Cross-national Studies: Interdisciplinary Research and Training Program (CONSIRT)
The Ohio State University and the Polish Academy of Sciences

Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences

Volume 5, Issue 1
Spring/Summer 2019

Editors
Irina Tomescu-Dubrow
and
Joshua K. Dubrow
CONSIRT

consirt.osu.edu/newsletter
ISSN 2392-0858

Discovery

Welcome to the latest issue of *Harmonization: Newsletter on Survey Data Harmonization in the Social Sciences*. Discovery, as defined by the Oxford English Dictionary (online), is "an act or the process of finding somebody/something, or learning about something that was not known about before." The ever growing community of scholars, institutions, and government agencies, via intellectual debate and manifold collaborations, continue their discovery in the rich field of data harmonization.

This issue features articles and community news. The first article, by **Marco Fattore** and **Filomena Maggino,** is on major conceptual challenges to the creation of society-level indicators. Next, **Joonghyun Kwak** and **Kazimierz M. Slomczynski** present a concrete example of survey data aggregation into macro-level indicators, using cross-national data on trust in public institutions for their illustration. We then feature a report on the **GESIS Roundtable** on ex-post harmonization and announce a forthcoming conference & workshop on data harmonization at the Polish Academy of Sciences. We present news about harmonization projects: the **American Opportunity Study**, **HaSpaD** at GESIS, and **Linguistic Explorations of Societies** at the University of Goteborg. We round out the issue with news of recent publications.

As with every issue of *Harmonization*, we welcome your articles and news. Please send them to the newsletter co-editor Josh Dubrow at dubrow.2@osu.edu.

In This Issue…

The editors thank Ilona Wysmulek, Joonghyun Kwak, and Kazimierz M. Slomczynski for their assistance with this newsletter.

Articles

Synthetic Indicators for Modern Societies: Conveying Information, Preserving Complexity

by Marco Fattore and Filomena Maggino

Synthetic indicators and rankings are indeed two of the products that decision-makers most demand from socio-economic statisticians, to the extent that we may say there is an indicator (and often more than one) for just about anything: education, poverty, inequality, well-being, democracy, freedom, health, crime, and sustainability, to name a few. Notwithstanding a long tradition in indicator construction, applied statistics is currently facing major challenges that lead to reconsidering what social indicators are, how they should be designed and built, and even how they should be communicated. These issues are technical, but they are also conceptual. In this short note, we try to highlight, from the point of view of "indicator builders," the most relevant issues. To do so, we consider some critical keywords.

Complexity

Our societies are increasingly complex and somehow chaotic, i.e. affected by unpredictable dynamics. Socio-economic traits, like deprivation, well-being, or sustainability, are intrinsically multidimensional and may show complex patterns, both spatially and temporally. They cannot be any longer measured by "averages," as if their variability were due just to fluctuations around a prototypical value to be considered as representative of the entire population. The consequence is that the statistical processes used to construct indicators must be, so to say, "complexity preserving," to provide faithful representations of economies and societies, of their nuances and vagueness, of their difference, in homogeneities and divergent dynamics (Fattore et al 2012).

What is Synthesis?

The above issue leads to a fundamental question: what does it mean for an indicator to be "synthetic"? In statistical practice, synthetic indicators are usually built as aggregations of elementary attributes or variables by using various kinds of weighted means. But aggregations collapse diversity into compensative objects, losing a great deal of information on the internal structure of the data. As soon as the underlying phenomenon is truly multidimensional, any "compressed" version of the input data is intrinsically unsuitable to provide realistic, and policy-making oriented, representations of it. In a

complex setting, synthesis must be more properly understood as the capacity to "stylize" socio-economic features, i.e. to preserve the essential relations defining and characterizing them. For example, when studying deprivation, it turns out that different people may suffer from different forms of poverty and fragility, which require different policies to be implemented. Deprivation shapes cannot always be ordered in terms of intensity; they are structurally different or, in technical terms, *incomparable*. Building deprivation indicators thus requires both accounting for a "vertical" dimension (intensity) and a "horizontal" one (incomparability).

New Data, New Algorithms

The challenge of complexity emerges also under a different form: the need to exploit new kinds and sources of data. Most information on societal features is non-numerical and ordinal. This leads research towards the importation of new mathematical tools in statistical practice. A relevant example is that of discrete mathematics and, in particular, Partial Order Theory. In many cases, statistical units are described by multidimensional score profiles on a set of ordinal variables, such as found in studies on material deprivation or subjective well-being. As a result, neither can we rank units univocally, due to "incomparabilities" among them, nor can we build synthetic indicators by applying algorithms designed for numerical scores. A new data structure must be dealt with, that of a "partially ordered set," and new algorithms are currently being developed to extract information out of it for ranking purposes, such as to measure multidimensional inequality, and to evaluate multidimensional deprivation, just to mention some reference examples (Arcagni et al 2019; Fattore and Arcagni 2018; Fattore 2017; Fattore 2016)

On a different side, big data open new avenues to look at socio-economic features with levels of timeliness and spatial resolution that cannot be achieved by using standard sampling and surveys. At the same time, the use of such large amounts of data that were not designed for statistical purposes requires a lot of methodological care and the development of proper statistical procedures that are mostly yet to come. And in general, as more complex data sources are involved, more complex algorithms are applied, like neural networks, machine learning, and so on. In the effort to overcome the limitations of linear procedures such approaches may not be effective when applied to complex data structures.

The Shape of Indicators

In this setting, are synthetic indicators still sets of numbers measuring clearly defined socio-economic entities? Sometimes yes, but increasingly often they get more articulated: bi-dimensional maps, set of measures complemented by uncertainty quantifications, dashboards, etc. Broadly speaking, indicators are turning from "points" to "shapes," with the aim to condense and convey faithful information on the complexity of our societies and economies.

A Communication Issue

As long as indicators become increasingly complex, a communication issue arises. We are currently facing a "trivialization" of the public – and often political – communication about socio-economic facts. In most cases, complex issues are reduced to simple rankings or "fragmented data," and the mass media convey statistical indicators more in an infotainment spirit rather than to effectively help people in their decision-making processes and to heighten their awareness. The challenge is to condense complexity in easy-to-understand communicative objects without giving up scientific soundness and information reliability and also avoid an overload of cognitive effort for final users. Perhaps this is not a task for applied statisticians, but certainly it requires an alliance among them, visual designers, and communication experts (Maggino 2018).

References

Arcagni, Alberto, Elisa Barbiano di Belgiojoso, Marco Fattore and Stefania Rimoldi. 2019. "Analyzing deprivation and fragility patterns of migrants in Lombardy, using partially ordered sets and self-organizing maps." *Social Indicators Research* 141(2): 551–579. DOI:10.1007/s11205-018-1856-9.

Fattore, Marco and Alberto Arcagni. 2018. "F-FOD: Fuzzy First Order Dominance analysis and populations ranking over ordinal multi-indicator systems." *Social Indicators Research* 144(1): 1-29. DOI: 10.1007/s11205-018-2049-2.

Fattore, Marco. 2017. "Functionals and synthetic indicators over finite posets." Pp. 71-86 in *Partial Order Concepts in Applied Sciences*, edited by Marco Fattore and Rainer Bruggemann. Berlin: Springer.

Fattore, Marco. 2016. "Partially ordered sets and the measurement of multidimensional ordinal deprivation." *Social Indicators Research*. 128(2): 835-858. DOI:10.1007/s11205-015-1059-6.

Fattore, Marco, Filomena Maggino and Emilio Colombo. 2012. "From composite indicators to partial orders: evaluating socio-economic phenomena through ordinal data." Pp. 41-68 in *Quality of Life in Italy: Researches and Reflections*, edited by Filomena Maggino and Giampaolo Nuvolati. Berlin: Springer.

Maggino Filomena. 2018. "Dealing with Syntheses in a System of Indicators." Pp. 115-137 in *Complexity in Society: From Indicators Construction to Their Synthesis*, edited by Filomena Maggino. Berlin: Springer.

Marco Fattore is from the Department of Statistics and Quantitative Methods, University of Milano – Bicocca, Italy - marco.fattore@unimib.it and Filomena Maggino is from the Department of Statistical Sciences – University of Rome "La Sapienza" - filomena.maggino@uniroma1.it, and editor-in-chief of the journal, Social Indicators Research.

