Modeling public opinion over time and space: Trust in state institutions in Europe, 1989-2019

Marta Kołczyńska[1] · Paul-Christian Bürkner[2,5] · Lauren Kennedy[3,6] · Aki Vehtari[4]

[1]Institute of Political Studies of the Polish Academy of Sciences

[2]Cluster of Excellence SimTech, University of Stuttgart

[3]Econometrics and Business Statistics, Monash University

[4]Department of Computer Science, Aalto University

[5]Department of Statistics, Technical University of Dortmund

[6]Mathematical Sciences, University of Adelaide

Combining public opinion data from different sources enables new cross-national and longitudinal research, but is accompanied by unique challenges related to the comparability of the source survey data. The analytic strategy we propose relies on Bayesian explanatory item response theory models to address differences in the measurement of attitudes, and poststratification that uses administrative population data to improve the quality of estimates and correct for differences in sample representativeness. Partially pooled models with data from all countries would be prohibitively slow, so we estimate separate by-country models in a way that maintains comparability of estimates across countries. We apply this strategy to data from 13 cross-national research projects from 27 European countries to estimate trajectories of political trust between 1989-2019.

Keywords: political trust; Bayesian models; splines; item response theory; poststratification

1 Introduction

Combining public opinion data from different sources creates new opportunities for social science research by increasing the scope of possible analyses, both in terms of geographical and time coverage. Of particular interest are studies that cover extended periods of time enabling research on the macro-level causes and consequences of public opinion. Such studies face challenges associated with joint analyses of data that have not been collected with comparability in mind and differ with regard to measurement of the key variables and representativeness of the survey samples.

Responding to these challenges, our paper proposes an analytic strategy for estimating country trajectories in public opinion that combines Bayesian explanatory item response theory (IRT) models to address the differences in the measurement of attitudes in different surveys, with poststratification that – by using population data from administrative sources – corrects for discrepancies in the repre-

sentativeness of survey samples. The strategy we propose makes several contributions. First, our model constitutes an alternative to the group-level IRT models proposed earlier (e.g., Claassen, 2019; Caughey et al., 2019; Solt, 2020), and more faithfully reflects the data generation process by incorporating individual (respondent) level random effects that account for the fact that the same respondents provide answers to more than one survey question – in this case regarding trust in different institutions. Second, we demonstrate that it is possible to obtain same-scaled estimates of country trends from separate by-country models. The need for substantial computational resources is one of the barriers to broader applications of Bayesian models to cross-national public opinion research, and splitting up the analysis into smaller chunks while maintaining the comparability of the resulting estimates helps lower this barrier. Third, the poststratification procedure we use accounts for an additional source of uncertainty from socio-demographic statistics, which – if one wants to poststratify by other characteristics than age and sex – are often incomplete and themselves rely on (large-scale and high quality) surveys. Our modeling strategy – in its entirety or selected components – is broadly applicable to aggregating measures of attitudes and opinions from diverse collections of survey datasets characterized by different quality, and exemplifies

Corresponding author: Marta Kołczyńska, Institute of Political Studies of the Polish Academy of Sciences, Warsaw, Poland (Email: mkolczynska@isppan.waw.pl)

the advantages of combining survey data with administrative or other population data.

We apply this strategy to data on political trust, measured with items about trust in the national parliament, political parties, and justice system, from 13 cross-national survey projects from 27 European countries between 1989 and 2019. Our findings are broadly consistent with previous research on political trust, but provide stronger evidence of patterns in a framework that facilitates further rigorous investigation of the correlates of trust – both overall and of differences in trust between social groups – in a longitudinal perspective. The data support the thesis about trendless fluctuations in political support (Norris, 2011; Van de Walle et al., 2008) rather than any clear long-term tendency. Specifically, there is no evidence for a consistent decline in political trust in the last 30 years in Western Europe. In each of the studied countries the trajectory of political trust is unique, which supports the conceptualization of political trust as primarily driven by national rather than international developments. Our results also confirm prior findings about the stability of country and regional rankings of political trust. In addition to overall trust levels, we examine differences between population groups by sex, age, and education, expecting that differences in the policy preferences, values, and political representation between these groups manifest in differences in political trust levels.

The paper is organized as follows. We start with an overview of longitudinal research on public support, highlighting its limitations resulting from reliance on single data sources, such as single cross-national survey projects, which are better suited for cross-national than for longitudinal analyses. Next, we present our analytic approach starting with the source data that consist of selected survey items from 13 survey projects. In describing the models for estimating average political trust, we highlight the differences and advantages of our approach compared to earlier studies. Further, we discuss poststratification and the decisions we made to address the limitations of the survey and population data. Following the presentation of results, i.e., country trajectories in societal levels of political trust as well as the changes in trust gaps by sex, age, and education, we conclude with a summary and an outline of opportunities for future methodological and substantive research.

2 Longitudinal research on political support

Owing to its link to political legitimacy and democratic governance, trust in state institutions is considered a source of stability and legitimacy of political systems (Easton, 1975; Hetherington, 1998; Klingemann, 1999; Norris, 2002; Seligson, 2002). Low levels of political trust have been interpreted as threatening the legitimacy of democratic regimes (Dalton, 2004), exhausting the reservoir of support for institutions and authorities, and lowering compliance with government regulation and civic duty (Letki, 2007; Tyler, 1990). Most recently, political trust proved instrumental in explaining levels of compliance with measures aimed at preventing the spread of infections in the COVID-19 pandemic (Oksanen et al., 2020). Despite the rich literature on micro- and macro-level correlates of political trust, there is a glaring scarcity of empirical tests beyond cross-sectional analyses (Torcal, 2017; van der Meer and Zmerli, 2017), to a large extent due to the lack of reliable data on levels of political trust over time.

Much of comparative research on political trust uses survey data, of which the vast majority rely on cross-national comparisons. This literature is extensive and there is little consensus on how political trust should be measured. Conventionally, substantive research on political trust measures it either with a single item, most often trust in parliament (cf. Catterberg, 2013; Dalton et al., 2010; Závecz, 2017), or multi-item scales created using factor analysis, principal component analysis or a simple sum of items (for a review see Breustedt, 2018), reflecting the debate about the dimensionality of political trust. Strict measurement equivalence tests of unidimensional models with broad sets of items such as trust in the government, police and civil service, in addition to trust in political institutions, typically fail to find measurement invariance (van der Meer and Ouattara, 2019; Breustedt, 2018), which indicates that items for the trust scales need to be selected carefully. Other authors have used less restrictive alternatives to strict measurement invariance testing, such as the alignment method (Asparouhov and Muthén, 2014), and concluded that for selected items, trust scales can be meaningfully compared (Coromina and Bartolomé Peral, 2020). It is worth noting that longitudinal research on political trust in the United States is somewhat distinct, in that it often uses single items from the American National Election Study, which has included trust questions since the late 1950s (Citrin and Stoker, 2018; Hetherington, 1998). A separate strand of typically single-country research also studies political trust with panel data, which typically cover relatively short periods of time (de Blok and Kumlin, 2021; Torcal and Carty, 2022; Kołczyńska and Sadowski, 2022).

Cross-country studies of political trust are typically limited by the scope of survey projects, because methodological differences, among others in the measurement of political trust, complicate the joint analysis of data from different projects. In Europe, the two most popular academic survey projects used in research on political trust are the European Social Survey (carried out biennially since 2002) and the European Values Study (every 8–9 years since 1981). Consequently, even the few longitudinal cross-national studies on political support in Europe are limited by relying either

Table 1

Description of survey projects included in the analysis

Project	N observations	N countries	N surveys	N waves	Year from	Year to
ASES	8430	9	9	1	2000	2000
CCEB	43,591	10	50	5	2001	2004
CDCEE	15,398	11	16	2	1990	2001
EB	918,724	25	1000	44	1996	2019
EQLS	78,665	26	76	3	2007	2016
ESS	312,049	27	193	9	2002	2019
EVS	123,696	27	95	4	1990	2018
INTUNE	13,845	16	16	1	2009	2009
ISSP	66,351	22	51	3	1990	2010
LITS	40,339	16	38	3	2006	2016
NBB	14,118	3	12	4	1993	2004
NEB	34,609	8	37	6	1992	2004
WVS	46,311	20	44	5	1989	2013
Total	1,716,126	–	1637	90	1989	2019

Survey refers to data collected from the same sample within the same country as part of the same fieldwork, e.g. ESS Round 1 Poland.
*ASES* Asia-Europe Survey, *CCEB* Candidate Countries Eurobarometer, *CDCEE* Consolidation of Democracy in Central and Eastern Europe, *EB* Eurobarometer, *EQLS* European Quality of Life Survey, *ESS* European Social Survey, *EVS* European Values Study, *INTUNE* Integrated and United, *ISSP* International Social Survey Programme, *LITS* Life in Transition Survey, *NBB* New Baltic Barometer, *NEB* New Europe Barometer, *WVS* World Values Survey

on dense time series since 2002 or sparse time series since 1981, thus failing to fully exploit the potential of available cross-national survey data. In the case of more volatile types of political support, such as trust in state institutions, using sparse time series is associated with a risk of overlooking short-term fluctuations.

Model-based estimates of societal levels of political trust based on data from multiple cross-national survey projects, which we present in this paper, help overcome the challenges resulting from the sparsity of survey coverage, and enable rigorous analyses of political trust, its causes and consequences. In addition to overall political trust, we analyze its level by age, sex, and education. Examining the relative levels in political trust—and their changes—between societal groups characterized by different values, policy preferences, forms and levels of political engagement, labor market positions, as well as degrees of political representation, is important both theoretically—to understand the nature of political trust and its driving factors, and practically—to monitor trust as an aspect of political legitimacy and engagement in society. Specifically, differences in political trust by sex may shed light on the extent to which trust reflects differences in status in politics and in society between men and women—an aspect of political trust that has so far received little researcher attention. Analysis of differences in political trust by age speaks to the debate

about the decline in political legitimacy of democratic governments among younger cohorts, which—if true—could threaten the future of democracy (Foa and Mounk, 2016, 2017).

