Estimating Subnational-level Political Trust in National and International Institutions

Lisa Dellmuth and Evelina Jonsson*

Stockholm University

Prepared for the 15th Annual Conference on the
Political Economy of International Organizations (PEIO)
4-6 May 2023, San Diego

Abstract

Subnational political trust measures are needed in numerous research areas in political science, but comparative datasets including measures of subnational-level political trust are scant. Using existing data on trust in national and international institutions from ten cross-national surveys worldwide, we build the Subnational Trust Database, 2001–2022, covering political trust measures for more than 1700 subnational units in 114 countries worldwide. The article compares different estimation methods, assesses their validity, and presents select results on subnational political trust in select countries and international institutions. The key result suggests that in the presence of small-*n* and unrepresentative survey data, Bayesian additive regression trees (BART) with classical or synthetic poststratification perform better than multilevel regression with classical poststratification (MrP) or synthetic poststratification (MrsP). We sketch implications for political science subfields relying on subnational political trust measures, in particular political economy, social legitimacy, and peace and conflict research.

Keywords

Bayesian additive regression trees, international institutions, legitimacy beliefs, political trust, subnational public opinion, synthetic poststratification

---

* Lisa Dellmuth is Professor of International Relations, Stockholm University, Sweden (lisa.dellmuth@su.se). Evelina Jonsson holds a Master's in International Relations with a specialization in Global Political Economy and is a research assistant at the Department of Economic History and International Relations at Stockholm University, Sweden (evelina.jonsson@ekohist.su.se).
We are grateful to Dominik Schraff and Lucas Leeman for helpful comments on an earlier draft. We also thank participants of the Global and Regional Governance seminar series in October 2021 at Stockholm University, Sweden, and the workshop on "Regional Inequality and the Political Geography of EU Support" in April 2021 at the ETH Zürich, Switzerland, for generous and valuable feedback. We thank Elisa Funk for helpful research assistance and the statistical offices that provided the underlying data making this research possible. This research is supported by the Legitimacy in Global Governance (LegGov) project, funded by Riksbankens Jubileumsfonds from 2016-2021, and Glocalizing Climate Governance (GlocalClim), funded by Formas from 2019-2023.

Political trust is crucial to a variety of aspects of successful modern life, including economic growth, law compliance, peaceful societies, political participation, transnational cooperation, and the functioning of democracy. When people trust political institutions, they tend to believe in their legitimacy (Dellmuth and Tallberg 2015; Marien and Hooghe 2011) and so feel it is worthwhile to engage with them, taking notice, discussing, participating, and complying with their norms and rules. Conversely, when political institutions do not enjoy public trust, they tend to be seen as coercive and might have difficulty to enforce compliance (Tallberg and Zürn 2019; Tyler, 1990), but they might also face constructive pressures for reform (Norris 2022; Sommerer et al. 2022).

A large number of research areas in political science has shown that political trust is consequential for the ability of political institutions to govern effectively. In political economy, studies have associated subnational political trust with subnational socioeconomic inequalities, which can make people feel left behind in a globalizing world (Lipps and Schraff 2021). In social legitimacy research, the success of subnational peacekeeping efforts has been shown to crucially depend on subnational-level trust in involved political institutions, as this trust shapes the social legitimacy of United Nations (UN) peacekeeping missions (Whalan 2017). In peace and conflict research, the mitigation of disaster impacts on communal conflict is conditional upon whether local populations trust the subnational, national and international institutions engaged in disaster management (Petrova 2022).

It is noteworthy that these debates are conducted in the absence of valid and reliable *subnational-level* measures of political trust in institutions across subnational, national and international levels. A certain "methodological nationalism" dominates political trust and legitimacy research which views nation states as the primary reference point for political socialization and for context effects on trust. Only in the European Union (EU), where researchers have over the past three decades increasingly recognized the importance of subnational units for European politics (typically referred to as "regions" in the European context), there are estimates of subnational support for the EU (Mayne and Katsanidou 2022; Lipps and Schraff 2021; Schraff et al. 2022). We therefore still know little about the patterns, causes and consequences of subnational trust in national and international institutions.

To address this limitation in earlier research, this article introduces the Subnational Trust Dataset, which is covers subnational-level measures of trust in national and international institutions covering 1747 subnational units in 114 countries worldwide.[2] The Subnational Trust Dataset is based on data from 10 major cross-national datasets spanning the years 2001-2022. To underline the validity of the subnational trust measures, we use different estimation methods, assess their validity, and discuss the results by using the examples of trust in national governments, the UN, and the World Bank. We foreground these institutions in this article to provide illustrations from both the national and international level, as well as from international institutions in different issue areas. While the UN is a multi-issue organization involved in human and state security governance, the mandate of the World Bank is more narrowly focused on economic governance.

This article details how we have sought to generate trust measures representative for subnational populations. Toward this end, we gathered census data from national and cross-national databases to generate new survey weights, and used these weights to estimate subnational measures of political trust aimed to be subnationally representative. For these estimations, we used four methods that are prominent in the literature on subnational public opinion, which has predominantly focused on voting behavior (e.g., Levendusky et al. 2008; Warshaw and Rodden 2012; Leeman and Wasserfallen 2017; Schraff et al. 2022): multilevel regression and classical poststratification (MrP), multilevel regression with synthetic

[2] All code, interim datasets, and final datasets will be made available as a replication package on the Harvard Dataverse upon publication of this paper.

poststratification (MrsP), Bayesian additive regression trees (BART) with classical poststratification (hereinafter referred to as classical BART), and BART with synthetic poststratification (hereinafter referred to as synthetic BART). These methods are applied to generate two distinct datasets from the same source, one for all available 1747 subnational units (across 114 countries), and another for all 541 units (across 70 countries) that we estimate to be subnationally representative.[3] We then chose a subset of 328 (across 47 countries) from the World Values Survey 7 (WVS7) and European Values Survey (EVS5) where representative data is available for comparison to conduct several validity tests.

Taken together, the Subnational Trust Dataset is designed to be used in research pushing forward theory and empirical research on the patterns, causes, and consequences of subnational-level trust in various political institutions. While the data can be easily downloaded and used, the methodology we propose can also be replicated for other surveys for different points in time to generate valid subnational-level trust measures comparable across countries worldwide.[4] One of the main innovations here is that this dataset spans a large number of countries worldwide. The most valid results are produced by the Bayesian approach when compared to the other methods used. In the presence of small-*n* and unrepresentative survey data, Bayesian additive regression trees (BART) with classic or synthetic poststratification perform better than multilevel regression with classical poststratification (MrP) and synthetic poststratification (MrsP). Synthetic and classical poststratification perform similarly well. By way of conclusion, this article discusses implications for political science research using subnational trust measures, foregrounding the examples of political economy, social legitimacy, and peace and conflict studies.

[3] Asia Europe Survey (2001) is omitted from this dataset that is used for validation exercises due the difference in time between the other surveys.
[4] All code, interim datasets, and final datasets will be made available at the Harvard Dataverse (see Appendix C for access to the database and Appendix D for instructions for replication).

Creating the Subnational Trust Database

In this section, we elaborate how we selected surveys, operationalized political trust, created subnational weights, and estimated weighted subnational trust by using different methods. We discuss relevant methodological challenges and how we addressed them.

Selecting Surveys and Measures of Political Trust

Given our ambition to create a dataset on political trust in national and international institutions for comparative research, we made the decision to rely on cross-national opinion polls, rather than national poll series. There are several national public opinion or household surveys including political trust measures, but such measures are typically not comparable across countries. Therefore, we selected the 10 most far-reaching and widely used cross-national opinion polls which include trust measures for 114 countries worldwide (out of 119 countries there is survey data for in total) (Table 1).

Table 1 Cross-National Surveys Containing Measures of Legitimacy

Survey	Time Period	Number of Surveyed Countries
Asia Europe Survey (ASES)	2001	17
Americas Barometer (LAPOP)	2016/2017	8
Afrobarometer	Round 4 (2008)	18
Arab Barometer	Wave 5 (2018-2019)	11
European Social Survey (ESS)	Round 9 (2018)	27
Eurobarometer 93	July 2020	29
European Values Study (EVS)	Wave 5 (2017-2021)	33
EU Neighbourhood Barometer	Wave 6 (2014)	14
Latinobárometro	2020	18
World Values Survey (WVS)	Wave 7 (2017-2022)	55

Note: The number of surveyed countries include countries with subnational-level data for the most recent data collection or “round”/“wave” available. For details on these polls’ sampling methodologies, subnational units covered, and measures included, see Appendix A.

Political trust refers to “citizens’ assessments of the core institutions of the polity and entails a positive evaluation of the most relevant attributes that make each political institution trustworthy, such as credibility, fairness, competence, transparency in its policy-making, and openness to competing views” (Zmerli 2014, 4887). Political trust overlaps with measures such as “diffuse support” (Easton 1975), which captures a willingness to accept decisions of a governing power although one might not be a beneficiary of these decisions, or “social legitimacy” (Dellmuth and Tallberg 2015), which is understood as the public perception that a governing power exercises its authority appropriately. A number of studies have used measures of trust or confidence in political institutions to assess attitudes toward national and international institutions (e.g., Harteveld et al. 2013; Marien and Hooghe 2011; Tyler 2006; Tyler and Huo 2002; Dellmuth and Tallberg 2015, 2023; Lipps and Schraff 2021).[5]

Based on these insights, we operationalize political trust by using a variety of measures based on the answers to survey questions about trust or confidence in political institutions. Examples are measures of confidence in government in the World Values Survey (WVS) or trust in the UN in the European Social Survey (ESS). In addition, we include various measures of support in our dataset which are closely related to measures of trust (Thomassen et al. 2017; Dellmuth and Schlipphak 2020). For example, relevant survey questions ask whether respondents think their country’s membership in the EU is good or bad (Eurobarometer), or whether respondents perceive institutions to do a good job or having improved various outcomes (Afrobarometer and EU Neighborhood Barometer) (see Appendix A for a list of included measures).

[5] However, it should be noted that those concepts are not synonymous and their complex relationship has been discussed (cf. Hooghe and Zmerli, 2011; Thomassen et al. 2017).

Creating Subnational Weights

Based on the selection of surveys and measures, we developed an approach to create subnationally representative trust measures. This is a key challenge as the selected cross-national databases are not always subnationally representative, and even if they aim to be, the number of survey participants is often low within subnational units which compromises subnational representativeness (Leeman and Wasserfallen 2020). However, existing opinion polls do not provide subnational survey weights to correct for potential imbalances, but typically include age-sex weights for the estimation of nationally but not subnationally representative measures.