Aggregating Survey Data on the National Level for Indexing Trust in Public Institutions: On the Effects of Lagged Variables, Data-Harmonization Controls, and Data Quality

by Joonghyun Kwak and Kazimierz M. Slomczynski

The purpose of this note is to share ongoing work on aggregating data from national surveys, as carried out in the Survey Data Recycling (SDR) project.[1] By aggregate data, in the context of survey research, we understand summary statistics on the level of countries in a given time that are obtained from respondent-level observations. Such statistics are commonly used in cross-national analyses where the country-years are units of observation. Often, simple mean values are reported without paying any attention to temporal validity, methodological differences in the original questionnaires, and overall quality of survey data. We are trying to account for these issues.

After exploring various aggregate functions, we apply linear regression models to provide the best estimates of the mean value of trust in three public institutions: parliament, legal system, and political parties. We deal with the temporal reliability issue by assessing the effects of lags. Since the SDR project introduces variables that have been harmonized ex-post, we also deal with data-harmonization controls that capture methodological differences in the original, survey-specific questionnaires. In addition, to account for variation in the quality of the source surveys, we use indicators of the quality of survey documentation and of data records in the original data files, and processing error measures. Both harmonization controls and survey quality indicators are available in the SDR database.[2]

Analyses presented in this note are based on data from the SDR Masterfile version 1.1. Altogether, we analyze national surveys stemming from 18 international survey projects, for the period of 2008-2013. We chose this period for two reasons: data coverage in the SDR Masterfile, and our preliminary analyses indicating relative time-stability of institutional ratings.

Trust in Institutions: Creating Common Measures in the SDR Database

In the SDR database, items on trust in public institutions differ with respect to three formal properties of rating scales that newly constructed harmonization control variables measure:

[1] The Survey Data Recycling project is supported by the US National Science Foundation Grant (SMA-1738502), PIs: J. Craig Jenkins, Irina Tomescu-Dubrow, and Kazimierz M. Slomczynski. For a general description of the SDR approach, see Tomescu-Dubrow and Slomczynski (2016), and Slomczynski and Tomescu-Dubrow (2018).

[2] The SDR database v.1.1 and its corresponding documentation are available at Harvard Dataverse (Slomczynski, Jenkins, Tomescu-Dubrow, et al. 2017).

L – Scale length, measures the number of points/answer options in the survey-specific response scales, with values ranging from 2 to 11;

D – Scale direction, coded 1 if the scale is ascending (responses are ordered from least trust to most trust), and coded 0 if the scale is descending;

P – Scale polarity, coded 1 if the concept is measured along one dimension (from no trust to a lot of trust), and coded 0 if the concept is measured along two dimensions (distrust vs. trust).[3]

To create measures of institutional trust that are common across all national surveys (i.e. target variables harmonized ex-post), one approach we use in the SDR framework involves the following two steps (cf. chapter 3 in Slomczynski, Tomescu-Dubrow, and Jenkins, et al. 2016): First, taking into account the information contained in variables L, D, and P, we assign consecutive positive integers k ($k = 1,\ldots, n$) to each answer set in the national surveys. Here, number 1 (as we assign it) corresponds to the lowest intensity of trust, and n corresponds to the highest intensity of trust. It is important to note that each value of k indicates a different intensity, depending on n. For example, in a 2-point scale, number 2 corresponds to the answer "tend to trust," while in a 4-point scale, it corresponds to "not very much [trust]".

Next, to give values of k more comparative meaning, we perform a second step: we transform k integers into a scale that ranges from 0 (lowest intensity of trust) to 10 (highest intensity of trust), with many different scores in-between.[4] Table 1 provides the scores of the new, harmonized, scale for the different number of response options in original survey-specific trust items.

Table 1. Rescaling Rating Scales of Trust in Public Institutions into Common Scores

Original scale	Recodes	Mean of scores	Standard 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

[3] See Kołczyńska and Słomczyński (2018). A variety of formulations of institutional trust items in a number of languages is extensively discussed by Cole and Cohn (2016).

[4] The linear transformation formula is: Target scale = (10/n×2) + (k–1)×(10/n)

Basic Equation

To deal with temporal validity, we take into account the "previous" measure of the same concept. After exploring how different lengths of the lag affect the means of rating scores of subsequent indicators of trust in parliament, legal system, and political parties, we decided to rely on a one-year interval lag. We considered also the possibility that the "previous" measure would correspond to any immediately preceding survey (i.e. allowing for lags longer than one year), but this decision led to too much varying timing in terms of calendar years. Since parliament, legal system, and political parties are relatively stable pillars of democracy, one-year lag in trusting these institutions should reflect temporal reliability quite well.[5]

In our basic equation, we postulate that the dependent variable – the mean of rating scores of a given public institution for a country-year – depends on the corresponding mean obtained for a survey taken a year earlier, and on scale length (L), scale direction (D), and scale polarity (P) of original scores from which the dependent variable was constructed. Thus, we apply the following regression equations to estimate the expected country-year mean values:

$$\widehat{y_t} = a + b_1 y_{t-1} + b_2 L_t + b_3 D_t + b_4 P_t$$

where:

$\widehat{y_t}$= predicted mean value of trust in a given public institution (at the national survey level),
y_{t-1}= mean value of trust (at the national survey level) from the preceding national survey (lagged),
L, D, P – harmonization controls, defined earlier.

We run this regression for each of the three types of institutional trust measures for national surveys as units of observation. The equation stipulates that the constant mean value for all surveys (parameter *a*) is modified by: (1) a one-year lagged mean value (y_{t-1}) being governed by parameter b_1, and (2) the control variables L, D, and P, with their parameters b_2, b_3, and b_4, respectively. Table 2 shows that the lagged variable adds significant points—from 0.654 (trust in political parties) to 0.952 (trust in legal system)—to the constant.

To better understand the effect of the lagged variable, let us consider the case where the mean value of the lagged trust in parliament (y_{t-1}) is 4.8, the scale length (L) is a 5-point scale, the scale direction (D) is descending, and the scale polarity (P) is not unipolar. According to the regression results in Table 2, the estimated value of the mean trust is: 1.396 + 4.8 × 0.800 + 5 × (-0.145) + 0 × 0.622 + 0 × (-0.012) = 4.511. In this case, the impact of the lagged effect 3.840 (4.8 × 0.800) accounts for 85% of the total estimated mean ([3.840/4.511] × 100).

[5] This strong assumption will be tested in further research.

Table 2. Regression of Mean Values of Trust in Parliament, Trust in Legal System, and Trust in Political Parties on Lagged Means and Harmonization Control Variables

	Trust in parliament[a]		Trust in legal system[b]		Trust in political parties[c]	
Variables	Coefficients (standard errors)					
Constant	1.396	(0.214)	0.308	(0.259)	1.955	(0.293)
Lagged variable, Yt-1	0.800	(0.047)	0.952	(0.050)	0.654	(0.076)
Scale length, L	-0.145	(0.039)	-0.182	(0.042)	-0.202	(0.037)
Scale direction, D	0.622	(0.211)	1.069	(0.258)	0.934	(0.261)
Scale polarity, P	-0.012	(0.168)	0.553	(0.237)	0.006	(0.256)
Constant	1.396	(0.214)	0.308	(0.259)	1.955	(0.293)
Adjusted R^2	0.626		0.719		0.461	

[a] 211 surveys, 57 countries
[b] 148 surveys, 53 countries
[c] 129 surveys, 41 countries

The regression analysis shows that the scale length is negatively related to the estimated mean of trust for all three public institutions, while the scale direction has a positive effect on the estimated mean. These results are consistent with the literature, suggesting that longer scales result in lower mean values (e.g. Dawes 2008) and the ascending direction of scales tend to increase mean values (e.g. Kamoen, Holleman, van den Bergh, and Sanders 2013). Fewer studies have been conducted on scale polarity effect, with one investigation illustrating mixed results (Schuman and Presser 1981). In our analysis, scale polarity has a positive and significant effect, but only on trust in legal institutions.

