Modeling public opinion over time: A simulation study of latent trend models

Marta Kołczyńska ^{1*+}, Paul-Christian Bürkner ^{2*},

¹ Institute of Political Studies of the Polish Academy of Sciences, Poland

² Cluster of Excellence SimTech, University of Stuttgart, Germany

* Authors have contributed equally

⁺ Corresponding author, e-mail: mkolczynska@isppan.waw.pl

Abstract

Analyzing trends in public opinion is important for monitoring social change and for testing theories aimed at explaining this change. With growing availability of multi-wave surveys, social scientists are increasingly turning to latent trend models applied to survey data for examining changes in social and political attitudes. With the aim of facilitating this research, our study compares different approaches to modeling latent trends of aggregate public opinion: splines, Gaussian processes, and discrete autoregressive models. We examine the ability of these models to recover latent trends with simulated data that vary with regard to the frequency and magnitude of changes in the true trend, model complexity and data availability. Overall, we find that all three latent trend models perform well in all scenarios, even the most difficult ones with frequent and weak changes of the latent trend and sparse data. The two main performance differences we find include the relatively higher squared errors of autoregressive models compared to the other models, and the under-coverage of posterior intervals in high-frequency low-amplitude trends with splines. For all models and across all scenarios performance improves with increased data availability, which emphasizes the need of supplying sufficient data for accurate estimation of latent trends.

1 Introduction

Measures of mass public opinion are of interest to social science researchers for monitoring social change, and as hypothesized causes or consequences of political, social, and economic developments. National and

cross-national surveys constitute a rich data source describing public attitudes, values, and preferences from multiple countries over many decades. Combining these data into country time series of aggregate public opinion creates new opportunities for research, and at the same time requires a number of methodological choices that the literature has thus far not systematically discussed.

To guide researchers through these choices, we propose a framework for analyzing public opinion over time by pooling existing survey data sets. The framework includes three main components: the measurement model, the latent trend model, and representativeness adjustments. The present paper focuses on the statistical validation of the middle component of the framework, namely the model for the latent trend. Prior studies proposed solutions primarily relying on discrete autoregressive (AR) models (Caughey and Warshaw, 2015; Claassen, 2019; Solt, 2020). The paper by Kołczyńska et al. (2020) used penalized splines. Gaussian processes (GP) provide yet another alternative, that – to the best of our knowledge – have not yet been applied or studied in this context. Our analysis constitutes the first systematic comparison of these three approaches, and unifies previously disconnected model classes that researchers have proposed.

We compare the performance of models with the latent trend modeled as a spline, GP, or AR process, in scenarios that differ with regard to the shape of the true latent trend, presence or absence of between-person variation, and data availability. We find that all three models performed well in all scenarios, even in the most challenging condition that combines small fluctuations of the true trend with data sparsity. Our results point to relatively weaker performance of AR models, in terms of root mean squared error, posterior standard deviation, and absolute bias, as well as substantial under-coverage of posterior credible intervals implied by the spline models. In scenarios with sparse data coverage, performance increases with higher data availability in the given year and in adjacent years.

2 Framework for modeling aggregate public opinion

The framework we propose consists of three components: the measurement model, the latent trend model, and sample representativeness corrections. The focus of this paper is on the second element, the latent trend model, so we address the remaining two only briefly. The measurement model links the latent variable to the observed survey responses. Common approaches in models estimating dynamic public opinion include various kinds of ordinal models (e.g. Caughey and Warshaw, 2015; Solt, 2020; Kołczyńska et al., 2020), binary models applied to dichotomized data (e.g. McGann, 2014; Claassen, 2019), and linear models applied to ordinal data treated as continuous (e.g. Durand et al., 2021). In the present analysis we choose the variant of modeling individual-level data with ordinal item response theory models, which reduces information loss,

respects the ordinal nature of the data, accounts for the different scale lengths, and enables the inclusion of individual level predictors. Future studies may examine the relative advantages and disadvantages of these measurement models more systematically. The most common strategy of correcting sample representativeness is weighting the data, and alternatives include multilevel regression and poststratification (Gelman and Little, 1997). For a poststratification approach applicable within our modeling framework, see Kołczyńska et al. (2020). Since our simulations do not involve non-response or individual-level covariates, we leave the relative merits of weights and poststratification to future research.

