPOR) webinars on YouTube: https://www.youtube.com/@worldassociationforpublico783
- Bayesian statistics: Bayesian Rethinking by Richard McElreath (book and lectures on Youtube): https://github.com/rmcelreath/stat_rethinking_2023
- Bayesian Data Analysis by Aki Vehtari - course at Aalto University, videos online: https://avehtari.github.io/BDA_course_Aalto/
- Andrew Gelman's blog: https://statmodeling.stat.columbia.edu/

Acknowledgements

NATIONAL SCIENCE CENTRE
POLAND

The content was developed in course of the project *Causes and Consequences of Political Trust: Polarization and Democratic Utility of Trust in Cross-national Perspective* funded in the Sonatina program (2019/32/C/HS6/00421).