Adding Variables on the Quality of Surveys

In the SDR analytic framework, we account for the quality of surveys as reflected in their documentation, processing errors, and errors or inaccuracies in computer data files.[6] Specifically, we use three indexes:

V – Data Documentation Index, a sum of binary variables describing whether the survey documentation provides information about the sampling process, response rates, control of the quality of the questionnaire translation, pretesting of the field instruments, and control of the fieldwork (Kołczyńska and Schoene 2018);

W – Processing Errors Index, a weighted sum of errors in labeling variable values (illegitimate, misleading, contradictory, discrepant values, or lack of value labels) for the following seven items: gender, age, year of birth, education level, years of schooling, trust in parliament, and participation in demonstrations (Oleksiyenko, Wysmułek, and Vangeli 2018);

S – Computer Files Quality Index, a sum of binary variables describing whether the survey contains missing case IDs, duplicated (non-unique) records, and a large amount of missing data (above 5%) on either gender or age.

[6]As a general rule, the SDR project uses the English-language documentation (technical reports, codebooks, and questionnaires) that the selected international survey projects provide.

We compare the results of our basic model presented in the previous section with more complex models, by adding:

(a) three variables at occasion *t* pertaining to the quality of surveys, that is:
V = quality control index: data documentation
W = quality control index: processing error
S = quality control index: data records

(b) the same three variables at occasion *t-1*, together with their interaction with Y_{t-1}.

Table 3 indicates that adding quality controls increases the explained variance of the survey-level mean values for each political institution.

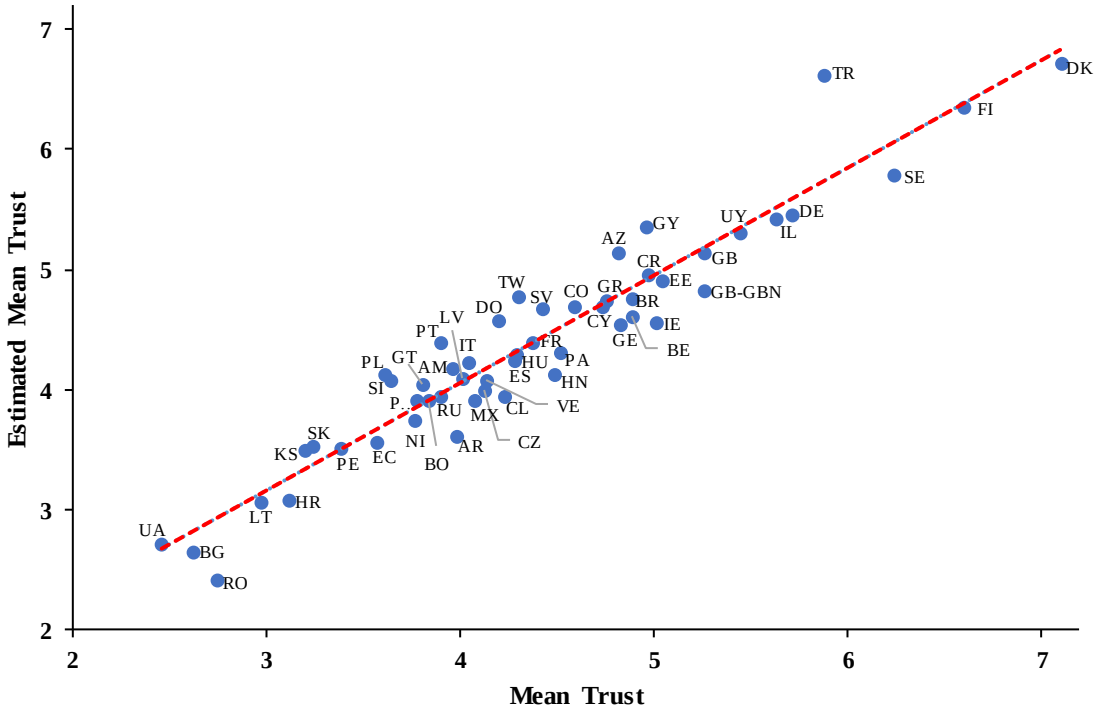
Table 3. Explained Variance by Models that Account for Source Survey Quality

Models	Trust in parliament	Trust in legal system	Trust in political parties
	Adjusted R^2		
Basic model, with variables L, D, P: Table 2	0.626	0.719	0.461
Extended model, with added variables V, W, and S for occasion *t*	0.644	0.738	0.474
Final model, with added variables V, W, and S for both occasions *t* and *t-1*, and with interactions of V, W, S for occasion t-1 with Y_{t-1}	0.684	0.781	0.523

We notice that adding the survey quality variables V, W, and S, yields a substantive increase in explained variance (above 5%). Thus, it makes sense to use estimated means of trust in public institutions that are free of the linear influence of the differences in quality of surveys.

In Figure 1 we present an example of the relationship between raw means and estimated means (derived from the final model in Table 3) for trust in legal system ($N_{countries} = 53$, $r = 0.957$). The list of countries analyses in this research note cover, including those in Figure 1, are listed in the Appendix.

Figure 1. The Association between Mean Trust and Estimated Mean Trust in the Legal System, 53 Countries, 2008-2013 ($r = 0.957$)



Let us recall here that in the SDR database, target variables for trust in public institutions also include transformations of binary source variables that we "stretched" into a common 11-point scale (see Table 1). Strictly speaking, a binary variable does not measure intensity well, and may distort the results if it is combined with scales of greater length.[7] In the 2008-2013 subset of the SDR database v.1.1, 44 surveys measure trust in parliament with a dummy, no survey does so for trust in the legal system, and 26 surveys contain dummy measures of trust in political parties.

Does it matter if the binary variable is removed from the new indices that we construct? Not particularly: for each of the institutional trust measures, the correlation between the index with the binary variable, and without it, is very strong ($r \geq 0.970$).

Bootstrapping

A main problem is the extent to which estimates of trust in public institutions obtained via linear regression, with lagged variable, harmonization controls and survey quality indexes, are accurate in terms of acceptable standard errors (for example 10% of the estimates). To look into this issue, we

[7] This was an assumption in Marta Kołczyńska's (2016) dissertation, where she used aggregate measure of trust in parliament, omitting binary responses (and still controlling for scale length).

applied bootstrapping for the estimated institutional means of trust for each country. We performed 1,000 bootstrap replications of averaging survey-level predicted means to estimate precise standard errors of the country-level estimated mean values of trust. The results for selected countries are presented in Table 4.

Table 4. Mean Values (MEAN) for Trust in Public Institutions, with Standard Errors (SE) Estimated by Bootstrapping

	Parliament		Legal system		Political parties	
	MEAN	SE	MEAN	SE	MEAN	SE
SE for min MEAN	2.332	0.279	2.646	0.097	2.639	0.258
SE for max MEAN	5.958	0.124	6.714	0.254	4.609	0.203
MEAN for min SE	4.718	0.006	4.085	0.008	3.967	0.017
MEAN for max SE	5.185	0.401	4.830	0.411	2.685	0.269
10 countries with the smallest SE (for at least one MEAN)[a]						
Austria	4.718	0.006(1)	-	-	-	-
Venezuela	4.313	0.159	4.085	0.008(2)	3.967	0.017(3)
Germany	4.221	0.186	5.456	0.031(4)	-	-
Bolivia	4.146	0.034(5)	3.913	0.091	3.105	0.164
Dominican Republic	4.812	0.039(6)	4.572	0.064(13)	3.793	0.040(7)
Guatemala	3.777	0.061(11)	4.038	0.075	3.489	0.042(8)
Turkey	5.445	0.059(9)	-	-	3.765	0.127
Colombia	4.522	0.061(10)	4.701	0.130	3.883	0.075
Chile	4.312	0.076	3.953	0.162	3.425	0.062(12)
Ecuador	3.723	0.200	3.571	0.064(14)	2.906	0.066
10 countries with the largest SE (for at least one MEAN)[b]						
Great Britain	3.820	0.217	4.830	0.411(1)	3.731	0.127
Costa Rica	5.185	0.401(2)	4.957	0.226	3.626	0.127
Poland	3.251	0.146	4.135	0.362(3)	-	-
Slovenia	3.667	0.211	4.078	0.321(4)	-	-
Spain	3.873	0.311(5)	4.241	0.104	3.181	0.158
Uruguay	5.477	0.288(6)	5.304	0.219	4.609	0.203
Romania	2.332	0.279(7)	-	-	-	-
Ukraine	2.581	0.182	2.718	0.159	2.685	0.269(8)
Croatia	2.916	0.143	3.089	0.096	2.639	0.258(9)
Italy	3.885	0.254(10)	4.231	0.204	3.563	0.127

[a] Countries are ordered according to the size of SE on either of the trust indices, from the lowest (1) to the highest (14).
[b] Countries are ordered according to the size of SE on either of the trust indices, from the highest (1) to the lowest (10).