To calculate new subnational weights, we collected and pre-processed census data at the subnational level in five steps (see Appendix B for more details).

First, we collected the most recent available subnational population census data for 114 countries from the following databases: Integrated Public Use Microdata Series (IPUMS) International, Eurostat, and individual national statistics databases. We collected joint distribution data on age groups and gender at the subnational level. Age and gender are variables that are available in population census data for most countries. Five-year age groups from the age 15 are used as age groups are more commonly used in the census data that is publicly available. Where five-year age groups are not available, the smallest intervals of age groups are collected, such as 7-year— or 10-year age groups.

Second, the collected census data from Eurostat and IPUMS data were pre-processed in Python to create a consistent global census dataset. The data that was downloaded from individual country statistics websites were first extracted from their various forms. A main difficulty that arose was the lack of data structure and format consistency. The census data was available through various Application Programming Interface softwares (API), online data portals, CSV files, Excel files, microdata files or found in tables included in extensive country statistical yearbook PDF files. We structured these files first in a semi-consistent way in Excel, which enabled us to then process the data further in Python to create the global census dataset.[6]

Third, we created a mapping file that connects the geographical units of each survey to the units in the census dataset. Each survey structured their data differently, used varying regional unit levels, and uniquely labelled the geographical units, so this step was necessary to avoid data mishaps. Although time-consuming, this step was executed manually as attempts at algorithmic matching of subnational labels proved unreliable and programmatic geographical matching using coordinates was unavailable.

A central challenge when working with subnational geographical data is correctly matching the varying hierarchical levels of regional units for each county, such as aggregate non-administrative units (e.g., macro regions and region groupings), first-level administrative units (e.g., states, provinces, districts, and governates), and second-level administrative units (e.g., municipalities). As the subnational unit level was not always consistent between the survey and census, the regional units of each survey database were matched to the level available in the national census datasets and aggregated to the census data level if it was higher than the survey data and vice versa.

Thus survey and census regions can map one-to-one (e.g., the same subnational unit levels are available in both databases), many-to-one (e.g., the subnational units are at a lower level —i.e., smaller geographical units—in the census than the survey data), or one-to-many (e.g., the subnational units at a higher level —i.e., larger geographical units— in the census data than the survey data). In a many-to-one mapping, when a survey has data on the provincial level, but the census data is on the municipality level, we aggregated the census data to the

[6] The data and pre-processed files, as well as the Python script, will be uploaded to the Harvard Dataverse together with the replication package. However, we do not have permission to share the census data from China and Colombia.

provincial level. Conversely, in a one-to-many mapping census data can be available on the macro-regional level, comprised of state groupings, but survey data is available on the state level, in which case we would aggregate the data to the macro-regional level used in the census.

Fourth, custom census datasets were generated to harmonize subnational units between the census and survey datasets according to the mapping file discussed above. For more details, see Appendix A for the subnational unit levels available in each original survey database and the levels used for our database. Additionally, see Appendix B (Table B1) for the subnational unit levels available in our Subnational Census Database.

Fifth, to match this census dataset, custom survey datasets were created in Python. In this step, the age groups were created to match the survey data age groups as closely as possible with the census data age groups. Survey data usually includes exact ages, which simplifies this process. To simplify poststratification, the survey data was processed to output the same columns and variable names for every survey as well as to match to the global census dataset: country, region, gender, age groups, and each of the survey questions. We used the variable names from the source data. The resulting database is the basis for the application of several methods to estimate subnational opinion averages.

Estimation Methods for Poststratification

Next, we identify and discuss a number of estimation methods to create subnational-level trust measures with poststratification. Previous literature on subnational opinion formation and voting behavior provides useful insights in this regard. This literature has discussed how to deal with the challenge that subsamples in subnational units are small, and therefore, or in addition, not representative for the subnational target population (cf. Warshaw and Rodden 2012; Ghitza and Gelman 2013; Toshkov 2015; Georgiadou et al. 2018; Bisbee 2019; Lipps and Schraff 2021; Leeman and Wasserfallen 2017, 2020; Broniecki et al. 2022; Schraff et al. 2022).

However, the evidence is inconclusive regarding which method performs best. Four methods are commonly used: MrP, MrsP, and BART (with and without synthetic poststratification). To begin with, MrP uses classical poststratification and BART can be estimated with classical poststratification as well. "Classical poststratification" means that weights are applied to the survey responses to ensure that the responses from each subnational unit match the responses of the representative strata of the unit. MrP can extrapolate sample inferences to a target population with either probability or nonprobability samples. A key strength of MrP is that the method has been shown to produce valid measures even if samples are not representative (e.g., Lax and Philipps 2009; Wang et al. 2015; Warshaw and Rodden 2012; see Leeman and Wasserfallen 2020 for a detailed introduction to MrP). At its core, this method is a multi-level model that regresses an outcome measure, such as voting intention, on a set of demographic characteristics (e.g., age, gender, marital status, and education) and group variables (e.g., subnational unit) to predict an outcome by subnational unit. The estimates are then calculated by weighting the predictions of citizens with the true population structure of the same demographic characteristics from subnational census data.

To illustrate, we estimate political trust with a MrP response model with two individual-level variables as random effects—gender (female/male) (α_j) and age groups (typically five-year age groups from 15 years onward with the last category being "85 years and older")[7] ($\alpha_{m)}$—and with random effects for the subnational unit (α_s), the subnational level (level

[7] The age groupings depend on the granularity of the data available in the census data. Usually, five-year age groups are available. Sometimes, 7-year, 10-year or even 15-year age groups are only available. The age groupings are unique to each country and not standardized across countries. We use the smallest age intervals that are available ranging to the last available age group. Age groups typically range to 85 or 100 years and older, but there are cases in which they range to only 60 years.

1/level 2), (α_l), and the country (α_c). We write the equation as a hierarchical probit model (cf. Leeman and Wasserfallen 2020, 374):

$$\Pr(y_i = 1) = \Phi(\beta_0 + \alpha_{j[i]}^{gender} + \alpha_{m[i]}^{age\ groups} + \alpha_{s[i]}^{subnational\ unit} + \alpha_{l[i]}^{subnational\ level} + \alpha_{c[i]}^{country})$$

$$\alpha_j^{gender} \sim N\left(0, \sigma_{gender}^2\right), \qquad \text{for } j = 1, \dots, J$$

$$\alpha_m^{age\ groups} \sim N\left(0, \sigma_{age\ groups}^2\right), \qquad \text{for } m = 1, \dots, M$$

$$\alpha_s^{subnational\ unit} \sim N(0, \sigma_{subnational\ unit}^2), \qquad \text{for } s = 1, \dots, S$$

$$\alpha_s^{subnational\ level} \sim N(0, \sigma_{subnational\ level}^2), \qquad \text{for } s = 1, \dots, L$$

$$\alpha_c^{country} \sim N\left(0, \sigma_{country}^2\right), \qquad \text{for } c = 1, \dots, C$$

There are 30 survey respondent group types in each subnational unit as $J = 2$ for gender and $M = 15$ for the 15 five-year age groups. The model estimates are used to calculate predictions $\hat{\pi}_{csjm}$ for all combinations of *j* and *m* in each subnational unit *s,* subnational level *l,* and country *c*.

To perform classical poststratification, we collected the "true" joint distributions in each subnational unit with the frequency of each survey respondent group type $(N_{1,1}, N_{2,1}, \dots, N_{15,2})$. A true joint probability distribution, in the context of this paper, describes the probability that a given person in a subnational unit takes on a set of specific values according to a number of demographic characteristics (e.g., a person living in a subnational unit that is woman *and* aged 45-50). Table 2 illustrates the requirements for classical poststratification using age groups and gender as an example, where the exact number of people living in a subnational unit according to each combination of demographic characteristics is required.

Table 2 Census Data Requirement for Classical Poststratification

Gender / Age Groups	Female	Male	Total
15-19	$N_{1,1}$	$N_{1,2}$	N_1
20-24	$N_{2,1}$	$N_{2,2}$	N_2
25-29	$N_{3,1}$	$N_{3,2}$	N_3
30-34	$N_{4,1}$	$N_{4,2}$	N_4
35-39	$N_{5,1}$	$N_{5,2}$	N_5
40-44	$N_{6,1}$	$N_{6,2}$	N_6
45-49	$N_{7,1}$	$N_{7,2}$	N_7
50-54	$N_{8,1}$	$N_{8,2}$	N_8
55-59	$N_{9,1}$	$N_{9,2}$	N_9
60-64	$N_{10,1}$	$N_{10,2}$	N_{10}
65-69	$N_{11,1}$	$N_{11,2}$	N_{11}
70-74	$N_{12,1}$	$N_{12,2}$	N_{12}
75-79	$N_{13,1}$	$N_{13,2}$	N_{13}
80-84	$N_{14,1}$	$N_{14,2}$	N_{14}
85+	$N_{15,1}$	$N_{15,2}$	N_{15}
Total	N_1	N_2	N

Note: *N* represents the number of people in a subnational unit according to each variable combination.

True joints between all predictor variables at the subnational level are needed to weigh each prediction by the joint distribution data of each specific survey respondent group divided by the number of people living in the subnational unit with the same demographic characteristics: (cf. Leeman and Wasserfallen 2017, 3):

$$\hat{\pi}_{c,s} = \frac{\sum_j \sum_m \hat{\pi}_{jm\in c,l,s} N_{jm\in c,l,s}}{N_{n\in c,l,s}}$$

$$= \frac{\sum_j \sum_m \Phi(\hat{\beta}_0 + \hat{\alpha}_m + \hat{\alpha}_j + \hat{\alpha}_s + \hat{\alpha}_l + \hat{\alpha}_c) N_{jm\in c,l,s}}{N_{n\in c,l,s}}$$

In the previous section, we introduced our subnational census dataset which includes demographic data with true joints between five-year age groups and gender at the subnational level for 114 countries. These two demographic variables were selected to carry out poststratification as they tend to be available within national census databases worldwide, except for select developing countries.

An alternative method shown to create valid measures is to use a Bayesian additive regression tree (BART) model with classical poststratification. The Bayesian backfitting algorithm estimates many decision trees to learn from the residuals of the previous tree, which are combined to describe the structure of the data. The model is comprised of two parts: a sum-of-trees model (i.e., a multivariate additive model) and the regularization prior, followed by a Bayesian back-fitting Monte Claro and Markov Chain (MCMC) algorithm for posterior computation. Poststratifying BART predictions, in the context of EU public support measures, yield better predictions than MrP (Lipps and Schraff, 2019).