Differences in political trust by education level reflect the democratic utility of political trust. Prior research found that differences in political trust between education groups depend on the country's democratic quality and on the pervasiveness of public-sector corruption: in more democratic and less corrupt countries education tends to be positively associated with trust (Hakhverdian and Mayne, 2012; Kołczyńska, 2020). Educated individuals are better equipped to assess the integrity and performance of the political regime; higher education is also associated with more liberal values and democratic orientations (Bobo and Licari, 1989; Hyman and Wright, 1979). The longitudinal version of the trust-education-democracy hypothesis states that as countries become more democratic, the positive association between trust and education becomes stronger, while in countries that become less democratic, the association becomes weaker or negative. Our analysis addresses both the cross-national and longitudinal hypotheses about the link between democraticness and the education gap in political trust.
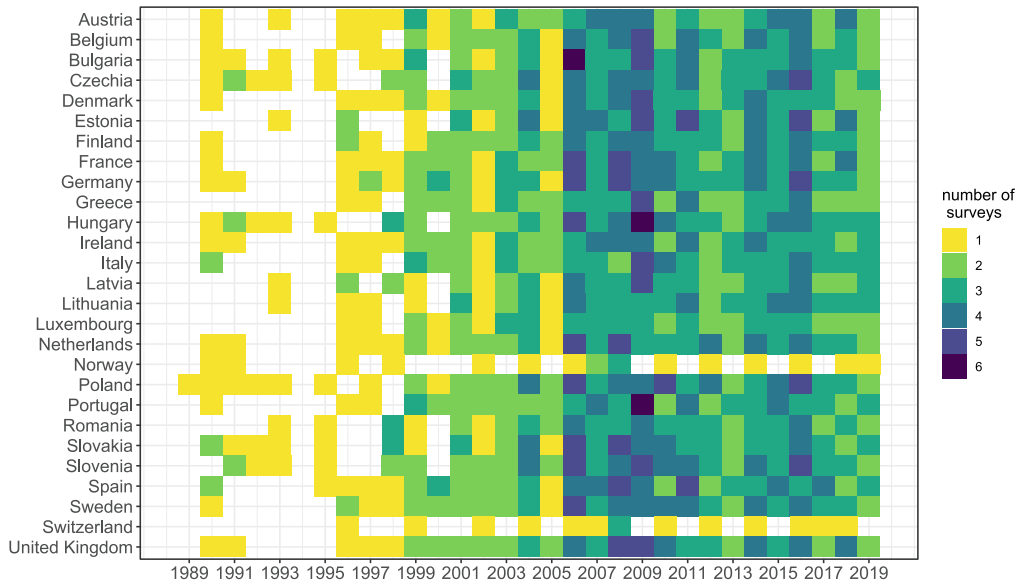
Fig. 1

Number of surveys included in the analysis by country and year

3 Method

Contributing to the literature on long-term trends in mass public opinion, we propose the following analytic strategy. First, we combine data from multiple cross-national survey projects carried out in Europe between 1989 and 2019 which feature questions on political trust. Second, we model responses to trust items as manifestations of underlying latent variables using IRT models (Embretson and Reise, 2013; van der Linden and Hambleton, 2013) in a way that accounts for the ordinal character of trust measurement and accommodates varying scale lengths across survey projects. Third, we use non-linear multilevel models to describe the latent process of the ordinal responses. Finally, we apply poststratification to improve the quality of the estimates of the country-year levels of political trust and to correct for the differences in representativeness of the survey samples. We elaborate on these elements in the sections below.

We adopt a fully Bayesian framework, which offers the necessary modeling flexibility. The analysis is performed in R (R Core Team, 2018) with the brms package (Bürkner, 2017), which provides a user-friendly and flexible interface to the probabilistic programming language Stan (Stan Development Team, 2020). The analysis constitutes an extension of multilevel regression and poststratification (MRP) to non-linear models. MRP (Gelman and Little, 1997) is rarely applied in cross-national research, but it can be particularly useful when analyzing surveys collected with different sampling strategies that have varying representativeness.

3.1 Data

We use data from 13 cross-national survey projects carried out in Europe between 1989 and 2019 which included questions about trust in institutions. The list of projects is presented in Table 1, while the references to the specific datasets are available in the web-based supplementary materials. All projects use samples that aim to represent entire adult populations of the respective countries, with data collected primarily via face-to-face interviews (with the exception of a few surveys in ISSP, which used self-completion questionnaires). The survey data we use vary with regard to their quality (e.g. Jabkowski et al., 2021). In the proposed framework, differences in sample representativeness are corrected via poststratification, as described in the section on using the model to estimate population trends.

The analysis covers 27 European countries, for which both survey data and population statistics necessary for poststratification are available: Austria, Belgium, Bulgaria, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, the Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, and the United Kingdom. Fig. 1 presents the availability of surveys with any of the necessary trust items for each country and year, indicating very good coverage of most countries since 2000 owing primarily to ESS, EQLS, and EB (with the exception of Norway and Switzerland, which do not participate in EB) and sparse coverage in the 1990s.

These gaps could be filled with national surveys which we have not attempted.

We use survey questions about the respondents' trust in the national parliament, justice system, and political parties—three key institutions of contemporary democracies that have been most frequently used as indicators of generalized political trust (Breustedt, 2018) and are the most common political trust items in cross-national surveys. We decided to not include trust in government, because by definition governments only directly represent part of the electorate and trust in government is closer in meaning to support for incumbent office-holders than for regime institutions (cf. Norris, 2017). The wording of trust items as well as information about sample types of the surveys included in the analysis are provided in the web-based supplementary materials; documentation of data processing is part of the replication materials.

3.2 Reconciling different response scales

Studies that rely on data on political attitudes from different cross-national survey projects most commonly deal with the different response scales by resorting to dichotomization or linear re-scaling of the responses to a common range (cf. Christmann, 2018; Dassonneville, 2021; Griffin et al., 2021; Závecz, 2017).[1] While practical, both approaches come at a cost. Dichotomization entails information loss and requires a decision about middle points on odd-numbered scales. Linearization relies on strong assumptions about the equidistant character of ordinal rating scales thus overestimating the amount of information in the data, and may lead to inflated error rates, distorted effect sizes, and even inversions of effects (Liddell and Kruschke, 2018; Bürkner and Vuorre, 2019; Cichocki and Jabkowski, 2022). The latter is particularly problematic when aggregating characteristics in order to explore the variation in group-level means, and the situation is likely exacerbated by combining surveys items with response scales of varying lengths and directions, which—as the rich literature on survey methodology shows—have different measurement properties (see e.g. Yan et al., 2018).

Our analysis relies on IRT models, whose application to modeling mass public opinion based on diverse survey questions was proposed by McGann (2014) as a superior alternative to the dyad ratios algorithm (Stimson, 1999, 2018), and further extended by Caughey et al. (2019), Claassen (2019), and Solt (2020) to model policy liber-

---

[1] Recently, de Jonge et al. (2021) and Singh (2022) have proposed new ways of dealing with changing verbal response scales in order to reconstruct time trends, but their applications are yet to be extended to account for changing scales in a cross-national context.

alism and democratic support. The next section describes the model we develop and its benefits over previous approaches.

3.3 Modeling

Our applied Bayesian statistical model consists of multiple components: (1) an ordinal response model with special restrictions on the ordinal thresholds to ensure identification, (2) a flexible predictor term including hierarchical splines to model the change in trust over time while taking individuals' characteristics (age, sex, education) as well as project bias into account, and (3) weakly informative prior distribution to help preventing overfitting and improving convergence. For reasons detailed later, we specify a separate model for each country.

To model individuals' survey responses, we used cumulative ordinal model (e.g., Samejima, 1997; Bürkner and Vuorre, 2019), which is commonly applied in various fields of the social and natural sciences to model ordinal data, for example, in IRT or multilevel modeling (e.g., van der Linden and Hambleton, 2013; Bürkner, 2021). The cumulative model assumes that the observed ordinal response y (leaving out indices for now to improve readability) originates from the categorization of a latent continuous variable $\widetilde{y}$. That is, there are latent thresholds $\tau_k (1 \leq k \leq K)$ which partition the values of $\widetilde{y}$ into the $K + 1$ observable, ordered categories of y. More formally, this can be expressed as

$$y = k \Leftrightarrow \tau_{k-1} < \widetilde{y} \leq \tau_k \tag{1}$$

for $-\infty = \tau_0 < \tau_1 < ... < \tau_K < \tau_{K+1} = \infty$. We denote $\tau = (\tau_1, ..., \tau_K)$ as the vector of thresholds. Assuming an additive model on the latent scale and a certain cumulative distribution function F of $\widetilde{y}$, this translates to

$$\Pr(y = k \mid \mu) = F(\tau_k - \mu) - F(\tau_{k-1} - \mu) \tag{2}$$

(e.g., see Bürkner and Vuorre, 2019), with μ being the additive predictor term to be specified later. The choice of F is mostly arbitrary and here we choose the cumulative distribution function Φ of the standard normal distribution (also known as probit response function). This has the advantage that the latent scale is interpretable as standardized z-values, a scale we assume is familiar to most readers.

Clearly, it is neither possible nor sensible to estimate one separate threshold vector τ for each observation, so typically τ is assumed constant over the whole dataset effectively replacing the model's overall intercept (e.g., Bürkner and Vuorre, 2019). However, for the present data, such an overall τ is not sensible either as surveys vary in the number of response categories of their administered trust items.

Also, response processes may vary as a function of item properties such as specific item formulations. For these reasons, we model one threshold vector τ_{ji} for every combination of survey j and political institution i. We also investigated more restrictive threshold choices, that is, thresholds assumed constant across surveys within the same project or even constant across all projects administrating items with the same number of response categories. However, these more restrictive models had worse predictive performance in almost all countries and were thus not considered further.