Assuming a t-distribution of sampling means, we can compute the confidence intervals for each country. For example, 95% confidence interval around the estimated mean value of trust in parliament for Austria stretches from 4.706 to 4.731, which is obtained from the estimated mean value of 4.718 and the bootstrapped standard error of 0.006. For the maximal value of the bootstrapped standard error of the same variable (0.401), 95% confidence interval around the estimated mean value (5.185) is from 4,399 to 5.971. However, usually the bootstrapped confidence intervals do not exceed one point on the 11-point scale.

Discussion

This report shows that applying an aggregate function in the form of linear regression leads to reasonable estimates of the means of trust in public institutions, with standard errors not exceeding 10% of the estimated mean values, with two exceptions: Romania (for trust in parliament) and Ukraine (for trust in political parties). We obtained such results by estimating the country means of institutional trust as a function of the lagged mean of institutional trust, harmonization controls, survey quality controls and their interactions with the lagged mean.

Discussing the validity of estimated means for groups in the context of using them in exploratory analysis, Croon and van Veldhoven (2007: 50) note that "Such an analysis will yield stable and interpretable results only if the number of research units is sufficiently large—in any case, substantially larger than the number of explanatory variables—and the groups means show sufficient variation that is not entirely due to within-group variation." In our case these minimal criteria are fulfilled. The number of countries is well above 30, while the number of variables does not exceed 10, and inter-country variation of institutional trust is larger than within-country variation.[8] It would be interesting to compare the method that we introduced here and that we plan to develop further, with other methods suggested in the literature (Croon and van Veldhoven 2007; Becker, Breustedt, and Zuber 2018).

Our analysis is based on the assumption that Likert-type scales can be used as metric scales without major distortions in the real distribution properties of the variables involved. However, ongoing debates exist about this assumption: some researchers advocate using Likert-type scales as matric scales (Bollen and Barb 1981; Labovitz 1967; Traylor 1983), whereas others are skeptical and against this practice (Liddell and Kruschke 2018; Long 1997, Wakita, Ueshima and Noguchi 2012; Winship and Mare 1984).

We are aware that treating the Likert-type scales as metric can systematically lead to Type I and Type II errors (Liddell and Kruschke 2018) in cross-national research. In our view, the effects produced by recalibrating ordinal scales into interval metric can be assessed for specific data by examining the results between the metric and non-metric models. In particular, one can compare ordered-probit models, suggested by Bürkner and Vuorre (2018) within a Bayesian framework, with "metric" model presented in this note. Would the order of countries according to the country-level intensity of trust in political institutions be almost the same or significantly different? We plan to explore such issues as we develop this research.

[8] The one way analysis of variance provides F statistics from 5.3 to 12.8, with $p<0.001$

Appendix

List of Countries in the Analysis, 57 Countries

Argentina (AR)	Iceland (IS)
Armenia (AM)	Ireland (IE)
Austria (AT)	Israel (IL)
Azerbaijan (AZ)	Italy (IT)
Belgium (BE)	Kosovo (KS)
Bolivia (BO)	Latvia (LV)
Brazil (BR)	Lithuania (LT)
Bulgaria (BG)	Luxembourg (LU)
Chile (CL)	Malta (MT)
Colombia (CO)	Mexico (MX)
Costa Rica (CR)	Nicaragua (NI)
Croatia (HR)	Panama (PA)
Cyprus (CY)	Paraguay (PY)
Czech Republic (CZ)	Peru (PE)
Denmark (DK)	Poland (PL)
Dominican Republic (DO)	Portugal (PT)
Ecuador (EC)	Romania (RO)
El Salvador (SV)	Russian Federation (RU)
Estonia (EE)	Slovakia (SK)
Finland (FI)	Slovenia (SI)
France (FR)	Spain (ES)
Georgia (GE)	Sweden (SE)
Germany (DE)	Taiwan (TW)
Great Britain (GB-GBN)	Turkey (TR)
Greece (GR)	Ukraine (UA)
Guatemala (GT)	United Kingdom (GB)
Guyana (GY)	Uruguay (UY)
Honduras (HN)	Venezuela (VE)
Hungary (HU)	

Acknowledgements

Our initial idea of estimating mean trust in public institutions accounting for the effects of lagged variables, data quality, and data-harmonization controls, was discussed with, among others, Francesco Sarracino, Frederick Solt, Irina Tomescu-Dubrow, and Marta Kołczyńska, whom we thank for their comments. We thank Irina for substantial revisions to this note. We also acknowledge Marta Kołczyńska's contribution to identifying some relevant papers on survey data aggregation, which she did as part of her post-doc tasks in the POLINQ project (National Science Centre, Poland 2016/23/B/HS6/03916, PI: Joshua K. Dubrow). This article is also based on work supported by the US National Science Foundation (Grant No. PTE Federal award 1738502). The SDR database v 1.0 was created with funding from the National Science Centre, Poland (2012/06/M/HS6/00322).

Joonghyun Kwak is a post-doctoral scholar at The Ohio State University for the Survey Data Recycling (SDR) project, funded by the National Science Foundation.

Kazimierz M. Slomczynski, director of CONSIRT, is a Professor at the Polish Academy of Sciences and Professor Emeritus at The Ohio State University.

References

Becker, Dominik, Wiebke Breustedt, Christina Isabel Zuber. 2018. "Surpassing simple aggregation: Advanced strategies for analyzing contextual-level outcomes in multilevel models." *Methods, Data, Analyses* 12(2): 233-264.

Bollen, Kenneth A. and Kenney H. Barb. 1981. "Pearson's R and coarsely categorized measures." *American Sociological Review* 46 (2): 232-239.

Bürkner, Paul-Christian and Matti Vuorre. 2018. Ordinal regression models in psychology: A Tutorial. Research Gate: DOI: 10.31234/osf.io/x8swp.

Cole, Lindsey M. and Ellen S. Cohn. 2016. "Institutional trust across cultures: Its definitions, conceptualizations, and antecedents across Eastern and Western European Nations." Pp. 157-176 in *Interdisciplinary Perspectives on Trust: Towards Theoretical and Methodological Integration*, edited by Ellie Shockley, Tess M. S. Neal, Lisa M. PytlikZillig, and Brian H. Bornstein. Cham, Switzerland: Springer.

Croon, Marcel A., Marc J. P. M. van Veldhoven. 2007. "Predicting group-level outcome variables from variables measured at the individual level: A latent variable multilevel model." *Psychological Methods* 12(1): 45-57.

Dawes, John. 2008. "Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales." *International Journal of Market Research* 50 (1): 61–77.

Kamoen, Naomi, Bregje Holleman, Huub van den Bergh, Ted Sanders. 2013. "Positive, negative, and bipolar questions: The effect of question polarity on ratings of text readability." *Survey Research Methods* 7 (3): 181-189.

Kołczyńska, Marta. 2017. *Stratified Modernity, Protest, and Democracy in Cross-national Perspective*. PhD dissertation. Department of Sociology, The Ohio State University. Committee: Kazimierz M. Slomczyński and J. Craig Jenkins (advisors), Edward Crenshaw, and Vincent Roscigno.

Kołczyńska Marta and Matthew Schoene. 2018. "Survey data harmonization and the quality of data documentation in cross-national surveys." Pp. 963-984 in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer. Hoboken, NJ: Wiley.

Kołczyńska Marta and Kazimierz M. Slomczynski. 2018. "Item metadata as controls for ex post harmonization of international survey projects." Pp. 1011-1033 in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer. Hoboken, NJ: Wiley.

Labovitz, Stanford. 1967. "Some observations on measurement and statistics." *Social Forces* 46 (2): 151–160.

Liddell, Torrin M. and John Kruschke. 2018. "Analyzing ordinal data with metric models: What could possibly go wrong?" *Journal of Experimental Social Psychology* 79: 328–348.

