Workshop

Building Multi-Source Databases for Comparative Analyses

Survey Data Recycling as an Analytic Framework for Survey Data Reprocessing (II)

The SDR Team

Harmonization process



Target variables

- Technical variables
- Variables measuring respondents' characteristics:
 - (a) socio-demographics:
 - (b) reported behavior
 - (c) attitudes and opinions

Technical variables

Interview date - date of the interview (where available)

Interview year – year of the interview (where available)

- Survey year uniform within national survey (or when the majority of interviews were conducted)
- **Country** ISO code: country or territory (plus: Germany: DE-E & DE-W, Israel: IL-JEW & IL-ARB)
- Weight individual case weight: design, post-stratification or combined weights, as provided in the original survey data set (where available, otherwise = 1) and recomputed weight (with corrections)

Socio-demographics

Target variable	SDR 1	SDR 2
Gender	+	D
Age & Year of Birth	+	D
Rural/Urban Locality	+	D
Education: Level and Years of Schooling	+	D
Household/Family Income	S	Ρ
Employment Status	-	Р
Non-manual/Manual Type of Work	-	Р
Weight/Height (BMI Index)	-	X

Behavioral variables

Target variable	SDR 1	SDR 2
Attending demonstration all combined	+	-
Attending demonstrations (ever)	-	D
Attending demonstrations (last year)	-	D
Signing petition	+	D
Boycotting	-	D
Voting	S	Р
Membership in organizations	-	Р
Talking about politics with friends	-	D

Attitudes and opinions

Target variable	SDR 1	SDR 2
Trust in parliament	+	D
Trust in legal system	+	D
Trust in political parties	+	D
Trust in people	+	D
Interest in politics	+	D
Life satisfaction – self rated	-	Р
Happiness – self rated	-	Р

Health

•

Health - self-rated General health Mental health

Standardization in survey data harmonization: general comments

Standardization in an *ex-ante* harmonization process:

- preparing a model (standard) of questionnaire items, including response categories
- assuring that this model has been applied by all relevant parties

Resulting variables in the dataset are considered as formally equivalent (e.g. ESS)

Standardization in an *ex-post* harmonization process is quite different:

- no a priori model (standard) to follow

Standardization in the SDR Project

Process by which survey data received in various formats of questionnaire items are transformed to a <u>common format of the resulting target variable</u>.

→standardization aims to achieve <u>optimum degree of uniformity</u> in a given target variable.

This means that:

(a) the essential meaning of the questionnaire items(s) in source data is the same;

(b) the values of the target variable are expressed in the same metric

Implication of common format

Variability with respect to the **meaning** of the questionnaire items

Semantic analysis of questionnaire items: <u>does particular item differ from the</u> <u>essential meaning of the target variable</u> with respect to:

- specified or implied <u>time</u> (last year, last 5 years, ever)
- specified or implied <u>space</u> (in your place of living, your community)
- <u>things and events</u> covered (marches, ralies, mass meetings)
- <u>attributes</u> mentioned (legal, illegal; national, international)?

In SDR, each diversion of national survey with respect to the semantics of the questionnaire items is accounted for \rightarrow

 \rightarrow <u>Control variables</u> for the meaning of the questionnaire items

Time span during which respondent might have participated in demonstrations

	1	2-3	4-5	10 (8)	ever
N-surveys	334	62	42	25	664
N-people	525,857	81,989	56,126	29,220	832,108

Implication common format

Variability with respect to the **categories** of answers

Formal analysis of answers. <u>Does the particular set of answers differ from the scale of the target variables?</u>

Four different types of scales in the SDR project:

1. Nominal scales **Control variables reflect any deviation from the assumed nominal scale**

2. Dummy scales (Yes=1, No = 0) Yes is affirmative Control variables reflect fuzziness of "No" response (",rather no" ",I am not sure")

3. Rating scales Control variables: Length (number of points), Direction (ascending, descending), Polarity (unipolar, bipolar)

4. Percentile-rank scales **Control variables: specific for given target variable**

Illustration: target variables and harmonization controls for institutional trust

<u>Target variables</u> on the respondent's level of trust in three basic public institutions:

- national parliament
- legal system
- political parties

Constructing target variables is accompanied by description of <u>source variables</u> in terms of <u>harmonization controls</u> accounting for methodological variability among national surveys.