The choice of assuming separate threshold vectors per survey implies that thresholds are allowed to vary over time, which is a reasonable assumption per se. However, without further constraints, this leaves the mean changes of trust over time (detailed below) to be collinear with the means of the threshold vectors. To avoid this identification issue, we apply a sum-to-zero constraint to each threshold vector τ_{ji}, that is, enforce

$$\sum_{k=1}^{K_{ij}} \tau_{jik} = 0 \qquad (3)$$

for all surveys j and institutions i. Intuitively, this constraint can be understood as the assumption that (latent) mean changes of trust responses over time are attributable to actual changes in trust and not to changes in item content or item interpretation. It also ensures comparability of the overall trust levels across countries. Equivalent constraints with the same goal have been imposed also by others in the context of modeling ordinal survey responses over time (cf. Caughey et al., 2019, p. 8; Solt, 2020, eq. 11). While needed here for identification, this assumption implies a strong equivalence of the overall item difficulties across countries and time, a fact that we will come back to in the Discussion.

In the following, we are going to index projects by q, persons by p, and time (in years) by t. In our model, we compute the latent mean mean $\mu_{iqp}(t)$ as

$$\mu_{iqp}(t) = b_0 + b_q + b_i + b_{q:i} + b_p + f_1(t) + \\ f_2\left(t,\ \text{age}_p\right) + f_3\left(t, \text{sex}_p\right) + f_4\left(t, \text{education}_p\right). \qquad (4)$$

Here, b_0 is the mean trust level across time, b_q and b_i are project- and political institution-specific deviations from the mean (fixed effects with a hard sum-to-zero constraint across projects and institutions for identification, $\sum_{q=1}^{Q} b_q = 0$ and $\sum_{i=1}^{I} b_i = 0$), $b_{q:i}$ is the deviation from the mean due to the project-institution interaction (fixed effects with a hard sum-to-zero constraint, $\sum_{q=1}^{Q} \sum_{i=1}^{I} b_{q:i} = 0$), b_p is the person-specific deviation from the mean (random effects with a soft sum-to-zero constraint via a hierarchical normal

prior, $b_p \sim \text{normal}(0, \sigma_P)$ such that σ_P is also estimated from the data; Bürkner, 2021).

Further, f_1 to f_4 are unknown functions of time to be estimated as part of the model, where f_1 represents the average changes of latent political trust across groups, while f_2, f_3, f_4 represent the corresponding differences across age, sex, and education groups, respectively. We used a thin-plate spline (Wood, 2003) for f_1 and hierarchical factor-interaction splines (Wood, 2017; see also Pedersen et al., 2019) for f_2 to f_4, with the latter varying across age, sex, and education categories, respectively. We chose to approximate the latent trend components with such smoothing splines because of their favorable (linear) scaling with the number of observations, their built-in regularization to reduce the danger of overfitting (Wood, 2004), as well as their efficient implementations available via the R packages mgcv (Wood, 2017) and brms (Bürkner, 2017). In more detail, thin-plate splines constitute a class of smoothers that can be considered optimal solutions to the variational problem of balancing accuracy (fit to the observed data) and smoothness of the approximating function (Wood, 2003). The hierarchical splines can be thought of as a functional generalization of random effects terms in multilevel models, that is, allow for poststratification on the estimated functions of time after model fitting (see the section on estimating population trends). All applied splines are parameterized in a way that their overall flexibility can be controlled by one or more parameters representing standard deviations over the non-linear coefficients of the splines (Wood, 2004). Note that the latent mean $\mu_{iqp}(t)$ does not vary by survey j to prevent it from being collinear with the survey-dependent threshold vector τ_{ji}. Together, these assumptions imply the following model for our observed, ordinal responses $y_{ijqpt} \in \{1, ..., K_{ji}\}$, where K_{ji} is the number of response categories in survey j for institutions i:

$$\Pr\left(y_{ijqpt} = k \mid \mu_{iqp}(t)\right) = \\ F\left(\tau_{jik} - \mu_{iqp}(t)\right) - F\left(\tau_{ji(k-1)} - \mu_{iqp}(t)\right). \qquad (5)$$

On all parameters of the model we specified weakly-informative prior distributions designed to make implausibly large parameter values (for the given scale of the parameters) unlikely while having only a small influence on the posterior distribution in the range of plausible parameter values (Gelman et al., 2008; Evans et al., 2011). Weakly-informative priors can help prevent overfitting (i.e. fitting overly complex models) and improve convergence of sampling algorithms (e.g., Stan Development Team, 2020). For the thresholds τ, the latent mean b_0 and the project- and institution-specific coefficients b_q and b_i as well as their interaction $b_{q:i}$, we choose normal (mean = 0, sd = 1.5) priors, while we used normal (mean = 0, sd = 2) priors truncated at a lower bound of 0 for the standard deviation parameters

of the splines and the person-specific coefficients. Since all parameters are defined on the probit scale (with a latent standard deviation of 1), these priors can be thought of as weakly-informative. Sensitivity analyses with priors of varying widths, which we performed as we were developing and testing the models and do not report here, indicate that our results are essentially invariant to the specific choice of these priors as long as they are not extremely narrow.

We estimate separate models for each country instead of the more common strategy of having a single model for the pooled data. We chose single-country models because our approach involves modeling individual-level responses rather than grouped data (i.e., aggregating over individuals by response option) as was the case in prior research (Caughey et al., 2019; Claassen, 2019; Solt, 2020). Group-level IRT models are less computationally intensive than individual-level models, but do not enable taking into account individual-level characteristics or the fact that multiple survey responses (to various survey items) may come from the same respondent. Our individual-level modeling strategy more faithfully reflects the data generation process, but is far more computationally demanding. Each single-country models runs for several days, so a pooled model combining data from all or even a selection of countries would have been computationally infeasible given the existing per-job run-time limits of university computing clusters of around five days. We discuss the implications of this choice, in relation to other elements of the modeling approach, in the Discussion.

We hard-centered the latent scale around 0 by design of our models, so that estimates are still on the same scale for all countries even though each country was modeled separately. Equivalent identification constraints were also imposed in the pooled-data models with ordinal outcomes (cf. Caughey et al., 2019, p. 8; Solt, 2020, eq. 11). At the same time, we do not constrain thresholds to be equal across countries or even across projects within the same country. In addition to likely being unjustified statistically, this would also not have been possible in the first place for response scales of different lengths. To demonstrate that our approach indeed represents relative country positions similarly to the pooled group IRT model by Solt (2020), we also fit the latter to our data. Results and comparison of approaches are available in the web-based supplementary materials, section 5.

While ensuring that estimates are on the same scale across countries is an important aspect to achieving their full comparability, the latter additionally requires measurement invariance of item discriminations across surveys, projects, and institutions across countries. In the context of the present analysis, we assume such measurement invariance after having performed additional sensitivity analysis presented in the web-based supplementary materials, sec-

tion 4. While we do account for a lot of potential sources of variation in our models, more work would be required to systematically study and validate measurement invariance and the consequences for the lack of invariance on final model estimates, as we highlight in the Discussion.

Taken together, our new modeling approach offers the following advantages over existing approaches:

1. We model data on the level of the individual subjects instead of on level of surveys (after aggregating over individuals), thus taking into account individual-level differences.
2. We explicitly handle different response scales via different sets of thresholds varying by survey.
3. We model changes over time continuously via flexible penalized splines instead of using a discrete approach via latent autoregressive effects of order 1 (a detailed comparison of these approaches is out of scope of the present paper).
4. We use a fully Bayesian approach to model estimation. This not only enables estimation of such complex models in the first place, but also allows for much more sophisticated post-processing, for example, posterior predictions preserving the uncertainty in the model estimates (Gelman et al., 2013).
5. We make fewer assumptions regarding shared parameters across countries, assumptions we do not think are justified for the present data (see above).
6. By-country models are smaller, hence enable us to fit more complex models. Partially pooled models with data from all countries would be prohibitively slow in light of the complex modeling strategy already employed for each country, especially when estimated using a fully Bayesian approach.
7. Our models are easily extendible, for example, to include individual-level predictors or to further investigate item invariance assumptions (see the web-based supplementary materials, section 4, for additional models testing invariance assumptions and their influence of the estimated latent trend).

3.4 Using the model to estimate population trends

The surveys used in this manuscript were collected over 30 years in different countries. Some had weights available, but not all, and of those that did have weights available, the weights did not consistently adjust for the same variables. Of particular note, the weights were not always poststratified to education. For example, the largest project included in this study, the Eurobarometer, does not use education when constructing poststratification weights (GESIS, n.d.). For more information about the diversity in the weights available in cross-national surveys see e.g. Zieliński et al.

(2018). One of our concerns with this is that there is potential for variability in the survey weighting procedure (particularly any poststratification) to artificially suggest differences between countries or indeed between survey years or different surveys.

One potential solution to this problem is multilevel regression and poststratification (MRP) (Gelman and Little, 1997; Park et al., 2004). With this technique the focus moves from adjusting the sample to represent the population of interest to instead modeling an outcome by a number of demographic quantities, and then predicting this outcome in the population. Traditionally MRP has used a multilevel model for the modeling stage. This model works well because it allows smaller cells to be regularized towards the overall model. Although this introduces bias, it also reduces error in prediction. However, recent work (Gao et al., 2021; Bisbee, 2019; Ornstein, 2020) suggests that regularizing models other than a multilevel regression can be used. In particular, Gao et al. (2021) suggests that in some cases a smoother regularizing tool (in their work auto-regressive (1), random walk or spatial models) can outperform a simple MRP model. We use these findings in our work when predicting the change in political trust over time by modeling political trust using a spline for the time component. This isn't the first attempt at using MRP to model surveys over time. Gelman et al. (2016) models support for marriage equality over time. They model time using a linear and quadratic term that varies over state, as well as allowing the demographic terms to vary across years. We take a slightly different approach, instead modeling the time trend with a spline, controlling for question level differences and allowing the splines to vary by person-specific demographics.