2.1 Measurement model

The measurement model is formulated as an ordinal cumulative model (e.g., Samejima, 1997; Bürkner and Vuorre, 2019) of individuals' (ordinal) survey responses on a given item, that is, on the most fine-grained level of the survey data. In the context of large-scale survey research, this model was put forward by Kołczyńska et al. (2020) and we briefly introduce its most important aspects below. The cumulative model assumes that the ordinal response y originates from the categorization of a latent continuous variables $\tilde{y}$ according to an ordered latent threshold vector τ :

$$y = k \Leftrightarrow \tau_{k-1} < \tilde{y} \leq \tau_k, \quad (1)$$

where $k = 1, \dots, K+1$ denotes the ordinal response categories implied by K internal thresholds $\tau = (\tau_1, \dots, \tau_K)$ and two external thresholds $\tau_0 = -\infty$ and $\tau_{K+1} = \infty$. Assuming the latent factorization $\tilde{y} = \mu + \varepsilon$ into a latent mean μ to be predicted and a standard normally distributed error term ε implies that the probability of y being equal to k is given by

$$p(y = k|\mu) = \Phi(\tau_k - \mu) - \Phi(\tau_{k-1} - \mu), \quad (2)$$

where Φ is the cumulative distribution function of the standard normal distribution. This distributional assumption has the advantage that the latent scale is interpretable as (standardized) z -values, a widely known and understood scale, but other choices are possible as well (Bürkner and Vuorre, 2019). The design of survey questions, in particular the length of response scales as well as question wording, vary across projects, and there are also possible differences between surveys¹. Because of this, we assume a separate threshold vector

¹We use the term project to refer to the organization that coordinates the survey process and publishes the survey data under the same brand, for example the European Social Survey or the World Values Survey. Within projects, individual surveys include data collected from the same sample of respondents in the same fieldwork process. Multi-wave projects include many waves of data collection, i.e. multiple surveys from the same countries collected at different points in time. Within each project both the fieldwork process and the questionnaire design are coordinated and standardized, which is a condition to achieve comparability of surveys from different countries and years. Including project bias in models is intended to account for the various differences in the design of the items of interest and the entire questionnaire, as well as other aspects of the survey

for each survey-item combination following Kołczyńska et al. (2020). Further, to ensure that the latent trend model (see below) is jointly identified with the threshold vectors, we apply a sum-to-zero constraint to each of these threshold vectors. This constraint implies that changes in the latent trend over time are assumed to reflect actual attitude changes instead of changes to the content or interpretation of items (Kołczyńska et al., 2020).

The latent mean μ_i for observation i now is modeled by

$$\mu_i = b_0 + b_{\text{project}_i} + b_{\text{item}_i} + b_{\text{person}_i} + f(t_i), \quad (3)$$

where, b_0 is the mean level of the attitude of interest across time, b_{project} and b_{item} are project- and item-specific deviations from the mean (with a sum-to-zero constraint across projects and items for identification), b_{person} is the person-specific deviation from the mean (with a soft sum-to-zero constraint via a hierarchical normal prior, Bürkner, 2020), f is an unknown function of time (i.e., the latent trend model), and t_i represents the time point to which observation i belongs. This model can be understood as an additive multilevel model on the latent scale (Bürkner, 2018). Importantly, the model of μ_i can easily be extended to more additive terms that might be relevant in a given survey context, thus providing researchers with a lot of flexibility to adjust the presented framework to their specific modeling challenges.

2.2 Latent trend models

As latent trend models f , we consider three different approaches: (a) thin-plate smoothing splines (Wood, 2003), (b) zero-mean Gaussian processes with a smooth covariance kernel (Rasmussen and Williams, 2006; Williams and Rasmussen, 1996), and (c) discrete autoregressive effects of order 1 (Akaike, 1969). Approaches (a) and (c) have been applied already in earlier research of modeling latent trends in public opinion across time with survey data (Caughey and Warshaw, 2015; Claassen, 2019; Solt, 2020; Kołczyńska et al., 2020), but have never been compared with each other. Approach (b) has, to our knowledge, not yet been applied in this context.

process, including interviewer training, which may affect respondents' answers.