Recent years have seen increased efforts to examine the performance of MrP and BART and how to leverage machine learning procedures for these models (e.g., Ornstein 2019; Broniecki et al. 2022). In this article, we take cues from this literature but do not rely on machine-learning which would require the use of context variables, which are not available at global scale (Broniecki et al. 2022).

To illustrate, we explore a BART probit model for binary classification (Y = 0 or 1), which will be used to estimate trust in international organisations worldwide:

$$\mathbb{P}(Y = 1 \mid X) = \Phi(T_1^M(X) + T_2^M(X) + \cdots + T_m^M(X)),$$

where the cumulative density function of the standard normal distribution is denoted as Φ. Each classification probability of x is a function of the sum of regression trees. T denotes a binary tree structure and the leaves, in other words, a set of parameter values associated with each of the terminal nodes, denoted as M. T^m denotes a tree structure and its set of leaf parameter values. Following the default parameter settings for a BART classification model outlined in Chipman et al. (2012), we apply the training set with the default setting of 200 trees, where m denotes the number of trees for the algorithm to iterate over. Using 200 trees has been shown to provide reliable predictions and to avoid overfitting (Chipman et al. 2012).

For the final BART model specification, we impose regularization of the prior on the sum-of-trees model to keep the tree components small to avoid overfitting the data. The prior on σ^2 is not required as the classification model implicitly assumes $\sigma^2 = 1$. For the choice of the variance hyperparameter σ_μ^2, we set the default choice of k = 2, which corresponds to the interval with 95% coverage of the provided response values in the training set. This prior helps regularize the model by shrinking the tree parameters towards the center of the response's distribution, in other words, shrinking towards the mean 0.5 (see Chipman et al. 2012 and Kapelner and Bleich 2013 for detailed descriptions of non-default applications).

Next, the BART model is estimated with Bayesian backfitting MCMC algorithm by generating samples from a posterior and creating predictions based on prior and likelihood (see Chipman et al. 2012 and Kapelner and Bleich 2013 for implementation and a thorough discussion). At a general level, estimations are obtained by drawing successive samples from the induced posterior probability, fitting each tree iteratively to predict the outcome. Small incremental changes are made to the tree structures at each iteration by growing, collapsing, or splitting terminal nodes until the trees evolve to learn the structure of the data. The MCMC process generates a sequence of draws until convergence to the posterior distribution of the true model is reached.

As a final step, the BART estimates from the previous step are poststratified with classical poststratification. The procedure for classical poststratification is carried out on BART estimates as described above for MrP.

MrP and classical BART have some disadvantages. A widely known challenge of working with census data at the subnational level is the limited granularity of the census data. This highlights a notable weakness of MrP and classical BART as detailed census data in the form of true joints between all predictor variables are required. Subnational census data with true joints of other potentially useful poststratification variables such as marital status, education level, and income are often less available, particularly in developing countries. This limits researchers to the use of a few demographic variables, given the strict data requirements for MrP and classical BART. Further, the extension of potentially powerful predictors (e.g., partisanship and social trust) from non-census data, which are typically only available in the form of marginal distributions at the subnational level, is largely unavailable.

One additional drawback of poststratified BART is that relative to MrP, BART is more challenging to understand, and researchers might need to adjust parameters, such as defining the number of trees and calibrating the priors (Lipps and Schraff, 2021). However, the default model specification typically performs well and hyperparameter tuning (e.g., the practice of running multiple trials to select a set of optimal parameters for a learning algorithm to achieve the best performance) is not always necessary (Chipman et al. 2012).

Next, we turn to two alternative methods with synthetic poststratification that help overcome the challenge of sparse subnational census data: MrsP and BART with *synthetic* poststratification. Instead of relying on census data with true joint distributions for poststratification, these methods only require "synthetic" joint distributions, computed using data on the marginal distributions (see Table 3 for an overview of data requirements; Leeman and Wasserfallen 2017; Lipps and Schraff, 2021). A marginal distribution is the distribution of each individual demographic variable, such as the gender distribution of a subnational unit where 45% of the population is female and 55% is male. In other words, the probability that a given person in a subnational unit takes on a specific value according to one demographic characteristic (e.g., a person living in a subnational unit that is aged 45-50). MrsP and synthetic BART are executed the same as their classical counterparts, except predictions are poststratified using synthetic joints on a set of demographic variables at the subnational level by multiplying the marginal distributions for those variables.

Table 3 Data Requirement for MrP, MrsP, Classical BART and Synthetic BART

	Classical/ True Joint Distribution	**Synthetic/ Marginal Distributions**
MrP	✓ Required	✕ Not sufficient
Classical BART	✓ Required	✕ Not sufficient
MrsP	✕ Not required	✓ Sufficient
Synthetic BART	✕ Not required	✓ Sufficient

Adding potentially powerful predictors of subnational differences might increase the predictive power of the models. For instance, age proved to be a powerful individual-level predictor in enhancing the prediction precision of estimating state-level political preferences in same-sex marriage questions in the US (Leeman and Wasserfallen 2017). Using MrsP and synthetic BART facilitates this, as researchers only need the regional marginal distributions of predictive variables to create synthetic distributions between all variables for poststratification, calculated by multiplying the regional margins for each category of predictors. The difference between true and synthetic joints is illustrated in Table 4, which displays the distributions of two binary variables, v1 and v2, as an example.

Table 4 Stylized Example of True and Synthetic Joint Distributions

v1 \ v2	*i*=1	*i*=2	Total
j=1	**40%**	**20%**	60%
j=2	**30%**	**10%**	40%
Total	70%	30%	100%

(a) True Joint Distribution

v1 \ v2	*i*=1	*i*=2	Total
j=1	42%	18%	**60%**
j=2	28%	12%	**40%**
Total	**70%**	**30%**	100%

(b) Synthetic Joint Distribution

Note: The values in percentages represent the relative share of each variable combination. (a) The value for each combination of variables is a true joint distribution. (b) The marginal distributions can be found in the margins, where the total column represents the marginal distribution of variable one and the total row represents the marginal distribution of variable two.

Table 3a illustrates an example of a true joint distribution, where the values in percentage represent the relative share of each variable combination derived from data on the exact number of values per combination. Table 3b demonstrates an example of a synthetic distribution where the values in percentage represent the relative share of each variable combination as a product of the marginal distribution of each variable combination: v1 (60% and 40%) and v2 (70% and 30%). For example, multiplying *j*=1 which has a marginal distribution of 60% and *i*=1 of 70% produces a synthetic joint between the variables of 42%.

When comparing Table 3a and 3b, it is apparent that the synthetic distribution only slightly deviates from the true distribution. Even in extreme cases, when the synthetic distribution deviates strongly from the true distribution, research finds that MrsP prediction using synthetic joints differ only slightly from MrP prediction (Leeman and Wasserfallen 2017). The extent to which the estimates deviate depends on the degree to which the individual-level variables are correlated. MrsP relies on the assumption that individual predictors are independent from one another, thus the higher the variables are correlated the greater the deviation. This points to a potential disadvantage of MrsP. However, Leeman and Wasserfallen

(2017) find that in applied work, synthetic joints should not produce a large deviation in prediction unless the correlation is very strong (e.g., a level of 0.6 and over).

Additionally, there is an alternative technique to accommodate for correlation between individual-level predictors by computing synthetic joints based on the correlation structure of predictors from the survey data. Researchers can estimate "adjusted" synthetic joints by extending the available joint distributions from subnational census data for all available predicators to include information on an additional predictor where only the marginal distributions in subnational units are available.

For this technique, the first step is to correct the distribution of the additional predictor in the survey data according to the known marginal distribution at the subnational level. As a second step, these corrected relative shares are combined to the available true joint distribution data of the predictors with available census data to create synthetic joints between all variable combinations. Adjusted synthetic joints assume that the correlations between variables are the same across subnational units (see Leeman and Wasserfallen 2017 for a detailed description). We compute synthetic joints for our MrsP and classical BART estimations as improvements in prediction precision as adjusted synthetic joints tend to produce marginal effects (Leeman and Wasserfallen 2017). Studies comparing the performance of MrsP and synthetic BART to their classical counterparts find that the methods relying on synthetic poststratification tend to perform better than—or at least on par with—classical poststratification (Leeman and Wasserfallen 2017; Lipps and Schraff, 2021; Hoover and Dehghani 2019).

To illustrate, Leeman and Wasserfallen (2017) estimated public opinion at the canton-level in Switzerland and the state-level in the US comparing the relative performance of MrP and MrsP finding that the prediction precision of MrsP using simple synthetic joints was at least identical to, if not better than, MrP. The elaborate version of MrsP performed only marginally better than MrsP with simple synthetic joints. Likewise, in analysis of subnational public opinion across Europe comparing the relative performance of the four methods, Lipps and Schraff (2021) find that MrsP with left-right political preferences as an additional predictor and MrP without the additional predictor led to similar performance results. Additionally, their study revealed that synthetic BART with left-right political preferences as an additional predictor led to moderate improvements over classical BART without the additional predictor. In general, these studies suggest the relative advantage provided by methods with synthetic poststratification by way of leveraging additional predictors with predictive capacity or at least matching the performance of classical methods without a reliance on detailed data.

Turning to the relative performance of multilevel regression and Bayesian additive regression trees with or without synthetic poststratification, the prediction precision gains of BART with additional regional predictors are substantial in comparison to MrP and MrsP (Lipps and Schraff 2021). One explanation for improvements in prediction precision of BART compared to multilevel regression with poststratification is that the former requires minimal researcher intervention to avoid overfitting the data, whereas the latter requires more thoughtful model specification. BART is more flexible in terms of model building in that researchers can add several—at times potentially irrelevant—variables and BART will usually ignore irrelevant predictors. Additionally, BART can manage complex data and account for nonlinearities and discontinuities as well as deep interactions between many variables allowing for more nuanced analyses of subnational public opinion, such as vote outcomes and turnout of small demographic subgroups (Montgomery and Olivella 2018; Ghitza and Gelman 2013).

Taken together, this overview suggests that the relative performance of these four methods is still being debated. Additionally, previous research has not yet systematically compared the four methods by using the same number of predictors in all models. Thus we will test the performance of these four methods to create the Subnational Trust Database, and compare models that have the same set of individual-level predictors. This approach has the dual advantage to contribute to the debate on methods performance, and to provide a systematic approach to estimate subnational political trust.