We deal with variablity of the questionnaire items, i.e. categories of precoded answers – scales:

- length of scales (L)
- direction of scales (D)
- polarity of scales (P)

Trust in institutions: Examples of the wording

- 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... [country]'parliament? the legal system? ...political parties? (ESS) [**11-point scale**]
- 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?... parliament...the justice system... political parties (EVS) [5-point scale]
- In order to get ahead, people need to have confidence and to feel that they can trust themselves and others. To what degree do you think that you trust the following totally, to a certain point, little, or not at all? ... political parties... the parliament (CDCEE 2) [**3-point scale**]

Linear transformation

		Mean of	Standard
Original scale	Recodes	scores	deviation
11-points	0, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0	5.0	3.16
10-points	0.5, 1.5, 2.5, 3.5, 4, 5.5, 6.5, 7.5, 8.5, 9.5	5.0	2.87
7-points	0.71, 2.14, 3.75, 5.00, 6.43, 7.86, 9.29	5.0	2.86
5-points	1.0, 3.0, 5.0, 7.0, 9.0	5.0	2.83
4-points	1.25, 3.75, 6.25, 8.75	5.0	2.79
2-points	2.5, 7.5	5.0	2.50

Percentile-rank scales

Formulas

$[(c_e + 0.5 f_i) / N] \times 100\%$

where c_e is the count of all scores less than the score of interest, f_i is the frequency of the score of interest, and N is the number of examinees in the sample.

<The score shows what proportion of cases are below a given point on scale>

Target variable: Trust in parliament

	Variable label	Variable name	Variable values*
Target variable	Trust in parliament (11-point scale)	T_TR_PARLI_11	0 = lowest trust or distrust 10 = completely trust
	Trust in parliament (distribution- preserving scale)	T_TR_PARLI_DISTRIB	0 = lowest point in distribution 100 = highest point in distribution

Harmonization controls: coding

Source: scale length 2 = 2-point scale 4 = 4-point scale 5 = 5-point scale 7 = 7-point scale 10 = 10-point scale 11 = 11-point scale Source: scale direction 0 = descending

1 = ascending

Source: scale polarity 0 = bipolar 1 = unipolar

Correlation of harmonization controls (H) with target variables (T)

	Trust, T		
Harmonization controls, H	Parliament	Legal system	Political Parties
	11-point scale		
Length of original scale	-0.011	0.022	-0.032
Direction of original scale	0.050	0.050	0.021
Polarity of original scale	-0.012	0.023	-0.018
	Distributional scale		
Length of original scale	-0.011	0.014	0.014
Direction of original scale	0.017	0.021	0.019
Polarity of original scale	-0.006	0.013	-0.006
Mean inter-scale correlation	0.869	0.875	0.872
Ν	1,676,289	1,499,173	1,232,684

Startegies of using harmonization controls

- Selection of surveys
- Weighting of surveys
- Controling for effects of harmonization controls

Effects of selecting surveys

What are the consequences of eliminating national surveys having:

- very short scales (e.g. dichotomies)?
- with ascending scales (in contrast to descending scales)?
- scales other than unipolar (e.g. bi-polar or nominal)?

Effects of weights for groups distinguished according to harmonization controls

<u>First group (standard, "the best"): 11-point scale, Ascending scale, and Unidirectional</u> <u>Second group ("the second best"): From 4-point to 10-point scale, Ascending scale, and Unidirectional</u> <u>Third group (with methodological disadventages): All other combinations of harmonization controls</u> Analysis for data from 13 international survey projects, 2007-2013, N = 448,557

	Weights	Impact of interest in politics on trust in parliament
	For groups	b
No weights	-	0.534
Progression toward standard	1.0, 0.7, 0.5	0.629
Strong progression toward standard	1.0, 0.5, 0.25	0.643

TRUST = $a + b_1^*INTREST_IN_POLITICS + b_2^*GENDER + b_3^*AGE + b_3^*EDUCATION + b_4^*RURAL + e$

Using harmonization controls when constructing latent variables

L = scale length, D = scale direction, P = scale polarity for PA = trust in parlaiment, LE = legal system, PO = political parties



Control variables for Household/Family Income - Planning

Definition

- 1 definition includes components
- 0 otherwise
 - not identifiable
 - missing

Taxation

- 1 before tax
- 2 after tax
 - not identifiable
 - missing

Time frame

- 1 year
- 2 quarter
- 3 month
- 4 week
- 5 other
- missing

<u>Unit</u>

- 1 household
- 2 family
- 3 unclear
- missing

<u>Scale</u>

- 1 currency without intervals
- 2 currency intervals
- 3 deciles
- 4 quintals or quartiles
- 5 non-distributional ordinal scale*
- 6 other
 - not identifiable
 - missing