2017 CSDI Workshop, Mannheim, March 16-18

Item-level Controls for Ex-post Harmonization of Cross-national Survey Data:

Theoretical Arguments and Empirical Illustration

Kazimierz M. Slomczynski

Irina Tomescu-Dubrow

Marta Kołczyńska

Cross-National Studies: Interdisciplinary Research and Training Program (*CONSIRT*)

Background and prsentation's aim

Long standing project: Survey Data Recycling (SDR). The project is about an analytic framework for integrating information from extant survey and non-survey sources to create multi-country multi-years datasets that enable comparative, cross-national research.

SDR data set: 22 international survey projects, 89 waves (i.e., project*wave) and 1,721 national surveys (i.e. project*wave*countries); 142 countries/territories, from 1966 to the first quarter of 2014. N= 2, 289,060 respondents.

SDR data set does not contain original data but only constructed (harmonized) variables and created metadata. Available at dataverse.

SDR project involves, among other things, ex-post harmonization of substantive variables, with control indicators describing features of the source data.

Presentation's aim: to discuss strategies of using harmonization controls in empirical analyses.

Harmonization process

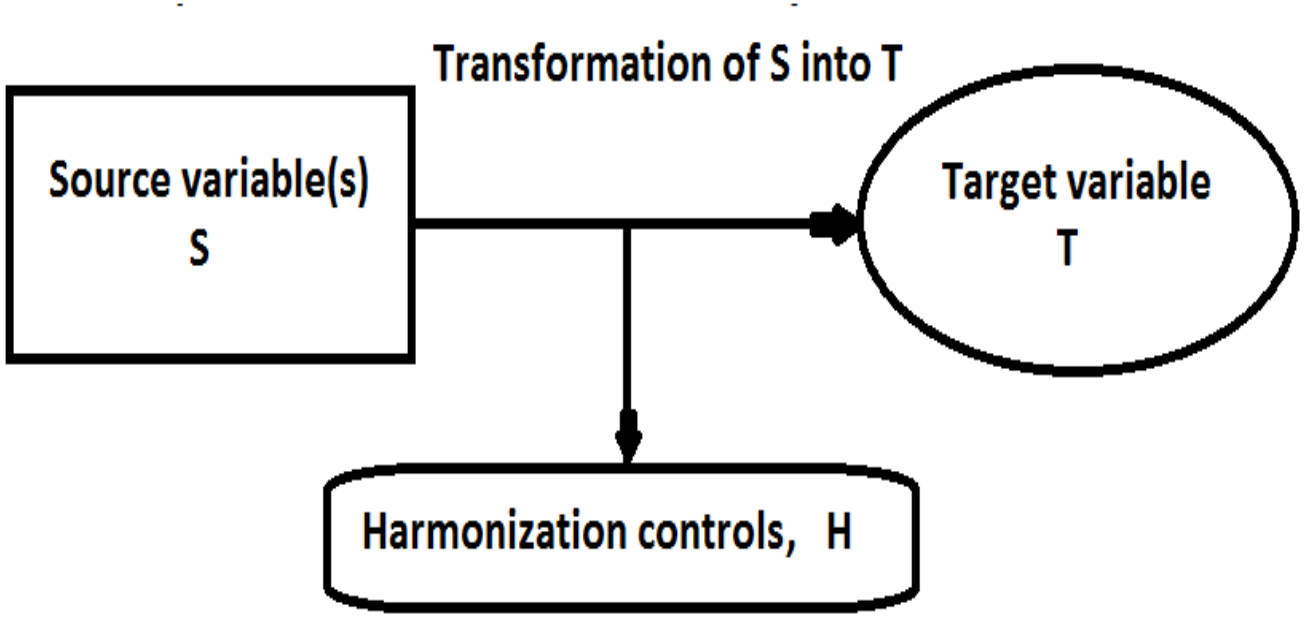
$$T = f(S)$$

<u>Transformation of S into T</u>:

Values of S are recoded into values of T so that under specified assumptions they are comparable across surveys.

Harmonization controls, H, account for methodological variability of S across surveys.