Estimating Subnational Political Trust

To estimate these methods, we need to choose predictors as indicated in the previous section. Previous studies with the geographical focus on western Europe have used five to six predictor variables in analyses using MrP and MrsP and as many as ten predictors in analyses using BART with or without synthetic poststratification (Lipps and Schraff 2019). Here we use five predictor variables available: two individual-level predictors (typically five-year age groups from 15 years onward with the last category usually being "85 years and older" and two genders), two subnational-level predictors (subnational unit and level), and one country-level predictor (country). While the use of a larger number of predictors could potentially improve predictions, as BART tends to favor the most relevant predictors to improve its fit and exclude those that are unrelated to the outcome (Chipman et al. 2010), the availability of predictors for a global dataset is limited, particularly in the developing world.[8]

The estimates for all four methods are generated by coding a binary variable, excluding missing values (see Appendix A for details on the recoding of responses to each survey question).[9] This dichotomization has three main advantages in the context of our study. First, we are conceptually interested in subnationally aggregated measures of either trust or distrust, rather than covering the full span of survey questions responses. Second, the use of a binary variable helps to standardize the varying lengths of response scales across surveys. Third, the

[8] Although marginal distributions of predictor variables on the subnational level could be used for synthetic poststratification, even such data is lacking, and where data is available, it would to be too labor-intensive and time-consuming to collect and process for analysis of a global dataset on subnational trust.
[9] The statistical software R is used to compute the estimates for each method. For MrP and MrsP, the Generalized Linear Mixed-Effects Model "glmer" function from the "lme4" package is used. For BART, the Bayesian Additive Regression Tree Model "bart" function from the "BayesTree" package is used.

dichotomization enables us to use the four discussed methods without extensions, and thus to add cumulative knowledge on model performance.

We thus recoded to estimate the positivity and negativity of the response. Neutral and mid-points were always coded zero. As an example, for a response scale of no trust (0), a little (1), somewhat (2), or a lot of trust (3), this would be recoded as no trust to a little trust (0) and somewhat to a lot of trust (1) for positivity. For negativity, it is coded as follows: trust to a little trust (1) and somewhat to a lot of trust (0). There are survey questions with a binary response scale 0-1, response scales ranging from 0-3, and other surveys with larger scales that range from 0-7.

Validating the Subnational Trust Measures

Next, we select two particular surveys from our dataset to test the validity of the data. We choose the WVS7 and EVS5, which are two widely used cross-national datasets with global reach. We report select test statistics on model fit, which estimate how well a given model specification with independent variables. We use five predictors (age, gender, subnational unit, subnational level, and country) to fit the data, including subnational units in which samples are representative, and report select test statistics on model fit with marital status as an additional predictor for comparison.

Before engaging in model validation, we illustrate our full dataset, including unrepresentative subnational samples, in a descriptive mapping of confidence in select national and international institutions across two countries: Brazil and the US. While both countries are important players on the world stage, we wanted some variation in terms of including the US which qualifies as a liberal democracy, while Brazil is an electoral democracy (V-Dem 2021). Depending on the state of democracy, trust in national and international institutions might vary.

Figure 1 illustrates the results from all four estimation methods, depicting box plots. Here we use our dataset including subnational units in which samples are both representative and unrepresentative. Box plots illustrate a sample using the 25th, 50th and 75th percentiles – also known as the lower quartile, median and upper quartile – and the interquartile range, which covers the central 50% of the data. We chose to estimate the percentage of Brazil and US citizens having a great deal or quite a lot of confidence in national government, the UN, and the World Bank, averaged across states, to determine whether the aggregated estimations have face value.

The results have face validity. When compared to Brazilians, US citizens have more confidence in their national government, while confidence ratings in the UN and in the World Bank are more similar, which is broadly in line with previous comparisons of national-level citizen confidence in these institutions (Dellmuth et al. 2022, ch. 3). Furthermore, the larger interquartile range in the case of US citizens' confidence in national government indicates a greater spread of confidence values across US states than across Brazilian states. In contrast, there appears to be a greater spread of confidence in international institutions across states in Brazil than across states in the US. When it comes to the differences between the estimations, these are substantial. While MrP and MrsP estimations are similar to each other, these estimations differ from classical and synthetic BART[10]. Classical and synthetic BART differ slightly from each other. The MrP and MrsP estimations appear to produce estimates with a greater spread than BART. Moreover, MrP and MrsP estimations appear to generate values with lower values.

In Figures 2 (US) and 3 (Brazil), we use variation at GADM 2. GADM levels do not always map onto administrative divisions used in the census or survey, but are useful for depictions of the spatial variation. These figures underline that there is considerable variation

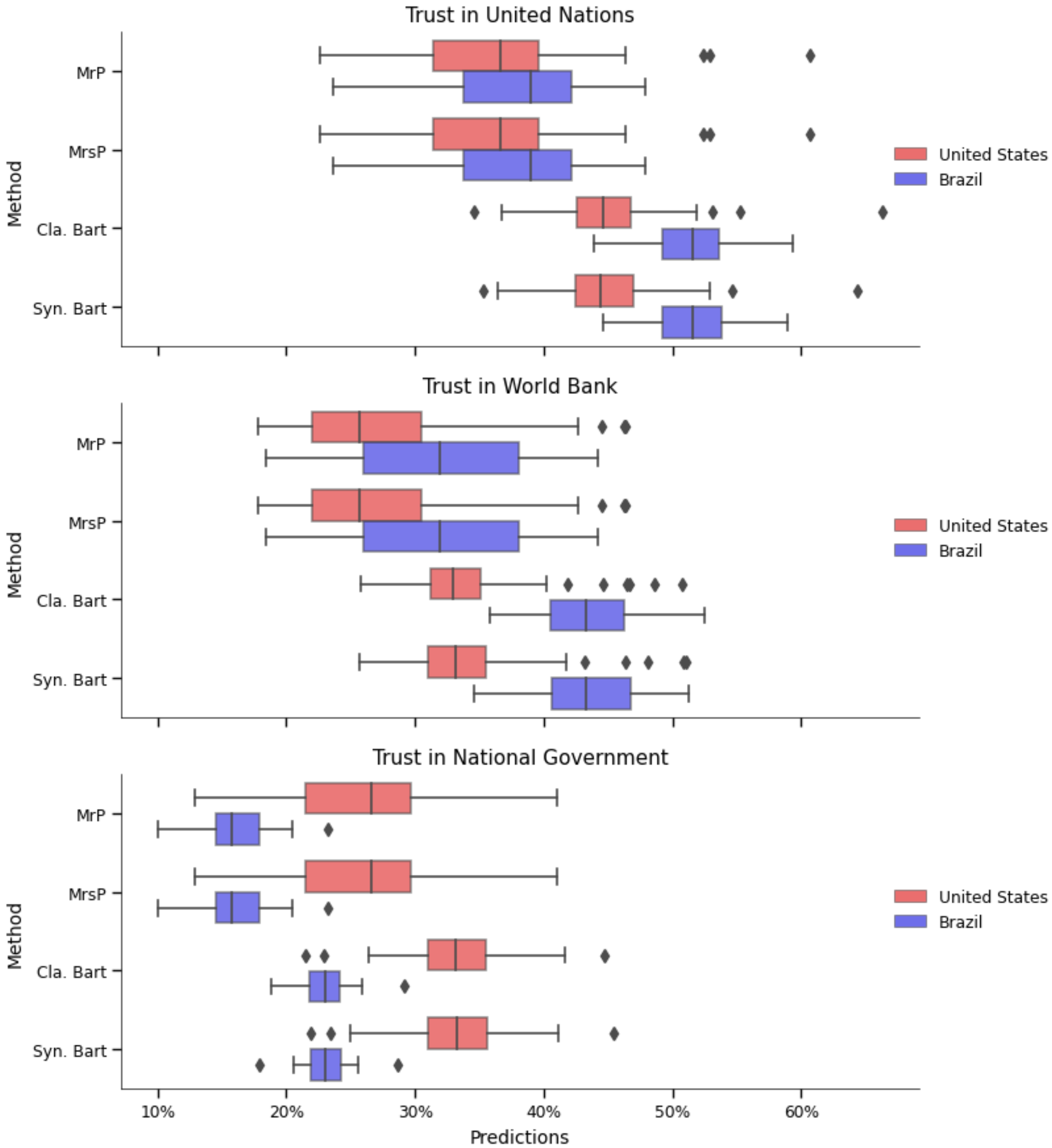
[10] Although the estimates look similar, there are differences in the decimals (see Appendix E).

in confidence in political institutions across subnational units in Brazil and the US. Moreover, BART does not generate consistently higher estimates than MrP or MrsP across all states. Taken together, these descriptive results suggest that the four estimation methods used generate substantially different results. Thus, we evaluate them more systematically in terms of their relative performance in turn.

Previous studies have compared the prediction precision of MrP, MrsP, and BART estimation results using subnationally unrepresentative public opinion survey data to subnationally representative public opinion survey data with congruent questions and predictor variables. The predicted values of each method are compared to the mean survey values of subnationally representative survey data to evaluate how far the predicted values deviate from the subnationally representative data (e.g., Leeman and Wasserfallen 2017). This option is unavailable to us as a global sub-nationally representative public opinion poll does not exist at this point in time. To analyze the validity of our global dataset, we select subnational units where representativeness can be assumed from 9 (out of 10) of our selected surveys.