2.2.1 Thin-plate splines

As our first approach to modeling f , we consider (penalized) smoothing splines expressed via a basis function approach:

$$f(t) = \sum_{m=1}^M \alpha_m \phi_m(t) + \sum_{l=1}^L \beta_l \psi_l(t), \quad (4)$$

where ϕ_m and ψ_l are simple analytic basis functions (e.g., polynomials), and α_m and β_l are the corresponding spline coefficients (model parameters) to be estimated from the data. That is, for any given set of basis functions, the shape of f is defined by the spline coefficients. The conceptual difference between the coefficients sets $\{\alpha_m\}$ and $\{\beta_l\}$ is that one of them is penalized while the other is not (Wood, 2004). Suppose α_m are the penalized coefficients, then, from a Bayesian perspective, this means that they have a joint hierarchical prior

$$\alpha_m \sim \text{prior}(\lambda) \quad (5)$$

with hyperparameters λ that are themselves model parameters estimated from the data; whereas β_l have mutually independent, potentially flat priors (Wood, 2004). The here considered splines are parameterized in a way that the priors of α_m are zero-mean normal distributions with hyperparameters λ representing standard deviations over the non-linear space of the splines (Wood, 2004). Importantly, λ are estimated from the data, along with all other model parameters. There is a wide range of basis functions to choose from implying different properties for the corresponding splines. In this paper, we focus on (low rank) thin-plate splines (Wood, 2003) as they provide a general 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 (for more details, see Wood, 2003). These thin-plate splines have also been used in the study of Kołczyńska et al. (2020) to model the latent trend of political trust over time.

2.2.2 Gaussian processes

As our second approach to modeling f , we consider (zero-mean) Gaussian processes. From this perspective, f is a realization of an infinite dimensional Normal distribution:

$$f \sim \text{normal}(0, C(\lambda)) \quad (6)$$

where C is a covariance kernel with hyperparameters λ that defines the covariance between two function values $f(t_1)$ and $f(t_2)$ for two time points t_1 and t_2 (Rasmussen and Williams, 2006). Similarly to the

different choices of the basis function for splines, different choices of the covariance kernel lead to different Gaussian processes. In this paper, we consider the squared-exponential kernel defined as

$$C(\lambda) := C(t_1, t_2, \sigma, \gamma) := \sigma^2 \exp\left(-\frac{(t_1 - t_2)^2}{2\gamma^2}\right) \quad (7)$$

with hyperparameters $\lambda = (\sigma, \gamma)$, expressing the overall scale of Gaussian process and the length-scale, respectively (Rasmussen and Williams, 2006). The advantages of this kernel is that it is computationally efficient and (infinitely) smooth making it a reasonable choice for the purposes of the present paper. Here again, λ are estimated from the data, along with all other model parameters.

2.2.3 Discrete autoregressive models

As our third approach to modeling f , we consider a discrete autoregressive model of order 1 such that

$$f(t) = \theta_t \sim \text{normal}(\theta_{t-1}, \sigma) \quad (8)$$

for $t = 2, 3, \dots$ and $f(1) = \theta_1 \sim \text{normal}(0, \sigma)$, where θ_t are the autoregressive coefficients (model parameters) and σ is a standard deviation (hyperparameter) controlling the size of the discrete ‘jumps’ from time t to time $t + 1$. As for the other latent trend models, this hyperparameter is estimated from the data along with all other model parameters. In the context of modeling latent trends over time in survey research discrete autoregressive models of order 1 have been applied in Caughey and Warshaw (2015), Claassen (2019), and Solt (2020).

3 Simulation design

The simulations envision data from a single country from five projects each with surveys administered every year for 25 years. In all projects each survey includes three questions on the same three issues, but the length of response scales varies by project. This design, and various parameters of the simulations described in the following sections, are modeled after an analysis of trajectories of average political trust in Europe in Kołczyńska et al. (2020). Data simulation scenarios represent combinations of (1) the shape of the true latent trend, (2) the presence of between-person variation, and (3) data availability.

3.1 Shape of the true latent trend

We examine six different types of the latent trend, representing combinations of three variants of the frequency of the trend, and two variants of the long term tendency, as presented in Figure 1. The three frequencies we examine include a pattern with low frequency and large amplitude (which covers much of the range of the latent scale), medium frequency and medium amplitude, and high frequency and small amplitude. All three trajectories represent realistic variation patterns of aggregate societal values or attitudes, of which each particular case of a country’s trajectory likely constitutes a mix. Given that there is still relatively little long-term longitudinal research of mass public opinion, the true volatility of the latent trend may not be known a priori. The two long-term tendencies include the lack of a long-term tendency, i.e. stationarity, and a weakly declining long-term linear trend, i.e. non-stationarity. Since we found no substantive differences between those two long-term scenarios, we focus only on the former in the presentation of results. The results for the non-stationary scenario are presented in the Appendix.

