Obstacles in conducting valuable analysis based on data from multiple sources. A case study of methodological and technical challenges in multilevel cross-national analyses on attitudes toward the environment
--- distributed paper ---

Piotr Jabkowski
Faculty of Sociology; Adam Mickiewicz University, Poznan, Poland

Abstract:
Ex-post data harmonization usually involves a wide array of procedures that integrate diverse survey data sets into one meta-base and promises to incorporate information mitigating time-series or country-coverage lapses. We hoped for similar benefits when merging data from two surveys of established quality with other survey data and administrative records.

Methods & Data: Our presentation is based on data from the 2017 wave of the European Values Study (EVS). When conducting multilevel cross-national regression analyses focused on the impact of the ideological orientation of political parties close to the respondents on attitudes toward environmental protection, we realized that the necessity to include measures corresponding to the level of respondents, political parties, and countries, demands going beyond data provided by the EVS. Thus, we incorporated two external sources of data: (1) the Chapel Hill 2019 Expert Survey (CHES) describing political parties and (2) World Bank Data on cross-country differences in the GDP per capita.

Results: We faced many obstacles from the specificity of external data we tried to merge with the EVS data. These challenges included not only technical limitations but also methodological complications resulting from the different quality of multiple sources of external data and its inconsistency with survey data, e.g., resulting from under- or over-coverage of political parties included in the EVS 2017 and CHES 2019, and time gaps in the Word Bank Data.

Lesson learned: Our analyses demonstrate that even though we failed to produce data of supreme quality and full comparability, merging different data sources provides promise for going much deeper through analyses based on multiple different sources rather than relying on data from a single cross-national project designed to be comparable.

Keywords*:*
multiple sources of data, data integration, cross-country analyses, attitudes toward the environment

Introduction

Ex-post data harmonization, a process that aims to integrate disparate survey data sets into a unified meta-base, offers a comprehensive solution to address the challenges arising from time-series gaps and country-coverage limitations (Dubrow & Tomescu-Dubrow, 2016; Granda et al., 2010). By merging data from two surveys known for their established quality and administrative records, we anticipated reaping similar benefits and achieving a more robust and comprehensive dataset in this study.

Ex-post data harmonization involves a wide range of procedures that require careful consideration and meticulous execution (Cichocki & Jabkowski, 2022; Singh, 2021; Tomescu-Dubrow & Slomczynski, 2014). First and foremost, integrating different survey data sets necessitates establishing a common framework and standardized variables. This step ensures that the data from various sources can be easily merged and analyzed cohesively while minimizing any discrepancies or inconsistencies between the datasets (Singh, 2020).

Furthermore, ex-post data harmonization also entails addressing potential issues such as missing data, varying data collection methodologies, and differences in data quality (Jabkowski et al., 2021; Jabkowski & Kołczyńska, 2020; Kołczyńska & Schoene, 2018). Imputation techniques, statistical adjustments, and data validation procedures are employed to enhance the accuracy and reliability of the merged dataset (Kołczyńska & Slomczynski, 2018; Slomczynski et al., 2021). These measures improve the overall data quality and enable the incorporation of valuable information that may have been missed due to time-series gaps or limited country coverage.

By merging data from two surveys of established quality with other survey data and administrative records, we aimed to unlock new insights and enrich our understanding of the phenomena under study. The combination of diverse data sources allows for a more comprehensive exploration of complex relationships, patterns, and trends that might not have been discernible when analyzing individual datasets in isolation.

Moreover, the augmented dataset resulting from ex-post data harmonization has the potential to support more robust statistical analyses, facilitate cross-country comparisons, and provide a solid foundation for evidence-based policymaking. Researchers and policymakers can leverage this harmonized dataset to gain deeper insights into socioeconomic trends, make informed decisions, and design targeted interventions to address pressing societal challenges.

Data and Methods

Our analysis is built upon the 2017 wave of the European Values Study (EVS, 2022), which serves as the primary data source for our research. The main aim was to assess the individual-level covariates of people's attitudes toward environmental protection. However, as we conducted multilevel cross-national regression analyses to examine the impact of the ideological orientation of political parties on attitudes toward environmental protection, we recognized the need to expand our dataset beyond the confines of the EVS. To achieve this, we integrated two additional external sources of data: the Chapel Hill 2019 Expert Survey (Bakker et al., 2020; Jolly et al., 2022) and the World Bank Data on cross-country differences in GDP per capita.