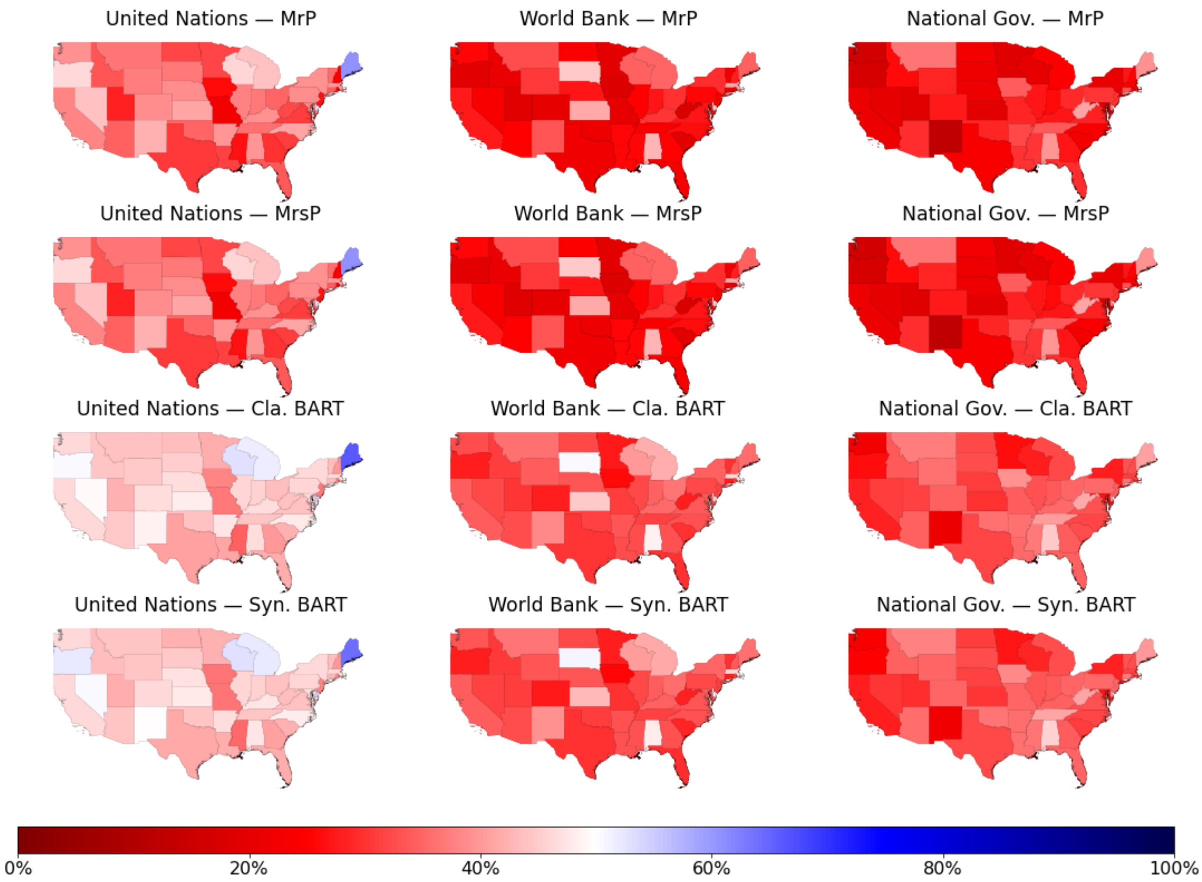
We calculate the degree of over- or underrepresentation at the subnational level by dividing the proportion of survey respondents in a subnational unit by the proportion of population living in a subnational unit. We do this for the sample size, five-year age groups and gender for each regional unit. Using the age group 40-44 years as an example, the degree of over-or underrepresentation is calculated by dividing the proportion of survey respondents within the age group 40-44 years living within a subnational unit by the true proportion of people aged 40-44 years living within a subnational unit. Values less than 1.0 underrepresent the true population distribution, values greater than 1.0 overrepresent the true population distribution and 1.0 matches the true population distribution perfectly. We assume representative samples at the subnational level for all observations with degrees of over-or underrepresentation between 0.5 and 2.0.

Figure 1 MrP, MrsP, Classical BART and Synthetic BART Estimates on Trust in Brazil and the United States



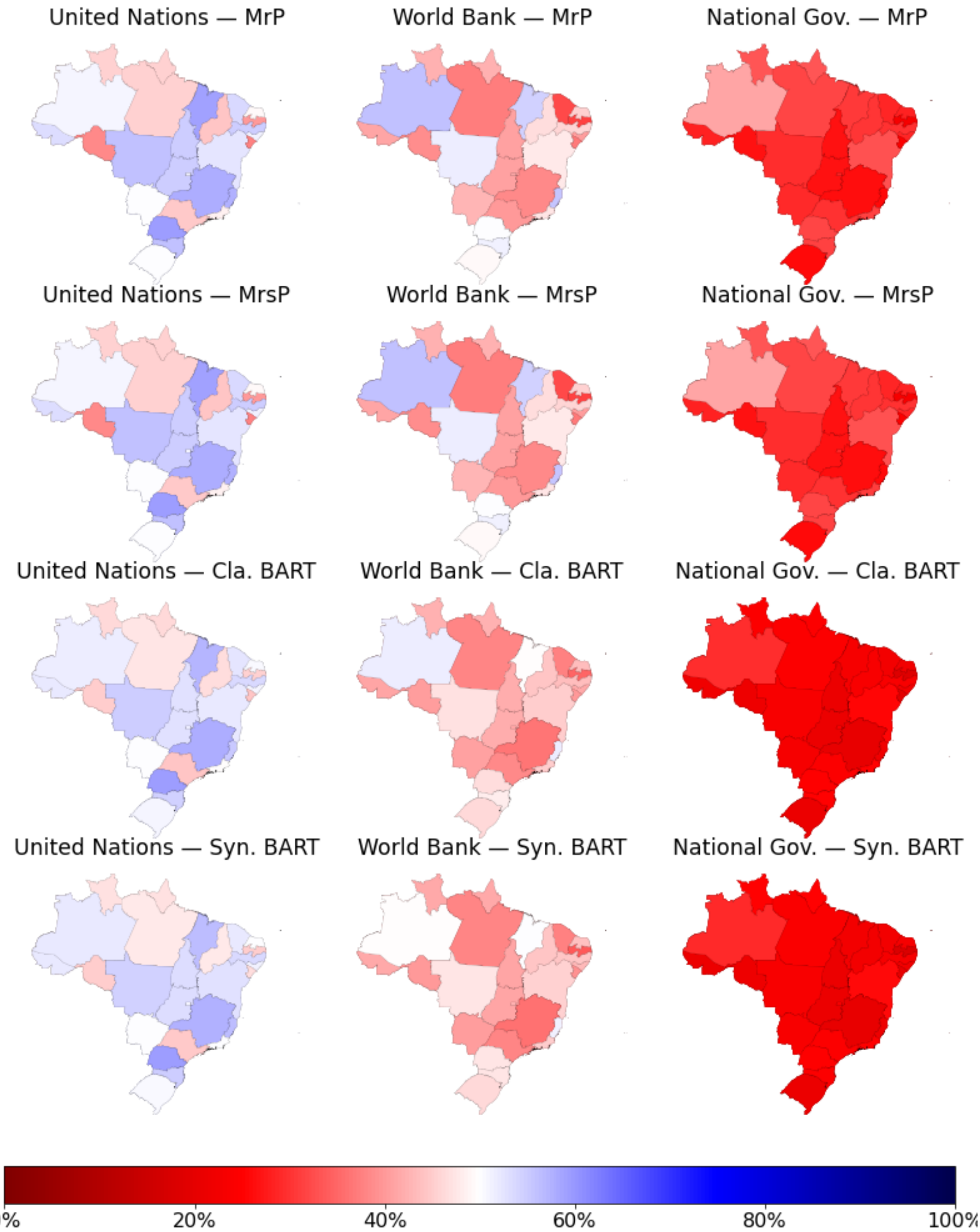
Note: Metric = Positivity (% of having a great deal or quite a lot of confidence). Database = Full Subnational Trust Database, predictions using WVS7 survey data and five-year age groups, gender, subnational unit, subnational level, and country as predictors. United States subnational units = 50 states. Brazil subnational units = 23 states.

Figure 2 MrP, MrsP, Classical BART and Synthetic BART Estimates on Trust in the US



Note: Cla. = Classic. Syn. = Synthetic. Metric = Positivity. Database = Full Global Trust Database, predictions using WVS7 survey data and five-year age groups, gender, subnational unit, subnational level, and country as predictors. United States subnational units = 50 states.

Figure 3 MrP, MrsP, Classical BART and Synthetic BART Estimates on Trust in Brazil



Note: Metric = Positivity. Database = Full Global Trust Database, predictions using WVS7 survey data and five-year age groups, gender, subnational level, subnational unit, and country as predictors. Brazil subnational units = 23 states.

Units that are sufficiently representative across all three categories (regional population distribution, regional age group distribution, regional gender distribution) and where at least half of the age groups are represented are coded 1 as sufficiently representative and 0 otherwise. We retain only the subnational units that are assumed to be representative for our analysis below.

To compare the relative performance of the four estimation methods, we show the mean subnationally representative values and the mean predicted values for levels of trust in national government, the United Nations, and the World Bank (Figure 4), and the prediction precision of each method for all trust indicators in 328 subnational units (Figure 5). In Figures 4 and 5, we focus on the regions surveyed in the WVS7 and EVS5 where representative subnational data is available using the dataset created and described above.[11]

To produce these figures, we calculate the following test statistics: Mean Absolute Error (MAE), Rooted Mean Squared Error (RMSE), Pearson's correlation and Kendall's correlation (following Toshkov 2015; Lipps and Schraff 2021; Leemann and Wasserfallen 2017; Hoover and Dehghani 2019). When using only age groups and gender as individual-level predictor variables, we find no real differences in predictive performance between the methods that use classical versus synthetic poststratification. MrP and MrsP both have a Mean Absolute Error (MAE) of 11.94 and 11.95 percent and a Rooted Mean Squared Error (RMSE) of 12.87 and 12.88 percent respectively. Whereas classical and synthetic BART both have a MAE of 6.1 percent and a RMSE of 7.15 and 7.16 percent respectively.

This demonstrates the predictive power and precision that MrsP and synthetic BART have compared to methods that rely on detailed data.[12] When only age and gender are used as individual-level predictors, BART with or without synthetic poststratification outperform both

---

[11] See the Appendix F for test statistics our predictions of trust measures using the other public opinion surveys where regions are assumed to be representative.