Typically MRP is conducted instead of a weighted analysis, with variables that would have been used to create the weights instead included in the prediction model. However, including design based information into a model can be challenging. To balance these challenges, we present the results from a model with a design weighted (where design weights were available) pseudo-likelihood. The design weights were included if they were provided separately in the survey (e.g., ESS and EQLS, 4th round). In addition, certain countries (Germany and the United Kingdom) had previously been poststratified into separate geographic regions (former East and West Germany, Great Britain and Northern Island). For these countries we constructed weights through the inverse probability of inclusion given the region. The weights were normalised so the weights for each survey-year summed to the sample size of the survey in that year. The purpose of this is to reduce the impact of the weights on the estimated variance (Savitsky and Williams, 2019). We also ran the models with full poststratification weights and no weights with little difference

in final predictions (see sections 6 and 7 in the web-based supplementary materials). Although weights do not impact the predictions in this case, they could in other cases. Further study is needed to better understand weighted-spline models. Alternative methods for including weights in MRP analyses include those proposed by Si et al. (2015) and area level models such as those used by Caughey and Warshaw (2015).

We intended to adjust for other demographic variables through the model in an MRP fashion. As we had to choose demographic variables for which there is census information available across the countries and years, we choose a relative concise list. We choose three age groups (20–34, 35–54, 55–74), two sexes (male and female), and three education groups (less than secondary education, secondary and post-secondary non-tertiary education, and completed tertiary and above), which combined result in 18 categories. EB, EVS/2 and WVS/2 do not provide education measured as completed levels, but as the number of schooling years or the age at completion of education. In these cases, individuals with fewer than 12 years of education were classified in the lowest education category, those with between 12 and 15 years of schooling in the middle education category, and 16 or more years in the highest category. Where necessary we further assumed that school starting age is 6 (cf. UNESCO, 2013). We use population data from the Eurostat: population counts by sex, age and educational attainment (Eurostat, 2020a) and population statistics by age and sex (Eurostat, 2020b). Of the characteristics that are known to be strongly associated with political trust, education is the only one for which reliable and consistently coded data are available for the countries and period covered by this analysis. Researchers interested in shorter time spans or smaller country coverage may consider poststratifying using other socio-demographic predictors of political trust, such as income, labor market status, migrant status, region or urban/rural residence.

Even though we use a minimalistic set of background characteristics, we still faced difficulties in constructing the poststratification tables for all countries starting in 1989 (or later if the survey data cover a shorter period), especially in the early 1990s. We were able to fill in some gaps in Eurostat's coverage with data from country censuses from IPUMS International (Minnesota Population Center, 2019). To combine the two, we fit Dirichlet models with hierarchical factor interaction splines (Wood, 2017) estimating the yearly proportions of the population in each education category given their age group and sex (details on the imputation model are available in the replication materials). As there is not sufficient education data in the early 1990s, this technique does involve extrapolating the observed trends. While extrapolation isn't as preferred as simply knowing these values from a national census, it is more preferable
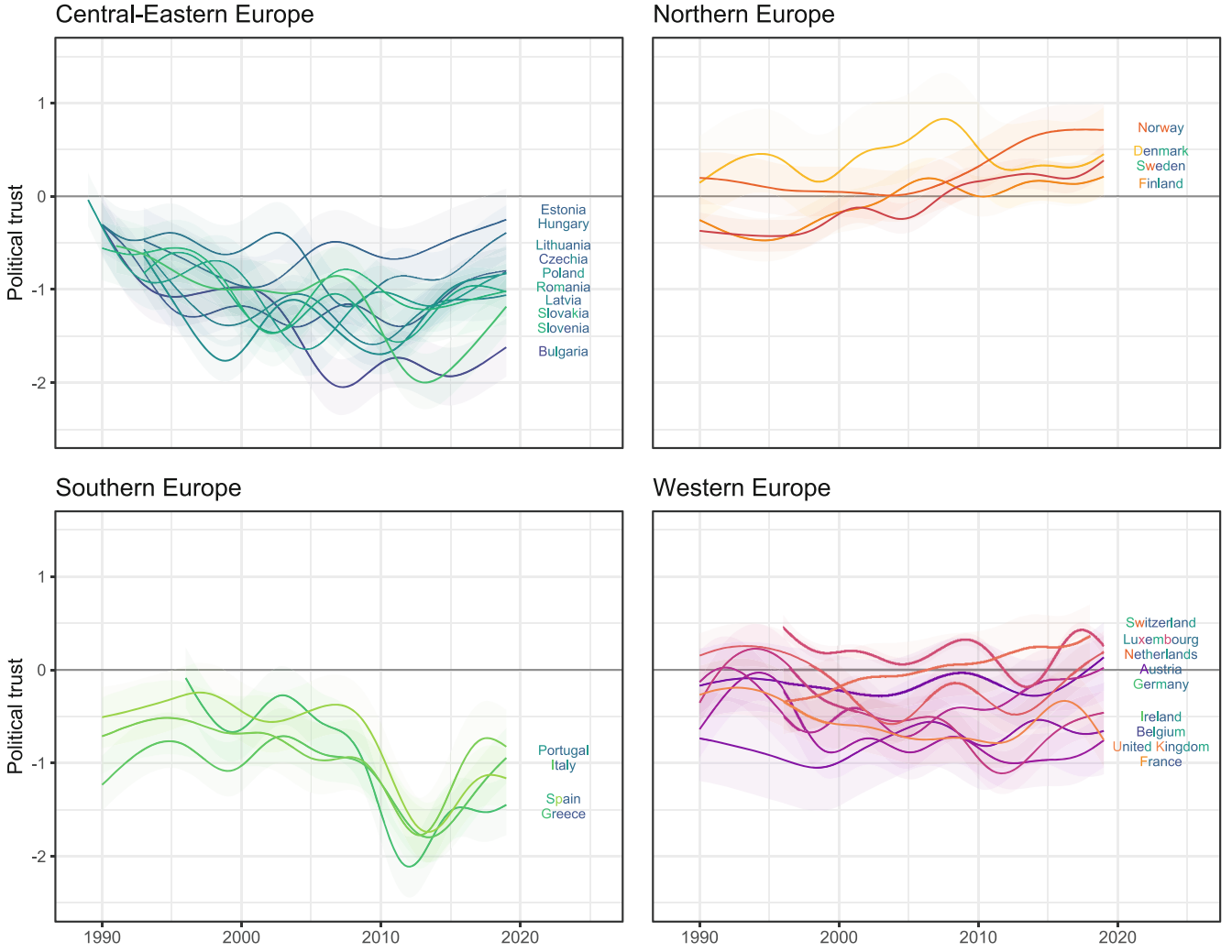
Fig. 2

Poststratified estimates of overall levels of political trust by region: Posterior medians and 95% credible intervals

than not adjusting for education at all. In addition, one of the benefits of our technique is that the uncertainty of predicting the proportion of education can be accounted for.

To do this, we use a technique briefly mentioned in Kastellec et al. (2015). Taking the age group and sex numbers in the population in a given year as known, our poststratification matrix would be 6 row by 3 columns (age group, sex and N_c), where N_c is the number of people in the population in the corresponding cell c. Using the imputation model, we take 100 samples predicting the proportion of education levels in each age group, sex, year and country combination. We multiply each of the proportions by the N_c of the corresponding cell, which allows us to maintain the uncertainty of our predictions.

The prediction component of our MRP analysis hinges on the constructed population. For each cell in our poststratification matrix, we predict the outcome, which in this

case is mean political trust for the mth posterior draw. If we had a simple poststratification matrix and a single posterior ($M = 1$), we could use the standard formula to estimate latent mean political trust for each country-year

$$\mu = \frac{\sum_{c=1}^{C} \mu_c * N_c}{\sum_{c=1}^{C} N_c}. \tag{6}$$

In our context, for each time point t, the cell means $\mu_c(t)$ for post-stratification cell c are computed as

$$\mu_c(t) = b_0 + f_1(t) + f_2(t, \text{age}_c) + f_3(t, \text{sex}_c) + f_4(t, \text{education}_c). \tag{7}$$

We refer to $\mu_c(t)$ as cell-specific latent political trust and to $\mu(t)$ as overall (post-stratified) latent political trust. Post-
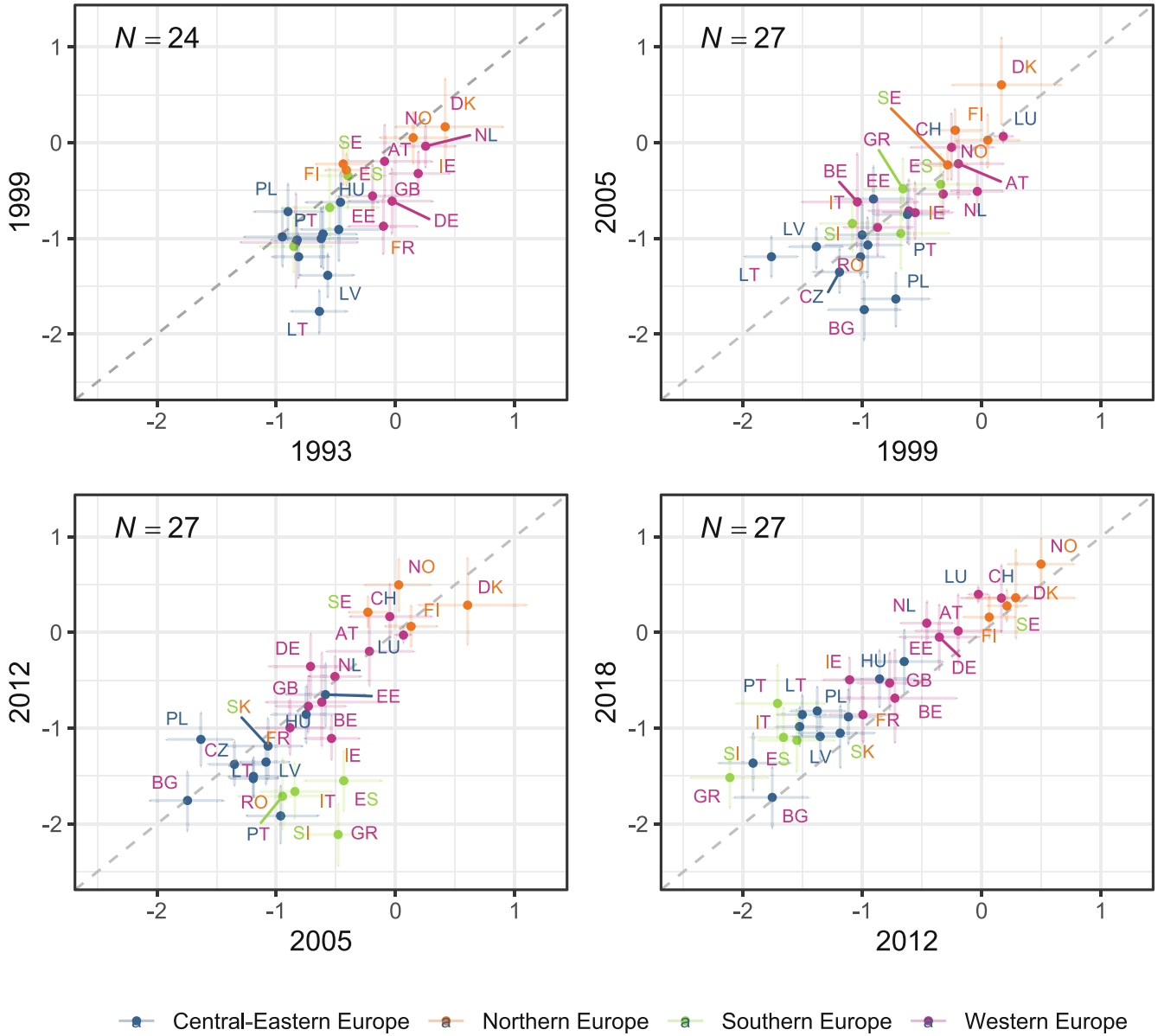
Fig. 3