Long, Scott J. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage Publications.

Munoz, Jordi. 2017. "Political trust and multilevel government." Pp. 68-88 in *Handbook on Political Trust*, edited by Sonja Zmerli and Tom W. G. van der Meer. Northampton, MA: Edward Elgar.

Oleksiyenko, Olena Ilona Wysmulek, and Anastas Vangeli. 2018. "Identification of processing errors in cross-national surveys." Pp. 985-1010 in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer. Hoboken, NJ: Wiley.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude surveys. Experiments on Question Form, Wording and Context*. London, UK: Academic Press.

Slomczynski, Kazimierz M. and Irina Tomescu-Dubrow. 2018. "Basic principles of Survey Data Recycling." Pp. 937-962 in *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop, and Brita Dorer. Hoboken, NJ: Wiley.

Slomczynski, Kazimierz M., Irina Tomescu Dubrow, J. Craig Jenkins, with Marta Kołczyńska, Przemek Powałko, Ilona Wysmułek, Olena Oleksiyenko, Marcin Zieliński, and Joshua Dubrow. 2016. *Democratic Values and Protest Behavior: Harmonization of Data from International Survey Projects*. Warsaw, PL: IFiS Publishers.

Slomczynski, Kazimierz M., J. Craig Jenkins, Irina Tomescu-Dubrow, Marta Kołczyńska, Ilona Wysmułek, Olena Oleksiyenko, Przemek Powałko, and Marcin W. Zieliński. 2017. "SDR Master Box", https://doi.org/10.7910/DVN/VWGF5Q, Harvard Dataverse, V1, UNF:6:HIWud4wueVRsU8wTN+lySg==

Tomescu-Dubrow, Irina and Kazimierz M. Slomczynski. 2016. "Harmonization of cross-national survey projects on political behavior: developing the analytic framework of survey data recycling." *International Journal of Sociology* 46 (1): 58–72.

Traylor, Mark. 1983. "Ordinal and interval scaling." *Journal of the Market Research Society* 25 (4): 297–303.

Wakita, Takafumi, Natsumi Ueshima, and Hiroyuki Noguchi. 2012."Psychological distance between categories in the Likert scale: Comparing different numbers of options." *Educational and Psychological Measurement* 72 (4): 533–546.

Winship, Christopher and Robert D. Mare. 1984. "Regression models with ordinal variables." *American Sociological Review* 49 (4): 512–525.

Conferences & Workshops

GESIS Roundtable on Ex-Post Harmonization of Rating Scales in National and International Surveys

by Ranjit K. Singh and Natalja Menold

The GESIS – Leibniz Institute for the Social Sciences hosted an expert roundtable on "Ex-post Harmonization of rating scales in national and international surveys" on May 9 and 10, 2019. Large national and international survey programs collect a wealth of data. However, they often measure the same constructs with different types of scales, which poses a challenge to ex-post harmonization projects. The roundtable addressed the issue of data comparability when different rating scales are used to measure the same concept. It also addressed the question of how comparability can be increased via ex-post harmonization, and which pitfalls should be avoided during the harmonization of rating scales.

The roundtable brought together experts on rating scales, psychometry, survey data harmonization, researchers involved in ex-post survey data harmonization, as well as representatives of several large scale survey programs, such as the German General Social Survey (ALLBUS), the European Values Study (EVS), the European Social Survey (ESS), and the German Family Panel - Panel Analysis of Intimate Relationships and Family Dynamics (pairfam).

The presentations and discussions resulted in some salient points that we summarize here. The discussions particularly emphasized the high complexity that ex-post harmonization of survey data entails. Various factors feed this complexity. Data harmonization involves many different stakeholders, on the side of data producers and data users, and often touches upon several scientific disciplines. It presupposes working with source data that often are of different quality, with quality itself being a multidimensional concept. The participants of the round table noted that data quality in ex-post harmonization projects depends on both the total survey error (resulting from errors of sampling and errors of measurement) associated with all source surveys used for harmonization, and on errors of comparability that often follow when rating scales with different properties are used to measure a given behavior, attitude, or opinion. Thus, comparability itself can be understood as a functional concept that depends on the research questions under investigation. However, comparability can be increased by increasing the measurement quality of the individual instruments and decreasing differences in their presentation formats (i.e. rating scale formats).

Concerning the ex-post harmonization of rating scales, the roundtable explored potential pitfalls of scale harmonization that can easily lead to spurious findings if not properly addressed. Valid rating scale harmonization requires, as a baseline, that all instruments under investigation measure the same concept, which is a question of ensuring validity. Next, measurement equivalence should be established, which means that measures with different rating scales should not differ in their position relative to the underlying concept, and individual response options of different rating scales have to cover comparable portions of the continuum of the underlying concept.

The roundtable also weighed the merits and limitations of different harmonization approaches, such as the discrete recoding of response options, the linear transformation of scores, distributional approaches, the usage of linking studies to link scales, and approaches that attempt to correct scale biases.[9]

The participants also discussed ways in which data producers may facilitate usage of their data for ex-post harmonization. The importance of survey documentation for data reprocessing was a common theme. Documentation needs to be easier to access and search, more comprehensive, and more reliable. There is also a pressing need for more clarity concerning the theoretical concepts underlying specific questions and variables. Lastly, ex-post harmonization would benefit greatly from the usage of validated scales and multi-item scales. We intend to continue the roundtable discussion in a special issue, which we will announce later on.

[9] By linking study we mean a study specifically conducted to make two or more scale variants comparable. In a linking study participants usually answer both variants of a scale so that answers on one scale can be "translated" into answers on the other scale and vice versa.

Building Multi-Source Databases for Comparative Analyses: International Conference and Workshop in Warsaw

In Winter 2019, from the 16th to the 20th of December, the Institute of Philosophy and Sociology, Polish Academy of Sciences, will host the international event *Building Multi-Source Databases for Comparative Analyses*. The event comprises two days of conference-style presentations, followed by a 3-day workshop.

The Conference (December 16-17) will feature presentations on survey data harmonization in the social sciences and contribute to a book that Christof Wolf (University of Mannheim, and GESIS) and the PIs of the Survey Data Recycling project (asc.ohio-state.edu/dataharmonization) are co-editing. Sociologists, political scientists, demographers, economists, and researchers in health and medicine are invited to give voice to both discipline-specific and interdisciplinary views on the challenges inherent in harmonization and how these challenges have been met.

The Workshop (December 18-20) will focus on substantive and methodological considerations that building multi-source databases for comparative analyses call for. A special session is devoted to missing data imputation. Stef van Buuren, professor of Statistical Analysis of Incomplete Data at the University of Utrecht and statistician at the Netherlands Organisation for Applied Scientific Research TNO in Leiden (stefvanbuuren.name), will deliver the lectures on missing data imputation for survey datasets with a multi-level structure, focusing on solving comparability problems by multiple imputation. The Workshop will also discuss the SDR analytic framework and issues relevant for constructing datasets stemming from the SDR database, among others.

This international event is organized jointly by the Survey Data Recycling (SDR) Project (NSF 1738502) and the project Political Voice and Economic Inequality across Nations and Time (POLINQ) (politicalinequality.org). Cross-national Studies: Interdisciplinary Research and Training program - CONSIRT.osu.edu of The Ohio State University and the Polish Academy of Sciences provides organizational support. Attendance is free of charge.

IFiS PAN hosted the 2019 CSDI International Workshop

In spring, the annual Comparative Survey Design and Implementation Workshop (CSDI, csdiworkshop.org) took place at the Institute of Philosophy and Sociology of the Polish Academy of Sciences (IFiS PAN), Warsaw, Poland (March 18-20, 2019). Funding and organizational support came from the Polish Academy of Sciences (pan.pl), IFiS (ifispan.pl), CONSIRT of the Ohio State University (OSU) and PAN (consirt.osu.edu), and CSDI.

CSDI enjoys high scientific recognition in the field of comparative survey methodology as they provide guidelines and best practices for all elements that form the lifecycle of multicultural surveys (ccsg.isr.umich.edu). CSDI annual workshops constitute a forum and platform of collaboration for scholars involved in research relevant for comparative survey methods. This year, over 45 scholars

from different disciplines and countries participated. The 2019 CSDI Program and Abstracts of presentations are available on CSDI's website.