Incorporating the Chapel Hill 2019 Expert Survey (CHES) into our analysis gave us a comprehensive description of political parties. This dataset offers valuable insights into the ideological positioning of political parties, allowing us to capture a more nuanced understanding of their orientations and the potential influence they exert on environmental attitudes. By merging the CHES data with the EVS dataset, we were able to bridge the gap

between party-level information and individual-level attitudes, enabling us to examine the relationship between the two more effectively.

Additionally, we recognized the importance of accounting for cross-country economic differences when analyzing attitudes toward environmental protection. To address this, we turned to the World Bank Data on GDP per capita, which offers a reliable measure of a country's economic status. By integrating this data into our analysis, we gained a macro-level perspective to explore how economic factors may interact with party ideologies to shape individuals' environmental attitudes. This integration further enhanced the robustness and comprehensiveness of our analysis, as we could examine the influence of both micro- and macro-level factors simultaneously.

Results

One of the main challenges we encountered was the varying quality of the external data sources. Each source had its own methodology and data collection process, which resulted in differences in data quality. These disparities posed challenges when attempting to harmonize the data and ensure its compatibility with the EVS dataset. We had to carefully evaluate and address these discrepancies through rigorous data cleaning and validation procedures to minimize the potential biases that could arise from such variations.

Moreover, issues related to the coverage of political parties included in the EVS 2017 and CHES 2019 datasets also emerged as a significant hurdle. The under or over-coverage of certain political parties in either dataset create imbalances and gaps in the merged dataset. We had to employ imputation techniques and statistical adjustments to account for these gaps and mitigate any potential biases arising from incomplete representation of political parties. This meticulous process was crucial to ensure the integrity and representativeness of the final merged dataset.

Table 1. A fraction of EVS (2017) respondents excluded after merging EVS (2017) and CHES (2019) data (due to the inconsistency of the party list in both data sets)

Country	The fraction of EVS respondents excluded from analyses
Albania	1.7
Austria	1.9
Bosnia & Herzegovina	4.9
Bulgaria	6.8
Croatia	7.7
Czechia	6.9
Denmark	0.0
Estonia	6.3
Finland	0.0
France	9.8
Germany	0.0
Hungary	7.0
Iceland	4.9
Italy	8.5
Lithuania	4.7
Montenegro	10.1
Netherlands	0.0

Country	The fraction of EVS respondents excluded from analyses
North Macedonia	8.9
Norway	1.8
Poland	0.0
Portugal	3.5
Romania	1.4
Serbia	8.9
Slovakia	4.6
Slovenia	7.3
Spain	1.8
Sweden	0.0
Switzerland	4.8
United Kingdom	0.0

Table 2. List of EVS parties excluded from analysis after merging EVS and CHER data (due to inconsistency of party list in both projects)

Country	EVS data: name of the party	Number of EVS respondents
Albania	AL: Communist Party of Albania	3
Albania	AL: Equal List, LIBRA	3
Albania	AL: Environmental Party	2
Albania	AL: The Human Rights Union Party	1
Albania	AL: Republican Party	1
Albania	AL: Movement of Legality Party	1
Albania	AL: National Front Party	1
Albania	AL: New Democrat Party	1
Albania	AL: Agrarian Environmentalist Party	1
Albania	AL: The People's Alliance Party	1
Albania	AL: Green Party	1
Austria	AT: List Peter Pilz	24
Bosnia & Herzegovina	BA: People and Justice	16
Bosnia & Herzegovina	BA: Independent Bloc	16
Bosnia & Herzegovina	BA: Bosnian Party	13
Bosnia & Herzegovina	BA: Croatian republican party	2
Bosnia & Herzegovina	BA: Croatian Party of Rights	2
Bulgaria	BG: Union of Democratic Forces	20
Bulgaria	BG: Party The Greens	17
Bulgaria	BG: Alternative for Bulgarian Revival	12
Bulgaria	BG: Movement Bulgaria of the Citizens	12
Bulgaria	BG: Movement 21	5
Bulgaria	BG: DOST Democrats for Responsibility, Freedom, Tolerance	4
Bulgaria	BG: Yes, Bulgaria	2
Croatia	HR: Smart	42
Croatia	HR: Croatian party of rights	25
Croatia	HR: Gorski-Kotar alliance	7
Czechia	CZ: SZ (Green Party)	46
Czechia	CZ: Svobodní (Party of Free Citizens)	20
Czechia	CZ: Realisté (Realists)	9
Czechia	CZ: DSSS (Worker's Party of Social Justice)	8
Czechia	CZ: Úsvit-NK (Dawn- National Coalition)	5
Czechia	CZ: SPO (Party for the Rights of Citizens)	5
Estonia	EE: Estonian Greens	33