Comparisons of country levels of political trust in selected years: Points indicate posterior medians, while error bars indicate 95% credible intervals.

stratification is performed separately for every time t. Due to continuity of all the involved functions with respect to t, the resulting function $\mu(t)$ is continuous as well. For readability, we will suppress the dependency on t in the following.

To obtain an estimate of uncertainty, we use $M = 1000$. We obtain M posterior estimates of μ_c, with the mth denoted $\mu_{c,m}$. For each m we estimate μ, resulting in 1000 posterior estimates. We then take the median over μ_m as the estimate for μ and lower 0.025 and 0.975 quantiles as our uncertainty.

However, we also have uncertainty in our estimate of N_c. To incorporate this uncertainty into our estimate of μ, we take $L = 100$ posterior estimates of N_c, with the lth denoted

as $N_{c,l}$. We define for the lth posterior estimate of $N_c(N_{c,l})$ and the mth posterior estimate of $\mu_c(\mu_{c,m})$

$$\mu_{l,m} = \frac{\sum_{c=1}^{C} \mu_{c,m} * N_{c,l}}{\sum_{c=1}^{C} N_{c,l}} \tag{8}$$

We then have a matrix of size 100×1000 possible estimates for μ, where the columns represent posterior uncertainty estimating μ_c and the rows represent posterior uncertainty estimating N_c. We stack these estimates and take the median, lower 0.025 and upper 0.975 quantiles. This pro-
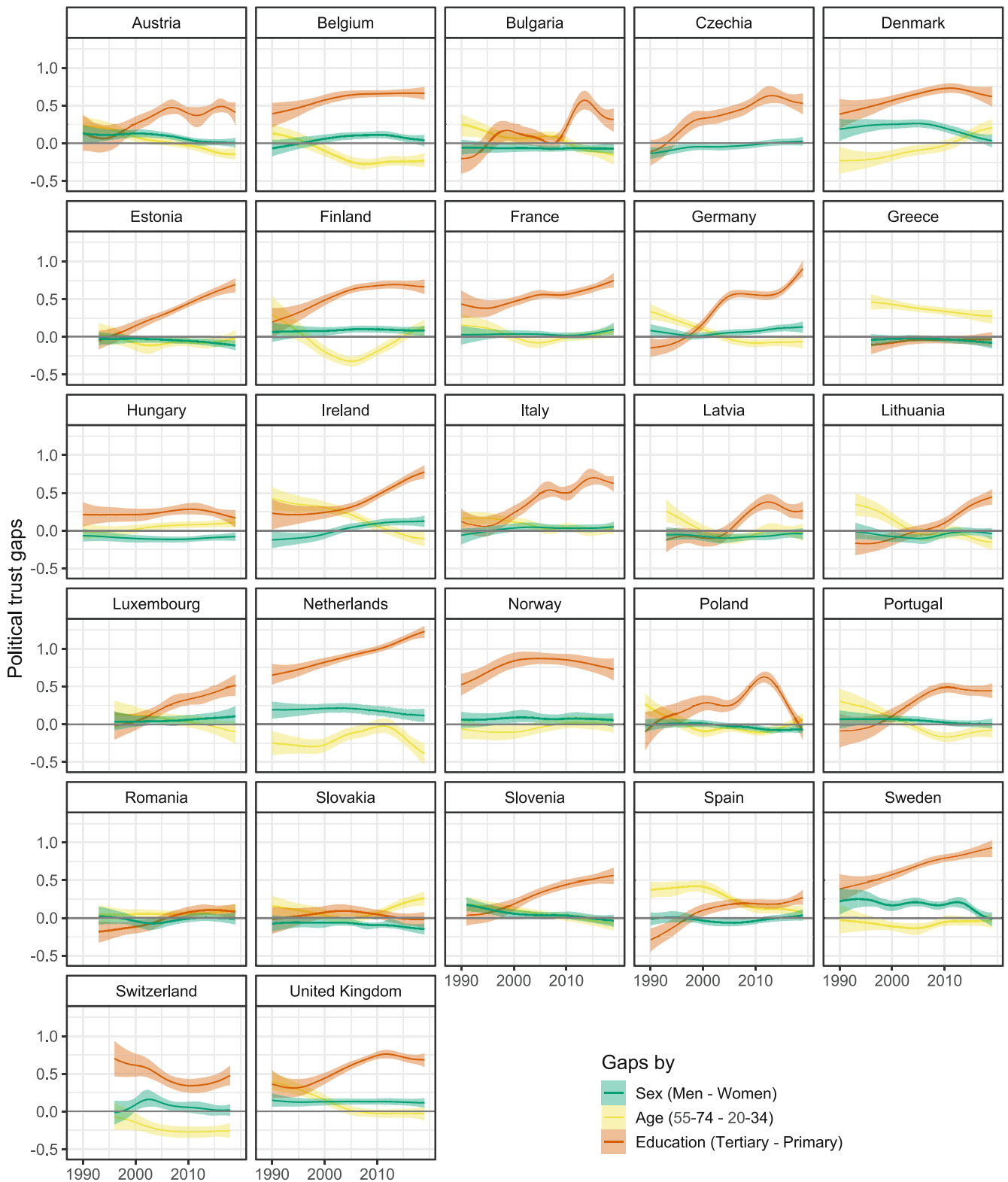
Fig. 4

Differences in political trust by sex, age, and education: The lines represent differences (posterior medians and 95% credible intervals) between men and women, the oldest and the youngest age group, and between people with tertiary and primary education. Positive values indicate higher trust among men than women, in the oldest age group compared to the youngest, and among people with tertiary education compared to those with primary education

cedure allows the incorporation of uncertainty of education level in the population.

4 Results

4.1 Overall trends in political trust

We start by describing overall levels and changes in latent political trust and then turn to differences between societal groups. Fig. 2 presents political trust in 27 countries across European regions (plots with separate facets per country are available in the web-based supplementary materials, section 8). The division into regions is not strictly geographic, but rather takes into account common historical experiences as well as social and political environments and—as prior research has shown and our results confirm—is meaningful from the point of view of comparing trajectories of political trust.

As shown in Fig. 2, and consistent with other cross-national studies, political trust is highest in Northern Europe, which also enjoyed a slight increase in trust in the course of the studied period. In the recent years, average values of political trust in all four Northern European countries exceeded 0 on the latent scale at least some of the time, which means that positive evaluations of state institutions dominated among survey respondents. In most of Western Europe, trust has been hovering around the neutral point of 0, indicating that positive and negative assessments of state institutions have been relatively balanced. The exceptions are Switzerland, where trust has increased from below 0 and below the region's average in the 1990s to levels characteristic for Northern Europe in 2018, and Belgium, which remained below 0 throughout the analyzed period.

Countries in Central-Eastern and Southern Europe have lower levels of trust, with negative evaluations dominating throughout the studied period, as indicated by the average trust levels below 0. In these regions trust is also characterized by greater volatility than in Northern and Western Europe, not only due to declines but also to occasional increases in trust (Marien, 2011). While in Austria, Belgium, and the United Kigdom political trust has oscillated within a range below 0.6 units, in Greece, Slovenia and Poland the range exceeds 1.6 units on the probit scale (cf. Torcal, 2017).

Overall levels of trust in institutions in Fig. 2 demonstrate the distinctiveness of the countries' trajectories, in line with the interpretation of political trust as driven primarily by national events. However, there is evidence of some commonalities among countries that shared certain experiences. First, most post-communist countries saw a decline in trust in the 1990s, reflecting the fading honeymoon effect of

exposure to pluralist politics and capitalist markets (Catterberg and Moreno, 2006). The initial optimism is particularly pronounced in Poland, the only country with a survey available from 1989. Bulgaria, Estonia, Latvia, Lithuania, and Slovenia also saw a drop in trust in the 1990s, but the duration and depth of the decline varied across countries. In all countries in this region, with the exception of Estonia, Hungary, and Lithuania, trust in the late 2010s was lower than in early 1990s.

Interestingly, countries in Central-Eastern Europe entered the 1990s with very similar levels of political trust, which contradicts Mishler and Rose's (1997) hypothesis about political trust in the first transition years being related to the level of repression in the communist period. The subsequent decline in political trust observed in most countries of the region may reflect either the return to country-specific baseline political trust levels inherited from the decades long communist rule (Rose-Ackerman, 2001), the assessment of the deficient performance of new institutions (Mishler and Rose, 2001), unfavorable comparisons with Western Europe or—likely—a mix of all three mechanisms. After these early declines, each country followed its unique trajectory throughout the studied period, much like countries in Western Europe, but with higher volatility. An analysis of data on elections could clarify to what extent this volatility is associated with the electoral cycle.