Figure 1: Shapes of the true latent trend.

3.2 Between-person variation

In modeling latent trends in public opinion, the data include multiple responses obtained from the same person, which can be accounted for by including person specific parameters (also known as person random effects) in the latent model (3). Specifically, in our simulations between-person variation is generated from a normal distribution with mean 0 and standard deviation τ . The two possibilities we consider include no between-person variation (i.e., $\tau = 0$) and $\tau = 1.2$, which is a realistic value based on results from Kołczyńska

Table 1: Summary of data simulation scenarios and latent trend model types. On data simulated within each data simulation scenario, corresponding to combinations of the conditions in first three columns, we estimated three types of latent trend models, listed in the fourth column.

Data simulation scenarios			Latent trend models
Shape of the true trend	Between-person variation	Data sparsity	
High frequency, stationary	Zero	Full data	Spline
Medium frequency, stationary	Non-zero	Sparse data	Gaussian process
Low frequency, stationary			Autoregressive model
High frequency, non-stationary			
Medium frequency, non-stationary			
Low frequency, non-stationary			

et al. (2020). In both conditions, the estimated models were correct, i.e. they modeled between-person variation if included in the data generation process, and did not otherwise.

3.3 Sparsity of the data

In the full data scenario, data are available for all projects, items, and all 25 years. In the sparse data scenario, data are available for selected years in each project, as presented in Figure 2. Whenever data are available, they are available for all three items.

This sparsity pattern mimics the availability of data from multi-wave survey projects for many European countries, for which there are survey projects that (1) span the entire period but with infrequent waves, (2) cover a small part of the studied period, (3) cover a large part of the studied period with regular and frequent waves, (4) cover an already rather well surveyed part of the studied period with annual surveys, and (5) provide infrequent surveys that do not cover the entire analyzed period.

The sparse scenario includes data for 31 out of the 125 project-years (25%), for 19 out of 25 years (76%), and without gaps in coverage longer than 1 year. In the light of the availability of survey data worldwide, this is a realistic but relatively comfortable situation characteristic for well-surveyed countries.

Overall, we have $6 \times 2 \times 2 = 24$ data scenarios, which correspond to all combinations of the conditions in the three first columns in Table 1. For each data scenario, we simulated 100 data sets.

Figure 2: Data availability per project and year in the sparse scenario.

3.4 Data simulation

The simulated data include five survey projects, which – in the full data scenarios – provide data every year for 25 years. In each year’s survey there are 100 respondents who provided responses to three items that can be treated as indicators of an underlying latent variable. There is no unit or item non-response. Altogether, each simulated dataset in the full data scenarios has $5 \times 25 \times 100 \times 3 = 37500$ observations. In the sparse data scenarios, each dataset has $31 \times 100 \times 3 = 9300$ observations.

Data are simulated with the following procedure. For each of the true latent trends described in Section 3.1:

- (1) Draw project bias (five projects) and item bias (three items) from $\text{normal}(0, 1)$, and center the project and item biases to each have a mean of 0.
- (2) For each survey, draw 100 person-specific deviations from $\text{normal}(0, \tau)$, where τ equals 0 in the scenario without between-person variation, and τ equals 1.2 in the scenario with between-person variation.
- (3) Each respondent’s latent mean μ_i per item is the sum of project and item bias, person-specific deviation, and latent trend $f(t)$, following Equation (3).
- (4) Draw per-observation random errors ε_i from $\text{normal}(0, 1)$.
- (5) Each respondent’s latent response $\tilde{y}_i$ per item is the sum of the latent mean and random error.
- (6) For each project, draw the number of thresholds from the set of options $\{1, 3, 4, 6, 9, 10\}$ corresponding to response scale lengths represented in surveys $\{2, 4, 5, 7, 10, 11\}$ with probabilities of $\{0.1, 0.2, 0.2, 0.2, 0.2, 0.1\}$, which roughly reflects the prevalence of the different response scale lengths in cross-national surveys.
- (7) Draw the threshold values τ from $\text{normal}(0, 1)$.
- (8) Center the thresholds to have a mean of 0 thereby enforcing the sum-to-zero constraint mentioned in Section 2.1.
- (9) Use the thresholds to cut the latent response into discrete categories following Equation (1).