News

"An Unassembled Panel": The American Opportunity Study

by Irina Tomescu-Dubrow and Joshua K. Dubrow

Although equality of opportunity and the promise of upward social mobility through hard work and grit are the latticework of the American Dream, what we know of the connection between social origins and social destinations in the United States needs an update. The last major mobility survey in the US was conducted in the 1970s, write Grusky and colleagues in the A*nnals of the American Academy of Political and Social Science* (2015). In the 46 years since the 1973 Featherman and Hauser study, the US has undergone a great restructuring, such as a rise in income inequality, the inclusion of many more migrants, women's mass entry into both the labor force and formerly male-dominated occupations, and the slow end of manufacturing jobs (and their rapid replacement by robots and service work). At the same time, divorce and cohabitation rates have risen, the number of people in jails has skyrocketed, and new educational forms have emerged.

Whereas there are competing claims that mobility is rising, falling, or staying the same, Grusky and colleagues argue that we lack the data - nay, the data infrastructure - to properly analyze these claims. Insufficient data have not clarified the mobility debate: "In a now-classic review," they write, "Lee and Solon (2009) concluded that available estimates on trends in intergenerational economic mobility are 'highly imprecise,' mainly because the available datasets (principally the Panel Study of Income Dynamics [PSID]) are extremely small" (66). Grusky et al (2015) argue that the PSID, like the General Social Survey (GSS), are not enough.

What we need, Grusky and colleagues argue, is a new research initiative on mobility that takes advantage of four decades of advances in social science methods, makes the most of the wealth of extant survey and non-survey data, and fulfills key requirements for sound mobility research (67): large samples that enable subgroup comparisons, an account of the new and complicated family arrangements along with the experiences of migrants and institutionalized populations, and new data on the influence of grandparents and other family members on mobility. Such an initiative should spur new theoretical and methodological models that can account for many different types of mobility (economic, political, and occupational). At the same time it would reckon with labor market turbulence, the influence of political power and civic and social engagement, and the emergence of diverse educational forms.

Grusky and his collaborators propose the American Opportunity Study (AOS) as an initiative to meet these criteria. At the core of AOS would be an innovative data infrastructure built on the US

Census, survey, and administrative data that can be assembled into a large-sample panel that contains a wide-ranging selection of intergenerational items. The data frame for the AOS initiative is represented in Figure 1 of Grusky et al. (2015, p. 69). It involves the following types of US data sources that would link to produce large-sample panels with coverage of class and stratification indicators: (i) Social Security Administration earnings records (1978-2013); Internal Revenue Service 1040 Tax data (1995-2013); Program data, such as Unemployment Insurance and Supplemental Nutrition Assistance Program; (ii) 1960-2010 censuses and 2008-2013 American Community Survey data, both of which record respondents' reported income, education, occupation, work status, family composition, etc.); and (iii) surveys with identifiers, such as the Survey of Income and Program Participation.

For AOS to function, it is necessary to link the units of observation across the different data sources and over time. To do so, Grusky et al (2015) propose the following strategy: (a) assign protected identification keys (PIK) "to the individual records in the 1960–1990 decennial censuses;" (b) use the identifiers to track the same individuals into the 2000–2010 censuses, the American Community Surveys (ACS) 2008–2013, and future census and ACS rounds (68 - 69); (c) use the same identifiers to link the census and survey data to administrative records; and (d) within AOS, create "intergenerational links between parents and children" by "drawing on existing databases that match the Social Security numbers of parents" and children (69).

Key to AOS would be on-demand capacity building: AOS panel datasets are to be assembled only for vetted research projects, will have project-specific composition (i.e. which data actually get combined), and last only for the project's duration.

Grusky and his collaborators (2015) extoll the many possible benefits of AOS, including its potential to increase the efficient use of existing data, and its long-term economic payoff. As they put it, "The AOS is... quite affordable because it exploits data that have already been collected for other purposes and adds value to those data by assembling the latent panel underlying them" (72).

They also acknowledge and discuss the main challenges that such an ambitious initiative poses, the obvious privacy risk of interlocking datasets, especially: "When the proposed project passes a stringent review, the AOS would allow the necessary linkages to then be implemented, with the resulting deidentified data passed on to the researcher only for the purpose of carrying out the pre-qualified research, presumably in Census Bureau research data centers (RDCs) or other secure venues" (69). They discuss these privacy and security concerns, including the possibility of a data breach and the public's possible discomfort with such interlocking government data (pp. 77 - 78).

Promising a continued open discussion of the merits and challenges of AOS, Grusky and colleagues (2015) conclude: "The United States has an unassembled panel that is standing unused and that, for a relatively small outlay, could be transformed into a major new infrastructural resource in the social sciences" (79). The AOS, they write, would "lead to a renaissance of labor market and mobility research" (79).[10]

[10] For more information, visit the Census webpage Data Linkage Infrastructure: American Opportunity Structure (AOS) at https://web.archive.org/web/20170524205441/https://www.census.gov/about/adrm/linkage/projects/aos.html. See also David B. Grusky, Michael Hout, Timothy M. Smeeding, and C. Matthew Snipp (2019).

Irina Tomescu-Dubrow and Johsua K. Dubrow are Professors at the Institute of Philosophy and Sociology, Polish Academy of Sciences.

References

Grusky, David B., Smeeding, Timothy M., and C. Matthew Snipp. 2015. "A New Infrastructure for Monitoring Social Mobility in the United States." *The Annals of the American Academy of Political and Social Science*, *657*(1): 63–82. https://doi.org/10.1177/0002716214549941

Grusky, David, B. Michael Hout, Timothy M. Smeeding, and C. Matthew Snipp. 2019. "The American Opportunity Study: A New Infrastructure for Monitoring Outcomes, Evaluating Policy, and Advancing Basic Social Science." *Russell Sage Foundation Journal* 5 (2): 20–39.

Lee, Chul-In, and Gary Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics* 91 (4): 766–772.

The HaSpaD (Harmonizing and Synthesizing Partnership Histories from Different Research Data Infrastructures) Project

by Anna-Carolina Haensch, Sonja Schulz, Sebastian Sterl, and Bernd Weiß

The HaSpaD (Harmonizing and Synthesizing Partnership Histories from Different Research Data Infrastructures) research project harmonizes, cumulates, and analyzes survey-based, longitudinal data compiled from relationship biographies. In doing so, the project aims to meet several goals.

First, HaSpaD aims to pool all relevant and available German surveys with information on relationship biographies, e.g., the beginning and the end of a relationship, the beginning and the end of cohabitation, or the beginning and the end of marriage. By now, nine survey projects have been included: SHARE - Survey of Health, Ageing and Retirement in Europe, The German Family Panel (pairfam), Generations and Gender Survey, ALLBUS, the German Life History Study, the German Socio-Economic Panel, the Mannheimer Scheidungsstudie (Mannheim Divorce Study), the Family and Fertility Survey, and the Familiensurvey (Family Survey). HaSpaD harmonizes and pools variables from these surveys that provide information on relationship biographies, together with selected anchor and partner variables (mainly socio-demographic variables). We intend to offer harmonization code files to the research community to allow others to create and use the harmonized dataset.

A further important goal of HaSpaD is to carry out methodological research. The datasets that HaSpaD selected for reprocessing can be characterized as heterogeneous, since they differ concerning various properties, such as data format, operationalization, or sampling. Hence, we study if and how strong these heterogeneities impact data harmonization and analyses conducted on the harmonized dataset.

Relatedly, we address methodological issues that research syntheses using non-experimental, survey-based data are likely to raise. Among others, we work on pooling survey weights for one-stage and two-stage individual participant data (IPD) meta-analysis (under review). To account for complex

sampling schemes or endogenous sampling, survey-based data often comes with survey weights ranging from design-based weights to nonresponse weights, as well as post-stratification weights. We systematically explore when and how to use survey weighting in regression-based analyses in combination with one-stage and two-stage IPD meta-analytical approaches. Building upon the work done for survey weighted regression analysis; we show through Monte Carlo simulations that endogenous sampling and heterogeneity of effects models require survey weighting to receive approximately unbiased estimates in the meta-analytical case. Even though most researchers primarily aim for approximately unbiased estimates, it is not recommended to use weights "just in case." Weights can increase the variance of meta-analytical estimates quite dramatically.

