



Research & Methods
ISSN 1234-9224 Vol. 23 (1, 2014): 103–114
The Ohio State University
Columbus, Ohio, USA
Institute of Philosophy and Sociology
Polish Academy of Sciences, Warsaw, Poland
www.askresearchandmethods.org

Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective

Irina Tomescu-Dubrow

Cross-national Studies: Interdisciplinary Research and Training program (CONSIRT)
and Polish Academy of Sciences

Kazimierz M. Slomczynski

Cross-national Studies: Interdisciplinary Research and Training program (CONSIRT)
and The Ohio State University

This article describes the research project *Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling*. This survey data harmonization project engages with the relationship between democracy and protest behavior in comparative, cross-national perspective by proposing a theoretical model that explains variation in political protest in light of individual-level characteristics, country-level determinants, and interactions between the two types of factors. Methodologically, the project requires data with information at both the individual- and the country-level that varies over time and across space. While the social sciences have a growing wealth of survey projects, the data are often not comparable. This project selects variables from existing international surveys for *ex post* harmonization to create an integrated dataset consisting of large number of variables with individuals nested in countries and time periods. Throughout this process, focus is on three important and well-defined fields of methodology, namely data harmonization, measurement comparability, and multi-level modeling.

Acknowledgements:

This project is conducted by the Institute of Philosophy and Sociology of the Polish Academy of Sciences (IFiS PAN) and the Mershon Center for International Security Studies at The Ohio State University (MC OSU), USA with funding from the National Science Centre, Poland (grant number 2012/06/M/HS6/00322). Cross-national Studies: Interdisciplinary Research and Training program (CONSIRT consirt.osu.edu) hosts the project in Poland.

DEMOCRATIC VALUES AND PROTEST BEHAVIOR: DATA HARMONIZATION, MEASUREMENT COMPARABILITY, AND MULTI-LEVEL MODELING IN CROSS-NATIONAL PERSPECTIVE

The project *Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling* (Harmonization Project, hereafter) is driven by closely linked substantive and methodological considerations. Substantively, the project engages with the relationship between democracy and protest behavior in comparative, cross-national perspective. Political protest can be of various types, such as participation in demonstrations or signing petitions. Regarding democracy, we consider both democratic practice – usually measured by “democratic indexes” characterizing countries – and democratic values that people hold. We propose a theoretical model that explains variation in political protest in light of individual-level characteristics, country-level determinants, and interactions between the two types of factors.

The methodological implications of this approach are clear: it requires data with information at both the individual- and the country-level that varies over time and across space. While the social sciences have a growing wealth of survey projects, the data are often not comparable. Regional specialization of international surveys - such as European Social Survey, Latino Barometer, Asia Europe Survey, among others - severely hinders research pertaining to world-relevant issues. To address these types of challenges, the Harmonization Project proposes that select variables from existing international surveys be harmonized *ex post* to create an integrated dataset consisting of large number of variables with individuals nested in countries and time periods. Throughout this process, focus is on three important and well-defined fields of methodology, namely data harmonization, measurement comparability, and multi-level modeling.

Problem Statement

Current studies on protest behavior in many countries, mainly European, focus on micro-determinants, such as gender, age, education, and interest in politics. However, limiting explanatory models to individual characteristics is not justified on empirical grounds, since over the world there is “marked variation in protest across nations, with a 20:1 ratio in protest mean- scores between the highest-ranking (Sweden) and lowest-ranking (Vietnam) nations” (Dalton Sickle and Weldon, 2009:14). In recent years researchers attention to obtaining a researchers paid between micro- and macro- determinants of political participation (Benson and Rochon 2004; Kriesi 2004; Dubrow, Slomczynski and Tomescu-Dubrow 2008; Marien, Hooghe and Quintelier 2010; Vrablikova 2013).

On the methodological side, most recent research on protest is based on some particular data sets involving a limited coverage of countries and topics; various kinds of protest activities are inconsistently combined, thus producing biased measures; the statistical techniques employed in analyses are not fully satisfactory, as oftentimes they do not explicitly deal with the hierarchical structure of the data (i.e. people nested in countries and time).