The second visible commonality across several countries is the sharp drop in trust after 2008–2009, when the European sovereign debt crisis followed the global financial crisis. The magnitude of the decline in trust, proportionally to the detrimental effect of the crisis on national economies, was greatest in Southern Europe, as shown in Fig. 2. Torcal's (2014) analysis of Spain and Portugal in 2011–2012 argued that the decreases in political trust in the years following the crisis were not only due to the austerity measures themselves, but also reflected a decline in the perceived responsiveness and integrity of state institutions. Trust would thus depend on the evaluation of both the product and the process of governance. In the recent years, in some countries (Ireland, Portugal) political trust has rebound, but in others it remains low (Greece, Spain). Some other Western and Northern European countries have seen some decline in political trust in the late 2000s, but of much smaller magnitude.

Despite these changes, the ranking of countries with regard to overall political trust has remained relatively stable. Fig. 3 presents changes in the level of trust between pairs of years 6–7 years apart: 1993, 1999, 2005, 2012, and 2018, with each point representing a country, error bars indicating 95% credible intervals, and colors corresponding to European regions. Countries tend to fall close to the 45-degree line of no change. The exception are declines in trust between 2005 and 2012, i.e. during the financial crisis, in

Southern Europe and in part of Central Europe. The following period 2012–2018 generally was a time of stabilizing or even increasing trust, correcting the earlier declines.

While the ranking remains rather stable, the spread across countries has increased markedly since the 1990s and reached highest levels in the 2010s. This change is primarily due to declines in trust in Southern Europe and in part of Central-Eastern Europe, when Northern Europe and most of Western Europe maintained high levels of trust. The stability of trust rankings may be interpreted as supporting theories that view persistent country characteristics as sources of political trust, with most prominent explanations including cultural and historical legacies (Mishler and Rose, 2001; Torcal, 2006) and institutional properties, such as aspects of the electoral and party systems (Criado and Herreros, 2007; van der Meer, 2010). The impact of economic conditions is most clearly visible in the effects of the economic crisis of 2008.

4.2   Political trust by age, sex, and education

Political trust is theorized as driven in large part by two factors: the satisfaction with the performance of state institutions, including the protection of one's rights and interests, and by political representation and the perception of the institutions' responsiveness to one's needs. Consequently, women would be expected to have lower political trust in countries where gender equality in politics and other spheres of life is not respected. We find the opposite pattern. As shown in Fig. 4, the greatest gaps are in the Netherlands, Denmark, and Sweden, where trust among women is lower, by around 0.1–0.2 units on the probit scale, compared to men (plots presenting group levels are available in the web-based supplementary materials, section 8). This is surprising, given that these very countries globally rank high in terms of gender equality. Two general patterns stand out. First, in many Central-Eastern European countries women have slightly higher trust compared to men, while the opposite is true in Western and Northern Europe. Second, in the majority of countries (with the exception of Germany and Ireland) the difference between trust among men and women has not increased but rather remained stable or declined, either by reducing men's trust bonus, as in Austria, Denmark, Sweden, and Slovenia, or by increasing women's advantage, as was the case in Estonia and Slovakia. This leads to the question: why are women in Central-Eastern Europe not more distrustful of institutions that cannot secure them an equal position in politics or on the job market? Potential explanations, including women's different expectations or benchmarks, attention to different aspects of state performance, or gender differences in re-

sponse behavior, constitute fruitful topics of future research (cf. Hodson, 1989; McDermott and Jones, 2020).

Age differences are also relatively small, and go in both directions. We note here that our analysis does not distinguish between age and cohort effects (cf. Wuttke et al., 2020). In Czechia (since around 2005), Finland (in the 2000s), the Netherlands (in the 1990s and since around 2015), and Switzerland, young people aged 20–34 had somewhat higher trust than the older groups. In Greece, Ireland, and Spain, trust has generally been lower among the young compared to the rest of the adult population. In all cases the differences between age groups did not exceed 0.5 units on the probit scale. Interestingly, in countries where the youth were hit particularly hard by the repercussions of the financial crisis, such as Spain and Greece, we detect no increase in the difference in trust levels between age groups in the period with peak unemployment between 2010 and 2015, as the trust-as-evaluation of performance hypothesis would predict. On the contrary, in these countries and periods the age gap in trust declined. To the extent that political trust is partly shaped by performance evaluations and value congruence, the differences between age groups and by sex are surprising and worthy of further investigation.

Education gaps in political trust are more consistent with theoretical expectations. In almost all countries, with the exception of Romania, higher education was associated with higher trust at least some of the time during the last three decades. Between-country differences are consistent with expectations about the link between the education gap in political trust and the democraticness of the political system: education gaps in trust tend to be larger among Northern and Western European countries than in Central and Southern Europe. The Netherlands, Norway, and Sweden have the highest average education gaps between trust among people with tertiary education and those with primary education, ranging between 0.6 and 1 units on the probit scale. In Greece, Romania, Slovakia, and Lithuania the differences in trust levels among the lowers and highest education groups are barely discernible.

There is also some evidence supporting the longitudinal hypothesis about the link of democratization to the education gap in trust. Among the analyzed countries, the most prominent increases in the level of democracy occurred in Central-Eastern Europe in the 1990s. Some post-communist countries have indeed seen an increase in the education gap in political trust in that period, including the Czech Republic, Estonia, Poland, and Slovenia, but others have not. During the last three decades two countries have seen substantial democratic backsliding, Hungary and Poland (Bakke and Sitter, 2020). Poland has seen a narrowing of the education gap in political trust during that time, from a relatively large difference in trust by education between 2010 and

2015, to no difference in 2019. In Hungary the gap has always been small, and has declined minimally between 2010 and 2019. At the same time, in almost all Western and Northern European established democracies the education gap in trust has increased since the 1990s although—according to the main democracy indexes—the quality of democratic governance in most of them has been high and stable. Future research may evaluate if the changes in the education gaps are systematically associated with changes in macro-level characteristics, including aspects of democratic performance, or with characteristics of incumbent governments, among others.

5   Conclusions

In this paper, we presented a strategy of estimating trends in political trust on the basis of items measuring trust in the national parliament, justice system, and political parties from 13 cross-national survey projects. The described approach combines ordinal response modeling and poststratification to address the two major sources of error in surveys: measurement and representation (Groves and Lyberg, 2010). The resulting dataset, covering 27 European countries, makes it possible to analyze changes in political trust over a period of three decades both on the societal level, as well as by age, sex, and education.

Descriptive results of country trajectories of political trust between 1989 and 2019 support the hypothesis of "trendless fluctuations" (Norris, 2011; Van de Walle et al., 2008) in political trust and indicate no long-term consistent tendency. Country profiles are largely unique, but there exist some common patterns that reflect events with international impact. First, in many post-communist countries trust declined markedly in the early 1990s and then stabilized at relatively low levels. This decline in trust, in the context of improving quality of governance, likely reflected a normalization of inflated expectations and optimism of the early transition years. Second, a strong decrease in trust in the wake of the global financial crisis in the most affected countries suggests a reaction to the deteriorating economic conditions combined with low perceived responsiveness and accountability of governments. Whether trust returned to pre-crisis levels in these countries may be related to the extent to which both the economic situation and perceptions of the government have rebound. Overall, our results suggest that political trust is relatively volatile, and estimating its trajectories requires dense time series.

Our analysis points to important differences in political trust between groups defined on the basis of sex, age, and education. Differences in trust between the sexes are generally small. In most countries where the differences exist, trusts tends to be higher among men than among

women except for Central-Eastern Europe, where women have slightly higher trust compared to men. The latter observation is surprising given the limited success in achieving gender equality in this region. Differences in trust between age groups are only slightly larger. We also observe no evidence of declines in political trust among the youngest age group compared to older adults.

Of the three socio-demographic characteristics we analyzed, education exhibits the strongest association with political trust. In line with earlier studies, we find that the trust gap between people with more education and those with less education is greater in Western and Northern European countries, regions with high democratic performance and low corruption, than in Southern and Central-Eastern Europe. We also find evidence of increases in the education gap over time in some of the post-communist countries included in the analysis in the decade following their transition to democracy.

The time series we created provide comparable data for investigating political trust in a longitudinal perspective. Future studies may analyze, for example, the responsiveness of trust to economic, social and political conditions, or to political events, as well as the consequences of political trust, in order to fill in gaps in trust research and to test whether the cross-national associations established in the literature hold when analyzing changes over time.

6   Discussion and limitations

Analyses with country-year levels of political attitudes estimated from survey data must take into account the uncertainty inherent in survey estimates. The Bayesian multilevel models we use—and Bayesian modeling in general—makes this straightforward to accomplish by providing (random draws of) the posterior distributions for all model parameters and the ability to propagate this uncertainty to all generated quantities, in particular to model predictions. As a result, our published dataset of country-year level trust estimates also contains measures of uncertainty, which can be used in subsequent models analyzing these time series (see Tai et al., 2022, for an illustration of the consequences of omitting uncertainty in time series models). A simulation study of the performance of different latent trend models showed that the thin-plate splines used in the present analysis tend to underestimate uncertainty, but the degree of the underestimation is minimal when sufficient data are provided (Kołczyńska and Bürkner, 2023). Given the generally high availability of data for the countries and period studied, as shown in Fig. 1, we can expect high accuracy of the credible intervals, with the possible exception of the more sparsely available surveys in the 1990s, where uncertainty is likely underestimated.