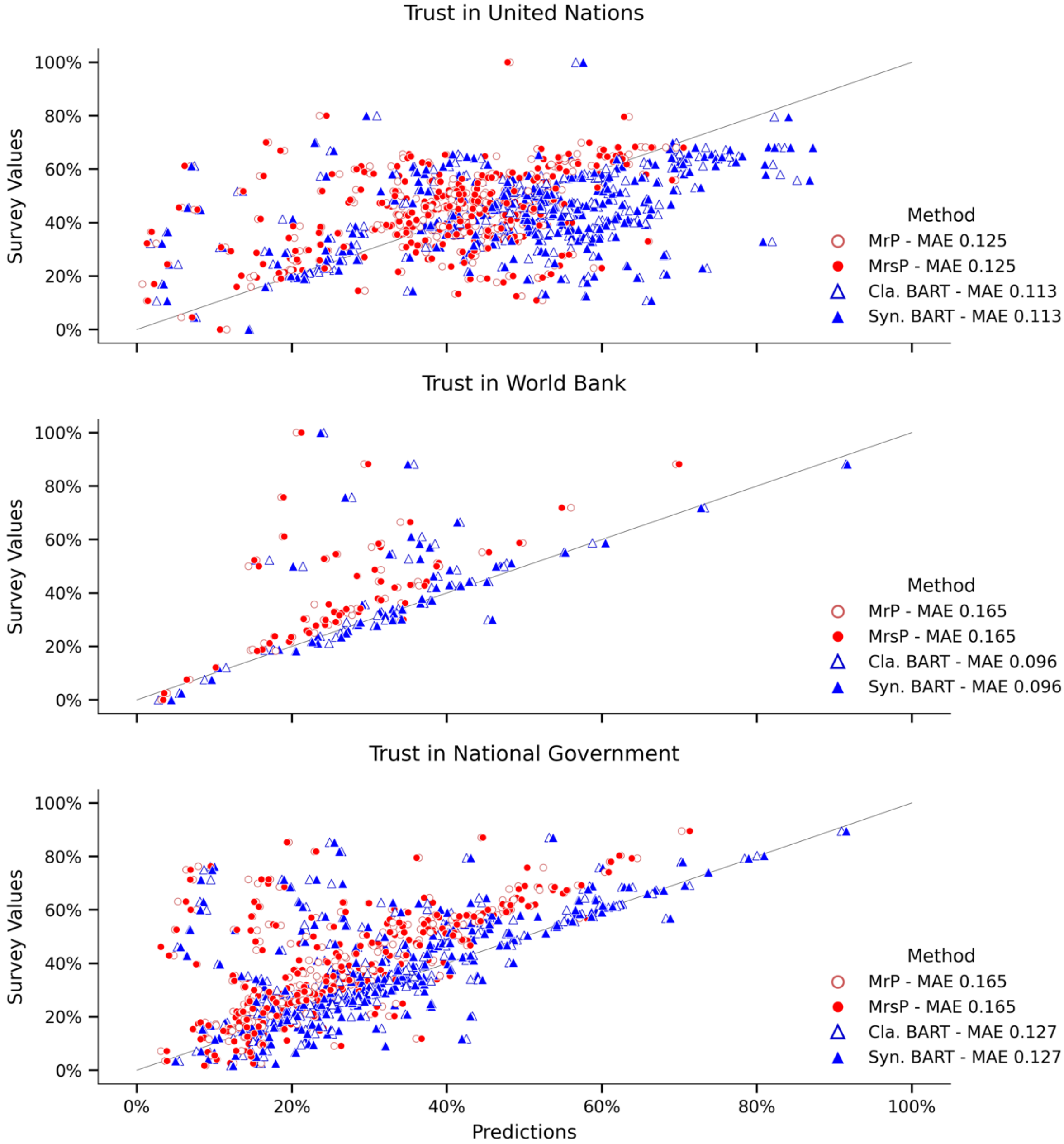
[12] Since we create joints between only two variables the deviation from the true joint distribution is small likely explaining this result.

MrP and MrsP, substantially increasing prediction precision by over 5 percent points. Here, estimations of correlations between covariates using Pearson and Kendall's range from 0.63 to 0.71, indicating strong correlation. The standard deviation of MAE is 9.59 and 9.57 percent and a RMSE of 10.91 and 10.90 percent for classic BART and synthetic BART, respectively. The standard deviation of MAE is 10.71 and 10.74 percent and a RMSE of 11.63 and 11.66 percent for MrP and MrsP, respectively.

Figure 4 shows the mean subnationally representative values for levels of trust in national governments, the UN, and the World Bank, compared to the mean prediction values for the each of the four methods—MrP, MrsP, classical BART and synthetic BART. The synthetic and classical methods each have nearly identical predications. Additionally, BART outmatches the predictive capacity of MrP and MrsP here. The BART estimations are closer to the mean survey values, visible in the shape of the plotted values which fit closer to the 45-degree reference line for each indicator.

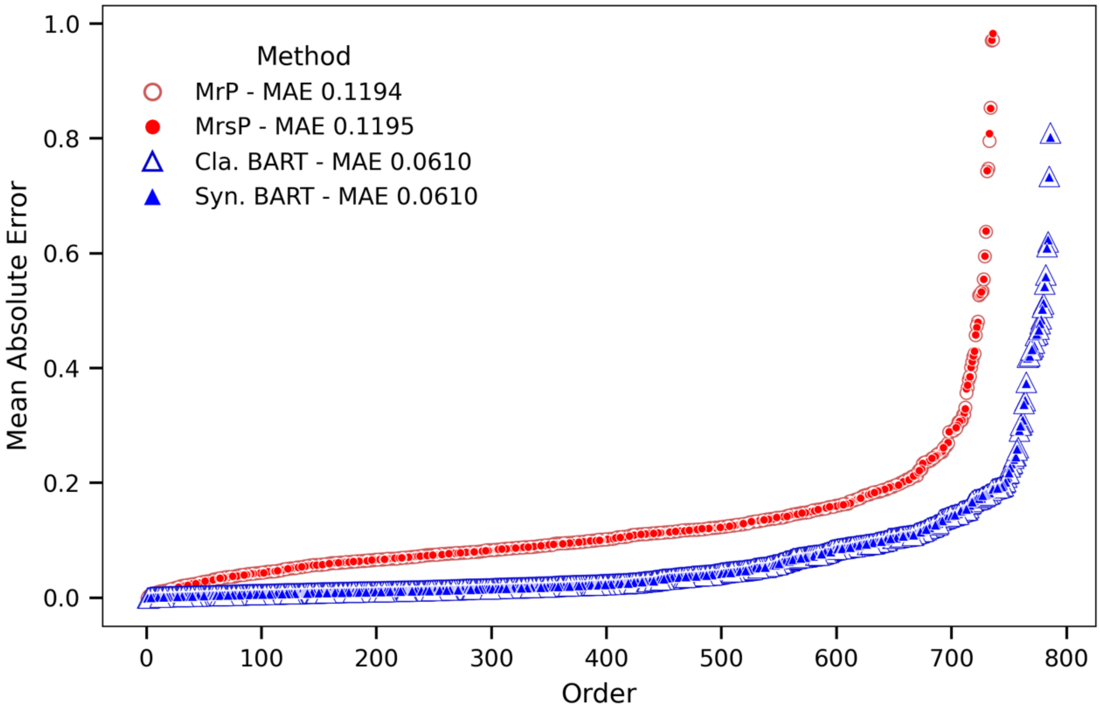
Figure 5 illustrates the prediction precision of each method with an ordered variable index, where MrP and MrsP have a MAE of 11.94 and 11.95 percent and BART with or without synthetic poststratification has a far lower MAE of 6.1 percent. This shows how well BART is able to maintain its good predictive performance relative to MrP and MrsP. Notice that the error values of MrP and synthetic BART overlap their classical counterparts, which illustrates their comparative performance.

Figure 4 MrP, MrsP, Classical BART and Synthetic BART Estimates of EVS5 and WVS7 Values



Note: Cla.=classical. Syn.=synthetic. Data=Subset of Subnational Trust Database, predictions using WVS7 and EVS5 survey data on 328 subnational units (across 47 countries) where subnationally representative data is available (see Section 4 Validation for a description). Predications are made using five predictor variables, country, subnational unit, subnational level, five-year age groups and gender. The 45-degree line is a reference line indicating perfect correspondence. Trust in United Nations: MrP — RSME = 0.141, r = 0.341, τ = 0.318; MrsP — RMSE = 0.141, r = 0.341, τ = 0.319; Classical BART — RSME = 0.127, r = 0.356, τ = 0.333; Synthetic BART —RMSE = 0.127, r = 0.35, τ = 0.315. Trust in World Bank: MrP — RMSE = 0.176, r = 0.549, τ = 0.526; MrsP — RMSE = 0.177, r = 0.549, τ = 0.526; Classical BART — 0.117, r = 0.572, τ = 0.569. Synthetic BART — RMSE = 0.117, r = 0.565, τ = 0.568. Trust in National Government: MrP — RMSE = 0.182, r = 0.473, τ = 0.408; MrsP — RMSE = 0.182, r = 0.473, τ = 0.407; Classical BART — RMSE = 0.153, r = 0.496, τ = 0.421; Synthetic BART — RMSE = 0.152, r = 0.497, τ = 0.431.

Figure 5 Prediction Precision of MrP, MrsP, Classical BART, and Synthetic BART



Note: Cla.= classical. Syn.= synthetic. Order=Ordered variable index of all estimations of trust included in WVS7. Data= Data=Subset of Subnational Trust Database, predictions using WVS7 and EVS5 survey data on 328 subnational units (across 47 countries) where subnationally representative data is available (see Section 4 Validation for a description). Predictions are made using five predictor variables, country, subnational unit, subnational level, five-year age groups and gender. MrP: RMSE = 0.13, r = 0.64; τ = 0.587. MrsP: RMSE = 0.13, r = 0.64; τ = 0.587. Cla. BART: RMSE = 0.075, r = 0.675; τ = 0.619. Syn. BART: RMSE = 0.075, r = 0.675; τ = 0.619.