The Harmonization Project develops innovative solutions in the substantive and methodological dimensions. On the theoretical side, the project proposes new hypotheses dealing with both individual and country characteristics. On methodological side, it creates comparable measurements of political protest, social values, and demographics via ex-post harmonization of selected variables from 22 international survey projects. Harmonized survey data are complemented with macro-level variables from external sources such as the World Bank, OSCE, UN agencies, Transparency International and others. Analyses involve structural equation modeling and multi-level modeling. The next section discusses these issues in more detail.

Proposed solutions

1. Developing a new theoretical model of protest behavior

The Harmonization Project proposes a multi-level model where political protest (individual-level) is explained by a set of theoretically-informed micro-level and macro-level (contextual) variables. Additionally, the model includes a cross-level interaction of micro-macro variables. The most general equation, for a given time t , is as follows:¹

$$Y_{ij} = \gamma_{00} + \gamma_{10}S_{ij} + \gamma_{01}E_j + \gamma_{11}S_{ij}E_j + u_{1j}S_{ij} + u_{0j} + e_{ij} \quad (1)$$

where Y , the dependent variable, varies among individuals (i) and countries (j). S is a vector of individual characteristics, E is a vector of macro-level characteristics, while the term SE (with appropriate subscripts) is a vector of cross-level interactions that model the varying regression slopes of individual level variables with the country level variables. Note that:

γ_{00} , γ_{10} , γ_{01} , γ_{11} are fixed coefficients/effects

and

u_{1j} , u_{0j} , e_{ij} are random coefficients/effects.

Under this specification, error terms are heteroscedastic instead of homoscedastic, as it is assumed in ordinary regression models where the residual errors are

considered independent of the values of explanatory variable. Dealing with the problem of heteroscedasticity is one of the main reasons for preferring multilevel models over the regular OLS models when analyzing hierarchical nested data (see Gelman and Hill 2007; Hox 2010).

Extending the model expressed by the general equation (1) for varying time t leads to more complicated models. Such models would involve interaction terms of the individual level variables with country-specific time-varying variables. Even more complicated models may include time-varying variables that are not country specific. In this case, the models are non-hierarchical multi-level models.

Beginning with the general equation (1), it is a rudimentary task to generate specific hypotheses related to particular sets of coefficients. A detailed presentation of the theoretical model is available on the Project's website.²

2. Creating a new, comprehensive, dataset via ex-post harmonization of well-known international survey projects

To increase geographic and temporal variability in the data, necessary for analyzing the relation between democracy and political protest in comparative perspective, the Harmonization Project selected 22 well-known international survey projects – listed in Table 1 – that span almost 50 years (1966–2013) and cover a total of 140 countries or territories (about two thirds of all states).³ The Harmonization Project includes information on more countries than any single survey project, yet noticeable gaps remain, especially in Africa (including whole Central Africa) and Oceania. Also, coverage is uneven. Fifty-seven countries, mostly African, South and South-East Asian ones, are surveyed up to five times. The rest of the world is fairly well represented, with especially high numbers of surveys in Europe and the Americas.

The international survey projects meet the following criteria: (1) they are non-commercial (mainly academic), (2) designed as cross-national (and preferably multi-wave) enterprises; (3) the samples are intended as representative of the entire adult population of a given country or territory; (4) they contain questions of substantive interest to the project, that is, about political attitudes and protest behaviors; (5) they are freely available in the public domain; and (6) sufficient documentation (study description, codebook and/or questionnaire) is provided in English.

In line with its substantive focus, the Harmonization Project selected 89 waves (i.e. project*year) from the 22 survey projects and pooled them into a relational database, which stores data in structured objects called tables (see Powalko 2014:5). This virtual database contains 81 data files with 1720 project*wave*countries (i.e. national samples in all surveys carried out in all waves and in all projects) and

a total of almost 2.3 million respondents.⁴ A single data file may contain from one country in one wave to many countries in many waves, hence the difference between number of waves (89) and the number of data files (81). The platform for data files of national surveys is organized such that in the future, any variable could be extracted and moved to the virtual integrated dataset.