Controls, H, deal with the content of items and response categories (scales)



Target variables and harmonization controls

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

- national parliament (PA)
- legal system (LE)
- political parties (PO)

Constructing target variables is accompanied by description of source variables in terms of harmonization controls accounting for methodological variability among national surveys. In this paper we deal with variablity of the questionnaire items, particularly with categories of precoded answers – scales:

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

Abbrev.	Survey Project	Time span	Waves	Data Sets	Cases	Trust in institutions
			Counts			
AFB	Afrobarometer	1999-2009	4	66	98,942	PA, LE, -
AMB	Americas Barometer	2004-2012	5	92	151,341	PA, LE, PO
ARB	Arab Barometer	2006-2011	2	16	19,684	PA, LE, PO
ASB	Asian Barometer	2001-2011	3	30	43,691	PA, LE, PO
ASES	Asia Europe Survey	2000	1	18	18,253	PA, LE, PO
CB	Caucasus Barometer	2009-2012	4	12	24,621	PA, LE, PO
CDCEE	Consolidation of Democracy (in CEE)	1990-2001	2	27	28,926	PA, -, PO
CNEP	Comparative National Elections Project	2004-2006	1	8	13,372	PA, LE, PO
EB	Eurobarometer	1983-2012	7	152	138,753	PA, LE, PO
EQLS	European Quality of Life Survey	2003-2012	3	93	105,527	PA, LE, PO
ESS	European Social Survey	2002-2013	6	146	281,496	PA, LE, PO
EVS/WVS	European Values Study / World Values Survey	1981-2009	9	312	423,084	PA, LE, PO
ISSP	International Social Survey Programme	1985-2013	13	363	493,243	PA, LE, -
LB	Latinobarometro	1995-2010	15	260	294,965	PA, LE, PO
LITS	Life in Transition Survey	2006-2010	2	64	67,866	PA, LE, PO
NBB	New Baltic Barometer	1993-2004	6	18	21,601	PA, LE, PO
Total		*1981-2013*	*83*	*1,677*	*2,225,365*	

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]