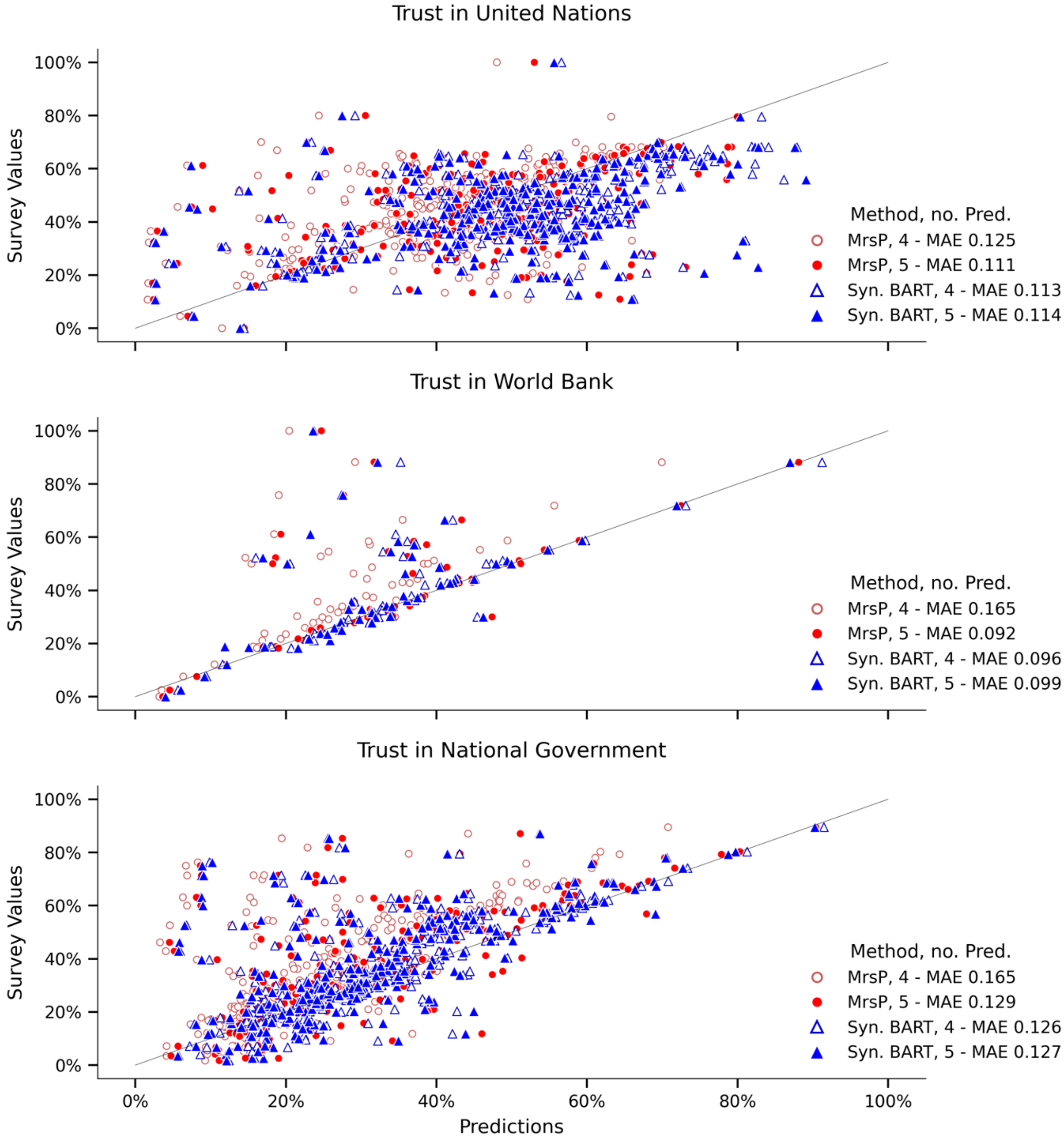
Next, we move forward with presenting the results of adding a third individual-level predictor, marital status, to estimate trust using WVS7 and EVS5 where representative subnational data is available for comparison. Adding marital status helps to test if it is an informative predictor for our models and it allows us to compare the relative performance of each method once again.

We do this to illustrate what the results are for the four methods when estimated with better data quality. To add marital status as a predictor variable, we create synthetic joints between marital status (with the following four categories: single/unmarried, married/union, divorced/separated, and widowed), gender (female/male) and five-year age groups (aged 15 to 85 and over) for all category combinations. We illustrate the mean subnationally representative values compared to WVS7 and EVS5 mean predicted values using MrsP and synthetic BART with and without marital status as an additional predictor when estimating levels of trust in the United Nations, World Bank and national governments (Figure 5) and the prediction precision of each for all trust measures in 328 subnational units (Figure 6).

Further, when adding marital status as an additional individual-level predictor, MrsP provides the average error rates of a MAE of 6.67 percent and a RMSE of 7.79 percent. This is a precision gain of roughly 5 percent points over MrsP without marital status as an additional predictor. BART with marital status as an additional predictor has an MAE of 6.4 percent and RMSE of 7.48 percent compared to BART without marital status as an additional predictor which has a MAE of 6.1 percent and RMSE of 7.15. That is, adding marital status as an additional individual-level predictor does not substantially improve predictive performance when applying synthetic BART, but does improve prediction precision when applying MrsP. Similarly, the correlation between covariates is as strong here as well, although marginally lower. The standard deviation of MAE is 10.73 and 9.57 percent and a RMSE of 11.66 and 10.90 percent for MrsP and synthetic BART without marital status as an additional predictor, respectively. The standard deviation of MAE is 9.91 and 9.66 percent and the standard deviation of RMSE of 11.27 and 10.95 percent for MrsP and synthetic BART with marital status as an additional predictor, respectively.

Figure 6 illustrates the mean values from the WVS7 and EVS5 for levels of trust in national governments, the UN, and the World Bank, compared to the mean prediction values for the each of the synthetic poststratified methods— MrsP and synthetic BART—with and without marital status as an additional predictor. BART without marital status as an additional predictor provides the best fit, while BART with marital status performs relatively similarly.

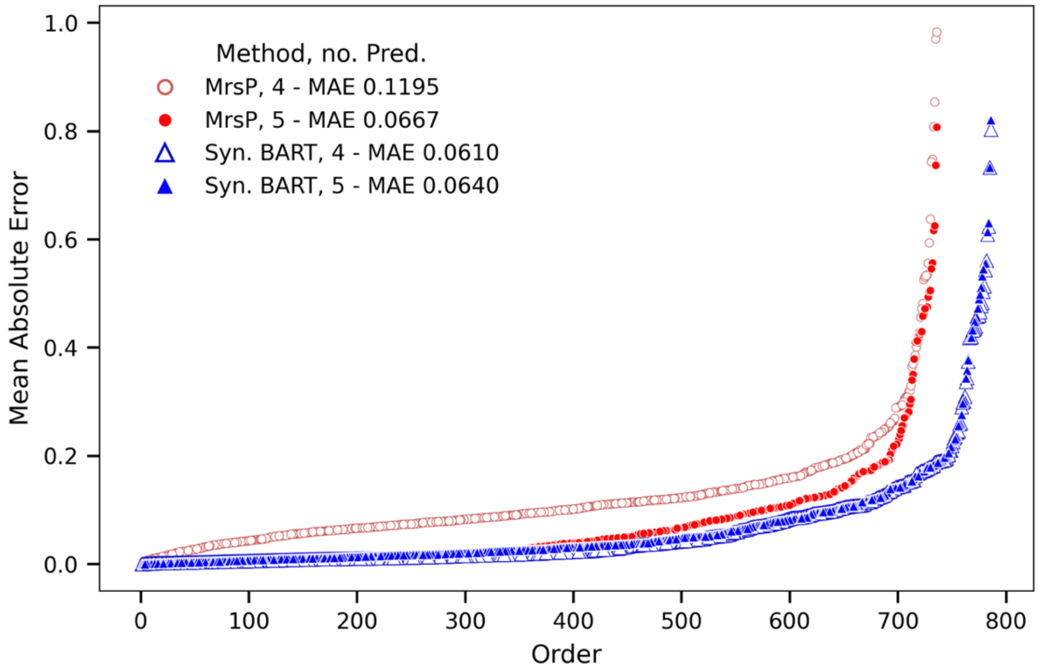
Figure 6 MrsP and Synthetic BART Estimates and WVS7 Values, with and without Marital Status



Note: Data=Subset of Subnational Trust Database, predictions using WVS7 and EVS5 survey data on 328 subnational units (across 47 countries) where subnationally representative data is available (see Section 4 Validation for a description). Predictions are made using five predictors (country, subnational unit, subnational level, five-year age groups and gender) and six predictors, where marital status is an additional predictor. The 45-degree line is a reference line indicating perfect correspondence. Trust in United Nations: MrsP with 4 predictors — RMSE = 0.139,, r = 0.369, τ = 0.333; MrsP with 5 predictors — RMSE = 12.67, ,r = 0.355, τ = 0.323; Synthetic BART with 4 predictors — RMSE = 12.83, r = 0.377, τ = 0.326. Synthetic BART with 5 predictors — RMSE = 0.1300, r = 0.377, τ = 0.326; Trust in World Bank: MrsP with 4 predictors — RMSE= 0.182 r = 0.616,, τ = 0.595; MrsP with 5 predictors—RMSE = 0.112, r = 0.586, τ = 0.612; Synthetic BART with 4 predictors—RMSE = 0.115, r = 0.611,, τ = 0.628; Synthetic BART with 5 predictors — RMSE = 0.119, r = 0.541, τ = 0.579. Trust in National Government: MrsP with 4 predictors — RMSE = 0.182, r = 0.453, τ = 0.385; MrsP with 5 predictors — RMSE = 0.150, r = 0.460, τ = 0.7384; Synthetic BART with 4 predictors — RMSE = 0.149, r = 0.481, τ = 0.412; Synthetic BART with 5 predictors— RMSE = 0.150, r = 0.453, τ = 0.389.

Figure 7 shows the prediction precision of MsP and synthetic BART with or without marital status as a third individual-level predictor with an ordered variable index. We find that adding marital status as an additional predictor has an improved performance effect for MrsP, allowing MrsP (MAE 6.67 percent and RMSE of 7.79 percent) to surpass MrsP without marital status as an additional predictor (MAE 11.95 percent and RMSE of 12.88 percent), Interestingly, adding marital status as an additional predictor does not improve prediction precision for synthetic BART (MAE of 6.4 percent and RMSE of 7.48 percent) compared to synthetic BART with only age and gender as individual-level predictors (MAE of 6.1 percent and RMSE of 7.15). This shows how well BART can sustain its predictive power with relatively few predictors and without exhaustive data.

Figure 7 Prediction Precision of MrsP and Synthetic BART with and without Marital Status



Note: Order=Ordered variable index. Data= Subset of Subnational Trust Database, predictions using WVS7 and EVS5 survey data on 328 subnational units (across 47 countries) where subnationally representative data is available (see Section 4 Validation for a description). Predictions are made using five predictors (country, subnational unit, subnational level, five-year age groups and gender) and six predictors, where marital status is an additional predictor. MrsP with 4 predictors: RMSE = 0.129, r = 0.678; τ = 0.626. MrsP with 5 predictors: RMSE = 0.078, *r* = 0.660; τ = 0.663. Syn. BART with 4 predictors: RMSE = 0.072, *r* = 0.709; τ = 0.656. Syn. BART with 5 predictors: RMSE = 0.075, *r* = 0.654; τ = 0.592.

