

**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 ([CONSIRT.osu.edu](http://CONSIRT.osu.edu))**

Aim of presentation: Discuss strategies of using harmonization controls in empirical analyses

Context: Longstanding project on Survey Data Recycling (SDR)

SDR is 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 survey dataset: 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 1<sup>st</sup> quarter of 2014. N= 2, 289,060 respondents.

- does not contain original data; only constructed (harmonized) variables and newly-created metadata. Available at: [dataverse](http://dataverse.org). See also: [dataharmonization.org](http://dataharmonization.org)

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

## The Harmonization process

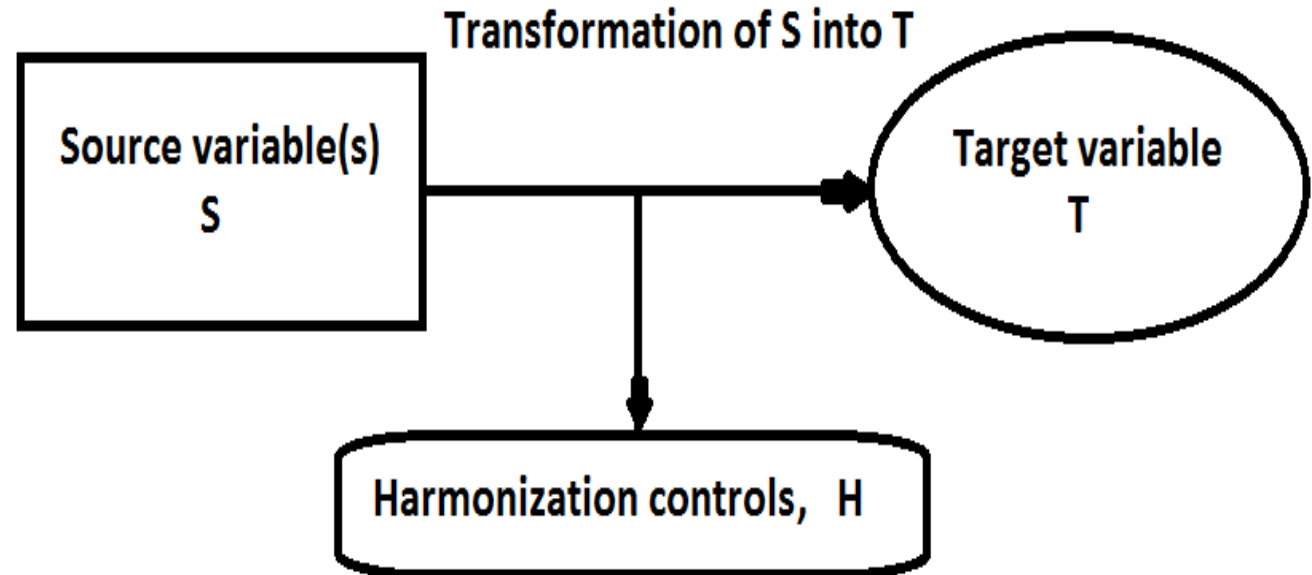
$$T = f(S)$$

### Transformation of S into T:

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 measuring respondents' level of trust in 3 main public institutions:

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

Constructing target variables is accompanied by describing source variables with harmonization controls that account for methodological variability among national surveys.

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

## Trust in Institutions: Examples of wording of the source items

- *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

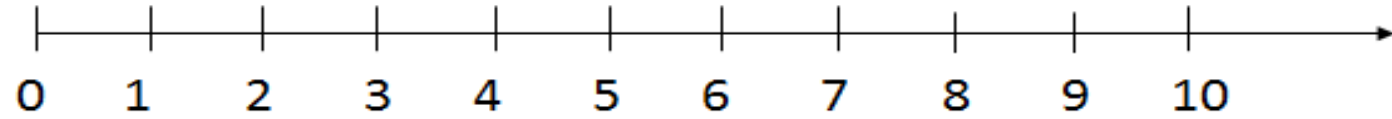
	Variable label	Variable name	Variable values*
Target variable	Trust in parliament (11-point scale)	T_TR_PARLI_11	0 = completely distrust 10 = completely trust
	Trust in parliament (distribution-preserving scale)	T_TR_PARLI_DISTRIB	0 = lowest point in distribution 100 = highest point in distribution

Missing value codes: SPSS (STATA).