- (10) In the sparse scenario, keep only the project-years as indicated in Figure 2 and discard the remaining simulated data.

On each of the 2400 data sets (100 data sets for each of 24 scenarios) we ran three types of latent trend models discussed in Section 2.2.

3.5 Model estimation

The analysis was performed using fully Bayesian estimation with Hamiltonian Monte Carlo as implemented in Stan (Carpenter et al., 2017). We set $\text{normal}(0, 3)$ priors on the thresholds and for the remaining parameters we used default priors as specified in the brms package (Bürkner, 2017). We used default settings for adapt delta of 0.8 and maximum tree depth of 10, and set initial values to 0 on the transformed, unbounded scale. Each model was run with one chain, 3000 iterations, of which 1000 were for warm-up. The model code, as well as all other scripts, are available in the replication materials.

The high total number of models made the whole task computationally intensive. Within each scenario, the 100 models were ran in parallel, one core per model, on a high-performance computing cluster. Each set of 100 models took up to 5 days for the full data scenarios, up to 24 hours for the sparse data scenarios.

We performed the analysis in R (R Core Team, 2018) using the brms package (Bürkner, 2017), which provides a high-level interface to the probabilistic programming language Stan (Carpenter et al., 2017). For parallelization we used the packages parallel, part of base R, foreach (Microsoft and Weston, 2020), and doParallel (Microsoft Corporation and Weston, 2019). We also used several tidyverse packages (Wickham et al., 2019) for data processing and visualization.

4 Results

The purpose of the models is to estimate average levels of the latent attitude in each year, so our analysis focuses on these estimates only. Within each data scenario, the crucial comparisons are between the performance of splines, GPs, and AR models, but we also discuss differences in overall performance between data scenarios. We report the results in terms of root mean squared error (RMSE), posterior standard deviation and absolute bias, as well as coverage by 66% and 90% posterior credible intervals. It is worth noting that in examining posterior credible intervals we are interested in frequentist properties of Bayesian credible intervals, although we acknowledge that these credible intervals are not necessarily designed to include the true value in the nominally indicated percentage of cases.

We analyze the results both descriptively, by examining the distributions of the performance metrics across 100 simulations per scenario, as well as by fitting Bayesian models to obtain (1) estimates and uncertainties of average performance per scenario, and (2) estimates and uncertainties of average performance depending on the number of available surveys within each sparse scenario. Details are provided in the Appendix, which also includes results for additional performance metrics.

First of all, we notice that across all datasets in all scenarios and for all latent trend models, RMSE remains below 0.5. The average RMSE per scenario barely exceeds 0.2 for only one of them (in the scenario with sparse data, presence of between-person variation, and the high-frequency trend) and remains substantially below 0.2 for the others. Given the scale of the latent trend these results indicate good performance of all models.

Comparing model performance across scenarios, models estimated on data without between person variation perform better compared to models estimated on data with between person variation in terms of RMSE, posterior standard deviation and absolute bias (compare facets in the first and second column of Figure 3, and those in the third and fourth column). Further, models estimated on full data perform better than models estimated on sparse data (compare facets in the first and third column of Figure 3, and those in the second and fourth column). In sum, the difficulty of the scenario has direct implications for model performance. At the same time, differences between the shapes of the true trends are in most cases relatively small and depend on the latent trend model.

We now turn to comparisons of latent trend models within each of the 12 data scenarios presented in Figures 3 and 4. AR models tend to have the highest RMSE (with the exception of the scenario with full data, no between-person variation, and high frequency trend; see facets in the top row of Figure 3). The difference between AR models and the other two is the biggest for the low frequency trend, smaller in the medium frequency scenario, and the smallest in the high frequency scenario. RMSE for spline and GP models is generally very similar.

The differences in RMSE between AR models and the other two are largely due to higher posterior standard deviations, and to a lesser extent due to absolute bias (Figure 3). In terms of absolute bias, AR models also perform worst with the exception of high frequency trends in all sparsity and between person variation combinations. In the high frequency scenario, spline models have the highest absolute bias.

In most cases, coverage is very close to what is expected based on the nominal width of the credible interval, with two main exceptions (cf. Figure 4). First, spline models exhibit lower coverage in models estimated on data with the high frequency trend. The differences are substantial, with average coverage by the 66%

credible interval not exceeding 50%, and as low as under 20% in the scenario with full data and no between person variation. Similar gaps are observed for other credible intervals. Second, coverage for AR models tends to be higher than expected in the scenarios with sparse data. There, in the low frequency scenarios, coverage by the 66% credible interval reaches 80%, and generally exceeds the credible intervals' nominal values.