Table 1 Selected International Survey Projects.

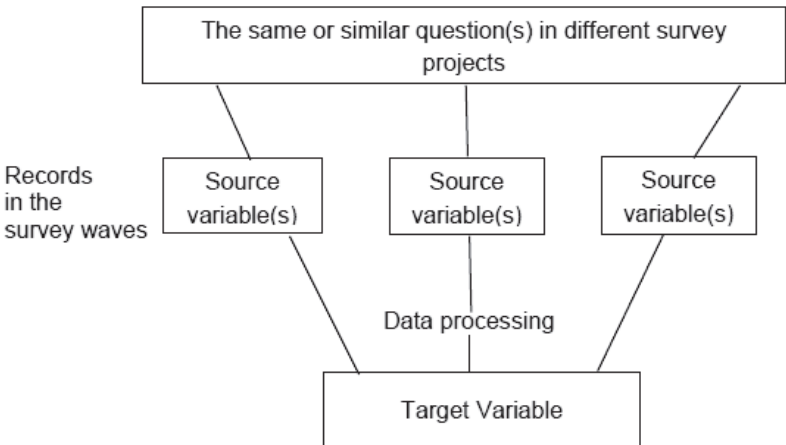
Abbrev.	Survey Project	Time span	Waves	Files	Data Sets	Cases
		Numbers				
AFB	Afrobarometer	1999-2009	4	4	66	98942
AMB	Americas Barometer	2004-2012	5	1	92	151341
ARB	Arab Barometer	2006-2011	2	2	16	19684
ASB	Asian Barometer	2001-2011	3	3	30	43691
ASES	Asia Europe Survey	2000	1	1	18	18253
CB	Caucasus Barometer	2009-2012	4	4	12	24621
CDCEE	Consolidation of Democracy in Central & Eastern Europe	1990-2001	2	1	27	28926
CNEP	Comparative National Elections Project	2004-2006	1	8	8	13978
EB	Eurobarometer	1983-2012	7	7	152	138753
EQLS	European Quality of Life Survey	2003-2012	3	1	93	105527
ESS	European Social Survey	2002-2013	6	2	147	283510
EVS/WVS	European Values Study / World Values Survey	1981-2009	9	1	310	423084
ISJP	International Social Justice Project	1991-1996	2	1	21	25805
ISSP	International Social Survey Programme	1985-2013	13	13	363	493243
LB	Latinobarometro	1995-2010	15	15	260	294965
LITS	Life in Transition Survey	2006-2010	2	2	64	67866
NBB	New Baltic Barometer	1993-2004	6	1	18	21601
PA2	Political Action II	1979-1981	1	1	3	6682
PA8NS	Political Action - An Eight Nation Study	1973-1976	1	1	8	12588
PPE7N	Political Participation and Equality in Seven Nations	1966-1971	1	7	7	16522
VPCPCE	Values and Political Change in Postcommunist Europe	1993	1	5	5	5769
Total		1966-2013	89	81	1720	2295351

In the combined dataset, the average number of cases per national sample is 1,332 respondents, the average number of variables per national sample is 265, and the size of the matrix for all national surveys if each sample is treated separately equals $2,291,040 \times 265 = 607,125,600$. To handle this large amount of data, as well as the variety of data formats deployed in so many sources, the project is developing in-house tools for extracting, transforming and loading the data into a common (relational) database. For a detailed discussion see Powalko 2014.

Ex-post harmonization of selected variables from the existing survey projects

The Harmonization Project identifies relevant variables that appear in at least five of the original survey waves and uses various data processing procedures to produce, in the relational database, a common variable with a unified measurement scheme. The literature refers to this process of integrating data from existing surveys as ex-post harmonization, to the original variables in the datasets of particular surveys as source variables, and to the harmonized, common, variable produced from the source variables as target variables (Gunther 2003; Ehling et al 2006; Granda and Blasczyk 2010, Granda, Wolf and Hadron 2010,). Figure 1 depicts, in a simplified way, the relation of source variables to the target variable.