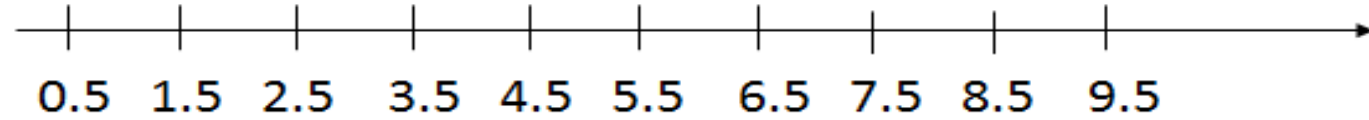
- 9 (.i) = missing data; -8 (.h) = question not asked in national survey; -7 (.g) = insufficient information for all response categories; -6 (.f) = insufficient information for single response category; -5 (.e) = variable not identified in data file; -4 (.d) = value not acceptable; -2 (.b) = not applicable;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

# Transformation of source values into the target 0-10 scale

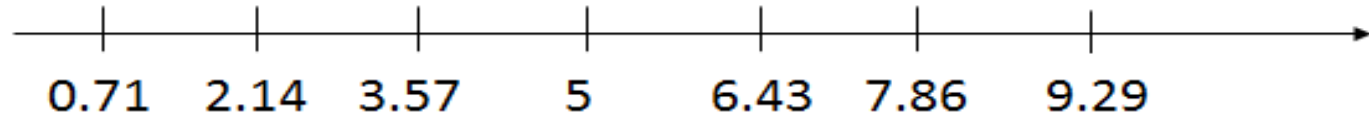
11-point



10-point



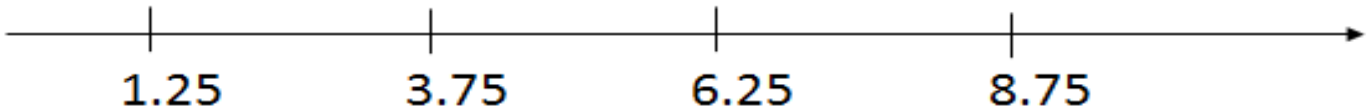
7-point



5-point



4-point



2-point





## Distributional scaling

For the source  $n$ -point scale, 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

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

## 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, L	-0.011	0.022	-0.032
Direction of original scale, D	0.050	0.050	0.021
Polarity of original scale, P	-0.012	0.023	-0.018
	Distributional scale		
Length of original scale, L	0.011	0.014	0.014
Direction of original scale, D	0.017	0.021	0.019
Polarity of original scale, P	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 strategies of using harmonization controls

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

## 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 theoretical consideration (e.g., education, EDU) are different than accumulated knowledge.

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

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

## Weights for groups distinguished according to harmonization controls

First group (standard, „the best“): 11-point scale, Ascending scale, and Unidirectional

Second group („second best“): From 4-point to 10-point scale, Ascending scale, and Unidirectional

Third group (with methodological disadvantages): All other combinations of harmonization controls

Analysis for data from 13 international survey projects, 2007-2013, N = 448,557

$$\text{TRUST} = a + b_1 \cdot \text{INTREST\_IN\_POLITICS} + b_2 \cdot \text{GENDER} + b_3 \cdot \text{AGE} + b_3 \cdot \text{EDUCATION} + b_4 \cdot \text{RURAL} + e$$