We also examined the role of data availability on model performance in the sparse scenarios. Data availability is measured as the number of surveys available in the given and adjacent years and ranges from 1 in years 3, 9, and 11, to 7 in year 20 (cf. Figure 2). In all scenarios having more surveys translates into lower RMSE, posterior standard deviation, and absolute bias (Figure 5). More data also reduces over-coverage in AR models, especially in the low frequency scenarios. Perhaps more importantly, a higher number of available surveys substantially improves coverage in the high frequency scenario in models with splines, which on average suffer from under-coverage (Figure 6).

Figure 3: Estimated average levels of root mean squared error, posterior standard deviation, and absolute bias by scenario. Each point represents the estimated mean of the given performance metric across 100 models in each simulation scenario. Error bars indicate 95% posterior credible intervals of the mean estimates. Gray horizontal lines indicate no error. BPV = between-person variation; Low frequency = low frequency trend; medium frequency = medium frequency trend; High frequency = high frequency trend; GP = Gaussian process; AR = autoregressive model.

Figure 4: Estimated average levels of coverage by scenario. Each point represents the estimated mean level of coverage by the given posterior credible interval across 100 models in each simulation scenario. Error bars indicate 95% posterior credible intervals of the mean estimates. Gray horizontal lines indicate the reference value corresponding to the nominal width of the credible interval. BPV = between-person variation; Low frequency = low frequency trend; medium frequency = medium frequency trend; High frequency = high frequency trend; GP = Gaussian process; AR = autoregressive model.

Figure 5: Estimated values of root mean squared error, posterior standard deviation, and absolute bias by the number of available surveys in the given and adjacent years for each sparse data simulation scenario. Gray horizontal lines indicate the reference value of 0. Uncertainty is indicated as 95% posterior credible intervals. BPV = between-person variation; Low frequency = low frequency trend; Medium frequency = medium frequency trend; High frequency = high frequency trend; GP = Gaussian process; AR = autoregressive model.

Figure 6: Estimated values of coverage by the 66% and 90% posterior credible interval by the number of available surveys in the given and adjacent years for each sparse data simulation scenario. Gray horizontal lines indicate the reference value corresponding to the nominal width of the credible interval. Uncertainty is indicated as 95% posterior credible intervals. BPV = between-person variation; Low frequency = low frequency trend; Medium frequency = medium frequency trend; High frequency = high frequency trend; GP = Gaussian process; AR = autoregressive model.

5 Conclusion

We introduced a framework to jointly model individual-level responses across items, surveys, projects, and time by combining ordinal measurement models with additive multilevel models on the latent scale. Within this framework, we specifically compared three approaches to modeling latent trends that can be used in analyses estimating changes in public opinion: splines, Gaussian processes, and discrete autoregressive models. Our simulations comprise scenarios combining the shape of the true latent trend, the complexity of the model, and data availability. Overall, our results show that all three types of latent trend models perform well even in the most difficult scenarios with between-person variation, sparse data, and a trend with frequent small fluctuations.

Our analysis suggests Gaussian processes as the most reliable method for estimating trends in the examined contexts if the variability of the latent trend is not known a priori. Splines perform equally well, except for cases when the true trend is weak with frequent small fluctuations. In that latter case, splines (at least the ones applied here) will likely underestimate the uncertainty, unless sufficient data are available – with enough data the under-coverage by spline models is small. AR models tend to overestimate uncertainty and on average have highest RMSE of all three latent model types. In addition to considering these differences, the ultimate choice of the latent trend model would likely depend on other factors as well, including the model’s ability to incorporate individual-level predictors and avoiding possible problems with model specification or estimation. Regardless of which model is chosen, we emphasize the need to account for the uncertainty associated with estimates of the latent trends in explanatory models.

The framework for analyzing public opinion trends with data from different survey projects conceptually connects to individual data meta-analysis performed in particular in medical or psychological research, where individual participants’ data from different studies are pooled and analyzed jointly (e.g. Ioannidis, 2017; Riley et al., 2010; Riley, 2010). In social survey research, methods for combining different existing survey data sets are studied under the label “ex-post survey data harmonization” (Wolf et al., 2016). Despite the differences in the types of data and analyses, both lines of research emphasize the need to take into account the clustering of participants within studies as well as to include study- and participant-level characteristics.