Figure 1 Relationship between Source and Target Variables



The specific data processing procedures and harmonization rules developed in the Harmonization Project form a complex, labor-intensive and multistage process that is still in progress. While their discussion is out of the scope of this article, we will comment on main elements in our approach.

In all survey projects, the units of observations are individuals. We select two types of source variables for ex-post harmonization: technical variables, provided by survey administrators and variables of substantive interest, such as types of political protest (e.g. participation in demonstrations), individuals socio-demographics (age, gender, education, urban/rural), measures of social attitudes (e.g. trust in parliament) and measures of democratic values (e.g. support for rule of law). As indicated earlier, the list of variables is not closed, thanks to the flexible set-up of the programming environment we are using.

We have created the following theoretically-informed target variables: gender, year of birth, age, participation in demonstrations and trust in parliament. Currently, the “new” technical variables in the Harmonization Project are (1) project ID; (2) Country ID, ISO-1, ISO-2, ISO-3; (3) Wave ID; (4) Year of the study; (5) Country’s administrative unit (6) Design and post-stratification weights; (7) Respondent ID within a survey and for all surveys..

Creating quality control variables for target variables represents a very important step in in the harmonization process, which aims at dealing with different types of errors. In our approach, the target variable T is considered a function of the source variables, S, and – in addition – three types of control variables, Q, V and R:

$$T = b_0 + b_1 Q + b_2 V + b_3 R + e \quad (2)$$

where

Q stands for Data Quality Controls (divergence from the survey standards)

V stands for Validity Controls (e.g. inter-survey differences in wording of the questionnaire items)

R stands for Reliability Controls (measurement inconsistency)

If $b_1, b_2, b_3 > 0$ some intervention is needed. A possible solution, which researchers in the Harmonization Project plan to assess empirically, is to partial out (control) the effects of Q, V, and R in statistical analyses.

Table 2 presents an example of the target variable participation in demonstrations and corresponding quality control variables. We constructed the latter from questions on participation in demonstration in the source surveys on (1) time frame of questions about demonstrations, (2) number of items, (3) legality of demonstrations, and (4) scope of action.

Table 2 Example of Quality-Control Variables for specific Target Variable

This example deals with *source variables on participating in demonstrations*. The *target variable* is a dummy showing whether the respondent participated in demonstration (yes = 1) or not (no = 0). The following quality-control variables were created:

1. Time frame as appears in the question*: 1 – one year or less, 3, 5, 7, 10, and 11 - years, and 12 – ever.
 2. How many items were used to create the target variable: 0 – one, 1 – more than one
 3. Does the question indicate participation in “illegal” demonstrations: 1 –yes, 0 – no
 4. Does the question include other (additional) words than demonstration (such as marches, boycotting, etc): 1 – yes, 0 - no
-

*Questions are of the type: Did you participate in a public demonstration in last six month?
Have you ever participated in street demonstration or political marches?

In addition, the Harmonization Project will examine the proportion of missing data on the items dealing with participation in demonstrations, as well as the correlation of the target variable with selected criteria variables. Results of these analyses may lead to constructing new variables pertaining to the reliability of the target variable.

3. Adding contextual variables to the harmonized dataset

The Harmonization Project is in the process of selecting, from official statistics and external sources such as Transparency International, the World Bank, OSCE, UN agencies, contextual variables that are relevant to the theoretical model of political protest. Especially relevant are level of democracy, economic development, and corruption. These macro-level variables will be added to the harmonized dataset and used in substantive analyses.

4. Implications for statistical analyses

There are two main types of statistical problems this project raises. The first type is related to constructing variables through ex-post harmonization and assessing their validity and reliability. To solve this type of problems we apply various statistical tools, mainly derived from confirmatory factor analysis and imputation of data with incomplete information. The second type of problems deals with causal analysis. The integrated data set has a complex structure: some variables vary among individuals, some variables vary only between countries, and some variables vary across time. In this case we will rely on hierarchical and non-hierarchical multi-level modeling. In some analyses we rely on multilevel structural equation modeling (Preacher, Zhang and Zyphur 2011).