Country	EVS data: name of the party	Number of EVS respondents
Estonia	EE: Estonian Free Party	20
France	FR: Left Wing Extremist Parties (New Anticapitalist Party, Workers' Struggle, Independent Workers' Party)	31
France	FR: Other Environmentalist Parties	28
France	FR: Union of Democrats and Independents	26
France	FR: Other Left Wing Parties (Radical Leftist Party, Republicain's and Citizen's Movement	25
France	FR: Act – The constructive right	6
France	FR: Other Right Wing Extremist Parties (The Patriots, National Republican Movement)	6
Hungary	HU: Hungarian Two-tailed Dog Party (MKKP)	59
Hungary	HU: Dialogue for Hungary (PM)	6
Hungary	HU: Hungarian Workers' Party (Munkáspárt)	4
Hungary	HU: Hungarian Liberal Party (Liberálisok)	1
Iceland	IS: Bright Future	38
Iceland	IS: The Icelandic National Front	10
Iceland	IS: The Humanist Party	9
Iceland	IS: Dawn - The organization of justice, fairness and democracy	6
Iceland	IS: The People's Front of Iceland	3
Italy	IT: Communist Party	53
Italy	IT: Power to the People!	17
Italy	IT: Us with Italy - UDC (Union of the Centre)	14
Italy	IT: Italy Europe Together	13
Italy	IT: The People of the Family	6
Italy	IT: Italy to the Italians	4
Italy	IT: CasaPound Italy	3
Italy	IT: Popular Civic Lorenzin	1
Lithuania	LT: Lithuanian Freedom Union (Liberals)	42
Montenegro	ME: True Montenegro	18
Montenegro	ME: Liberal party of Montenegro	10
Montenegro	ME: Albanian Alternative	9
Montenegro	ME: Democratic union of of Albanians	8
Montenegro	ME: Montenegrin (party)	2
Montenegro	ME: Positive Montenegro	1
Montenegro	ME: United Montenegro	1
North Macedonia	MK: Civil Option For Macedonia	9
North Macedonia	MK: RAM (Roma Aliance of Macedonia)	9
North Macedonia	MK: DPTM (Democratic Party of Turcs In Macedonia)	6
North Macedonia	MK: PEF (Party for European Future)	5
North Macedonia	MK: POPGM (Party of United Pensioners and Citizens of Macedonia)	5
North Macedonia	MK: SPM (Socialistic Party of Macedonia)	5
North Macedonia	MK: Democratic Renewal of Macedonia	4
North Macedonia	MK: PDT (Party for Movement of the Turks in Macedonia)	2
North Macedonia	MK: SPM (Serbian Party of Macedonia)	2
North Macedonia	MK: NSDP (New Social-Democratic Party)	1
Norway	NO: Pensioners' Party	6
Norway	NO: The Christians	5
Norway	NO: Democrats in Norway	4
Norway	NO: Coastal Party	3
Portugal	PT: Enough	14
Portugal	PT: Liberal Initiative	3