5.1 Limitations

The simulations we designed take into account some of the issues that researchers of aggregate public opinion encounter, most notably limitations to data availability. Others have remained unexamined, including possible problems with data quality, such as representation errors, as well as added complexity of the measurement model, such as varying discrimination parameters across items and/or projects. All these aspects will influence the performance of the estimated latent trend models.

We also examined just one sparsity pattern, due to space limitations, and one that is rather optimistic in the light of survey data availability on the issues that are of interest to social scientists (cf. Claassen, 2020). When including item and project bias, what will matter is not only the number of surveys overall, but the number of surveys per project and how the surveys from different projects are spread over the period of interest. In applications with real data, the combination of the sparsity pattern and the true latent trend will likely determine the difficulty of the modeling task, because detecting smaller fluctuations requires more data. Overall, when relying on sparse data coverage – which is the reality in this kind of research – caution

is advised when drawing conclusions about fluctuations smaller and more frequent than the ones in our high-frequency scenario.

6 Funding

This work was supported by the Polish National Agency for Academic Exchange in the Bekker programme [award number PPN/BEK/2019/1/00133].

7 Acknowledgements

We thank Aki Vehtari for discussions of the simulation design. We also acknowledge the computational resources provided by the Aalto Science-IT project.

8 Data Availability Statement

Replication code for this article is available in the Open Science Framework repository: <https://osf.io/j45er/>

References

- Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21(1), 243–247.
- Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80(1).
- Bürkner, P.-C. and M. Vuorre (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science* 2(1), 77–101.
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1), 395–411.
- Bürkner, P.-C. (2020). Bayesian item response modelling in R with brms and Stan. *Journal of Statistical Software*, 1–54.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software* 76(1).
- Caughey, D. and C. Warshaw (2015). Dynamic estimation of latent opinion using a hierarchical group-level IRT model. *Political Analysis* 23(2), 197–211.
- Claassen, C. (2019, jan). 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.
- Durand, C., L. P. Peña Ibarra, N. Rezgui, and D. Wutchiett (2021, jan). How to combine and analyze all the data from diverse sources: a multilevel analysis of institutional trust in the world. *Quality & Quantity*.
- Gelman, A. and T. C. Little (1997). Postratification into many categories using hierarchical logistic regression. *Survey Methodology* 23(2), 127–135.
- Ioannidis, J. (2017, oct). Next-generation systematic reviews: prospective meta-analysis, individual-level data, networks and umbrella reviews. *British Journal of Sports Medicine* 51(20), 1456–1458.
- Kołczyńska, M., P.-C. Bürkner, L. Kennedy, and A. Vehtari (2020). Modeling public opinion over time and space: Trust in state institutions in europe, 1989-2019. osf.io/preprints/socarxiv/3v5g7.

- McGann, A. J. (2014, jan). Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm. *Political Analysis* 22(1), 115–129.
- Microsoft and S. Weston (2020). *foreach: Provides Foreach Looping Construct*. R package version 1.5.0.
- Microsoft Corporation and S. Weston (2019). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.15.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rasmussen, C. E. and C. K. I. Williams (2006, January). *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. Cambridge, MA, USA: MIT Press.
- Riley, R. D. (2010, oct). Commentary: Like it and lump it? Meta-analysis using individual participant data. *International Journal of Epidemiology* 39(5), 1359–1361.
- Riley, R. D., P. C. Lambert, and G. Abo-Zaid (2010, aug). Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 340(feb05 1), c221–c221.
- Samejima, F. (1997). Graded response model. In *Handbook of Modern Item Response Theory*, pp. 85–100. Springer.
- Solt, F. (2020). Modeling Dynamic Comparative Public Opinion. *SocArXiv*.
- Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani (2019). Welcome to the tidyverse. *Journal of Open Source Software* 4(43), 1686.
- Williams, C. K. I. and C. E. Rasmussen (1996, June). Gaussian processes for regression. In *Advances in neural information processing systems 8*, Cambridge, MA, USA, pp. 514–520. Max-Planck-Gesellschaft: MIT Press.
- Wolf, C., S. L. Schneider, D. Behr, and D. Joye (2016). Harmonizing Survey Questions Between Cultures and Over Time. In C. Wol, D. Joye, T. W. Smith, and Y.-c. Fu (Eds.), *The SAGE Handbook of Survey Methodology*, pp. 502–524. 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.