At the moment, we are also working on using Multiple Imputation (MI) to account for missing data in covariates for our discrete-time survival analyses. When pooling data, we often face the problem of systematic missingness – one or more studies did not measure one or more variables of interest. However, specifying an appropriate imputation model is especially difficult in case of systematic missingness since appropriately incorporating the heterogeneity of studies is a daunting task.

Last but not least, the HaSpaD research project is driven by substantive research goals. We are interested in identifying determinants of relationship events that shape relationship biographies, such as mate choice, or separation and divorce. The pooled dataset that HaSpaD builds enables the investigation, from a historical and life-course perspective, of previously unanswered questions with regard to relationship stability. This includes potential reasons for the increase of divorce rates over the last decades, and whether risk factors for separation have changed over time. Also, rare populations (binational couples, same-sex partnerships) can be studied in greater detail as our harmonized dataset provides a higher sample size compared to single surveys in which the number of cases is often too low.

Our project is based at GESIS Leibniz Institute for the Social Sciences (gesis.org) in Germany. We are in contact with other harmonization projects, at GESIS and beyond, and exchange experiences concerning variable harmonization and providing user-tailored harmonization code files. Since the survey-based social sciences have little experience with this kind of data pooling and research synthesis, HaSpaD intends to collaborate with other projects to develop best practices, including for aggregating survey data.

Sonja Schulz is a PostDoc at the GESIS department Data Archive for the Social Sciences (DAS). Together with Bernd Weiß, she leads the HaSpaD project. Her research interests address issues in relationship stability, intergenerational transmission of behavior, and risk behavior.

Bernd Weiß is head of the GESIS Panel, and deputy head of the GESIS department Survey Design and Methodology, and serves as co-chair of the Campbell Collaboration training group. He is currently involved in systematic reviews and meta-analyses in the areas of survey methodology, economics, and educational sciences.

Anna-Carolina Haensch and Sebastian Sterl are doctoral students working at GESIS.

Linguistic Explorations of Societies – Using Language Technology to Assist Comparative Survey Research

by Sofia Axelsson and Stefan Dahlberg

Linguistics Explorations of Societies (LES) is an interdisciplinary research program seated at the University of Gothenburg that cuts across the disciplines of political science, computational linguistics, and computer science. LES consists of the two interlinked methodological research projects *Language Effects in Surveys* and *Studying Opinions and Populations in Online Text Data*, which draw on recent developments in natural language processing (NLP) to meet the challenges facing the changing landscape of comparative survey research.

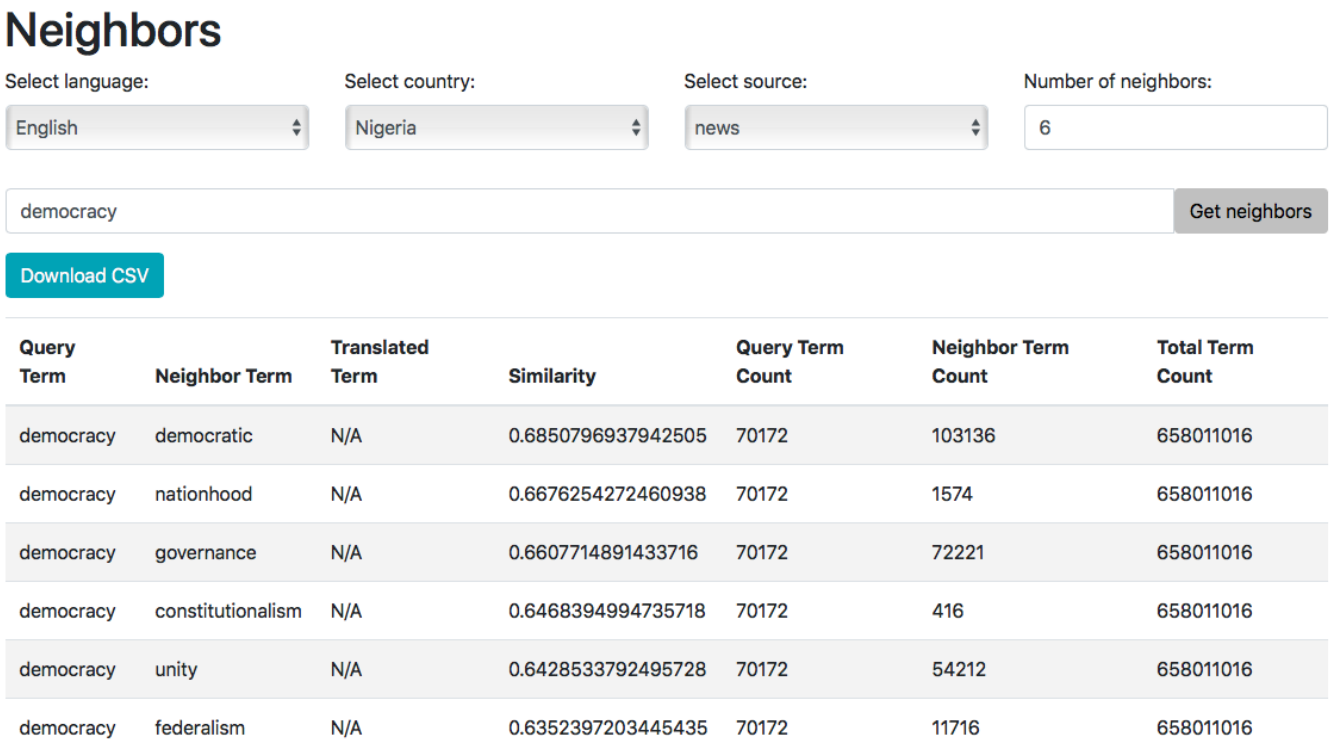
The first project, *Language Effects in Surveys,* addresses issues of survey item comparability and measurement equivalence in comparative survey research. Cross-national and/or cross-cultural survey research rests on the assumption that if survey features are kept constant to the maximum extent, data will remain comparable across languages, cultures, and countries. Yet complex political concepts are not easily defined and can invoke different meanings in different linguistic, cultural and institutional settings. The application of language technology – coupled with vast amounts of geo-coded online data – allow us to explore meaning and usage of important concepts across different languages and cultures to an extent that is unparalleled in the field of comparative survey research. Using NLP, the project aims to identify and possibly control for translational discrepancies in cross-cultural surveys that affect response patterns in survey items that are central to comparative politics and comparative public opinion.

The second project, *Studying Opinions and Populations in Online Text Data,* focuses on critical methodological issues for researchers applying online text data in a comparative survey setting. The application of Big Data technology in the social sciences has provided scholars with innovative ways and means to manage and analyze vast amounts of online text data, which potentially can be used as a complement to traditional polls and surveys. However, the use of online text data in social scientific research involves a number of methodological biases that remain unresolved. The project develops and applies NLP techniques in conjunction with large-scale survey experiments to assess the validity and reliability of online text data for the purpose of improving comparative survey research. Assessing possibilities and challenges related to data distribution and access, availability and representativity of data sources, as well as data ownership and privacy, we seek to provide a state-of-the-art account on how to use online text data in relation to existing survey data.

One of the major contributions of LES is a distributional semantic online lexicon that enables our researchers to collect and analyze relevant social scientific concepts across a large number of languages and countries. In distributional semantics, a field within NLP, semantic similarity is defined in terms of linguistic distributions. Distributional semantic models thus collect co-occurrence from large dynamic text data – often referred to as Big Data – in order to produce a multidimensional vector

space in which each word is assigned a corresponding vector. Word vectors are positioned in the vector-space such that words that share a common context are located in close proximity to one another. Put differently, the models function as statistically compiled lexica that, when probed with a target term – e.g. democracy – returns a given number of semantically similar terms – sometimes referred to as neighbour terms – including specified additional co-occurrence information (Figure 1).