Country	EVS data: name of the party	Number of EVS respondents
Portugal	PT: Alliance	2
Portugal	PT: Alternative Socialist Movement	1
Portugal	PT: FREE	1
Portugal	PT: Earth Party	1
Romania	RO: Great Romania Party	11
Serbia	RS: It's enough (DJB)	25
Serbia	RS: League of Social Democrats of Vojvodina (LSV)	12
Serbia	RS: Social Democratic Party (SDS)	8
Serbia	RS: Party of United Pensioners of Serbia (PUPS)	7
Serbia	RS: United Serbia (JS)	7
Serbia	RS: Roma Party (RP)	3
Serbia	RS: Party for Democratic Action (SDA)	3
Serbia	RS: Social Democratic Party of Serbia (SDPS)	3
Serbia	RS: Liberal Democratic Party (LDP)	2
Serbia	RS: Alliance of Vojvodina Hungarians (SVM)	2
Slovakia	SK: Slovak Green Party	24
Slovakia	SK: Communist Party of Slovakia	20
Slovenia	SI: SLS - Slovene people's party	22
Slovenia	SI: PS - Positive Slovenia	17
Spain	ES: We can together	4
Spain	ES: Compromise	3
Spain	ES: Yes to the Future	2
Spain	ES: Initiative For Catalonia	1
Spain	ES: In Tide	1
Spain	ES: Aragonese Party	1
Spain	ES: Asturias Forum	1
Switzerland	CH: The alternative Left	36
Switzerland	CH: Swiss Labour Party	28
Switzerland	CH: Federal Democratic Union	21
Switzerland	CH: Ticino League	18
Switzerland	CH: Pirate Party	11
Switzerland	CH: Movement of the Citizens of French-speaking Switzerland	7

Furthermore, time gaps in the World Bank Data posed additional challenges during the merging process. Temporal inconsistencies became apparent as the EVS and CHES data were collected at different time points than the World Bank Data. We needed to carefully address these time gaps and reconcile the data to ensure that the analysis accurately captured the relationship between economic factors and attitudes toward environmental protection. Through careful alignment and interpolation techniques, we overcame these time gaps and established a coherent temporal framework for our analysis.

Despite these obstacles, it was essential to acknowledge and navigate these challenges to leverage the potential insights and comprehensive perspective that merging external data with the EVS dataset offered. By implementing rigorous data cleaning, validation, imputation, and temporal alignment procedures, we were able to mitigate the limitations and enhance the overall quality and integrity of the merged dataset.

Discussion

Our analyses have discovered that while we may not have achieved data of supreme quality and full comparability, merging different data sources holds excellent promise for conducting in-depth analyses. By incorporating multiple data sources rather than relying solely on a single cross-national project designed for comparability, we have been able to delve deeper into our research questions and gain a more comprehensive understanding of the phenomena under investigation.

The decision to merge different data sources was driven by the recognition that relying solely on a single dataset, even one specifically designed for comparability, like the EVS, may have limitations and constraints. By incorporating additional data sources, we overcame some of these limitations and expanded the scope of our analysis. Although this approach introduced its challenges, as previously mentioned, it also offered unique opportunities to explore complex relationships and uncover new insights.

Merging multiple data sources allowed us to harness the strengths of each dataset and leverage their respective advantages. While the EVS provided a solid foundation for our research, including external sources enriched our analysis by offering complementary perspectives and additional dimensions of inquiry. This multi-source approach enabled us to compensate for gaps or limitations in a single dataset, ensuring a more robust and comprehensive exploration of the research questions at hand.

Furthermore, by merging different data sources, we also tapped into the potential for triangulation and validation. When multiple sources converge and support similar findings, it lends greater credibility to our results and strengthens the reliability of our conclusions. This approach goes beyond the limitations of a single dataset and provides a more nuanced and accurate portrayal of the phenomena under study.

While acknowledging the challenges encountered during the merging process and the compromises made regarding data quality and comparability, we emphasize the inherent promise of this approach. The ability to draw upon multiple data sources, each with its strengths and limitations, enhances the depth and breadth of our analyses and offers a more comprehensive understanding of complex social phenomena.

In conclusion, our analyses have demonstrated that despite not achieving data of supreme quality and full comparability, merging different data sources holds significant promise for advancing research. By embracing a multi-source approach, we overcame limitations inherent in a single dataset and delved deeper into our research questions. The combination of diverse data sources enriched our analysis and offered new insights, fostering a more nuanced understanding of the phenomena under investigation. Going beyond reliance on a single cross-national project designed for comparability opens up new avenues for research and expands the potential for generating valuable knowledge.