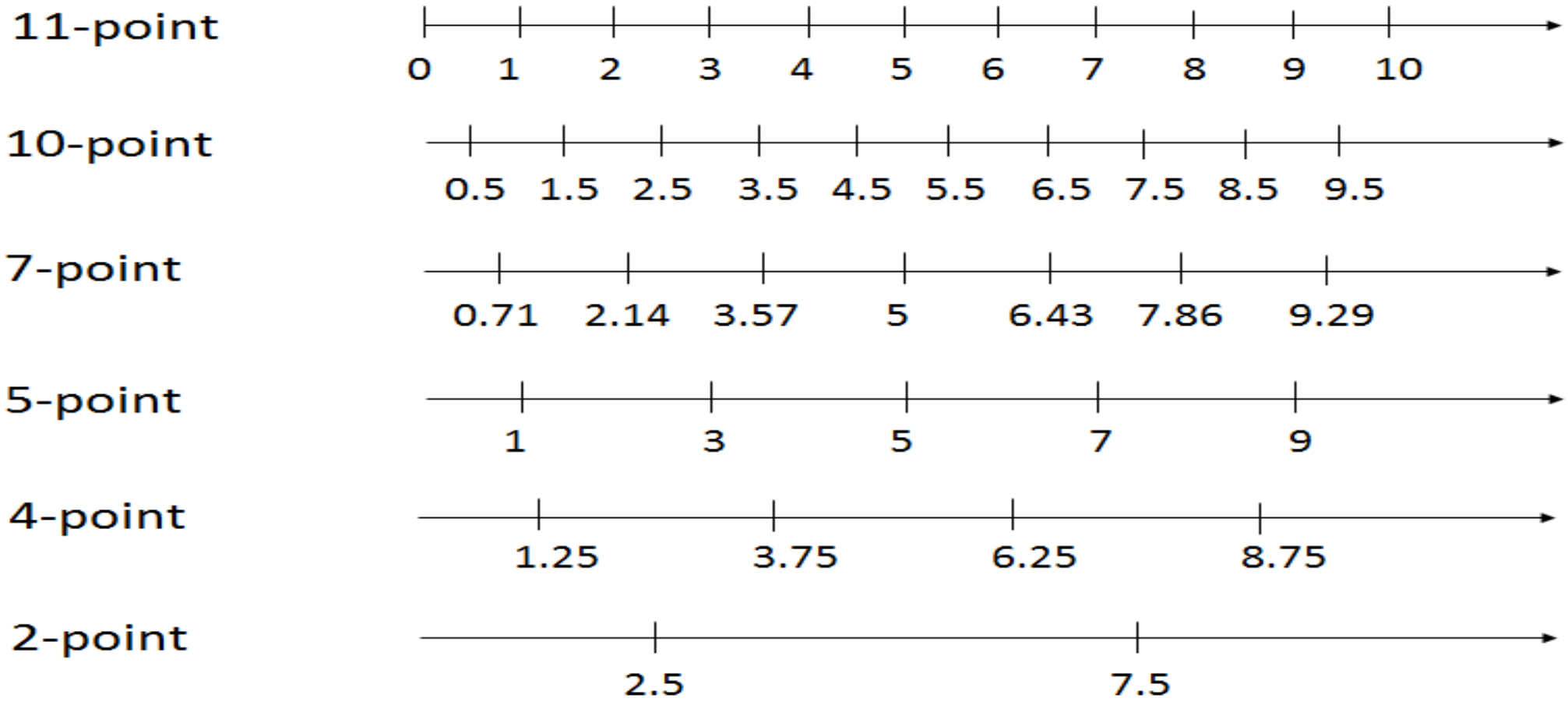
Target variable: Trust in parliament

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

Missing value codes: SPSS (STATA).

- 9 (.i) = missing data; -8 (.h) = question not asked in national survey; -7 (.g) = insufficient information for all response categories; -6 (.f) = insufficient information for single response category; -5 (.e) = variable not identified in data file; -4 (.d) = value not acceptable; - 2 (.b) = not applicable; -1 (.a) = don't know

Transformation of source values into the target 0-10 scale



Distributional scaling

For the source *n*-point scale the values, for values *k* ranging from 1 to *n*, where X_k is the distribution of the variable, *k* was recoded to *m* according to the formula:

$$m = \sum_{i=1}^{k-1} X_i + \frac{X_k}{2}$$

Distribution-based transformation. Example: TRUST IN PARLIAMENT, LITS/2/PL

Source value k	Distribution X_k	Cumulative distribution $\sum_{i=1}^{k} X_i$	$\sum_{i=1}^{k-1} X_i$	$\sum_{i=1}^{k-1} X_i + \frac{X_k}{2}$	Target value (rounded to integer)
1	10.68	10.68	0	= 10.68/2 = 5.340	5
2	32.75	43.44	10.68	= 10.68 + 32.75/2 = 27.055	27
3	32.11	75.55	43.44	= 43.44 + 32.11/2 = 59.495	59
4	21.69	97.23	75.55	= 75.55 + 21.69/2 = 86.395	86
5	2.77	100	97.23	= 97.23 + 2.77/2 = 98.615	99

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

Diversity of response scales in items about trust in the parliament

Length of scale	Direction of scale	
	Traditional (descending)	Reversed (ascending)
11		CNEP, ESS
10		EQLS
7		AMB, NBB
5	ISSP, VPCPEE	CB, LITS
4	ARB, ABS, ASES, CDCEE, EVS, LB, NBB, WVS	AFB
2	EB	

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

Harmonization controls	Trust		
	Parliament PA	Legal system LE	Political Parties PO
	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
Inter-scale correlation	0.869	0.875	0.872
N	1,676,289	1,499,173	1,232,684

Three 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)?

Proposed criterion: for surveys with a given scale property (e.g., dichotomies) correlations of the target variables (e.g., PA) with some substantive variables, postulated by a theoretical consideration (e.g., education, EDU) are much different than in other examined surveys and in contrast to our knowlege.

Is it justifable to eliminate surveys with dichotomies, if $r_{pa.edu}$ for dichotomies << $r_{pa.edu}$ for 4-point or longer scales?

Researchers using SDR data must have some argument for elimiating surveys on the basis of L, D, P. (E.g., dichotomies in comparison with longer scales introduce too much uncontrolable (random) error).

Effects of weights for groups distinguished according to harmonization controls

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

$$TRUST = a + b_1*INTREST\_IN\_POLITICS + b_2*GENDER + b_3*AGE + b_3*EDUCATION + b_4*RURAL + e$$

	Weights		**Impact of interest in politics on trust in parliament**	
	For groups	**Sample**	**b**	**Beta**
No weights	**-**	**-**	**0.534**	**0.205**
Equalizing groups	**1.0, 1.0, 1.0**	**2.06, 1.65, 0.51**	**0.604**	**0.219**
Progression toward standard	**1.0, 0.7, 0.5**	**2.06, 1.17, 0.25**	**0.629**	**0.224**
Strong progression toward standard	**1.0, 0.5, 0.25**	**2.06, 0,82, 0.13**	**0.643**	**0.227**

Effects of harmonization controls

Proportion of common variance for PA, LE, PO without harmonization controls (above diagonal) and with harmonization controls (below diagonal)

	PA	LE	PO
PA	1.00	0.373	0.394
LE	0.364	1.00	0.261
PO	0.386	0.264	1.00

Application of harmonization controls into construction of latent variables

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