Conclusions

From climate change to health pandemics, the challenges of solving political problems confronting contemporary society are daunting, particularly when the political institutions addressing them are not trusted by general publics. To push forward the study of political trust in national and international institutions, this article has introduced the Subnational Trust Database, which for the first time includes *subnational-level* political trust measures that are comparable across countries worldwide. We propose to estimate these measures by using BART, as this method performs better than MrP and MrsP in the presence of small-*n* and unrepresentative survey data. This dataset has implications for political science research relying on subnational political trust measures, and especially for ongoing discussions in political economy, social legitimacy, and peace and conflict research.

Political economy research has increasingly explored the consequences of subnational socioeconomic conditions for subnational-level attitudes toward national and international governing institutions (Rodriguez-Pose 2018; Stein et al. 2020), including populist attitudes and voting behavior (Monnat and Brown 2017; Gavenda and Umit 2016; Schraff 2020). As this literature has predominantly focused on the EU or the US, our global-scale dataset can be used to promote cross-national research at world-scale on the relationship between subnational socioeconomic structures and political trust and related attitudes.

The study of social legitimacy in comparative politics and International Relations has focused on a variety of national (Gilley 2006; Esaiasson et al. 2016; Tyler 2006) and international institutions (Johnson 2011; Marks and Hooghe 2005; Dellmuth et al. 2022). This literature has a tendency to view the nation state to be the appropriate reference point for individual political beliefs. However, we know from other strands of research on American states and EU support that subnational beliefs vary greatly within countries (Leeman and

Wasserfallen 2017; Boniecki 2019). Subnational trust measures are useful to advance existing theories to also include subnational level trust for outcomes such as the success of UN peacekeeping (e.g., von Billerbeck 2017; Whalan 2017).

Finally, our findings could contribute to peace and conflict research on the performance of aid and disaster management. For instance, local subnational trust has been shown to be central for the success of humanitarian operations after extreme weather events (Petrova 2022; De Juan and Hänze 2021). While it is generally believed that trust in political institutions matters, it remains understudied due to data limitations. Further research is needed that connects subnational trust to the rich subnational datasets available on aid, disasters, and conflict (see von Uexkull and Buhaug 2021 for an overview). Our Subnational Trust Database is intended to inspire future studies pursuing this aim.

References

Bisbee, James. 2019. BARP: Improving Mister P Using Bayesian Additive Regression Trees. *American Political Science Review* 113(4): 1060–1065.

Broniecki, Philipp, Lucas Leemann, and Reto Wüest. 2022. Improved Multilevel Regression with Poststratification through Machine Learning (autoMrP). *The Journal of Politics* 84(1): 597-601.

Chipman, Hugh A., George, Edward I., and Robert E McCulloch. 2012. "BART: Bayesian Additive Regression Trees." *Annals of Applied Statistics* 6(1): 266-298.

De Juan, Alexander, and Niklas Hänze. 2021. "Climate and Cohesion: The Effects of Droughts on Intra-Ethnic and Inter-Ethnic Trust." *Journal of Peace Research* 58(1), 151–167.

Dellmuth, Lisa, and Jonas Tallberg. 2015. "The Social Legitimacy of International Organisations: Interest Representation, Institutional Performance, and Confidence Extrapolation in the United Nations." *Review of International Studies* 41(3): 451-475.

Dellmuth, Lisa, Jan A. Scholte, Jonas Tallberg, and Soetkin Verhaegen. 2022. *Citizens, Elites, and the Legitimacy of Global Governance*. Oxford, UK: Oxford Academic.

Dellmuth, Lisa, and Bernd Schlipphak. 2020. Legitimacy beliefs towards global governance institutions: a research agenda. *Journal of European Public Policy*, 27:6, 931-943.

Easton, David. 1975. "A Re-Assessment of the Concept of Political Support." *British Journal of Political Science* 5(4): 435-57.

Gavenda, Mario, and Umit, Resul. 2016. “The 2016 Austrian Presidential Election: A Tale of Three Divides." *Regional & Federal Studies* 26, 419–432.

Georgiadou, Vasiliki, Lamprini Rori, and Costas Roumanias. Georgiadou V, Rori L and Roumanias C. 2018. “Mapping the European Ffar Rright in the 21st century: a Meso-Llevel Aanalysis.” *Electoral Studies* 54, 103–115.

Ghitza, Yair, and Andrew Gelman. A (2013.) “Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups.”. *American Journal of Political Science* 57, 762–776.

Gilley, Bruce. 2006. The Meaning and Measure of State Legitimacy: Results for 72 Countries. European Journal of Political Research 45(3): 499–525.

Harteveld, Eelco, Tom van der Meer, and Catherine E. De Vries. 2013. In Europe We Trust? Exploring Three Logics of Trust in the European Union. *European Union Politics* 14(4): 542–565.

Hooghe, Liesbet, and Gary Marks. 2005. “Calculation, Community and Cues: Public Opinion on European Integration.” *European Union Politics* 6(4)),: 419–443.

Hoover, Joe, and Morteza Dehghani. 2020. “The Big, the Bad, and the Ugly: Geographic Estimation with Flawed Psychological Data.” *Psychological Methods 25*(4), 412–429.

Johnson, Tana. 2011. Guilt by Association: The Link between States’ Influence and the Legitimacy of Intergovernmental Organizations. Review of International Organizations 6(1): 57–84.

Kapelner, Adam, and Justin Bleich. 2013. “Prediction with Missing Data via Bayesian Additive Regression Trees.” *Canadian Journal of Statistics*, 43(2), 224-239.

Lax, Jeffrey R., and Justin H. Phillips. 2009. “How Should We Estimate Public Opinion in the States?” *American Journal of Political Science* 53(1): 107–121.

Leemann, Lucas and Fabio Wasserfallen. 2017. “Extending the Use and Prediction Precision of Subnational Public Opinion Estimation.” *American Journal of Political Science* 4, 1003–1023.

Leemann, Lucas and Fabio Wasserfallen. 2020. “Measuring Attitudes – Multilevel Modeling with Poststratification (MrP).” In *Sage Handbook of Research Methods in Political Science and International Relations*, eds. Luigi Curini and Robert J. Franzese Jr. (California, US, SAGE Publications).

Levendusky, Matthew S., Jeremy C. Pope, and Simon D. Jackman. 2008. “Measuring District-Level Partisanship with Implications for the Analysis of U.S. Elections.” *Journal of Politics* 70(3): 736–53.

Lipps, Jana, and Dominik Schraff. 2021. “Regional Iinequality and Iinstitutional Ttrust in Europe.” *European Journal of Political Research*, 60, 892-913.

Marien, Sofie, and Marc Hooghe. 2011. “Does Political Trust Matter? An Empirical Investigation into The Relation Between Political Trust and Support for Law Compliance.” *European Journal oOf Political Research* 50, 267–291.

Mayne, Quinton, and Alexia Katsanidou. 2022. “Subnational Economic Conditions and Changing Geography of Mass Euroscepticism: A Longitudinal Analysis.” *European Journal of Political Research.*

Monnat, Shannon M., and Brown, David L. 2017. “More than a Rural Revolt: Landscapes of Despair and the 2016 Presidential Election.” *Journal of Rural Studies.* 55, 227–236.

Montgomery, Jacob M., and Santiago Olivella. 2018. "Tree-Based Models for Political Science Data." *Americans Journal of Political Science*, 62(3), 729-744.

Norris, Pippa (ed.). (1999). "Introduction: The Growth of Critical Citizens and its Cconsequences." In *Critical Citizens: Global Support for Democratic Government* (Oxford: Oxford University Press), 1-28.

Norris, Pippa. 2022. *In Pratice of Skepticism: Praise but Verify*. New York: Oxford University Press.

Ornstein, Joseph T. 2020. Stacked Regression and Poststratification. Political Analysis 28(2): 293–301.

Petrova, Kristina. 2022. Floods, communal conflict and the role of local state institutions in Sub-Saharan Africa. *Political Geography* 92, 102511.

Rodríguez-Pose, Andres. (2018). "The Revenge of the Places That Don't Matter (and What to do About It)." *Cambridge Journal of Regions, Economy and. Society.* 11, 189–209.

Schraff, Dominik, Ioannis Vergioglou, and Buket Buse Demirci. 2022. "The European NUTS-Level Election Dataset: A Tool to Map the European Electoral Geography, Party Politics." *Party Politics*.

Stein, Jonas, Buck, Marcus, and Hilde Bjørnå. 2019. "The Centre–Periphery Dimension and Trust in Politicians: The Case of Norway." *Territory Politics, Governance.* 1–19.

Tallberg, Jonas, and Michael Zürn. 2019. "The Legitimacy and Legitimation of International Organizations: Introduction and Framework." *Review of International Organizations* 14(4), 581–606.

Thomassen, Jacques, Rudy Andeweg, and Carolien van Ham. 2017. Political trust and the decline of legitimacy debate: a theoretical and empirical investigation into their interrelationship. In Handbook on Political Trust, edited by Sonja Zmerli and Tom W.G. van der Meer (Cheltenham: Edward Elgar, pp. 509–525).

Toshkov, Dimiter. 2015. "Exploring the Performance of Multilevel Modeling and Poststratification with Eurobarometer Data." *Political Analysis* 23, 455–460.

Tyler, Tom R. 1990. *Why People Obey the Law: Procedural Justice, Legitimacy, and Compliance*. New Haven, CT: Yale University Press.

Tyler, Tom R. 2006. "Psychological Perspectives on Legitimacy and Legitimation." *Annual Review of Psychology* 57, 375–400.

Tyler, Tom R., and Yuen J. Huo. 2002. *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. New York: Russell Sage Found.

V-Dem. 2021. *Autocratization Turns Viral*: *Democracy Report 2021.* Gothenburg: V-Dem Institute, University of Gothenburg.

Von Billerbeck, Sarah B. K. 2017. UN Peace Operations and Conflicting Legitimacies, Journal of Intervention and Statebuilding, 11:3, 286-305

Von Uexkull, Nina, and Halvard Buhaug. 2021. "Security Implications of Climate Change: A Decade of Scientific Progress." *Journal of Peace Research* 58(1).

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." *International Journal of Forecasting* 31(3): 980–991.

Warshaw, Christopher, and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74(1), 203–219.

Whalan, Jeni. 2017. "The Local Legitimacy of Peacekeepers." *Journal of Intervention and Statebuilding* 11(3), 306-20.

Zmerli, Sonja. 2014. "Political Trust". In: Michalos, A.C. (eds) Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht (pp. 4887-4889).