Conclusions

In conclusion, merging external data with the EVS dataset presented various obstacles, including technical limitations and methodological complications. Challenges such as differing

data quality, under or over-coverage of political parties, and time gaps in the World Bank Data required careful consideration and robust methodologies. By acknowledging and proactively addressing these challenges, we created a more comprehensive and reliable merged dataset, enabling us to conduct a more robust analysis of the impact of ideological orientation on attitudes toward environmental protection.

References

Bakker, R., Hooghe, L., Jolly, S., Marks, G., Polk, J., Rovny, J., Steenbergen, M., & Vachudova, M. A. (2020). *2019 Chapel Hill Candidate Survey." Version 2019.1*. chesdata.eu

Cichocki, P., & Jabkowski, P. (2022). Response scale overstretch: linear stretching of response scales does not ensure cross-project equivalence in harmonised data. *Quality & Quantity*. https://doi.org/10.1007/s11135-022-01523-5

Dubrow, J. K., & Tomescu-Dubrow, I. (2016). The rise of cross-national survey data harmonization in the social sciences: emergence of an interdisciplinary methodological field. *Quality & Quantity*, *50*(4), 1449-1467.

EVS. (2022). *European Values Study 2017: Integrated Dataset (EVS 2017), ZA7500 Datenfile Version 5.0.0.* GESIS. https://doi.org/10.4232/1.13897

Granda, P., Wolf, C., & Hadorn, R. (2010). Harmonizing survey data. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, B. E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 315-332). John Wiley & Sons, Inc.

Jabkowski, P., Cichocki, P., & Kołczyńska, M. (2021). Multi-Project Assessments of Sample Quality in Cross-National Surveys: The Role of Weights in Applying External and Internal Measures of Sample Bias. *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smab027

Jabkowski, P., & Kołczyńska, M. (2020). Sampling and Fieldwork Practices in Europe: Analysis of Methodological Documentation From 1,537 Surveys in Five Cross-National Projects, 1981-2017. *Methodology. European Journal of Research Methods for the Behavioral Science*, *16*(3), 186-207. https://doi.org/https://doi.org/10.5964/meth.2795

Jolly, S., Bakker, R., Hooghe, L., Marks, G., Polk, J., Rovny, J., Steenbergen, M., & Vachudova, M. A. (2022). Chapel Hill Expert Survey trend file, 1999–2019. *Electoral Studies*, *75*, 102420. https://doi.org/https://doi.org/10.1016/j.electstud.2021.102420

Kołczyńska, M., & Schoene, M. (2018). Survey Data Harmonization and the Quality of Data Documentation in Cross-national Surveys. *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC)*, 963-984.

Kołczyńska, M., & Slomczynski, K. M. (2018). Item Metadata as Controls for Ex Post Harmonization of International Survey Projects. In *Advances in Comparative Survey Methods* (pp. 1011-1033). https://doi.org/https://doi.org/10.1002/9781118884997.ch46

Singh, R. K. (2020). The sum and its parts: The benefits of combining data from different surveys. *GESIS Blog: Adventures in ex-post harmonization: Frankenstein's Creature*. https://doi.org/10.34879/gesisblog.2020.22

Singh, R. K. (2021). (Not) by any stretch of the imagination: A cautionary tale about linear stretching. *GESIS Blog: Adventures in ex-post harmonization: Frankenstein's Creature*. https://doi.org/10.34879/gesisblog.2021.30

Slomczynski, K. M., Tomescu-Dubrow, I., & Wysmulek, I. (2021). Survey Data Quality in Analyzing Harmonized Indicators of Protest Behavior: A Survey Data Recycling Approach. *American Behavioral Scientist*, 00027642211021623. https://doi.org/10.1177/00027642211021623

Tomescu-Dubrow, I., & Slomczynski, K. M. (2014). Democratic Values and Protest Behavior: Data Harmonization, Measurement Comparability, and Multi-Level Modeling in Cross-National Perspective. *ASK. Research and Methods*, *23*, 103-114.