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



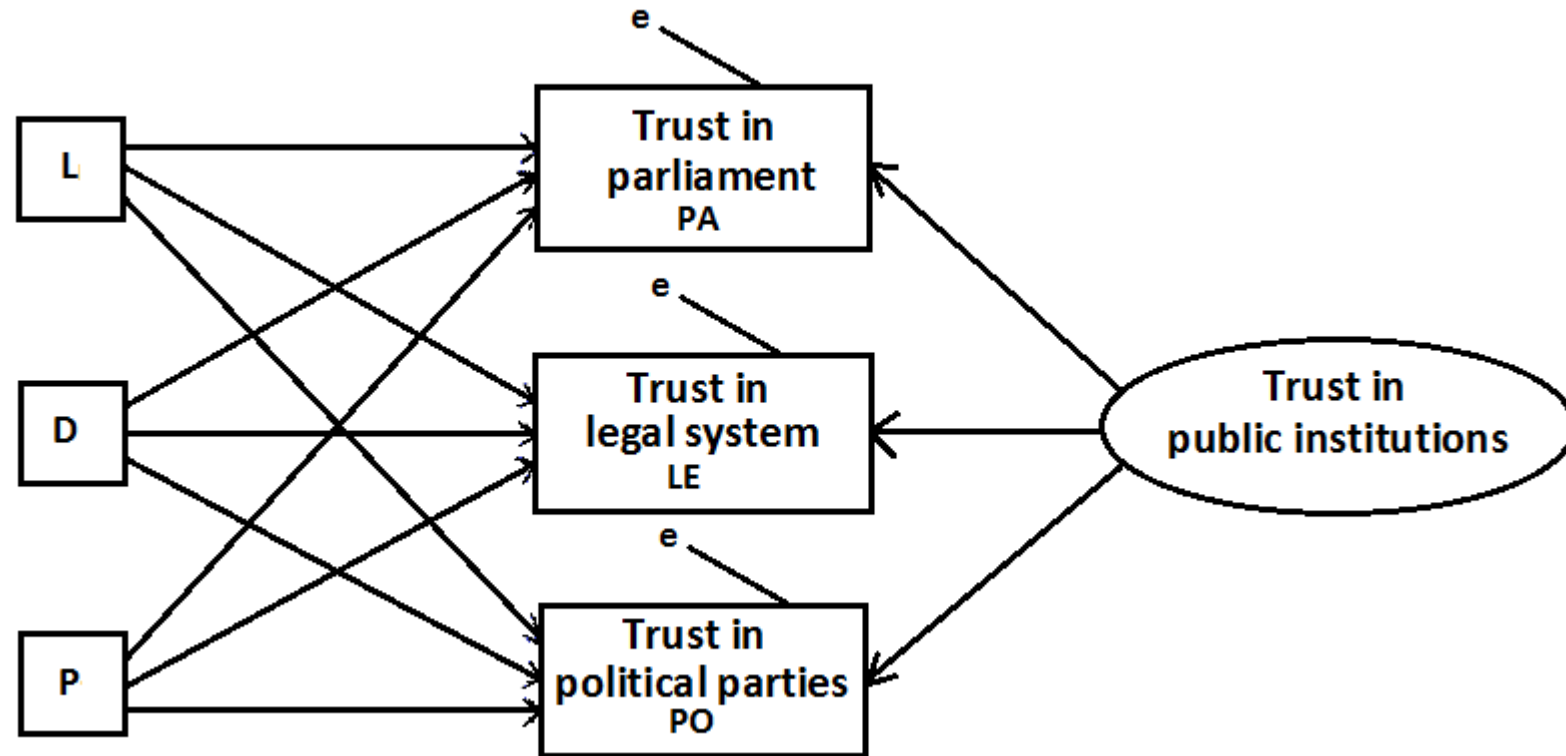
## Effects of harmonization controls

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

## Using harmonization controls in constructing latent variables

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



## Conclusions

- Harmonization process requires item-specific controls capturing inter-survey methodological variability. They should be used in substantive analyses.
- Harmonization controls can be used for
  - Selection of surveys
  - Weighting of samples
  - Accounting for effects