In terms of software, for constructing the integrated data file we use Structured Query Language (SQL), a database administration and query tool that provides a single consistent interface for various databases (Powalko 2014). For solving statistical problems mainly Stata and R are used.

Methodological Contribution of the Harmonization Project

Data management of major international surveys in combination with other non-survey data will lead to increased knowledge about qualities of these surveys as well as about techniques of harmonizing data from different sources. We will also work with missing data imputation techniques to account for possible deficiencies in the data. Generally, this would require advanced statistical methods. Table 3 provides an overview of the methodological contribution of the Harmonization project.

Table 3 Methodological Contribution of the Harmonization Project

	Current situation	Research content	Domains of advancing knowledge
Harmonization	Existing international datasets are not combined even if this would be possible for the same or similar indicators; combining survey data with official statistics data and other non-survey data is rarely performed	Data management of major international surveys in combination with official state statistics and other non-survey data	Knowledge of main international datasets archived in ICPSR, GESIS, and other institutions; knowledge of techniques of harmonizing data and combining datasets
Comparability	The problems of equivalence in national and international data on democratic values and protest behavior are not solved	Quantitative methods of establishing functional equivalence of indicators and constructs; regression data imputation	Practical knowledge of psychometric criteria of validity and reliability applying them to survey data; data imputation with incomplete information; functional equivalence analyses
Quantitative methods	Quantitative methods are rarely used for harmonized data; combining structural equation modeling (SEM) with hierarchical linear modeling (HLM) requires new efficient solutions	Structural equation modeling (SEM) combined with hierarchical linear modeling (HLM) applied to harmonized data	Applicable knowledge of Stata and appropriate routines in the environments of R for combining structural equation modeling (SEM) with hierarchical linear modeling (HLM)

CONCLUSIONS

The Harmonization Project is work in progress. As our work unfolds, it prompts us to reconsider how exiting survey data can best be used, in light of possibilities for standardization and ex-post harmonization, but also with regards to accounting for various methodological problems in the existing surveys. Regarding the latter, the Harmonization Project opens up three lines of research related to the source survey projects: on the general survey documentation, on the specific data description, and on the computer data files. Lack or inadequate information in documentation reduces confidence in the data, inconsistencies between the resources defining the meaning of variables and their values, and records on the computer data file (i.e. data description) decrease interpretability of the data, while errors in data files lead to possible distortion of the empirical results based on the data. The Harmonization Project suggests that quality-control variables for each of these three aspects are necessary; their relevance has to be checked in empirical analyses. A detailed discussion of survey data harmonization and the quality of data documentation in cross-national surveys is available in Schoene and Kolczynska (2014).

Together with harmonization, the option of constructing quality control variables to be included in substantive analyses ties into our view that data from different sources and of varying quality may be combined into a full-value product for researchers to use. While there are many methodological challenges to be addressed, the benefits of increasing the simultaneous use of existing international survey projects for comparative analyses are worth the effort.

NOTES

- 1 We assume that most of these variables could be operationalized in different ways.
- 2 See <http://dataharmonization.files.wordpress.com/2013/12/grant-proposal-democratic-values-and-protest-behavior-2012.pdf>
- 3 We refer to the selected projects as well-known on the basis of publication records and the impact that they have on the social-science disciplines. For practical reasons, we stopped adding new data in the second quarter of 2014.
- 4 Because of the thematic coverage criterion, we include only survey waves that contain relevant questions on protest behavior and/or democratic values; hence not all waves of ISSP, EB and CNEP are in our data.

REFERENCES

- Benson, M. and Rochon, T. R 2004. "Interpersonal Trust and the Magnitude of Protest: A Micro and Macro Level Approach." *Comparative Political Studies* 37 (4): 435-457.
- Dalton, R. J., A. van Sicle and S. Weldon. 2009. "The Individual-Institutional Nexus of Protest Behaviour." *British Journal of Political Science* 40 (1): 51-73.