Our modeling approach can also be applied to constructs other than trust, such as policy preferences or value orientations, if enough equivalent items are available in a sufficient number of surveys and years in the same country. While the method we propose enables joint analysis of variables measured with ordinal response scale of different lengths, it still relies on the assumption that all items measure the same construct. This leads to the question of measurement invariance, a highly relevant topic but also one that is not easy to tackle (Davidov et al., 2014, 2018). Of course, a dataset combining multiple surveys is more heterogeneous in various ways than data from a single survey project where country surveys are harmonized a priori. In order to estimate country trends in political trust with the available data, we made two important assumptions regarding measurement invariance. First, our models assumed item factor loadings (or, equivalently, item discriminations) to be comparable across surveys and items (corresponding to trust in different institutions). This strong assumption was required for these models to fit well in acceptable amounts of time (i.e., less than five days per model on a high performance computing cluster). For selected countries, we have estimated models assuming varying factor loadings across projects and institutions (see the web-based supplementary materials, section 4). Results indicate that, while factor loadings differ as expected, the prediction of the latent trajectories of trust are barely affected, which demonstrates the robustness of our main results and the conclusions drawn from them.

The second invariance-related assumption of our modeling approach is implied by the sum-to-zero constraint of each of the survey-specific threshold vectors (see the Modeling section). This is a common assumption in models estimating dynamic cross-national public opinion (Caughey et al., 2019; Solt, 2020) and it is necessary for identification of trust changes over time within each country. However, to this end, it assumes invariance of overall item difficulties across countries and time, while still allowing for non-invariance in relative threshold difficulties. This assumption might not necessarily hold. For example, it is possible that individual countries, or groups of countries, have unique characteristics that could impact overall item difficulties or that overall item difficulties change over time even within the same country. In the case of political trust, such changes could be due to regime change or institutional reform that would alter the role of certain institutions (whose trust items form the political trust measure) within the political system. In such cases our model would incorrectly interpret changes in overall item difficulties as changes in trust itself. Given the assumption of overall difficulty invariance, it is crucial to choose the items used for estimating country time trends with care, considering whether there is reason to believe that the content or interpretation of the items differ substantially between countries or over time. The literature

on cross-national and cross-time measurement equivalence provides results of invariance tests for various scales in different sets of countries, less often over time. These results may provide evidence in favor or against the assumption of overall difficulty invariance. Special attention is required when comparing mass attitudes across countries that vary strongly in ways relevant for the given attitude. The assumption of invariant overall difficulties with regard to political trust is thus more likely to hold within the European Union than globally. Given the difficulty in comprehensive verification whether the assumption in question holds across countries and time periods, it would be desirable to relax this assumption without breaking identification. This, however, requires further research and potentially incorporating additional data sources. We think that studying different aspects of measurement invariance of aggregated survey data will be an important area for future research.

The assumption of equal overall difficulties is related to information-sharing between countries. As mentioned earlier, our modeling approach consists in estimating separate models for each country, which has the great advantage of being computationally feasible in reasonable time, and does not require explicit assumptions about which parameters are to be shared across countries and how. However, fixing overall difficulties by constraining the sum of thresholds to zero, as mentioned in the previous paragraph, implicitly assumes that overall difficulties are equal across countries.

While our decision to run a separate model for each country was primarily driven by considerations of feasibility, we agree that a hierarchical multi-country model could be considered in the future. We would still not assume similar ups and downs in trust between countries, but we could have a hierarchical prior for the hyperparameters controlling the smoothness and magnitude of variation. However, we expect that such hierarchical prior would not change the results much for these data.

The decision to model countries separately necessarily restricts the applicability of our models to countries that are frequently included in cross-national survey projects or otherwise have enough national surveys with the necessary questions. On the other hand, recent work by Solt (2020) and Claassen (2020) argues for pooling data from various countries, including those with many surveys and ones with few surveys, to share information and increase country coverage. The choice of approach should depend on the nature of the studied phenomenon, in particular its volatility and sensitivity to national events versus global trends, as well as characteristics of the available survey data. More research is needed in this area.

The choice of sex, age, and education for poststratification was determined—in addition to substantive interests—by the requirement of the poststratification procedure that necessitates population joint distributions of all post-

stratification variables consistently measured for multiple years for 27 European countries. The availability of population data constitutes one of the challenges for applying poststratification, but—as we show—gaps in the population data can be filled by imputing the missing data points and then incorporating the uncertainty associated with the imputation into the poststratification procedure. Among factors that are known to be strong predictors of political trust, apart from other political attitudes, the strongest is party affiliation, especially electoral winner-loser status, but reliable population statistics on these characteristics by levels of education—which turned out to be important in shaping political trust at least in some countries—are not available. Future research may want to explore the possibilities of compiling different population data sources for purposes of poststratification.

Our study is part of a broader area of methodological research focused on combining data from different sources. This research includes individual data meta-analysis (or "integrative data analysis", Curran and Hussong, 2009) employed in medical or psychological research, which involves pooling and joint analysis of individual participants' data from different studies (e.g. Ioannidis, 2017; Riley et al., 2010; Riley, 2010). In the social sciences, the term "ex-post survey data harmonization" (Wolf et al., 2016) is used for methods of combining survey datasets that were not a priori designed with comparability in mind. Further research in this area, beyond enabling analyses that had been infeasible with single studies, contributes to improving the interoperability and reusability of research data, which are elements of data FAIRness (aside findability and accessibility).

Data Availability Statement Documentation of data processing and replication materials are available in the replication materials and at: https://osf.io/uyzfe/.

This article (https://doi.org/10.18148/srm/2024.v18i1.8119) contains supplementary material.

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495–508.

Bakke, E., & Sitter, N. (2020). The EU's enfants terribles: democratic backsliding in central europe since 2010. *Perspectives on Politics*. https://doi.org/10.1017/S1537592720001292.

Bisbee, J. (2019). Barp: improving mister p using bayesian additive regression trees. *American Political Science Review*, *113*(4), 1060–1065.

de Blok, L., & Kumlin, S. (2021). Losers' consent in changing welfare states: output dissatisfaction, experienced voice and political distrust. *Political Studies*. https://doi.org/10.1177/0032321721993646.

Bobo, L., & Licari, F. C. (1989). Education and political tolerance: testing the effects of cognitive sophistication and target group affect. *Public Opinion Quarterly*, *53*(3), 285.

Breustedt, W. (2018). Testing the measurement invariance of political trust across the globe. A multiple group confirmatory factor analysis. *methods, data, analyses*, *12*(1), 7–46.

Bürkner, P.-C. (2017). brms : an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v080.i01.

Bürkner, P.-C. (2021). Bayesian item response modelling in R with brms and Stan. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v100.i05.

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: a tutorial. *Advances in Methods and Practices in Psychological Science*, *2*(1), 77–101.

Catterberg, G. (2013). Intergenerational value change and transitions to democracy. Toward the consolidation of a third wave generation? *Revista Latinoamericana de Opinión Pública*, *3*, 53–80.

Catterberg, G., & Moreno, A. (2006). The individual bases of political trust: trends in new and established democracies. *International Journal of Public Opinion Research*, *18*(1), 31–48.

Caughey, D., & Warshaw, C. (2015). Dynamic estimation of latent opinion using a hierarchical group-level IRT model. *Political Analysis*, *23*(2), 197–211.

Caughey, D., O'Grady, T., & Warshaw, C. (2019). Policy ideology in European mass publics, 1981–2016. *American Political Science Review*, *113*(3), 674–693.

Christmann, P. (2018). Economic performance, quality of democracy and satisfaction with democracy. *Electoral Studies*, *53*, 79–89.

Cichocki, P., & Jabkowski, P. (2022). Response scale overstretch: linear stretching of response scales does not

ensure cross-project equivalence in harmonised data. *Quality & Quantity*, *57*(4), 3729–3745.

Citrin, J., & Stoker, L. (2018). Political trust in a cynical age. *Annual Review of Political Science*, *21*(1), 49–70.

Claassen, C. (2019). Estimating smooth country–year panels of public opinion. *Political Analysis*, *27*(1), 1–20.

Claassen, C. (2020). Does public support help democracy survive? *American Journal of Political Science*, *64*(1), 118–134.

Coromina, L., & Bartolomé Peral, E. (2020). Comparing alignment and multiple group cfa for analysing political trust in europe during the crisis. *Methodology*, *16*(1), 21–40.

Criado, H., & Herreros, F. (2007). Political support. *Comparative Political Studies*, *40*(12), 1511–1532.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychological Methods*, *14*(2), 8–100.

Dalton, R. J. (2004). *Democratic chal lenges, democratic choices: the erosion of political support in advanced industrial democracies*. Oxford: Oxford University Press.

Dalton, R. J., Van Sickle, A., Weldon, S., Sickle, A. V., & Weldon, S. (2010). The individual-institutional nexus of protest behaviour. *British Journal of Political Science*, *40*(1), 51–73.

Dassonneville, R. (2021). Change and continuity in the ideological gender gap a longitudinal analysis of left-right self-placement in OECD countries. *European Journal of Political Research*, *61*(1), 225–238.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75.

Davidov, E., Muthen, B., & Schmidt, P. (2018). Measurement Invariance in cross-national studies. *Sociological Methods & Research*, *47*(4), 631–636.

Easton, D. (1975). A re-assessment of the concept of political support. *British Journal of Political Science*, *5*(4), 435–457.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. New York: Psychology Press.

Eurostat (2020a). Population by sex, age and educational attainment level (1000): lfsa_pgaed. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsa_pgaed

Eurostat (2020b). Population on 1 january by age and sex: demo_pjan. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=demo_pjan

Evans, M., & Jang, G. H., et al. (2011). Weak informativity and the information in one prior relative to another. *Statistical Science*, *26*(3), 423–439.

Foa, R. S., & Mounk, Y. (2016). The democratic disconnect. *Journal of Democracy*, *27*(3), 5–17.

Foa, R. S., & Mounk, Y. (2017). The signs of deconsolidation. *Journal of Democracy*, *28*(1), 5–15.

Gao, Y., Kennedy, L., Simpson, D., Gelman, A., et al. (2021). Improving multilevel regression and poststratification with structured priors. *Bayesian Analysis*, *16*(3), 719–744.

Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, *23*(2), 127–135.

Gelman, A., Jakulin, A., Pittau, M. G., Su, Y.-S., et al. (2008). A weakly informative default prior distribution for logistic and other regression models. *The annals of applied statistics*, *2*(4), 1360–1383.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd edn.). London:: Chapman and Hall/CRC.