Figure 1. LES Online Distributional Semantic Lexicon



Query Term	Neighbor Term	Translated Term	Similarity	Query Term Count	Neighbor Term Count	Total Term Count
democracy	democratic	N/A	0.6850796937942505	70172	103136	658011016
democracy	nationhood	N/A	0.6676254272460938	70172	1574	658011016
democracy	governance	N/A	0.6607714891433716	70172	72221	658011016
democracy	constitutionalism	N/A	0.6468394994735718	70172	416	658011016
democracy	unity	N/A	0.6428533792495728	70172	54212	658011016
democracy	federalism	N/A	0.6352397203445435	70172	11716	658011016

LES uses the neural network word2vec in order to build models that are trained on different languages. Applying the models to vast amounts of geo-coded language data from online editorial and social sources provided by different data vendors, we have built an extensive online lexicon that covers approximately 45 languages and 120 countries across the world.

Our interdisciplinary research team includes researchers from the University of Gothenburg, the Research Institutes of Sweden (RISE), Södertörn University, the University of Bergen, GESIS Leibniz Institute for the Social Sciences, and the University of Toronto. Read more about our research at les.gu.se, or contact our Research Director Stefan Dahlberg (stefan@pol.gu.se), or Research Coordinator, Sofia Axelsson (sofia.axelsson@gu.se).

Sofia Axelsson is Deputy Chief Analyst at the Department of Political Science, University of Gothenburg.

Stefan Dahlberg is Professor at the Department for Comparative Politics, University of Bergen, and the Department of Political Science, University of Gothenburg

Publications

Using Public Opinion Surveys to Evaluate Corruption in Europe: Trends in the Corruption Items of 21 International Survey Projects, 1989–2017 by Ilona Wysmułek in *Quality & Quantity*. 2019. https://doi.org/10.1007/s11135-019-00873-x

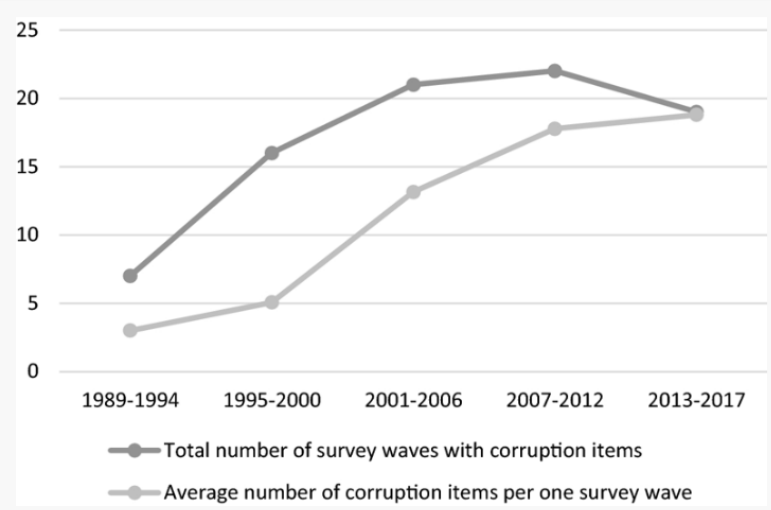
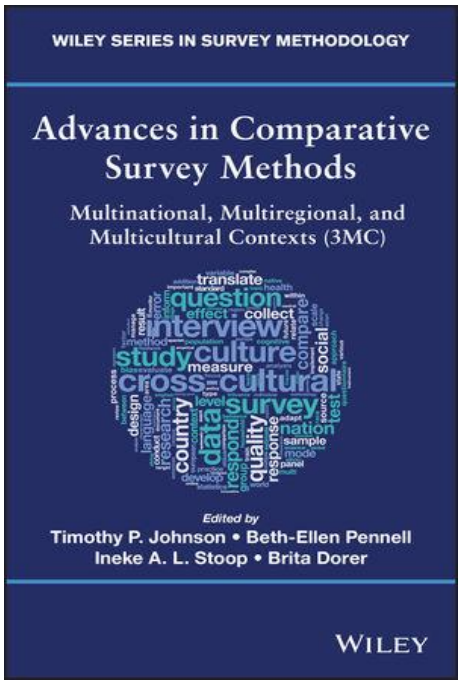


Fig. 1
The total number of survey project waves and the average number of corruption items per wave by year

"Many international survey projects contain items on corruption that facilitate comparative analyses of individual-level determinants of perceived and experienced corruption, yet such data remain under-used. To encourage more and better use of the wealth of available survey projects, this article presents a comprehensive review of the largest collection of extant cross-national data suitable for research on corruption in Europe. I examine a total of 1129 items on corruption stemming from 21 international survey projects and their 89 survey waves that cover 45 European countries during the period 1989–2017. Within three decades, the number of corruption items has grown remarkably, rising from just one in 1989 to nearly a hundred in 2017. This article shows the trends: a considerable increase in experiential items; greater differentiation between forms of corruption; a move from items on 'what government has done' to items on 'what ordinary people can do'; and inclusion of items on corruption in the private sector. Researchers interested in understanding perceptions and experiences of corruption, as they are shaped by social contexts, are offered an opportunity to explore the availability of corruption items in international survey projects in a systematic manner in order to analyze patterns of corruption, and its causes and consequences. Concluding part of the paper contains some remarks on the challenges of using survey data on corruption in a comparative framework."

The article received an honorable mention for the Harkness Student Paper Award 2019 competition of the World Association for Public Opinion Research (WAPOR). Full-text available here: https://rdcu.be/bxS7B.

Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC). 2018. Timothy P. Johnson, Beth-Ellen Pennell, Ineke A. L. Stoop, & Brita Dorer (Eds). Wiley Hoboken, New Jersey. ISBN: 978-1-118-88498-0



This volume provides expert contributions on current methodology of comparative survey research that is informative to academic survey researchers, market researchers, and students engaged in comparative projects. The book deals with essential dimensions of the survey lifecycle, resulting in rich and up-to-date discussions of issues pertaining to study design, sampling, questionnaire design, translation, data collection modes, the regulatory environment, quality assurance and control, data documentation, dissemination, and data reprocessing, among others.

Members of the Survey Data Recycling (SDR) team contributed a series of chapters devoted to methodological considerations for reprocessing cross-national survey data via ex-post harmonization, as developed in the SDR analytic framework. The following chapters deal directly with survey data harmonization:

Data Harmonization, Data Documentation, and Dissemination, by *Peter Granda*

Basic Principles of Survey Data Recycling, by *Kazimierz M. Slomczynski and Irina Tomescu-Dubrow*

Survey Data Harmonization and the Quality of Data Documentation in Cross-national Surveys, by *Marta Kołczyńska and Matthew Schoene*

Identification of Processing Errors in Cross-national Surveys, by *Olena Oleksiyenko, Ilona Wysmulek, and Anastas Vangeli*

Item Metadata as Controls for Ex Post Harmonization of International Survey Projects, by *Marta Kołczyńska and Kazimierz M. Slomczynski*

The Past, Present, and Future of Statistical Weights in International Survey Projects: Implications for Survey Data Harmonization, by *Marcin W. Zieliński, Przemek Powałko, and Marta Kołczyńska*

To see the full list of topics and the chapters that *Advances in Comparative Survey Methods* covers, please go to the book's Table of contents.

Harmonization would like to hear from you!

We created this Newsletter to share news and help build a growing community of those who are interested in harmonizing social survey data. We invite you to contribute to this Newsletter. Here's how:

1. Send us content!

Send us your announcements (100 words max.), conference and workshop summaries (500 words max.), and new publications (250 words max.) that center on survey data harmonization in the social sciences; send us your short research notes and articles (500-1000 words) on survey data harmonization in the social sciences. We are especially interested in advancing the methodology of survey data harmonization. Send it to the co-editor, Joshua K. Dubrow, dubrow.2@osu.edu.

2. Tell your colleagues!

To help build a community, this *Newsletter* is open access. We encourage you to share it in an email, blog, or social media.

Support

This newsletter is a production of Cross-national Studies: Interdisciplinary Research and Training program, of The Ohio State University (OSU) and the Polish Academy of Sciences (PAN). The catalyst for the newsletter was a cross-national survey data harmonization project financed by the Polish National Science Centre in the framework of the Harmonia grant competition (2012/06/M/HS6/00322). This newsletter is now funded, in part, by the US National Science Foundation (NSF) under the project, "Survey Data Recycling: New Analytic Framework, Integrated Database, and Tools for Cross-national Social, Behavioral and Economic Research" (SDR project - PTE Federal award 1738502). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The SDR project is a joint project of OSU and PAN. For more information, please visit asc.ohio-state.edu/dataharmonization.

Copyright Information