- Dubrow, J. K., K. M. Slomczynski and I. Tomescu-Dubrow. 2008. "Effects of Democracy and Inequality on Soft Political-Protest in Europe: Exploring the European Social Survey Data." *International Journal of Sociology* 38 (3): 36 – 51.
- Ehling, M. / Rendtel, U. et al. 2006. Synopsis. Research Results of Chintex - Summary and Conclusions. Last accessed November 28 2014, at www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/Synopsis.html
- Gelman, A. and J. Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
- Granda P and E. Blasczyk 2010 Corss-cultural Survey Guidelines. XIII. Data Harmonization. Last accessed November 28, 2014 at <http://ccsg.isr.umich.edu/harmonization.cfm>
- Granda P., C. Wolf, and R Hadorn, R. 2010. Harmonizing survey sata. In J. A. Harkness, M. Braun, B. Edwards, T. Johnson, L. E. Lyberg, P. Ph. Mohler, B-E., Pennell & T. W. Smith (Eds.), *Survey methods in multinational, multicultural and multiregional contexts* (pp. 315-332). Hoboken, NJ: John Wiley & Sons.
- Gunther, R. (2003). Report on compiled information of the change from input harmonization to ex-post harmonization in national samples of the European Community Household Panel – Implications on data quality (Working Paper #19). Last accessed November 28, 2014 at https://www.destatis.de/DE/Methoden/Methodenpapiere/Chintex/ResearchResults/Downloads/WorkingPaper19.pdf?__blob=publicationFile
- Hox, J. J. 2010. *Multilevel Analysis. Techniques and Applications*. New York: Routledge.
- Kriesi, H. 2004. "Political Context and Opportunity." Pp. 67-90 in D. Snow, S. Soule and H. Kriesi, eds. *Blackwell Companion to Social Movements*. Malden, Oxford Carlton: Blackwell.
- Marien, S., M. Hooghe, M. and E. Quintelier. 2010. "Inequalities in Non-Institutionalized Forms of Political Participation: A Multi-level Analysis of 25 Countries." *Political Studies* 58 (1): 187-213.
- Powałko, Przemek. 2014. "Working with Big Data: Experiences with the Cross-national Survey Data Harmonization Project." *CONSIRT Working Papers Series 2* at consirt.osu.edu.
- Preacher, K. J., Z. Zhang & M. J. Zyphur. 2011. Alternative Methods for Assessing Mediation in Multilevel Data: The Advantages of Multilevel SEM.. *Structural Equation Modeling: A Multidisciplinary Journal* 18 (2): 161-182
- Schoene, Matthew and Marta Kolczynska. 2014. "Survey Data Harmonization and the Quality of Data Documentation in Cross-National Surveys." *CONSIRT Working Papers Series 3* at consirt.osu.edu.
- Vrablikova, K. 2013. "How Context Matters? Mobilization, Political Opportunity Structure and Nonelection Political Participation in Old and New Democracies." *Comparativ;e Political Studies*.

Irina Tomescu-Dubrow received her PhD from The Ohio State University. She is Associate Professor at the Institute of Philosophy and Sociology, Polish Academy of Sciences (PAN) and Program Manager for *Cross-National Studies: Interdisciplinary Research and Training* program – CONSIRT of the Ohio State University and PAN. Her work includes research in social stratification, structural change, and cross-national methodology.

E-mail: tomescu.1@osu.edu

Kazimierz M. Slomczynski is Professor Dr. Hab. at the Institute of Philosophy and Sociology, Polish Academy of Sciences (PAN), and Emeritus Professor, Department of Sociology, The Ohio State University. He is the Director of *Cross-National Studies: Interdisciplinary Research and Training* program – CONSIRT of the Ohio State University and PAN. Professor Slomczynski directs the Polish Panel Study 1988-2013 (POLPAN.org), a unique longitudinal study on the transformation of the Polish stratification system after 1989.

E-mail: Slomczynski.1@osu.edu