Gelman, A., Lax, J., Phillips, J., Gabry, J., & Trangucci, R. (2016). *Using multilevel regression and poststratification to estimate dynamic public opinion*

GESIS (n. d.). Weighting overview. https://www.gesis.org/eurobarometer-data-service/survey-series/standard-special-eb/weighting-overview/. Accessed 2020-07-30.

Griffin, J. D., Kiewiet de Jonge, C., & Velasco-Guachalla, V. X. (2021). Deprivation in the midst of plenty: citizen polarization and political protest. *British Journal of Political Science*, *51*(3), 1080–1096.

Groves, R. M., & Lyberg, L. (2010). Total survey error: past, present, and future. *Public Opinion Quarterly*, *74*(5), 849–879.

Hakhverdian, A., & Mayne, Q. (2012). Institutional trust, education, and corruption: a micro-macro interactive approach. *The Journal of Politics*, *74*(3), 739–750.

Hetherington, M. J. (1998). The political relevance of political trust. *American Political Science Review*, *92*(4), 791–808.

Hodson, R. (1989). Gender differences in job satisfaction: why aren't women more dissatisfied? *The Sociological Quarterly*, *30*(3), 385–399.

Hyman, H., & Wright, C. R. (1979). *Education's lasting influence on values*. Chicago: University of Chicago Press.

Ioannidis, J. (2017). Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *British Journal of Sports Medicine*, *51*(20), 1456–1458.

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.

de Jonge, T., Kamesaka, A., & Veenhoven, R. (2021). How to reconstruct a trend when survey questions have changed over time. *Survey Research Methods*, *15*(1), 101–113.

Kastellec, J.P., Lax, J.R., Malecki, M., & Phillips, J.H. (2015). Polarizing the electoral connection: partisan representation in supreme court confirmation politics. *The Journal of Politics*, *77*(3), 787–804.

Klingemann, H.-D. (1999). Mapping political support in the 1990s: a global analysis. In P. Norris (Ed.), *Critical citizens: global support for democratic government* (pp. 31–56). Oxford: Oxford University Press.

Kołczyńska, M. (2020). Democratic values, education, and political trust. *International Journal of Comparative Sociology*, *61*(1), 3–26.

Kołczyńska, M., & Bürkner, P.-C. (2023). Modeling public opinion over time: a simulation study of latent trend models. *Journal of Survey Statistics and Methodology*. https://doi.org/10.1093/jssam/smad024.

Kołczyńska, M., & Sadowski, I. (2022). Seeing the world through party-tinted glasses: performance evaluations and winner status in shaping political trust under high polarization. *Acta Politica*, *58*(2), 380–400.

Letki, N. (2007). Institutional performance. In M. Bevir (Ed.), *Encyclopedia of governance* (Vol. 1, pp. 457–458). Thousand Oaks: SAGE.

Liddell, T.M., & Kruschke, J.K. (2018). Analyzing ordinal data with metric models: what could possibly go wrong? *Journal of Experimental Social Psychology*, *79*, 328–348.

van der Linden, W.J., & Hambleton, R.K. (2013). *Handbook of modern item response theory*. New York: Springer.

Marien, S. (2011). Measuring political trust across time and space. In *Political trust. Why context matters* (pp. 13–46). Colchester: ECPR Press.

McDermott, M.L., & Jones, D.R. (2020). Gender, sex, and trust in government. *Politics and Gender*, *18*(2), 297–320.

McGann, A.J. (2014). Estimating the political center from aggregate data: an item response theory alternative to the Stimson dyad ratios algorithm. *Political Analysis*, *22*(1), 115–129.

van der Meer, T. (2010). In what we trust? A multi-level study into trust in parliament as an evaluation of state characteristics. *International Review of Administrative Sciences*, *76*(3), 517–536.

van der Meer, T.W., & Zmerli, S. (2017). The deeply rooted concern with political trust. In S. Zmerli & T.W.G. van der Meer (Eds.), *Handbook on Political Trust* (pp. 1–16). Cheltenham: Edward Elgar.

van der Meer, T.W.G., & Ouattara, E. (2019). Putting 'political' back in political trust: an irt test of the unidimensionality and cross-national equivalence of political trust measures. *Quality and Quantity*, *53*(6), 2983–3002.

Minnesota Population Center (2019). *Integrated public use microdata series, international: version 7.2 [dataset]*

Mishler, W., & Rose, R. (1997). Trust, distrust and skepticism: popular evaluations of civil and political institutions in post-communist societies. *The Journal of Politics*, *59*(2), 418–451.

Mishler, W., & Rose, R. (2001). What are the origins of political trust? Testing institutional and cultural theories in post-communist societies. *Comparative Political Studies*, *34*(1), 30–62.

Norris, P. (2002). *Democratic phoenix: reinventing political activism*. Cambridge: Cambridge University Press.

Norris, P. (2011). *Democratic deficit: critical citizens revisited*. Cambridge: Cambridge University Press.

Norris, P. (2017). The conceptual framework of political support. In S. Zmerli & T.W. van der Meer (Eds.), *Handbook on Political Trust* (pp. 19–32). Cheltenham: Edward Elgar.

Oksanen, A., Kaakinen, M., Latikka, R., Savolainen, I., Savela, N., & Koivula, A. (2020). Regulation and trust: 3-month follow-up study on COVID-19 mortality in 25 European countries. *JMIR Public Health and Surveil lance*. https://doi.org/10.2196/19218.

Ornstein, J.T. (2020). Stacked regression and poststratification. *Political Analysis*, *28*(2), 293–301.

Park, D.K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*. https://doi.org/10.1093/pan/mph024.

Pedersen, E.J., Miller, D.L., Simpson, G.L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, *7*(5), e6876.

R Core Team (2018). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Riley, R.D. (2010). Commentary: like it and lump it? meta-analysis using individual participant data. *International Journal of Epidemiology*, *39*(5), 1359–1361.

Riley, R.D., Lambert, P.C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. https://doi.org/10.1136/bmj.c221.

Rose-Ackerman, S. (2001). Trust and honesty in post-socialist societies. *Kyklos*, *54*(2–3), 415–443.

Samejima, F. (1997). Graded response model. In *Handbook of modern item response theory* (pp. 85–100). New York: Springer.

Savitsky, T.D., & Williams, M.R. (2019). Pseudo bayesian mixed models under informative sampling. *arXiv*. https://doi.org/10.48550/arXiv.1904.07680.

Seligson, M. (2002). The impact of corruption on regime legitimacy: a comparative study of four latin American countries. *The Journal of Politics*, *64*(2), 408–433.

Si, Y., Pillai, N.S., & Gelman, A. (2015). Bayesian nonparametric weighted sampling inference. *Bayesian Analysis*, *10*(3), 605–625.

Singh, R.K. (2022). Harmonizing single-question instruments for latent constructs with equating using political interest as an example. *Survey Research Methods*, *16*(3), 353–369.

Solt, F. (2020). Modeling dynamic comparative public opinion. *SocArXiv*. https://doi.org/10.31235/osf.io/d5n9p.

Stan Development Team (2020). *Stan modeling language users guide and reference manual. Version 2.19*. https://mc-stan.org

Stimson, J.A. (1999). *Public opinion in america: moods, cycles, and swings* (2nd edn.). Boulder: Westview Press.

Stimson, J.A. (2018). The dyad ratios algorithm for estimating latent public opinion. *Bulletin of Sociological Methodology*, *137–138*(1), 201–218.

Tai, Y.C., Hu, Y., & Solt, F. (2022). Democracy, public support, and measurement uncertainty. *American Political Science Review*. https://doi.org/10.31235/osf.io/y5fdv.

Torcal, M. (2006). Political disaffection and democratization history in new democracies. In M. Torcal & J.R. Montero (Eds.), *Political disaffection in contemporary democracies. Social capital, institutions and politics* (pp. 157–189). London: Routledge.

Torcal, M. (2014). The decline of political trust in Spain and Portugal. *American Behavioral Scientist*, *58*(12), 1542–1567.

Torcal, M. (2017). Political trust in western and southern Europe. In S. Zmerli & T.W. van der Meer (Eds.), *Handbook on political trust* (pp. 418–439). Cheltenham: Edward Elgar.

Torcal, M., & Carty, E. (2022). Partisan sentiments and political trust: a longitudinal study of Spain. *South European Society and Politics*. https://doi.org/10.1080/13608746.2022.2047555.

Tyler, T.R. (1990). *Why people obey the law: procedural justice, legitimacy, and compliance*. New Haven: Yale University Press.

UNESCO (2013). *UIS methodology for estimation of mean years of schooling*. New Haven: Yale University Press.

Van de Walle, S., Van Roosbroek, S., & Bouckaert, G. (2008). Trust in the public sector: is there any evidence for a long-term decline? *International Review of Administrative Sciences*, *74*(1), 47–64.

Wolf, C., Schneider, S.L., Behr, D., & Joye, D. (2016). Harmonizing survey questions between cultures and over time. In C. Wol, D. Joye, T.W. Smith & Y. Fu (Eds.), *The SAGE handbook of survey methodology* (pp. 502–524). Thousand Oaks: SAGE.

Wood, S.N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(1), 95–114.

Wood, S.N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*(467), 673–686.

Wood, S.N. (2017). *Generalized additive models: an introduction with R*. Boca Raton: CRC press.

Wuttke, A., Gavras, K., & Schoen, H. (2020). Have europeans grown tired of democracy? New evidence from 18 consolidated democracies, 1981–2018. *British Journal of Political Science*. https://doi.org/10.1017/S0007123420000149.

Yan, T., Keusch, F., & He, L. (2018). The impact of question and scale characteristics on scale direction effects. *Survey Practice*, *11*(2), 1–10.

Závecz, G. (2017). Post-communist societies of Central and Eastern Europe. In S. Zmerli & T.W. van der Meer (Eds.), *Handbook on political trust* (pp. 440–460). Cheltenham: Edward Elgar.

Zieliński, M.W., Powałko, P., & Kołczyńska, M. (2018). The past, present, and future of statistical weights in international survey projects. In *Advances in comparative survey methods* (pp. 1035–1052). Hoboken: Wiley.